Scalable Computing:
Practice and Experience

Scientific International Journal
for Parallel and Distributed Computing

ISSN: 1895-1767

Universitatea de Vest din Timişoara

Volume 26(2) March 2025

SUBSCRIPTION INFORMATION: please visit `http://www.scpe.org`

Scalable Computing: Practice and Experience

Volume 26, Number 2, March 2025

TABLE OF CONTENTS

Papers in the special issue on Unleashing the power of Edge AI for Scalable Image and Video Processing:

Papers in the special issue on High-performance Computing Algorithms for Material Sciences:

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 503–516, DOI 10.12694/scpe.v26i2.3969

MODELING AN INTELLIGENT FRAMEWORK FOR OPTIMIZING UAV PATH PLANNING AND ANTI-COLLISION IN AGRICULTURE

VIJAYA CHANDRA RAO V[*], KARTHIKEYAN M P[†], DR SAVITA[‡], VENKATA SAIRAM KUMAR N[§], ANUBHAV BHALLA[¶], AND LAKSHYA SWARUP[‖]

Abstract. Unmanned aerial vehicles (UAV) are increasingly utilized for monitor expansive farming to their effectiveness in observation and data gathered. Efficient path planning and collision prevention are critical for optimizing UAV operation in such settings. The efficiency can be controlled by the quality and declaration of UAV sensed data, as well as the difficulty of real world ecological variables that could impact path optimization and collision prevention. This research introduce a novel technique, the customizable dung beetle search tuned random forest (CDBS-RF) method, designed to improve UAV route planning. The CDBS-RF method integrated the dung beetle search (DBS) algorithm, known for its robust optimization capabilities, with random forest (RF) to enhance optimal path security and effectiveness. This method dynamically adjusts path safety and effectiveness. This approach dynamically adjusts path planning parameter to make sure optimal route selection and collision prevention. The suggested technique was evaluated utilizing UAV-sensed data and implement in python based virtual environment. Experimental consequences demonstrate that the CDBS-RF approach considerably enhances UAV path planning performance, provide safer and more efficient navigation for self and more efficient navigation for self-sufficient tarp monitoring. The performance evaluation techniques include the planning time (0.789) and path length (21.526). By utilize advanced optimization and anti-collision algorithms, this technique offer a promising solution for improving UAV operations in agricultural surveillance.

Key words: Agriculture, UAVs (unmanned aerial vehicles), path planning, obstacles, anti-collision, customizable dung beetle search-tuned random forest (CDBS-RF)

1. Introduction.

Unmanned aerial vehicle s(UAVs) are suitable additional established and utilized in several different sectors, such as profitable deliveries, farming, and emergency relief. worldwide, there have been a lot UAV flying operation. The anti-collision approaches of UAVs have garnered important attention because of their wide utilizes, to avoid UAVs from collide with other objects [1].

1.1. UAVs Path Planning.

UAVs are superior demand between civilians for uses such as aerial investigation, search and rescue operations, and military uses. According to, UAVs have shown to be quite helpful in the agricultural sector, particularly for monitoring palm oil plantation, which gives farmers on how to supervise their plantation [2]. One additional advantage of using UAVs is that they can equipped with a range of measuring devices that can assist in eliminating fixed position constraints and enable real-time, unrestricted measurement in three dimensions.

The quick expansion of many aircraft types, particularly UAVs has been fueled by the increasing liberalization of low-altitude airspace and the quick growth of the general aviation sector [3].UAVs are becoming an important part of the worldwide commercial industry. Depending on their intended applications, civilian drones can be classified as patrol, agricultural, meteorological, exploratory, or disaster relief drones. Based on this, nations substantial sums of money have been invested globally in the study and creation of UAVs with

[*]Computer Science and Engineering, G. Pullaiah College Of Engineering and Technology, India (vijaychqndrarao.v@gmail.com)
[†]Department of Computer Science and Information Technology, Jain (Deemed to be University), Bangalore, Karnataka, India (karthikeyan.mp@jainuniversity.ac.in)
[‡]Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India (Savita@muit.in)
[§]Department of Civil Engineering. R.V.R & J.C College of Engineering, Guntur, India (sairam852@gmail.com)
[¶]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India (anubhav.bhalla.orp@chitkara.edu.in)
[‖]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103, India (lakshya.swarup.orp@chitkara.edu.in)

greater integration, improved processing efficiency, and faster reaction times [4].

1.2. Advancements in Agricultural UAV Applications. Plant protection, farmland information, precision agriculture, and agricultural insurance UAV use is increasing in businesses that include pesticide spraying, fertilization, and seeding. UAVs are currently being used to monitor and control illnesses and pests in local areas and are encouraged by the governments of various countries since agriculture is becoming more and more popular. Plant protection UAVs provide several benefits over traditional pesticide spraying methods, including superior labor efficiency, excellent mobility, consistent and modest pesticide dose per unit area and a vertical takeoff and landing capability [5].

Robotic systems have become more and more common in the past ten years as a means of handling laborious agricultural jobs in field operations. Numerous robotic systems have been created to address the complexity of agricultural field tasks, including harvesting, spraying, fertilizing, sowing, weeding, and harvesting [6].Civilizations centered on trees for many nations, europea L. represent a significant source of income. The quality of the product is affected when the fly female deposits in the fruits that developing larvae eat. Pesticides are used to combat this type of infestation, but environmental issues and the resulting financial expenses make the development of innovative strategies to minimize their usage necessary [7].Recent technical breakthroughs regarding farming have led to a rise in the usage of Robots in the industry. A cheap alternative is provided by the UAV, a mobile robot for conventional detecting technologies and data analysis methods. UAVs come in a range of variety, and inexpensive UAVs are competent of gathering high-resolution images from a lot of location in space. Even while UAVs aren't at the present utilized in the majority of accuracy farming application, are becoming additional concerned in the business in terms of beneficial and sustainable farming techniques. The UAV, which is utilized in farming, offers dimension accuracy and drastically lower the need for human resources. When the data from the UAV are appropriately assessed and analyze, they can increase crop efficiency and develop crop yield [8].

UAVs have a great possible to enhanced crop, water, and pest management effectiveness, and they can be especially useful for precision agricultural applications.They are also capable of doing a variety of agricultural tasks, such as monitoring soil health, applying fertilizer, and analyzing weather. A UAV with a camera and agile motions can supplement human labor in scenario evaluation and surveillance tasks by offering basic assistance. Additionally, the UAV may utilize numerous sensors at once, and sensor fusion can boost analysis. Farmers may continually monitor crop variability and stress levels by using UAVs to obtain vegetation indicators [9].

1.3. The challenges of path planning and collision avoidance. It is a critical challenge in UAV operations, particularly in complex agricultural environments. UAV must navigate dynamic and cluttered landscape, which can include varying terrain, tall vegetation, and unpredictable weather conditions. Effective path planning involves not only finding the most efficient route but also dynamically adjusting to obstacles that can appear suddenly such as other UAV. Collision avoidance further complicated the task by requiring real-time detection and response to potential threats, which can be hindered by limited sensor range oorr processing power. Additionally, the need to balance optimal path efficiency with safety considerations adds another layer of complexity. Recent advancements in algorithms and sensor technologies are addressing these challenges, but achieving robust and reliable performance remains a significant hurdle. As UAV are increasingly deployed for agricultural tasks, developing sophisticated path planning strategies and collision-avoidance systems is essential for enhancing operational safety and effectiveness [10].

Aim of the study. The principal aim of this undertaking is to produce a path collision avoidance and path planning system for a UAV that will be used to check apple fly traps.This study aims to improve an online adaptive method of UAV route planning based on the Customizable dung beetle search-tuned random forest (CDBS-RF) approach.

The remainder of the paper. Section 2 contains related works. There was a comprehensive methodology within Section 3. An analysis of the findings is provided in Section 4, and a conclusion is provided in Section 5.

2. Related work. According to the author [11]examined the variety of navigation situations with increasing complexity and static obstacle density are applied to evaluate two of the most popular search path planning methods for geometrical simulations: the (A* and D*) algorithms. It executes intricate situations with outstanding outcomes in terms of computation time, collision avoidance abilities, and length of path.

Study [12] explored the safe Blockchain-based communication between UAVs and wireless unmanned aerial vehicles (WUAV).Because the position is subjected to change in an unexpected environment, base station transfer will be delayed. UAVs are more likely to be the target of security breaches. The use of a flocking control method inspired by starling behavior, a large-scale UAV swarm working in a dynamic and unpredictable three-dimensional environment can increase the efficacy, security, and dependability of obstacle avoidance [13]. The motion model is constructed by examining the systematic and quick obstacle avoidance behavior of the starlings. Current examples of RF data transfer and visual processing to integrate leader-follower UAVs are provided in this study. This system uses embedded electronics, visual processing, and controls to provide the simple deployment of several quadrotor UAVs under the control of a single operator in search and rescue scenarios [14].According to [15] the paper provided a thorough overview of UAV anti-collision systems. To prevent collisions at the policy level, we must propose legislation and regulations on UAV safety. Furthermore, from the standpoint of quick obstacle detection and quick wireless networking in UAV anti-collision technology are discussed. The coverage issues that might come up while monitoring and when implementing agro-technical interventions are in this article [16]. The proposed method, mhCPPmp, utilizes a genetic algorithm to plan flight paths for various UAV types while they are refueling and charging on a moving ground platform. The suggested multigene, enhanced anti-collision RRT* and Iterative Adaptive Configuration-Rapidly Exploring Random Tree (IAC-RRT*) algorithms modify the probability of mutation and crossover. UAVs can efficiently cut down on energy use and task completion time because of their multigene and IAC-RRT* algorithms [17].

Study [18] suggested a way for autonomously assigning tasks and making decisions for numerous cooperative quad copters to design a coverage path. The best solution for the given issue was found by applying the Sequential Quadratic Programming (SQP) approach. Next, using the Stateflow approach, MATLAB Graphical User Interfaces (GUIs) construct a simulation platform, and numerous the real flying experiments were carried out using ZY-UAV-680 quadrotor UAVs . Article [19] proposed the use of an intelligent logistics UAV as a courier substitute for minor goods deliveries. The quad copter was controlled by a mobile application that was linked with a webcam and an ultrasonic ground proximity warning system (GPWS). Improved PID (proportion-integral-derivative) controllers and LQR(linear quadratic regulator) were used in its light control system. The article employs the ant colony method and dynamic route planning, which could quickly determine the ideal path of a UAV in challenging terrain when compared to the classic an algorithm and artificial potential field technique. Study [21] examined the use of ML techniques to predict crop yields using a large dataset from Indian agriculture that includes variables such as soil composition, seasonal variations, and fertilizer use. Ten machine learning methods were evaluated for performance. The paper examined the vital topic of enhancing agricultural decision-making and guaranteeing food security, which are essential components of the sustainable development goal (SDG).

2.1. Problem statement. The problem addressed in the reviewed studies is the optimization of path planning, collision avoidance, and efficient communication for UAV operating in dynamic and complex environments. The examination of various navigation algorithms, such as A* and D* reveals their effectiveness in different scenarios, yet challenges remain in enhancing computation time and collision avoidance capabilities. UAV face increased risks of security breaches and operational delays due to their variable positioning and environmental unpredictability. To address these issues, several advanced approaches are explored flocking control inspired by starling behavior to improve swarm coordination, integration of RF data transfer and visual processing for leader- follower UAV systems, and multi gene algorithms for optimizing flight path and reducing energy consumption. Additionally autonomous task assignment and converge path planning are achieved through sequential quadratic programming (SQP) and real flight experiments. The use of intelligent logistics UAVs for small deliveries highlights the need for improved route optimization to overcome challenging terrain and enhance overall UAV performance.

3. Methods. The quality and resolution of UAV sensed data, together with the complexity of real-world environmental factors that may affect path optimization and collision avoidance, might limit the efficacy. In order to improve UAV route planning, this research presents a unique method called the customizable dung beetle search tuned random forest (CDBS-RF) the technique displayed in Figure 3.1.

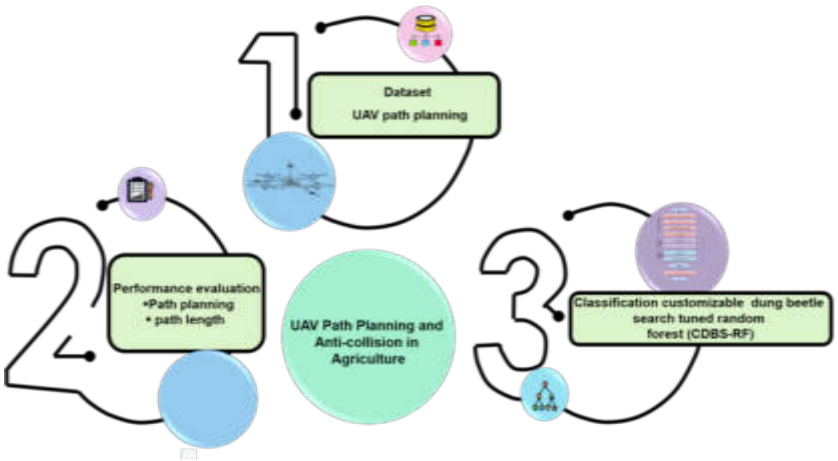


Fig. 3.1: Block schematic illustrating the intended workflow

3.1. Dataset. A UAV is utilized to images the fly traps that are attached to the tree crop along a growing zone to perform an apple as part of the fly trap inspection procedure. Due to its unstructured nature, this type of space needs the ability to identify and prevent collisions to ensure process security. An image of a region used for apple cultivation. To obtain the trap images, the UAV has to fly at a height that falls between the tree crops that is not much higher than the mean of the tree diameters.

The intricate flying zone, with a few branches that reach beyond the tree canopy, is depicted in the illustration. When the UAV is moving, other dynamic objects in this area, such as people, animals, and vehicles that can be seen. In certain cases, the flight controller must intervene to prevent them. This image depicts the UAV's possible behavior during its relocation to arrive at a point of inspection. The UAV makes use of light detection and ranging (LIDAR) to its surroundings as it travels toward its goal after obtaining the route program. Upon arriving at the trap location, the UAV takes an image before moving on to the next target point. The following criteria are used to evaluate the proposed solution in this paper: There are five traps per tree, arranged in a random pattern across the areas

• The minimum distance required to take an image of the trap is 3.0 meters.

• The UAV's planar LIDAR, which has a360-degree scanning angle, can be used to identify obstructions in space.

• Every trap has a preset position that is utilized to provide the route planner with target locations.

The LIDAR scanning methodology led the UAV to decide against using the vertical collision avoidance. Since the UAV lacks an integrated sensor to recognize objects on the aircraft's upper side, it would be difficult for the algorithm to work if vertical detection was included. To ensure a trajectory free of collisions, the RF will create the path between the rows of apple trees. The robot is believed to be able to cohabit with other robots, farm people, shifting tree branches, etc. when faced with dynamic impediments. The proposed CDBS-RF can handle extremely dynamic and time-varying situations while achieving a map exploration speed.

3.2. Quadrotor model design. The four fundamental motions were controlled by four separate cascaded proportional-integral (PI) controllers as shown in Figure 3.2. The inner loop and outer loop are two types of PI controllers. When disturbances impact a quantifiable secondary variable input in the middle that is directly related to influences the main outcome that has to be managed, there are benefits to this type of cascaded controller.

Disorders that reach the second variable can have an unwanted effect that the cascaded control system can lessen. This system's outer loop works at a lower frequency than the inner loop to manage the system bandwidth. Reference position data is received by the position control PI loop, and output data is received by the PI loop for speed control. The speed loop generates the command to drive the motors. The inertial

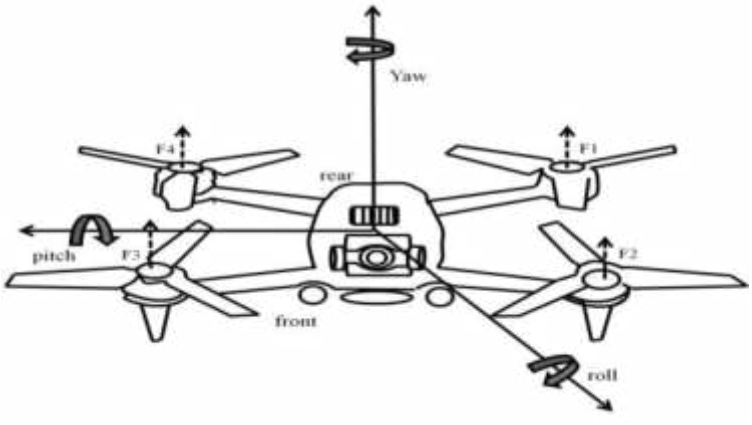


Fig. 3.2: Structure of UAV Quadrotor

measurement unit's (IMU) gyroscope provides instantaneous feedback to the speed loop. The controller's is to change the four propellers' speeds to obtain the intended quad-rotor orientation. One used a cascading connection to connect the foundations (1), where l is the separation between the rotors that are in opposition to one another.

$$\begin{bmatrix} \Omega_1^2 \\ \Omega_2^2 \\ \Omega_3^2 \\ \Omega_4^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4c}V_1 - & \frac{1}{2cl}V_3 - & \frac{1}{4d}V_3 \\ \frac{1}{4c}V_1 - & \frac{1}{2cl}V_2 + & \frac{1}{4d}V_3 \\ \frac{1}{4c}V_1 + & \frac{1}{2cl}V_3 - & \frac{1}{4d}V_3 \\ \frac{1}{4c}V_1 + & \frac{1}{2cl}V_2 + & \frac{1}{4d}V_3 \end{bmatrix} \tag{1}$$

3.3. Path planning using Customizable dung beetle search-tuned random forest (CDBS-RF). The Customizable dung beetle search-tuned random forest (CDBS-RF) is a path planning method that combines the CDBS algorithm's adaptive exploration capabilities with RF robust predictive modeling. This method optimizes navigation in complex environments by dynamically adjusting the search strategy based on environmental feedback. The algorithm refines path predictions through decision trees and accommodates real-time adjustments to changing conditions, resulting in a more accurate, resource-efficient, and adaptable path planning solution, offering significant improvements over traditional methods in computational efficiency and accuracy in dynamic and complex scenarios.

3.3.1. Customizable Dung Beetle Search (CDBS). The fundamental CDBS is population-based and principally inspired by the actions of the dung beetle including thieving, rolling balls, dancing, scavenging, and procreating. Four kinds of search agents little dung beetles, brood balls, ball-rolling dung beetles, and thieves—are used by the CDBS to divide the population. More specifically, every search agent has a unique set of updating guidelines. Figure 3.3 displays the CDBS flow chart.

A. Ball-rolling dung beetle. To maintain a straight course for the rolling dung ball, dung bugs must adhere to astronomical signals. Consequently, the rolling dung beetle's location has been modified and can be expressed as follows:

$$W_j\left(s+1\right) = W_j\left(s\right) + \alpha \times l \times W_j\left(s-1\right) + a \times \triangle w \tag{2}$$

$$\triangle w = \left|W_j\left(s\right) - W^x\right| \tag{3}$$

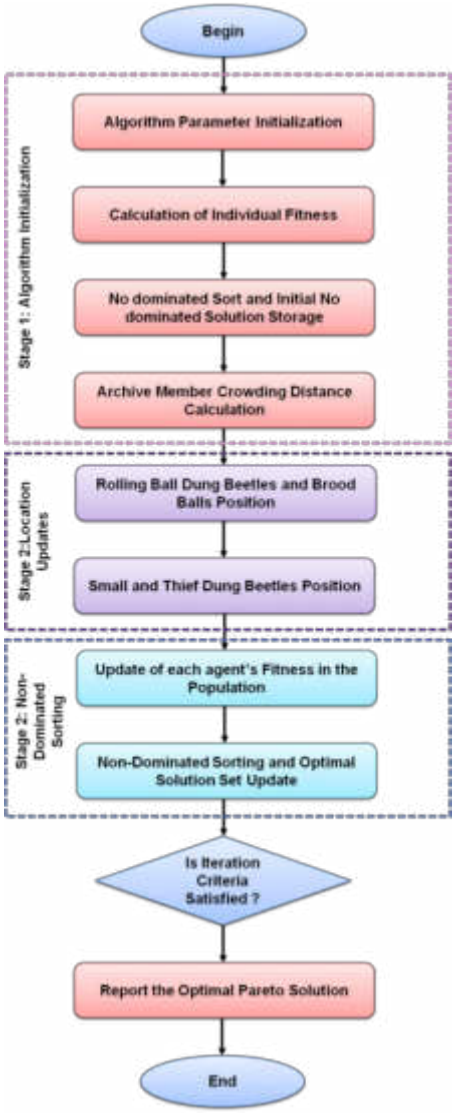


Fig. 3.3: Flow diagram of CDBS

where $s$is the number of iterations at this time$w_j(s)$is the location data of the j^{th}dung beetle across the t^{th} iteration, $a$is a fixed value that is a part of (0,1), $l \in (0,\ 0.2]$If the deflection coefficient represented by a constant value, $\alpha$is given a natural coefficient to 1 or -1, $W^x$is the worst position in the world, and $\triangle w$ mimics variations in light intensity.

When faced with obstacles that prevent from moving forward, a dung beetle will employ dance to discover a new route. A tangent function provides a rolling direction by imitating the dance behavior. Thus, this is the ball-rolling dung beetle was last seen:

$$W_j\left(s+1\right) = W_j\left(s\right) + tan(\theta)\left|W_j\left(s\right) - W_j\left(s-1\right)\right| \tag{4}$$

where $\theta \in [0,\ \pi]$ is the angle of deflection.

B. Brood ball. For dung beetles to give their young secure environment, selecting an appropriate spawning place is essential. To replicate the female dung beetle spawning region, a boundary selection approach

is suggested in CDBS and is described as follows:

$$lb^* = max(W^* \times (1 - Q), lb) \tag{5}$$

$$ub^* = min(W^* \times (1 - Q), lb \tag{6}$$

where $lb^*$and ub^* indicate the spawning area's top and bottom bounds, and W^* represents the best spot in the area at the moment:$R = \ 1 \ - s/S_{max}$ where $S_{max}$is the most iterations that can be made Lb and $Ub$are, respectively, the upper and lower bounds of the search space.

$$A_j(s+1) = W^* + a_1 \times (a_j(s) - lb^*) + a_2 \times (a_j(s) - ub^*) \tag{7}$$

where a_1 and $a_2$are two separate, size-dependent random vectors $1 \ \times C; C$is the size$a_j(s)$represents the j^{th} sphere's location at the s^{th} iteration

C. Tiny dung beetles. A special kind of adult dung beetle that burrows into the earth in quest of food is the tiny dung beetle. The following factors determine the boundaries of the optimal feeding region for little dung beetles:

$$lb^* = max(W^b \times (1 - Q), lb) \tag{8}$$

$$ub^* = min(W^b \times (1 - Q), ub \tag{9}$$

where Lb^a and $Ub^a$are the top of the perfect foraging area and lower bounds, correspondingly. andW^a is the ideal location on the planet. The little feces beetles have relocated to this area.

$$w_j(s+1) = W_j(s) + D_1 \times \left(W_j(s) - lb^b\right) + D_2 \times \left(W_j(s) - ub^b\right) \tag{10}$$

where $D_1$is a normal distribution applied to the random number, $D_2$is the arbitrary vector that falls between (0,1), and $w_j(s)$ is the location of at the s^{th} iteration the j^{th} dung beetle.

D. Thief. Some bugs pilfer other bugs' excrement balls; these are called thieves. The ideal food supply is W^a, as can be shown from Equation (5). $W^a$is the best place for competing food, therefore let's assume so. The following is an update to the thief's position information during the iteration process:

$$w_j(s+1) = W^a + T \times h \times ((W_j(s) - x^*| + |W_j(s) - x^b|) \tag{11}$$

There T is a fixed figure, h is a chance vector with a size of $1 \ \times C$ that is subject to a normal distribution, and $w_j(s)$ discloses the location of the j^{th} thief at the s^{th} repetition.

3.3.2. Random forest (RF). RF is an algorithm that uses random sampling searches inside a predetermined state space. Ensemble learning is the term used to describe the application of many models to a single classification issue. When it comes to RF, the forecasts are made by a voting procedure over the results and flow chart of RF as show in Figure 3.4

Each tree chooses at random a subset of the characteristics that are present in the data to produce different distributions in the models, as well as random samples taken from the dataset. The Gini Index (GI) produces greater precision in the experiments that are carried out, to assess the significance of the characteristics before generating a new node

$$GI = 1 - \sum_{j=1}^{D} o_j^2 \tag{12}$$

where o_j, the probability of class j, is determined by its frequency of presence on the split under consideration, and d is the number of classes. Using a Grid Search approach yields the RF architecture for training. We construct a set of potential parameters for each job (landing and mid-range flights), and we use all conceivable

Fig. 3.4: Structure of the RF

combinations to generate distinct forests. Every forest undergoes evaluation using a subset of the training data, known as the validation set and the optimal parameter configuration is kept for the whole training process.

By combining the clever navigational techniques of dung beetles with cutting-edge machine learning algorithms, CDBS-RF path planning effectively plans routes and implements anti-collision procedures as shown in Algorithm 1.

Algorithm 1: CDBS-RF
Initialize parameters:
Define the number of UAVs (N)
Define the environment map ($grid\_size$, $obstacles$, $waypoints$)
Set the maximum number of iterations (max_iter)
Initialize DBS parameters (*e.g.*, $population\_size$, $search\_radius$)
Initialize RF parameters (*e.g.*, $number\_of\_trees$, max_depth)
Generate initial population of paths:
For $i = 1$ *to* N:
Generate random path ($start\_point$ *to* end_point)
Evaluate path using RF model to calculate safety and efficiency
Perform DBS optimization:
For $iteration = 1$ *to* max_iter:
Evaluate fitness of each path in the population using the RF model
Update the best path based on fitness evaluation
Update paths using DBS algorithm:
Move each path based on Dung Beetle Search rules (e.g., attraction to better paths)
Apply random adjustments to paths for exploration
Refine paths using RF model:
For each path in the population:
Tune path parameters using RF model to optimize safety and efficiency

Table 4.1: Experimental setup

Category	Component	Details
Hardware	Operating System	64-bt Ubuntu18.04 bionic
	CPU	2.7 Ghz core-i5-5200
	RAM	8GB
	Desktop environment	LXDE
Software	Python version	Python 3.10
	Simulation software	Gazebo
	Description of software	Gazebo is comprehensive simulation software that enables the creation of realistic environments with dynamic features like gravity and wind, enabling the addition of visual components.

Choose the optimal course of action:
Select the route from the ultimate demographic that has the highest fitness score
Output the optimal path:
Display or save the best path for UAV navigation
End

With this novel method, dung beetle search patterns' flexibility and random forest models' predictive capacity are combined to create a system that can safely navigate over obstacles while dynamically adapting to changing circumstances. Through the imitation of natural navigation skills and the utilization of artificial intelligence's processing power, CDBS-RF provides a strong solution for efficient path planning in intricate situations, improving autonomy and security in self-governing systems.

4. Results and Discussion.

4.1. Configuration of Environment and Hardware. The Table 4.1 summarizes the hardware and software used for the simulations, including the specific versions and functionalities relevant to the experiments.

4.2. Simulation Resullts. The authors experimented with several CDBS-RF models, including the one recommended by others, to improve results for an environment where $[0, 0, 0] \leq Z \leq [10, 10, 10]$ and M=15. Model 3 has a steady reaction than others, as can be seen in Figure 4.1.Compared to the previous models, Model 3 appears to have a consistent reaction. Its exploration factor among the lowest, nevertheless. Though it differs significantly throughout the exploitation phase, Model 1 produces almost equal outcomes. The model is still exploring close states at that point.

These comparisons are explained, and a comparison based on the overall mean average is as shown in Figure 4.2. The model's performance is shown by the black line following eight training cycles, the blue line marking the violet line indicating modifications at the conclusion of the exploration phase in the stage of discovery during the utilization phase. Based on their Total Mean Reward and Exploration Phase End event, three models are compared in the table that is presented. At the twentieth exploration phase, Model 1 produced a Total Mean Reward of -10.26. Model 2 had lengthier exploration duration, as seen by its lower reward of -10.30 at the 800th phase. With a payout of -10.20 at the 60th phase, Model 3 landed in the middle.

It is crucial to remember that the CDBS-RF algorithm's path may not be collision-free because it knows information about static objects. The suggested approach accomplishes the goal in every experiment. The difference between the angular coefficient of CDBS-RF pathways and the header UAV angle throughout the validation testing is shown in Figure 4.3. These deviations, on average, range from -1^0 to 1^0, and they deviate when the agent must avoid moving impediments. The suggested technique was verified by the authors through a run-time comparison experiment between the CDBS-RF model and alternative path-planning algorithms.

4.3. Performance evolution. The suggested performance metrics for the CDBS-RF algorithm are shown in the Table 4.2. With a score of 90, the algorithm shows good path efficiency and successful route optimization. With a notable high of 95 for collision avoidance, it demonstrates its capacity to reduce impediments during

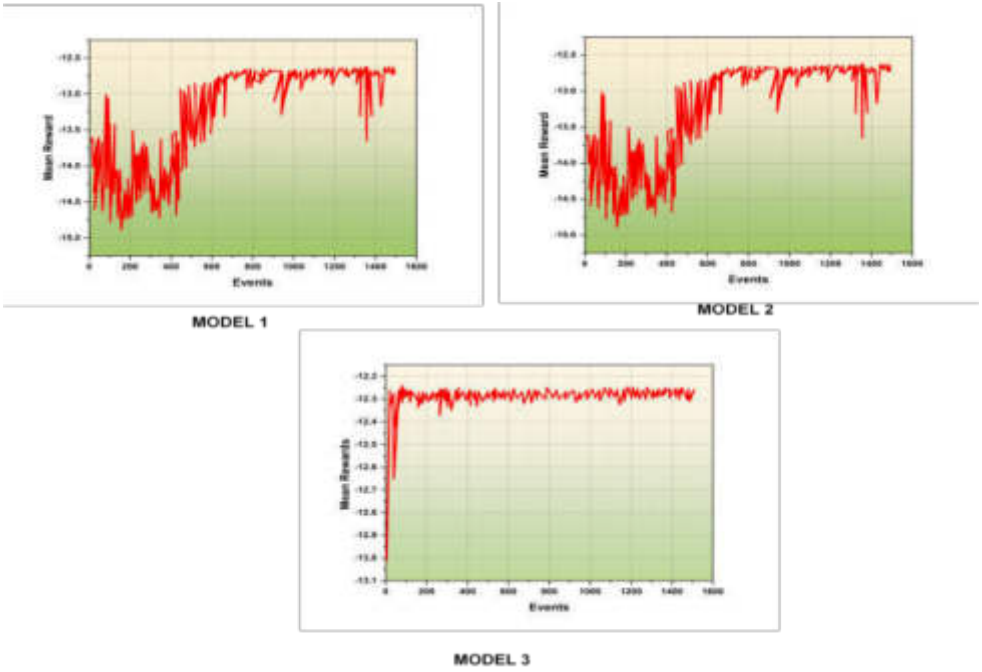


Fig. 4.1: Returns on investment for several models of CDBS-RF with 1500 events and 0.001 learning rate

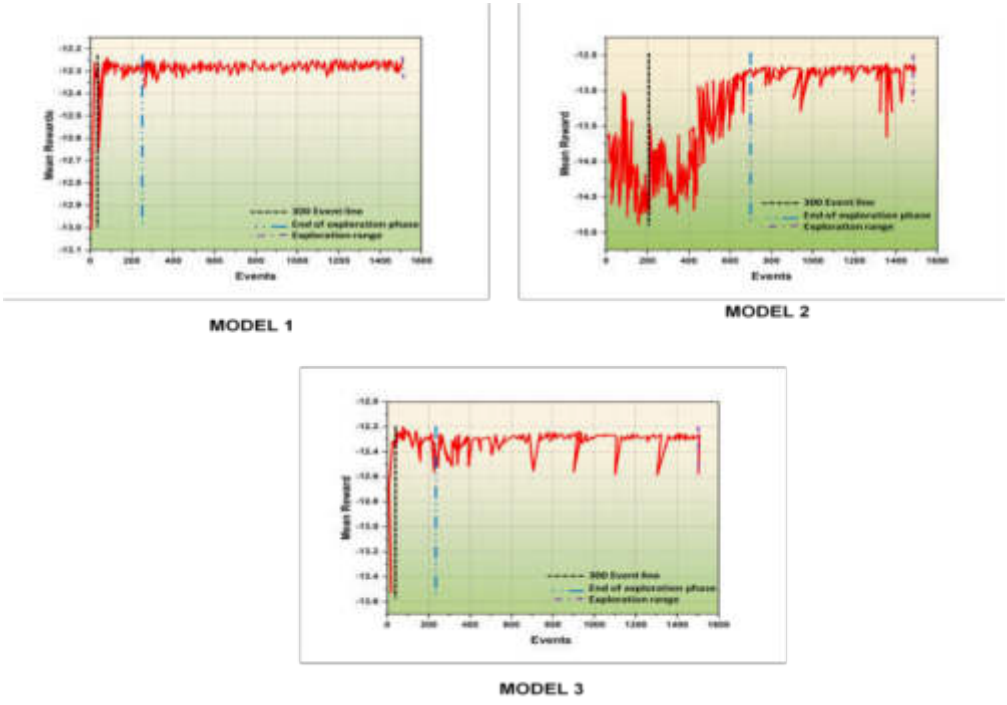


Fig. 4.2: Evaluating the top three models' performances in comparison

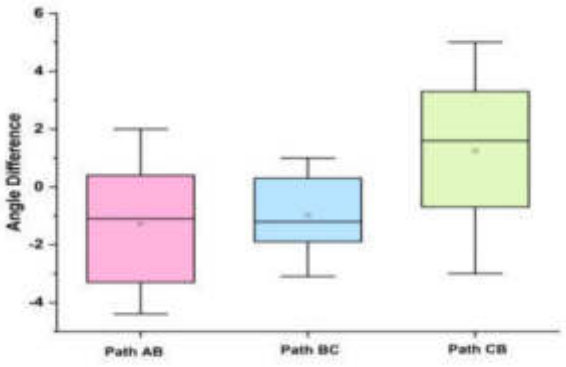


Fig. 4.3: Disturbance in angle between UAV header and CDBS-RF route

Table 4.2: Outcomes of performance

Evaluation parameter	**CDBS-RF [Proposed]**
Path Efficiency	**90%**
Collision Avoidance	**95%**
Computational Efficiency	**80%**
Accuracy	**92%**
Adaptability	**85%**

path planning. With an 80 score, computational efficiency indicates a well-balanced utilization of execution time and processor power. With a score of 92, accuracy emphasizes how precisely the algorithm produces correct results. The algorithm's strong performance in adapting to dynamic and changing surroundings is indicated by its adaptability score of 85. These numbers show the overall efficacy and dependability of the CDBS-RF algorithm across a range of assessment criteria.

4.4. Comparison phase. From the starting sites to the final location, the route that A* path planning algorithm creates often follows the edges of obstacles as it moves across the scene. It evaluates the performance of the CDBS-RF with existing methods like Optimized RRT [20], Generalized Wave front [20], and A* [20] Metrics like planning time and path length are used for assessment.

4.4.1. Planning time (s). The amount of time needed for any method to calculate a path for UAVs, representing its efficiency, is referred to as planning time. The planning timeframes for the different strategies are compared in Table 4.3 and Figure 4.4. The suggested CDBS-RF, A*, Generalized Wavefront, and Optimized RRT and the outcomes show that CDBS-RF is the most efficient in creating pathways rapidly, with the least planning time of 0.789 seconds. It shows the path created by the suggested method CDBS-RF is smoother and significantly farther from obstructions.

4.4.2. Path length (M). The distance of the route produced by each method for UAV navigation is represented by path length (M). The path lengths for the various techniques optimized RRT, Generalized Wavefront, A*, and the suggested CDBS-RF are displayed in Table 4.4 and Figure 4.5. According to the data, CDBS-RF produces the shortest path length (21.526 meters), suggesting a more direct, effective, and obstacle-avoidance method. It shows the path created by the suggested method CDBS-RF is smoother and significantly farther from obstructions.

Table 4.3: Results ofplanning time (S)

Methods	**Planning time (S)**
Optimized RRT [20]	0.857
Generalized Wavefront [20]	1.008
A* [20]	1.685
CDBS-RF [Proposed]	0.789

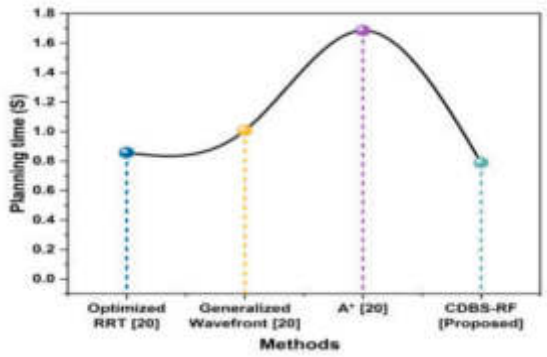


Fig. 4.4: Comparison of planning time (S)

Table 4.4: Path length (M)

Methods	**Path length (M)**
Optimized RRT [20]	27.85
Generalized Wavefront [20]	23.892
A* [20]	26.845
CDBS-RF [Proposed]	21.526

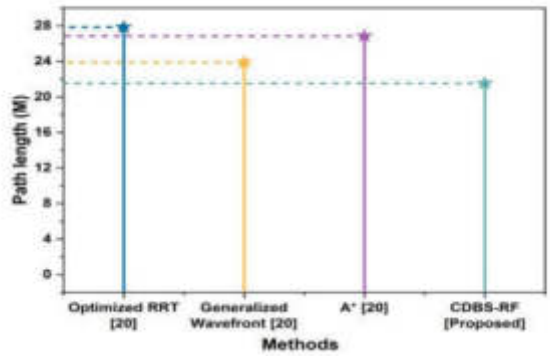


Fig. 4.5: Comparison of Path length (S)

4.5. Discussion. The limitations of the optimized RRT, generalized wave front, and A* algorithms are primarily related to their path efficiency and computational complexity. Optimized RRT, while effective for

high dimensional spaces, can produce suboptimal paths that can be longer and less smooth, generalized wave front offers smooth paths but can be computationally intensive and slow, especially in complex environments, A* is known for its optimally and completeness but can suffer from high computational costs and slower performance in large-scale or dynamic environments. Each algorithm faces challenges in balancing path length, smoothness, and computational efficiency, impacting their overall effectiveness in practical applications. The CDBS-RF method addresses these issues by integrating dung beetle search optimization with RF techniques, which enhances path efficiency and smoothness while reducing computational complexity. This approach ensures shorter, smoother paths with faster planning times compared to traditional algorithms.

5. Conclusion. The primary aim of the study was to optimize UAV path planning and collision avoidance in agricultural surveillance by introducing a novel technique, the CDBS-RF method. The objective was to enhance UAV route planning by integrating the DBS algorithm with the RF model, thus improving path safety and operational efficiency. The proposed method dynamically adjusted path planning parameters to ensure optimal route selection and effective collision avoidance, thereby addressing the complexities of real-world environmental variables that impact UAV operations. The CDBS-RF technique was implemented in a python based virtual environments and evaluated using UAV sensed data. Experimental results demonstrated that this approach significantly improves UAV path planning performance, leading to safer and more efficient navigation. The integration of advanced optimization and anticollision algorithms within the CDBS-RF framework offers a promising solution for enhancing UAV operations in agricultural surveillance, particularly in monitoring large, complex agricultural landscapes. The performance evaluation techniques include the planning time (0.789) and path length (21.526).

5.1. Limitations and future scope. *Limitations:* Despite the promising results, the study faced limitations related to the quality and resolution of UAV sensed data, which could constrain the effectiveness of the CDBS-RF technique in real-world applications. Additionally, the model performance can be influenced by environmental variability that was not fully captured in the virtual testing environment.

Future scope: Future research should focus on validating the CDBS-RF method in diverse real-worls agricultural settings with varying environmental conditions to assess its robustness and adaptability. Further improvements could involve integrating more advanced sensors and data fusion techniques to enhance data quality as well as exploring Ml models beyond RF to potentially increase path planning precision and collision avoidance capabilities.

REFERENCES

[1] Mohsan SAH, Othman NQH, Li Y, Alsharif MH, & M A Khan. Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends. Intelligent Service Robotics. 2023;16(1):109–137.

[2] Muntasha G, Karna N, & S Y Shin. Performance analysis on artificial bee colony algorithm for path planning and collision avoidance in swarm unmanned aerial vehicle. 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS). 2021;p. 1–6.

[3] Huang J, Luo Y, Quan Q, Wang B, Xue X, & Y Zhang. An autonomous task assignment and decision-making method for coverage path planning of multiple pesticide spraying UAVs,Computers and Electronics in Agriculture. 2023;212:108128–108128. Available from: https://doi.org/10.1016/j.compag.2023.108128.

[4] Diao Q, Zhang J, Liu M, & J Yang. A Disaster Relief UAV Path Planning Based on APF-IRRT* Fusion Algorithm,Drones. 2023;7. Available from: https://doi.org/10.3390/drones7050323.

[5] Hu W, Quan J, Ma X. Unmanned aerial vehicles for plant protection and precision agriculture: a study on low-altitude route planning method of unmanned aerial vehicles. Pak J Agri Sci. 2023;60(2):435–455.

[6] Ünal I, Kabaş Ö, Eceoğlu O, & G Moiceanu. Adaptive Multi-Robot Communication System and Collision Avoidance Algorithm for Precision Agriculture. Applied Sciences. 2023;13(15).

[7] Messina G, & G Modica. Twenty years of remote sensing applications targeting landscape analysis and environmental issues in olive growing: A review. Remote Sensing. 2022;14(21).

[8] Aslan MF, Durdu A, Sabanci K, Ropelewska E, Gültekin SS. A comprehensive survey of the recent studies with UAV for precision agriculture in open fields and greenhouses. Applied Sciences. 2022;12(3):1047–1047.

[9] Tahir MN, Lan Y, Zhang Y, Wenjiang H, Wang Y, Naqvi SMZA. Application of unmanned aerial vehicles in precision agriculture. Precision Agriculture. 2023;p. 55–70.

[10] Meng K, Wu Q, Xu J, Chen W, Feng Z, Schober R, et al. UAV-enabled integrated sensing and communication: Opportunities and challenges. 2023;Available from: https://doi.org/10.1109/MWC.131.2200442.

[11] Chattaoui S, Jarray R, & S Bouallègue. Comparison of A* and D* Algorithms for 3D Path Planning of Unmanned Aerial Vehicles. 2023 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE). 2023;p. 1–6.
[12] Kayalvizhi M, & S Ramamoorthy. Blockchain-based secure data transmission for UAV swarm using modified particle swarm optimization path planning algorithm. International Journal of Advanced Computer Science and Applications. 2021;12(11).
[13] Wu W, Zhang X, & Y Miao. Starling-Behavior-Inspired Flocking Control of Fixed-Wing Unmanned Aerial Vehicle Swarm in Complex Environments with Dynamic Obstacles. Biomimetics. 2022;7(4).
[14] Kamel B, & A Oussama. Cooperative Navigation and Autonomous Formation Flight for a Swarm of Unmanned Aerial Vehicle. 2021 5th International Conference on Vision, Image and Signal Processing (ICVISP). 2021;p. 212–217.
[15] Wei Z, Meng Z, Lai M, Wu H, Han J, & Z Feng. collision technologies for unmanned aerial vehicles: Recent advances and future trends. IEEE Internet of Things Journal. 2021;9(10):7619–7638.
[16] Mukhamediev RI, Yakunin K, Aubakirov M, Assanov I, Kuchin Y, Symagulov A, et al. Coverage, path planning optimization of heterogeneous UAVs group for precision agriculture. IEEE Access. 2023;11:5789–5803.
[17] Zhou Q, Feng H, & Y Liu. Multigene and Improved Anti-Collision RRT* Algorithms for Unmanned Aerial Vehicle Task Allocation and Route Planning in an Urban Air Mobility Scenario. Biomimetics. 2024;9(3).
[18] Huang J, Luo Y, Quan Q, Wang B, Xue X, Zhang Y. An autonomous task assignment and decision-making method for coverage path planning of multiple pesticide spraying UAVs Computers and Electronics in Agriculture. 2023;212:108128–108128. Available from: https://doi.org/10.1016/j.compag.2023.108128.
[19] Ahmed S, Lee CS, On R, Planning, Logistics, Unmanned, et al. International Journal of Robotics and Automation. 2024;39(6).
[20] Wu S, Du Y, & Y Zhang. Mobile robot path planning based on a generalized wavefront algorithm. Mathematical Problems in Engineering. 2020;2020:1–12.
[21] Kapoor R, Mittal S, Kumari AC, Srinivas K. Advancing sustainable agricultural development and food security through machine learning: a comparative analysis of crop yield prediction models in indian agriculture. 2025;3:101–108. Available from: https://doi.org/10.61552/JEMIT.2025.02.005.

Edited by: Vadivel Ayyasamy
Special issue on: Internet of Things and Autonomous Unmanned Aerial Vehicle Technologies for Smart Agriculture Research and Practice
Received: May 18, 2024
Accepted: Dec 19, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 517–530, DOI 10.12694/scpe.v26i2.3942

ENHANCING CLOUD DATA SECURITY THROUGH AN ENCRYPTED AND EFFICIENT CONNECTION MODEL BASED ON BLOCKCHAIN TECHNOLOGY

SONALI SHARMA*, SHILPI SHARMA† AND TANUPRIYA CHOUDHURY ‡

Abstract. Centralized file storage systems are widely used for sharing data and services, offering oversight capabilities to governing bodies. Despite their prevalence, these systems raise significant concerns about potential data misuse by authorities and control over internet information dissemination. The vulnerability of centralized systems has been highlighted by security breaches that exposed user data. To counteract these issues, researchers have proposed blockchain-based solutions for enhanced data protection. This research work highlights an innovative approach utilizing the services of Interplanetary File System (IPFS), a secure and encrypted connection framework accessible via web browsers that supports global data storage and distribution on a decentralized peer-to-peer network. This model ensures privacy, accuracy of information, and eliminates additional service fees or software requirements. However, even decentralized networks like IPFS have faced cyber threats compromising user data. In response to this challenge, the presented research outlines a model incorporating AES-256 encryption to protect files before uploading them onto IPFS nodes. Secure transmission is achieved through WebRTC technology which facilitates the exchange of encrypted file hashes and keys with recipients. The key takeaways from this study include an optimized method for file storage and distribution designed to thwart cyber-attacks targeting stored or shared files while also providing economic benefits by removing extra storage costs associated with large files in centralized systems. Overall, the research offers a robust solution aimed at safeguarding user privacy without sacrificing functionality or incurring high costs.

Key words: IPFS, AES-256, Web RTC, Blockchain Technology, Decentralized, Distributed

1. Introduction. The current version internet comprises of the widespread usage of services provided by the Web2 environment. The online transactions including- sharing of data, communication of information, confidential file storage and end to end transfer of content is governed by the centralized systems. The significant drawbacks of accelerating transactions of data and information through the Web2 environment include- services in exchange of personal data of end users, centralized data monitoring by the governing bodies where the end users often mistake themselves to be the owner of shared information and limited control of the metadata produced on the Web2 environment [1]. Web3 is the new generation of internet which attracts attention due to its decentralization abilities. This also means that the privileges of artificial intelligence and cyber security can be implemented to make the internet more secure and empower the individuals to have complete control over the content they create or share online [2]. Web3 is often termed as semantic or spatial web that leverages decentralization capabilities. A distinct concept from Web 3 which focuses on capabilities of 3D virtual realm and augmented reality is termed as Metaverse. Metaverse is an environment wherein the three dimensional objects interact and provide better interface for human engagement. Therefore, Web3 and Metaverse can be used together for providing a more interactive environment for the users but they are not interchangeable [2]. Decentralized applications (DApp) that operate on the concept blockchain technology and other distributed platforms play a crucial role in making the Web3 environment permission less, secure and private. Under the proposed approach, the files are uploaded on the web using the potential of the Web3 environment and are distributed using interplanetary file system (IPFS) to ensure security and ownership of the content by the creators. However, the modern day malicious software including the zero day attacks can infiltrate any device. To safeguard our system from the viruses, we can scan the files before uploading/downloading them by any efficient anti-virus software available. However, the malicious files cannot damage the file storage system

*Amity University, Uttar Pradesh, India (sonali.260893@gmail.com).
†Department of computer Science and Engineering, Amity University, Uttar Pradesh, India (ssharma22@amity.edu).
‡School of Computer Sciences, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand, India (tanupriya@ddn.upes.ac.in).

implemented using IPFS but can potentially harm any user system who downloads such files from IPFS. Our research aims to propose a model that can facilitate the transition between the Web2 and Web3 seamlessly.

1.1. Contribution. Cloud based platforms are crucial for files and data storage. However, they raise security concerns which impact the integrity and confidentiality of the data shared on the centralised platforms. They are expensive and susceptible to man in the middle attacks. The proposed and implemented model presented in this paper ensures data security of the stored and shared data over the network. To achieve the server less architecture we have designed a user interactive framework using React and Next.js. To establish a secure connection between the sender and receiver WebRTC has been used and decentralisation is achieved using IPFS. The framework incorporates the data sharing mechanism using AES-256 file encryption method. The research paper presents two novel algorithms, which have been implemented to ensure data confidentiality and integrity. The proposed framework has been compared with the existing frameworks to demonstrate its efficiency and cost-effectiveness.

The research paper is arranged as Literature Review in section 2, Motivation in section 3, Methodology in section 4, Proposed Framework in section 5, Implementation of Framework in section 6, Results and conclusion in section 7 and Future Work in section 8.

2. Literature Review. The centralised systems work on the capabilities of the client-server approach powered by HTTP (Hyper Text Transfer Protocol) due to its well implemented architecture to meet the industry standards for communication of data. Here are some drawbacks of accessing data and information using the potential of the HTTP architecture.

Limited Bandwidth: Channelising information through single server to multiple clients.

Paid services: Global access to the shared data requires a server to be setup or the services provided by the central storage application is paid.

Ephemeral: Failure of the centralized server, will lead to data loss.

Data Duplication: Files having similar information, name and metadata can be found at multiple servers.

IPFS on the other hand provides better access to the data by decentralizing it, so that even if one node breaks down, the data must be present with many other nodes. IPFS has high bandwidth as the data is not downloaded from single server, it is accessed by many nodes having that information [3]. Similar to the centralized data storage mechanisms IPFS also maintains a cache or a list of frequently accessed files, if the user wants a file to be available for quick access then cryptocurrencies can be used to incentivize the information or there is a cost free mechanism of file pinning which allows quick access to the frequently requested files. Using the mentioned method the file will be consistent and available in a network whenever required. To prevent data redundancy IPFS uses the concept of content based addressing where the files with same name and data cannot be stored even on different locations [4]. It is important to note that both HTTP and IPFS protocols provide data sharing and accessibility mechanisms and but have very distinct structure:

Content addressing: HTTP locates the stored information by the help of URL (Uniform Resource Locator), whereas IPFS uses content-based addressing for locating the stored resources through the cryptographic hash of the content regardless of its location.

Decentralization: HTTP relies on centralized servers to host and distribute content, while IPFS is a decentralized network where content is distributed across many different nodes [5]. This means that IPFS can be more resilient to network failures and censorship.

Performance: HTTP typically relies on a client-server architecture, where a client requests data from a server and the server responds with the requested data [5].

Persistence: To gather information through HTTP architecture, it relies on the centralized server up time and as long as it is available for data retrieval. Whereas, IPFS structure depends on the availability of nodes hosting the data on network.

While HTTP and IPFS both serve the purpose of sharing and accessing content over the internet, IPFS offers some unique advantages over HTTP, particularly in terms of decentralization, content addressing, and persistence [6]. However, HTTP is still widely used and has a well-established infrastructure that many websites and applications rely on. Fig. 2.1 shows average analysis on the research work published worldwide from the year 2018 to 2024 on IPFS protocol. The analysis has been done using VOSviewer with input data from the

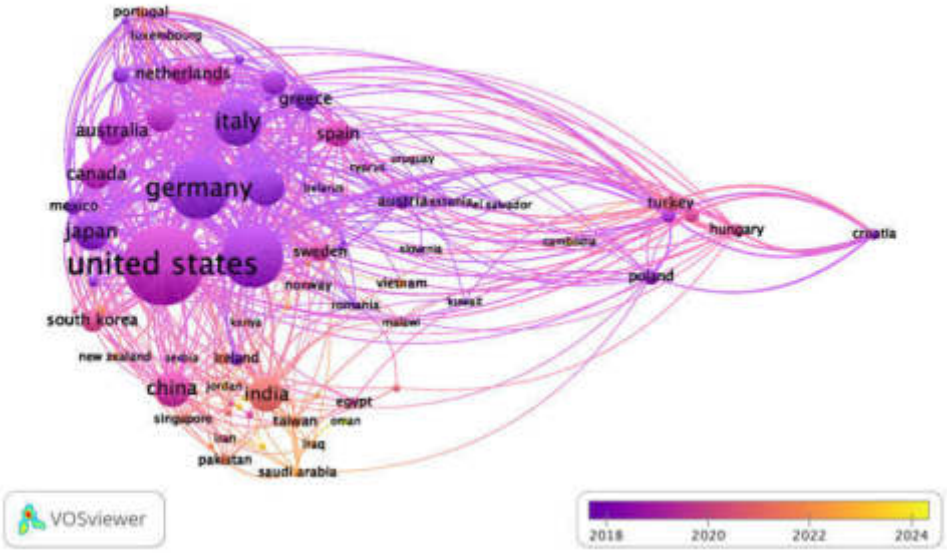


Fig. 2.1: Analysis of average documents published from 2018 to 2024 on the study of IPFS

OpenAlex API. The results showcase links for 3430 documents published worldwide. The maximum research work has been initiated in the United States followed by United Kingdom and Germany.

Service availability is one of the major concerns when we implement distributed/decentralised framework and mechanisms. Traditional methods of calculating the uptime of service is measured by percentage of available time, explained by the below mentioned equation [7][8].

$$Availability = \frac{MTTF}{MTTF + MTTR} \tag{2.1}$$

MTTF: Mean Time Failure of service
MTTR: Mean Time to Repair the service

However, the above equation does not explain the reason of service failure. In case of IPFS or decentralised frameworks, there can be many reasons for service failure like downtime of requested node, network congestion, and unavailability of requested of data. To better describe the data or service availability, Giwon On, et al. [7, 8] have mentioned an optimised definition of service and data availability of IPFS:

$$Avail_{Service} = Avail_{Data} X Avail_{System} \tag{2.2}$$

$$Avail_{System} = Avail_{Node} X Avail_{Link} \tag{2.3}$$

$$Avail_{Node} = Avail_{Dynamic} X Avail_{Intrinsics} \tag{2.4}$$

We elaborate the availability matrix on the basis of our proposed framework using IPFS.

$Avail_{Service}$**:** The available service of our proposed framework is directly proportional to data availability and system availability. Therefore, service completion depends on available data and system uptime.

$Avail_{Data}$**:** This refers to availability of updated data for user access and use. IPNS (Interplanetary Name System) and dynamic links explained in further sections are implemented in IPFS to ensure updated data availability.

Table 3.1: BitTorrent vs IPFS

BitTorrent	**IPFS**
BitTorrent does not have the capabilities of using DHT (Distributed Hash Table), therefore a torrent and meta-data file is required to start downloading the file.	Hash of the file is required to start downloading the file from IPFS. This is due to Kademlia- a lookup mechanism by IPFS which supports DHT.
There is no mechanism of leveraging the CID (Content Identifier) and DHT of the file content. Therefore, the content of the files cannot be compared while it is being uploaded from different sources. This leads to data redundancy.	Files having same CID will be detected as soon as they are uploaded on IPFS. Files having same metadata, content, file name, byte level compression and encryption will not be uploaded on IPFS due to their similar CID.

$Avail_{System}$**:** The system availability depends on availability of dynamic IPFS nodes, their configuration, and data-link availability.

$Avail_{Link}$**:** Available link means that the node link between the user and server is within the reachable limit with considerable delay to provide service using high bandwidth resource.

$Avail_{Node}$**:** This indicates the service node strength, bandwidth and performance to support the service completion and data delivery.

$Avail_{Dynamic}$**:** Stability in node uptime is dynamic node availability, to help and process the request within a tolerable time limit.

$Avail_{Intrinsics}$**:** Node Intrinsic availability refers to the bandwidth, storage capacity, system configuration and processing power of the IPFS nodes in the distributed framework.

WebRTC is based on open standards consisting of several APIs (Application Programming Interfaces) to enable standalone structure for video and audio calls, chat, and file sharing mechanism so that the developers can use it to build real-time communication applications. For fast, secure and efficient communication it uses peer-to-peer connections between the devices allowing the usage of advanced features such as encryption, which ensure private connection [9]. It therefore consulates as the building block of decentralized applications based on real-time connections. Due to the persistence and resilience of WebRTC structure there is no single point of failure, which enables its adaptability to the changing modern standards [10]. To enable low latency communication: WebSockets or signalling servers powered by WebRTC ensure device discovery and connection. Upon successful connection establishment, the information can be easily exchanged without the centralised servers. Overall, WebRTC is a powerful technology to ensure real-time communication for decentralized environment.

Another aspect of conflict is to understand the difference between IPFS and blockchain. They both serve the purpose of decentralisation with different use cases. IPFS however, showcases capabilities to overcome the limitations of blockchain [11]. One of the major limitations of blockchain include scalability, where the size of network may affect the synchronization between the blockchain nodes resulting in slow processing time and higher price. To overcome this issue IPFS works on the mechanism of storing the large files off-chain. This reduces the size of network making it more manageable, easy to verify and scalable [12].

IPFS enables the developers to create applications where blockchain based solutions to store large data can be provided without any additional cost as storing large amounts of data on-chain can be expensive. In summary, the limitations of cost, decentralization, scalability can be overcome by using the capabilities of IPFS [13].

3. Motivation. In the era of innovation, the decentralization mechanisms have led to invention of many file sharing protocols. One such popular file sharing mechanism is BitTorrent which enables the users to share the files through peer-to-peer mechanism over the internet. A group of swans (hosts) is used to download the information through BitTorrent instead of centralized servers [10]. Table3.1 shows differences between the functionalities of BitTorrent and IPFS. The file accessibility speed is similar to IPFS, where the most requested/visited file is available for faster access or download by the users.

BitTorrent and IPFS are based on data security while sharing between the users, although they do not use content based encryption. Additionally, the Content Identifier(CID) can be accessed in public domain due

Fig. 3.1: Properties of IPFS

to the usage of public Distributed Hash Tables by IPFS [14]. Similarly, the IP addresses are public in the Bit Torrent architecture. This affects the confidentiality of the data as the hackers can constantly monitor the sources and destination of file sharing. To overcome this Filecoin, SiaCoin, SafeCoin, Internxt and Storj along with many other DApp (Decentralized applications) are recently being used as peer to peer file sharing mechanisms which aim at addressing the issues of BitTorrent and IPFS.

IPFS being an extension to the traditional and popular HTTP protocol is efficient with high bandwidth and no single point of failure. Fig. 3.1shows the properties and advantages of using IPFS in our implemented framework for ensuring security and integrity of data. Our proposed framework does not use the capabilities of an Ethereum blockchain to implement the framework using the peer to peer network since data storage is expensive.

4. Methodology. Interplanetary File system, widely known as IPFS is a content based decentralized storage protocol which can be used to store, access and retrieve files and data of any form [15]. IPFS stores the contents in two main parts:

(i) An unstructured binary data block of size 256 KB. When the file size is more than 256 KB then a list containing links of file chunks small in size are maintained.

(ii) An array having links to IPFS objects associated with the same directory.

Fig. 4.1 shows the fundamental principles of understanding the working of IPFS. Let's closely look at these features.

4.1. CID (Content Identifier). CID is a hash unique to the content of the file. Generally used in the URL to access the file by its content. This also means that if the content of the file changes then the CID will also change. In some instances changing the URL every time the content of the file is updated might not be feasible, for this IPFS has a concept of mutable pointers also known as Interplanetary Name System (IPNS). In this case the address of the pointer which is pointing to a CID is shared and every time there is any update in the content of the file the CID is updated but the address of the pointer remains the same [14]. IPFS uses DHT (Distributed Hast table) for providing efficiency and to make the file retrieval process robust. DHT is responsible for the information of all the CIDs. Therefore when a CID is requested by the user, the IPFS nodes send queries to the DHT for finding the nearest nodes which can help you retrieve the requested CID [15]. The CIDs which are requested frequently are available in the DHT for a limited time and they can also be pinned if you never want them to be garbage collected. This information about which CIDs are being requested frequently and by which nodes is publicly available.

4.2. DAG (Directed Acyclic Graph). IPFS uses the advantages of DAG data structure to split the files into blocks. All files are broken down into blocks which can be later retrieved from different sources, each

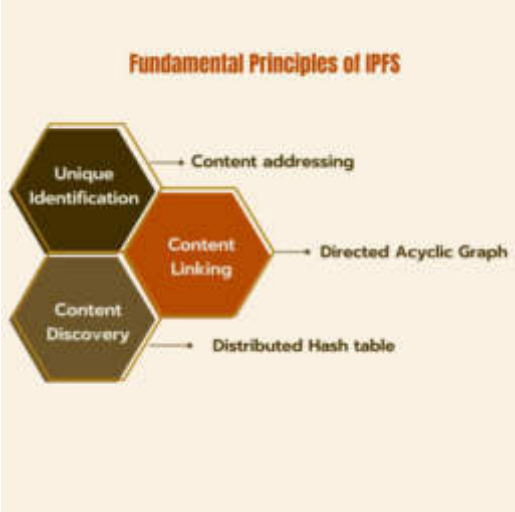


Fig. 4.1: Fundamental Principles of IPFS

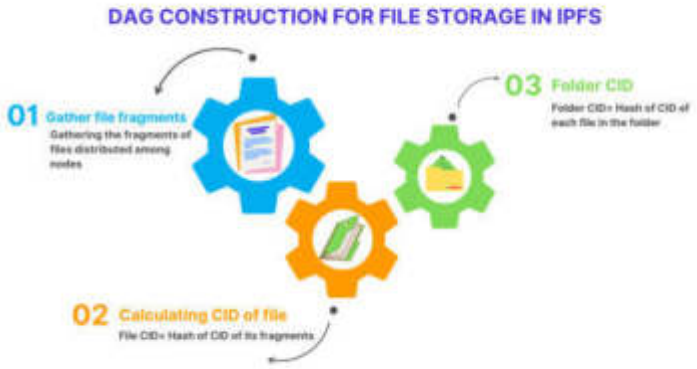


Fig. 4.2: DAG construction for file storage in IPFS

block receives a unique CID based on its content as discussed in the previous section. The CID of the file is a hash of the CID's of the blocks. Therefore, a folder containing many files has a CID which will be made by the hashes of the CID's of the files stored inside it [16]. Fig. 4.2 explains the working of the DAG.

The advantages of using DAG is that even if there are similar files in a system then different DAG's can reference to same subset of data [17]. This can be helpful if some part of the bigger folder is being updated then all other parts can continue to refer the same destinations and only the updated file can change its reference as per the requirement.

4.3. DHT (Distributed Hash Table). A hash table is the one which maps the keys to values. A distributed hash table however is split across the network and is used to connect different peers [18]. Once the content to be requested is identified using the previous two steps, location of the content and where to find the content is taken care by the DHT. Therefore DHT is responsible for:

- Identification of the peers who have the requested content
- Location of the Content

IPFS uses a data exchange protocol called Bitswap which is responsible for fetching the content from a peer and sending it to another [19]. The designing of this protocol aims at balancing the latency and throughput

Fig. 4.3: Working of BITSWAP

of the entire system. Bitswap is a simple protocol which can receive and transfer blocks up to size 2MB for achieving maximum compatibility.

Fig 4.3 shows the working of BITSWAP protocol which helps in fetching the data without any latency real time.

If the blocks are to be retrieved from the peers then a WantList can be created which has the details of the request blocks [20]. Hash of the received contents can be verified upon retrieval of blocks from the peers using Bitswap. Verification can be done by hashing the contents to fetch CID of the received blocks and match them with the CID of the request blocks. If both match then they can be sent to the other peers who have requested the blocks.

5. Proposed Framework. IPFS is a peer to peer protocol and a globally distributed server which is used for data storage and delivery. CIDs can easily be used to reference the distributed server and the nodes that are communicating with each other. Let's closely look at the attributes of IPFS which are publicly declared to be referenced by nodes.

- Although the communication between the nodes is encrypted but the metadata produced by nodes which is shared with the DHT is public.
- DHT contains a variety of information produced by the IPFS nodes which is public- CID of data and PeerIDs (Unique node identifier).
- Retrieval and request of CIDs by the nodes.

There are two types of encryption schemes that can be used for data protection: Transport encryption and Content encryption. IPFS uses transport encryption so that anyone cannot view the files/data while they are being transported between nodes but the content is not protected as anyone who as the CID can download and view the contents of the file. The additional security measures therefore are required to be taken for data protection. The framework proposed in this paper aims at achieving privacy, confidentiality and security of data while it is shared using IPFS.

1. *Browser Application*: Build using TypeScript because it has the capabilities of specifying the type of data being passed into the code but in JavaScript this facility is not there as the variables and parameters don't give any information. The compile time type checking is possible by using the TypeScript. For user interaction Next.js and React is used so that the browser application can be easily accessed using web and no extra server storage is required to run this.
2. *Encryption of file uploaded by the user*: AES-256 is used for the encryption of files on the user system before sending the file to IPFS. AES-256 encryption is not easy to decrypt because of the large key size. It is even quantum safe because it requires 6600 logical, error-corrected qubits to break the encryption of AES-256 but the latest quantum computer Osprey has 433 qubits achieved by IBM. The computational powers of the presently used computers also cannot decrypt the files encrypted by AES-

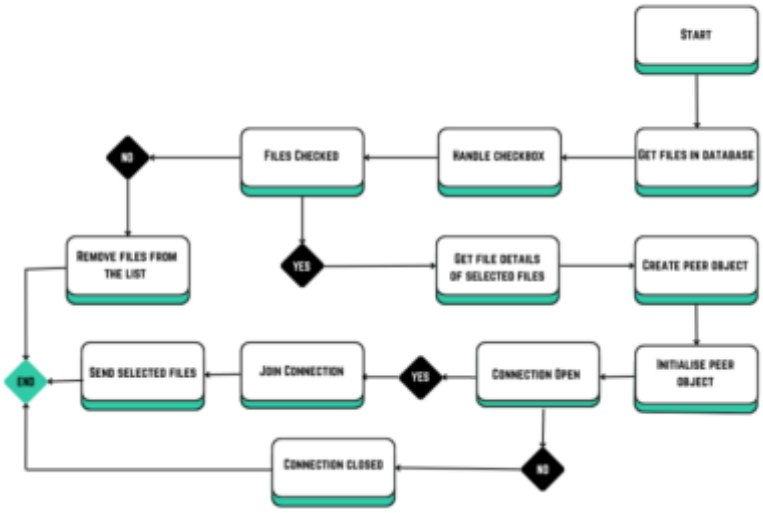


Fig. 5.1: Flowchart for file encryption and upload using the proposed framework

256 easily without the encryption key [21]. In this model the keys are being stored on the browser and a single key is used for the encryption of files [17]. Fig.5.1 shows the process of uploading the encrypted file using IPFS.

3. *Uploading encrypted files to IPFS*: For implementation and usage of IPFS environment we are using the HTTP-client library which can work with both the Go and JS implementations of IPFS without installing any extra software on the system [22]. As the files being shared with the IPFS nodes are encrypted and are being shared directly there are very remote chances of many attacks such as the man in the middle attack (MIMT)and brute force attacks. The API used for implementation of IPFS is responsible for splitting the files into blocks and create hashes of the contents. The CID is not shared with the nodes and the metadata is directly stored at the database and not in the server.
4. *File sharing*: Files uploaded to IPFS can be shared with the peers by creating and initializing a peer object. Create peer object with connection to shared peer server and initialize the peer object. Initialize an IPFS client to interact with the IPFS network. This is necessary for accessing and retrieving the uploaded files. After initialization share the data with peer object [23]. To write clear and concise instructions for sending files uploaded to IPFS to peers by creating and initializing a peer object, you can follow these steps:
   (i) Initialize a peer object to communicate with peers over a shared peer server. Make sure to provide the necessary configuration options, such as the server URL, port, and any other required settings.
   (ii) Set up callbacks to handle various events related to the peer connections, such as session timeout or data transfer events. This ensures you can react appropriately to these events.
   (iii) To share files or data with peers, you can use the connection object created in the event handler. Send the data to the connected peer using the 'send' method.
   (iv) If session timeout or other connection-related events are important for your application, you can implement callbacks for these events to handle them appropriately.
   (v) If needed, you can also close connections gracefully when you're done with them.
5. *File Retrieval:* The user request for file retrieval is processed by IPFS and information about the nodes storing the CID of the requested file is shared by referring to the DHT. After the nodes having the blocks of the requested file are identified, the encrypted file is send back along with its hash to the browser application at the user end.

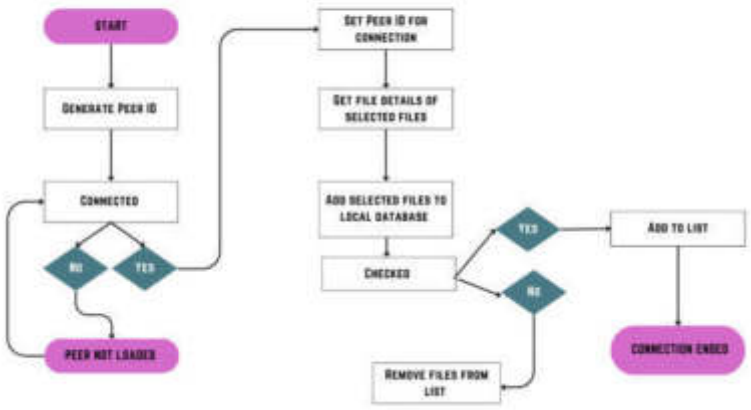


Fig. 5.2: Flowchart for file decryption and download using the proposed framework

6. WebRTC employs encryption protocols like Datagram Transport Layer Security (DTLS) and Secure Real-Time Protocol (SRTP) to ensure secure transmission of data. The user can decrypt the file using the key stored on the browser and retrieve the original readable file. Fig. 5.2 shows the retrieval process of the encrypted file.

6. Implementation of framework. The proposed framework aims to build a web based application which does not require any external server for file storage. The framework consists of the following components:
IPFS Node: A local instance of the IPFS daemon that handles file storage, retrieval, and distribution.
Encryption/Decryption Module: Encrypts files before uploading them to IPFS and decrypts them upon retrieval using a symmetric encryption algorithm.
Key Management Module: Generates, stores, and manages encryption keys.
WebRTC Signaling Server: Facilitates the exchange of signaling data between peers, enabling the establishment of a WebRTC connection.
WebRTC Data Channel: A secure communication channel for exchanging encryption keys between peers.

A user uploads a file to the system. The Encryption/Decryption module encrypts the file using a randomly generated encryption key. The encrypted file is added to the local IPFS node and distributed across the network [24]. The user shares the unique IPFS hash of the encrypted file and establishes a WebRTC connection with the recipient using the signaling server. Algorithm 1 explains the procedure to encrypt the file using a secure randomly generated password before sharing the file with the peers. The encryption key is securely transmitted to the recipient via the WebRTC Data Channel. The recipient downloads the encrypted file from IPFS using the hash and decrypts it using the received encryption key. Our proposed system ensures the following security properties:

1. *Confidentiality:* File content is encrypted using a symmetric encryption algorithm, making it unreadable to unauthorized parties.
2. *Integrity:* The content-addressed nature of IPFS ensures that the file content remains unaltered, due to a different hash value for any changes made to the uploaded file.
3. *Availability:* IPFS's distributed architecture ensures file availability even in the presence of network failures or censorship attempts.
4. *Secure Key Exchange:* WebRTC's encryption protocols (DTLS and SRTP) provide secure transmission of encryption keys, preventing eavesdropping and man-in-the-middle attacks.
5. *Cost effective:* There is no cost required for sharing the files using the proposed framework.

Table 6.1: Libraries, frameworks and technologies

Technology	**Version**
IPFS	Ipfs-desktop-0.24.1-mac.exe
React	18.2.0
Next.js	Next.js 13
IPFS-http-client	https://github.com/ipfs/js-ipfs.git
Web Crypto API	https://github.com/mdn/content/blob/main/files/en-us/web/api/web_crypto_api/index.md?plain=1

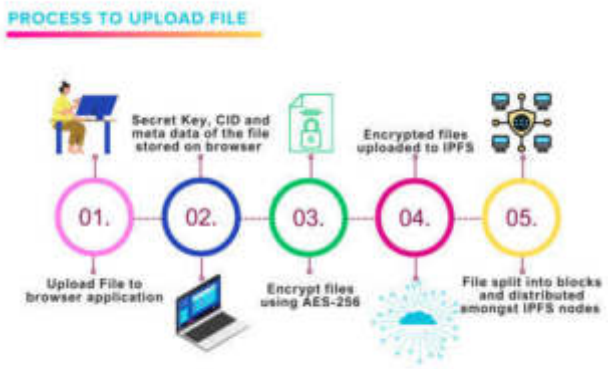


Fig. 6.1: Process diagram to upload the encrypted file using the proposed framework

6. *Low space utilization:* CID and hash are the only values stored for each file on each node, very limited space utilization is required by our system as compared to other blockchain based solutions proposed in the past.
7. *Not searchable on web:* The file is not searchable on the internet due to encrypted path information.
8. *Eavesdropping and man in the middle attack is not possible:* Unique peer ID is used for establishing the secure connection between the sender and receiver.
9. *Server less file upload/download and transportation:* No server space is required for running the proposed model.
10. *Two way security:* The established connection is verified through shared Peer ID communicated and sent using WebRTC data channel.

The technology stack used to build the web application is discussed in Table 6.1.

In the proposed web application, the files are encrypted by using AES-256 and then the encrypted file stored on the system by .enc extension can be uploaded on IPFS nodes. Fig. 6.1 and Fig.6.2 show the upload and download procedure of encrypted files using the proposed mechanism. There is no cost of managing the files as they are being uploaded on the web in a secured manner [25]. HTTPS is used for this purpose in the current searches and file storage but due to centralized client server mechanism it is considered to be prone to failure in the future. After installing the latest IPFS version on your system, for running the web application on the browser of a PC or mobile run the following commands on the terminal to start IPFS client. Algorithm 2 explains the procedure to share the encrypted files securely. The two-way security mechanism to establish the connection between sender and receiver ensures no cyber-attacks are possible while file transmission.

While establishing secure connection between the sender and receiver, a unique peer ID will be generated the similar way we generate the one time stamp to open connection for the transaction to be successful. To ensure the ease of use for the sender and receiver we have generated random IDs as a combination of adjectives + nouns + alphanumeric string. This will help in connecting and login to the connection without any delay. Additionally, this peer ID will be valid for sharing only one set of file. For another file sharing option, we have

Algorithm 1 Encrypt and upload file to IPFS

Input: IPFS API Address, File, User interaction through UI elements.
Output: Encrypted File, File CID, Stored file details in the browser.

Initialise Variables: Password, selectedFile, apiAddress, client, clientID, ipfsFile, encryptedFile, filesInDB.
Generate a password using alphanumeric characters
Create an IPFS client instance using the provided API address
while IPFS client is active **do**
  Set the client instance and client details (ID)
  for all selectedFile from input field do **do**
    Set the 'password' variable based on user input
    for encryptedFile do **do**
      Read the selected file as a Uint8Array
      Create salt, key, and IV for encryption
      Encrypt the file using AES-CBC and derived key
      $encrypteddata \leftarrow salt$
      $encryptedFile \leftarrow EncryptedData$
      Display an alert confirming successful encryption
    end for
  end for
  for all encrptedFile do **do**
    Upload the encrypted file to IPFS
    $ipfsFile \leftarrow FileCID$
    Display an alert confirming successful file upload
  end for
  if $UploadFile = PinFile$ **then**
    Pin the file to IPFS
    Display an alert confirming successful file pinning
  end if
  if ipfsFile, selectedFile, and password are available **then**
    Create a PouchDB instance
    Create a document with file details and store it in the local database
    Display an alert confirming successful file details storage
  end if
  Call getFilesInDB to refresh the file list
end while

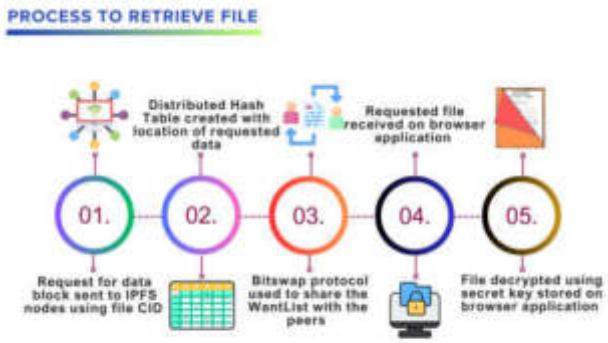


Fig. 6.2: Process diagram to download and decrypt the file from the proposed framework

to generate a different Peer ID. Due to this feature the model is secured from the brute force attacks and man in the middle attacks.

Algorithm 2 File sharing mechanism using the proposed framework

Input: Encrypted File, IPFS API address.
Output: File in database

Initialize Variables: lastPeerId, receiverId, peer, conn, status, filesInDB, selectedFiles
Create a PouchDB instance called db for the files database
Retrieve all documents from the files database
Handle Checkbox(event)
if The checkbox is checked **then**
    Add the checkbox value to the selectedFiles list
else
    Remove the checkbox value from the selectedFiles list
end if
Create an empty list selectedFilesDetails
for all Files in filesInDB **do**
    if file.id is in selectedFiles **then**
        $selectedFilesDetails \leftarrow files$
        return selectedFilesDetails
    end if
end for
Set receiverId to the value entered in the input field
Initialize a Peer Object
Create a peer object and connect to a shared PeerJS server
Create a new peer object and assign it to peer
while The peer object is successfully opened **do**
    Assign the peer ID
    Define event handlers for peer events, such as open, connection, disconnected, close, and error
    Enter the peer ID of the receiver
    Display the current connection status
    Join a connection to the destination peer based on the provided peer ID
    Select files to share
    Allow the user to refresh the list of files available in the local database
    Enable the user to select files for sharing using checkboxes
    for all Selected file do **do**
        if conn exists **then**
            Send the details of selected files to 'conn'
            Display a list of selected file CIDs that will be shared
            Send the selected file CIDs to the receiver
        else
            display an alert indicating that the connection is closed
        end if
    end for
end while
Close connection

7. Results and Conclusion. In this paper, we have presented a secure mechanism for file storage and peer-to-peer sharing which uses the decentralized capabilities of IPFS and the secure communication channel provided by WebRTC. The proposed system encrypts files before uploading them to IPFS and securely shares the encryption keys using WebRTC. Our approach provides confidentiality, integrity, and availability for shared files, while secure key exchange architecture between peers. The limitation of the proposed framework is that both the sender and receiver must have IPFS installed on their respective devices for communication of information. This can however, be made easily adaptable for use and practice as we are adapting to various new technologies for securing our file sharing systems. The robustness of the secure channel communication enables to share any size of files across the network without paying for the services like server space for storage and data encryption.

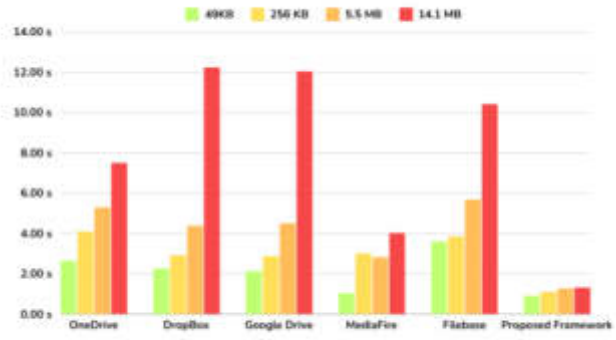


Fig. 7.1: Comparative analysis for file uploading on cloud based platforms

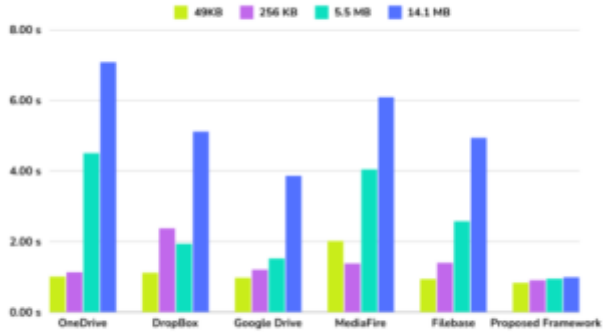


Fig. 7.2: Comparative analysis for file downloading from cloud based platforms

To compare our proposed framework with the popular cloud based solutions built on centralized server we use the React File Picker component that works with the Apideck (File Storage API). The results have been recorded based on the time parameter with 48 data points. System information – M2 MAC with 16GB RAM; tests were performed on fiber connection connected through ethernet on Safari browser. Files of different sizes have been compared for both the download and upload procedure- 49 KB, 256 KB, 5.5 MB and 14.1 MB. The proposed framework is the fastest as compared to OneDrive, DropBox, Google Drive, MediaFire and Filebase platforms as shown in Fig.7.1 and Fig.7.2.

8. Future Work. Our proposed framework combines the potential of IPFS and WebRTC to implement an encrypted, secure, decentralized peer-to-peer file sharing framework. The mechanism has the capability to serve as a foundation for further research in decentralized file sharing and content distribution. However, the further research and development can work around the following areas:

Scalability: Due to increase in the number of users, files shared across the nodes can increase and therefore its performance should be enhanced accordingly.

Access Control: These policies can be implemented to provide users with more control over the users who can access their shared files.

Key Management: Exploring more advanced key management techniques, such as key rotation, modification and regular updates to enhance security and mitigate potential key sharing risks.

Mobile Support: Mobile application for file sharing using similar technology can be implemented, which can extend the system to support mobile devices for a seamless cross-platform experience.

REFERENCES

[1] Kuai, Le, Mary Lacity, and Jeffrey K. Mullins. *Web 2 vs. Web 3 Paths to the Metaverse: Who Is Leading? Who Should Lead?*, The Journal of The British Blockchain Association, September 2023.

[2] W. C. Diehl *Artificial Intelligence, Web 3, and the Future of Distance Education*, American Journal of Distance Education, vol. 37, no. 2, pp. 83-84, 2023.

[3] C. Patsakis and F. Casino *Hydras and IPFS: a decentralised playground for malware*, International Journal of Information Security, vol. 18, pp. 787–799, 2019.

[4] S. Vimal and S. K. Srivatsa *A new cluster P2P file sharing system based on IPFS and blockchain technology*, Journal of Ambient Intelligence and Human Computing, 2019.

[5] A. A. Battah, M. M. Madine, H. Alzaabi, I. Yaqoob, K. Salah, and R. Jayaraman *Blockchain-Based Multi-Party Authorization for Accessing IPFS Encrypted Data*, IEEE Access, vol. 8, pp. 196813-196825, 2020.

[6] N. Nizamuddin, K. Salah, M. Ajmal Azad, J. Arshad, and M. H. Rehman *Decentralized document version control using Ethereum blockchain and IPFS*, Computers & Electrical Engineering, vol. 76, pp. 183-197, 2019.

[7] Zeng R., You J., Li Y., Han R. *An ICN-Based IPFS High-Availability Architecture*, Future Internet, 2022.

[8] On G., Schmitt J., Steinmetz R., *The Effectiveness of Realistic Replication Strategies on Quality of Availability for Peer-To-Peer Systems*, In Proceedings of the Third International Conference on Peer-To-Peer Computing (P2P2003),pp. 57–64, Linkoping, Sweden, 1–3 September 2003.

[9] M. Tomaiuolo, M. Mordonini, and A. Poggi *A P2P Architecture for Social Networking in Applying Integration Techniques and Methods in Distributed Systems and Technologies*, IGI Global, 2019.

[10] V. Pattanaik, I. Sharvadze, and D. Draheim *A Peer-to-Peer Data Sharing Framework for Web Browsers*, SN COMPUT. SCI., vol. 1, p. 214, 2020.

[11] W. P. scWardlaw *The RSA public key cryptosystem in Coding theory and cryptography*, Berlin: Springer, pp. 101–23, 2000.

[12] A. V. Sambra, E. Mansour, S. Hawke, M. Zereba, N. Greco, A. Ghanem, D. Zagidulin, A. Aboulnaga, and T. Berners-Lee *Solid: A platform for decentralized social applications based on linked data*, Tech. rep., MIT CSAIL & Qatar Computing Research Institute, 2016.

[13] A. Tenorio-Fornés, S. Hassan, and J. Pavón *Open Peer-to-Peer Systems over Blockchain and IPFS: an Agent Oriented Framework*, 2018.

[14] G. Alagic, J. Alperin-Sheriff, D. Apon, et al. *Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process*, NISTIR 8309, NIST, U.S. Department of Commerce, July 2020.

[15] S. El Adib and N. Raissouni *AES Encryption Algorithm Hardware Implementation Architecture: Resource and Execution Time Optimization*, International Journal of Information & Network Security (IJINS), vol. 1, no. 2, pp. 110-118, June 2012.

[16] A. K. Das *A random key establishment scheme for multi-phase deployment in large-scale distributed sensor networks*, International Journal of Information Security, vol. 11, pp. 189–211, 2012.

[17] C. Yang, T. Liang, N. Shi, B. Xu, Y. Cao, and K. Yu *AuthPrivacyChain: A Blockchain-Based Access Control Framework With Privacy Protection in Cloud*, IEEE Access, pp. 1-1, 2020.

[18] G. Alagic, J. Alperin-Sheriff, D. Apon, et al. *Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process*,NISTIR 8309, NIST, U.S. Department of Commerce, July 2020.

[19] X. Bonnetain, M. Naya-Plasencia, and A. Schrottenloher *Quantum Security Analysis of AES*,IACR Transactions on Symmetric Cryptology, Ruhr Universität Bochum, 2019.

[20] M. K. C. Ledda, B. D. Gerardo, and A. A. Hernandez *Enhancing IDEA Algorithm using Circular Shift and Middle Square Method*, ICT and Knowledge Engineering (ICT&KE) 2019 17th International Conference, pp. 1-6, 2019.

[21] Abbadini, M., Beretta, M., di Vimercati, S. D. C., Facchinetti, D., Foresti, S., Oldani, G., ... & Samarati, P. *Supporting Data Owner Control in IPFS Networks.*, In Proceeding of the IEEE International Conference on Communications (IEEE ICC 2024).

[22] Dwivedi SK, Amin R, Vollala S.*Smart contract and IPFS-based trustworthy secure data storage and device authentication scheme in fog computing environment.*, Peer-to-Peer Networking and Applications. 16(1):1-21; Jan 2023.

[23] Medina J, Rojas-Cessa R. *AMI-Chain: a scalable power-metering blockchain with IPFS storage for smart cities.*, Internet of Things.1:101097; Feb 2024.

[24] AlKhader W, Jayaraman R, Salah K, Sleptchenko A, Antony J, Omar M. *Leveraging blockchain and NFTs for quality 4.0 implementation in digital manufacturing.*, Journal of Manufacturing Technology Management. 24;34(7):1208-34; Oct 2023.

[25] Alam S, Bhatia S, Shuaib M, Khubrani MM, Alfayez F, Malibari AA, Ahmad S. *An overview of blockchain and IoT integration for secure and reliable health records monitoring.*, Sustainability. 23;15(7):5660; Mar 2023.

Edited by: Kavita Sharma
Special issue on: Recent Advance Secure Solutions for Network in Scalable Computing
Received: May 15, 2024
Accepted: Sep 1, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 531–539, DOI 10.12694/scpe.v26i2.3946

IMPROVEMENT OF THE ACO ALGORITHM FOR INTELLIGENT TASK SCHEDULING IN CLOUD SYSTEMS

ESMA INSAF DJEBBAR*AND GHALEM BELALEM†

Abstract. Cloud computing involves accessing and using computing resources, such as servers, storage, and software applications, over the Internet, enabling scalable access on demand. Cloud computing systems are becoming an essential platform for scientific applications. They enable task scheduling and IT resource allocation. When these resources are not enough to meet demand, planning techniques are necessary. We propose to apply an improved ACO algorithm for intelligent task scheduling and appropriate resource allocation in a cloud environment. This work proposes a modified version of the ACO algorithm that can quickly converge to the best solution to further optimize the total response time, the average response time and the total execution cost. The algorithm suggested based on artificial intelligence application of enhanced ACO is compared with the classical ACO algorithm using Cloudsim simulator. The results obtained after the experiments and simulation are very encouraging to adopt this technique.

Key words: Intelligent Task Scheduling, improved ACO Algorithm, Cloud computing Systems, Cloudsim simulator.

1. Introduction. Cloud computing is an area that brings together the distribution technologies, on demand and via the Internet, of computer software and hardware services. The central idea behind these techniques is to distribute computing resources as a common necessity, as envisioned by the founders of the revolutionary technology over 40 years ago [1]. This principle of public distribution of computing resources also drives the grid computing community, so it is sometimes difficult to distinguish the boundary between Grid and Cloud Computing. This difficulty is all the more real since cloud computing is a young concept, the first implementations of which date back to 2006, and whose development has accelerated in recent years.

Cloud computing reflects a recent model for the provision of computing resources. This model facilitates access to resources via the network in sequence to minimize the costs incurred in managing hardware and software resources. It constitutes the importance of viewing computing as a service in which principled reduction can significantly minimize the cost of computing resources. With the cloud, there is no point investing in an Infrastructure that would be overpriced to acquire. Since the contribution is menstrual, you have better surveillance of your budget and you only pay for what you use. You no longer have to worry about updates, storage and performance issues. Using the cloud, all this is managed by your service provider. The applications and services you use in the Cloud are accessible wherever you are as long as you have a terminal and an internet connection. If your needs change, it is possible to adapt your offer quickly and simply. The user can personalize the services present in his interface.

The concept of Scheduling tasks within cloud computing systems is garnering more attention as Cloud technology becomes increasingly popular. Task scheduling typically involves organizing tasks and assigning them to available resources based on their properties and requirements. This process is vital for ensuring the efficient operation of Cloud systems, as it involves considering various task modules to schedule tasks appropriately. It's essential to utilize the appropriate resources cost-effectively without changing the settings of the Cloud service.

Various algorithms can be employed to select the necessary resources in a cloud system. Resource sorting methods may include random selection, Round Robin, or greedy algorithms (based on resource processing capacity and waiting time), among others. Similarly, task sorting can be based on FCFS (First Come, First

*Department of Computer Science Systems Engineering, National Polytechnic School of Oran-Maurie Audin, Oran, Algeria (esma.djebbar@gmail.com).
†Department of Computer Science, University of Oran1-Ahmed Ben Bella, Algeria

Served), SJF (Shortest Job First), or priority-based approaches. The scheduling algorithm determines which task to execute and on which available resource. Each selection strategy offers distinct advantages, and by implementing them strategically, it's possible to harness their benefits and mitigate the drawbacks, ultimately aiming for an optimized solution.

The remainder of the article is structured as follows: Section 2 introduces the Ant Colony Optimization (ACO) algorithm. Section 3 reviews related studies on task scheduling utilizing the ACO algorithm in cloud computing. Section 4 details the proposed Intelligent Ant Colony Optimization Task Scheduling algorithm. Section 5 presents the experimental results conducted in the Cloudsim simulator. Lastly, Section 6 wraps up the article by summarizing our contributions and suggesting potential avenues for future research.

2. Fundamentals of Ant Colony Algorithm Programming. Ant colony algorithms are derived from the behavior of ants and constitute a collection of optimization meta-heuristics.
The ant colony optimization algorithm operates as an iterative process involving individuals that communicate similar findings, enabling them to guide their subsequent decisions and signal paths to follow or discard to subsequent individuals.

This concept originates from observing how ants search for food resources. Despite their limited individual cognitive abilities, ants collectively manage to discover the most efficient route between a food source and their nest.

The ACO organization chart is provided as a follow-up (Figure2.1). A framework describing this operation is:

1. An ant circulates almost randomly around its colony space.
2. If it detects a feeding orovenance, it usually returns immediately to the nest, releasing a trail of pheromones in its wake.
3. These pheromones being endearing, ants passing through the surrounding area will have the instinct to pursue the imprint, more or less directly

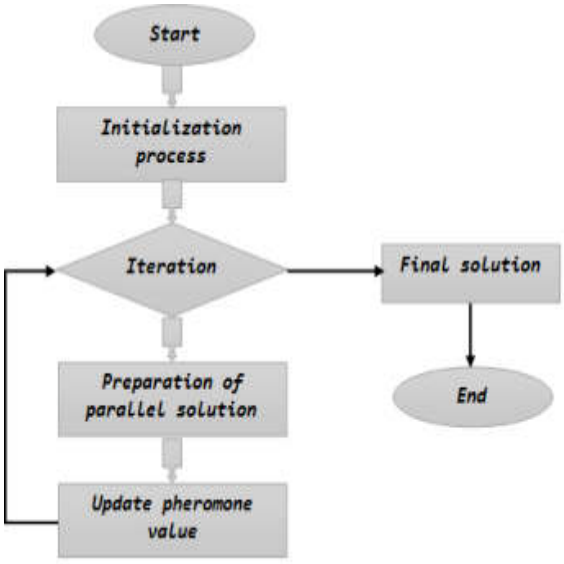


Fig. 2.1: Model of basic ACO Algorithm

3. Related works. Ant colony algorithms (ACO) for optimization can give approximate solutions to difficult problems. In this part, we summarize some works that have used the ACO algorithm for scheduling in cloud systems. Researchers have provided satisfactory algorithms and methods for task scheduling and resource allocation, but there is still room for imperfection since the methods and algorithms mentioned are complex

Table 3.1: Comparison of Task scheduliong algorithm with proposed work.

Work	Objective	Used algorithm	Key contribution	Results
[10].	Ant Colony Optimization for Scientific Workflow.	A Genetic Algorithm and Ant Colony Optimization Algorithm.	Implement Ant Colony Optimization (ACO), initializing the allocation of underutilized VMs with the Pareto distribution.	Reduce response time.
[11].	Optimizing Multi Objective Based Dynamic Workflow.	Black Hole Algorithm and Ant Colony Optimization algorithm.	It gives the hybrid approach for proposed algorithm.	Reduce makespan and total execution cost.
[12].	Enhancing the Ant Colony Optimization Algorithm for Improved Performance.	A* Algorithm and Ant Colony Optimization algorithm.	Introducing the ACOStar algorithm aims to augment ACO performance by integrating the evaluation function used in the A* algorithm.	Reduce execution time.
[13].	An optimized resource scheduling algorithm in fog computing.	A Genetic Algorithm and Ant Colony Optimization algorithm.	Propose New Genetic Ant Colony Optimization Algorithm (NGACO).	Reduce makespan, economic cost, and total cost.
[14].	A hybrid algorithm for deadline-constrained workflow scheduling.	A Genetic Algorithm and Ant Colony Optimization Algorithm.	Introduce a hybrid scheduling algorithm, Partial Critical Path–Ant Colony Optimization (PCP–ACO).	Reduce execution cost.
[15].	Ant colony optimization for parallel test assembly.	Ant Colony Optimization algorithm.	Showcase the versatility of the ACO to construct multiple parallel short scales.	Reduce parallel test assembly.
[16].	An enhanced algorithm for service composition based on ACO.	Ant Colony Optimization algorithm.	An Ant Colony Optimization (ACO) algorithm that incorporates a multi-pheromone approach.	Improve the quality of service (QoS).
[17].	An adaptive ACO algorithm with dynamic ant quantity for addressing the traveling salesman problem.	Dynamic adaptive ant quantity ACO Algorithm and Ant Colony Optimization algorithm.	Dynamically determine the number of ants to prevent them from falling into local optimization.	Reduce convergence time and increase solution quality.
Proposed.	Improvement of the ACO Algorithm for Intelligent Task Scheduling.	Ant Colony Optimization algorithm and Improved ACO Algorithm.	Select the best virtual machine for running tasks.	Reduce response time and execution cost.

class problems (NP-difficult/NP-complete). Table 3.1 provides a comparison of the studied works on task scheduling in cloud computing systems presenting their objectives, algorithms, key contributions and specific results.

The authors in [10] use the Ant Colony Algorithm (ACO) intelligent optimization, in which the allocation of underutilized virtual machines is initiated by the Pareto law repartition. The ACO program is chosen to lead to the determination of migration of virtual machines (VM) by its outcome towards minimum values of cost and response time.

In their study [11], the authors propose enhancing dynamic task scheduling in the Cloud by employing two distinct scheduling algorithms: the Ant Colony Optimization and the Black Hole Algorithm. Their objective is to reduce overall costs and total response time for users. They achieve this by integrating the Black Hole Algorithm with the ACO program, replacing certain steps.

In their research [12], the authors introduce the ACOStar algorithm, aiming to enhance the efficiency of ACO by incorporating an estimation function into its mathematical framework. To validate the performance of this algorithm, they apply it to the shortest path problem, demonstrating its effectiveness and reliability through practical tests.

In [13], the authors propose a resource scheduling model called Normalization Processing, which optimizes pheromone levels to minimize total costs. They also introduce the New Genetic Ant Colony Optimization (NGACO) Algorithm, combining enhancements from Genetic Algorithm and ACO. NGACO features a novel pheromone update approach and utilizes the roulette wheel algorithm for enhanced exploration and population diversity.

The hybrid scheduling algorithm, Partial Critical Path-Ant Colony Optimization (PCP-ACO), is introduced in [14] with the aim of reducing workflow execution costs in cloud environments while meeting user-defined deadlines. By merging the heuristic PCP algorithm with ACO, PCP-ACO achieves faster convergence.

In [15], the authors demonstrate ACO's adaptability in constructing multiple parallel short scales that fulfill various criteria simultaneously. They curate tests from a pool of knowledge items, ensuring gender fairness and precise measurement of factual knowledge.

In [16], a service composition strategy in a multi-cloud environment integrates an Ant Colony Optimization (ACO) algorithm with a multi-pheromone mechanism to enhance Quality of Service (QoS). Additionally, a mutation operation inspired by genetic algorithms is incorporated to mitigate the risk of local optima.

In [17], the authors present the dynamic adaptive ACO algorithm (DAACO), which utilizes a hybrid local selection strategy to enhance the quality of ant optimization and reduce optimization time. Experimental results on the Traveling Salesman Problem (TSPLIB dataset) demonstrate that DAACO outperforms existing ACO algorithms in terms of convergence time, solution quality, and average performance.

4. Proposition of modified Algorithm Ant Colony Optimization Task Scheduling. We propose applying an enhanced ACO algorithm for intelligent task scheduling and optimal resource allocation in a cloud environment. This study introduces a modified version of the ACO algorithm aimed at further optimizing total response time, average time, and overall execution costs. The proposed algorithm, leveraging enhanced ACO through artificial intelligence techniques, is compared against the traditional ACO algorithm using the Cloudsim simulator. A graphical representation of our approach is illustrated in Figure 4.1.

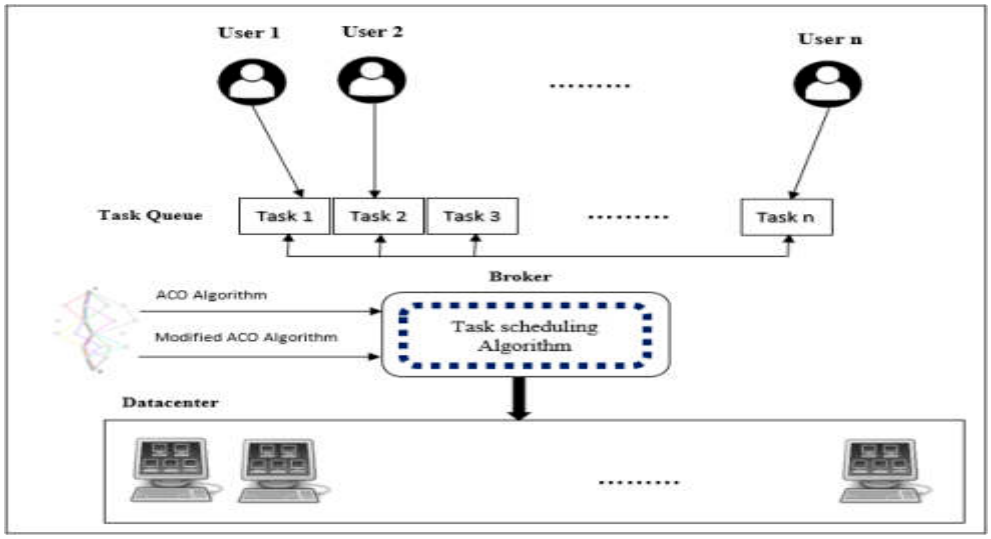


Fig. 4.1: The graphical representation of the proposed work

The improved ACO algorithm for selecting the best virtual machines for task execution and which allows rapid convergence towards the best solution and which is composed of three phases:

4.1. The choice of the VM for the task. For each repetition of the ACO algorithm, and for each individual ant k, $k = 1, ..., m$ (m is the quantity of ants), creates a round carrying out n steps (n is the set of tasks) of which a probability equation is used. The k-th ant selects Vm_j for the following task i with a probability calculated by equation 4.1:

$$P_{ij}^{k}(t) = \begin{cases} \frac{[\tau_{ij}(t)]\times\alpha^2\times[\eta_{ij}(t)]\times\beta^2}{\sum_{j\in authorized_k}[\tau_{ij}(t)]\times\alpha^2\times[\eta_{ij}(t)]\times\beta^2} & \text{if } j \in authorized_k \\ 0 & \text{else.} \end{cases} \tag{4.1}$$

$\tau_{ij}(t)$ indicates the concentration of pheromones at time t between task i and vm_j on the path.
η_{ij} represents visibility; the expected execution and transmission times of task i on the virtual machine vm_j.It is possibly estimated using the following equation 4.2:

$$\eta_{ij} = \ 1/C_{ij} \tag{4.2}$$

C_{ij} is the completion time. $authorized_k$ represents the virtual machines authorized for the ant k. $k = \{0, 1, ..., n-1\}$.
α and β are two parameters that look at weight for pheromone trail and visibility information respectively.

4.2. Pheromone Update. When a turn is made, equation 4.3 calculates the quantity of pheromones on each path (i, j), deposited by each ant k:

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \frac{Q}{L} & \text{if}(i,j) \in T_k(t) \\ 0 & \text{else.} \end{cases} \tag{4.3}$$

$T_k(t)$ represents for each iteration t the turn performed by the ant k. L_k calculated by equation 4.4, defines the length of the turn:

$$L_k = \ argmax_j \left\{\sum_{i\in IJ} C_{ij}\right\} \tag{4.4}$$

IJ is the totality of tasks assigned to virtual machine vm_j. Q is a adjustment coefficient.

The pheromone update employed at all paths is updated by equation 4.5 after each iteration:

$$\tau_{ij}(t+1) = \ (1-\rho)\tau_{ij}(t) + \rho\Delta\tau_{ij}(t) \tag{4.5}$$

ρ is the pheromone evaporation factor, $0 < \rho < 1$.

$$\Delta\tau_{ij}(t) = \ \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t) \tag{4.6}$$

4.3. General Pheromone Update. Once all the ants have completed an iteration, the pheromones of the bridges (i, j) corresponding to the preferable path are reinforced by a quantity Q/L_{best}, where L_{best} is the duration of the best path found T_{best}. This reinforcement is called general pheromone updating and is calculated using equation 4.7:

$$\tau_{ij}(t) = \ \tau_{ij}^{(t)} + \frac{Q}{L_{best}} \quad \text{si}(i,j) \in T_{best} \tag{4.7}$$

5. Experiments and Results. In a cloud computing environment, we opted to compare the basic ant colony optimization algorithm with a modified version for task scheduling and resource allocation using the Cloudsim simulator. The simulations were conducted with varying parameters including the number of cloudlets and VMs, utilizing 40 ants for the experiments.

5.1. Total response time compared to tasks. In the first simulation result, we estimated the average response time by varying the quantity of virtual machines (VM) and the effective of cloudlets. Figure 5.1 shows the simulation result for a number of cloudlets less than or equal to 100 with a number of virtual machines set to 10.

The result of the simulation for a number of cloudlets between 100 and 1000 on 50 virtual machines is given in figure 5.2.

For the simulations whose results are given in figure 5.3, we fixed the number of cloudlets to 500 and varied the number of virtual machines (between 20 and 50 respectively).

The first graphs (Figures 5.1, 5.2, et 5.3) concerning the total response time of queries show that the response time decreases remarkably with a high gain when using the modified ACO optimisation algorithm.

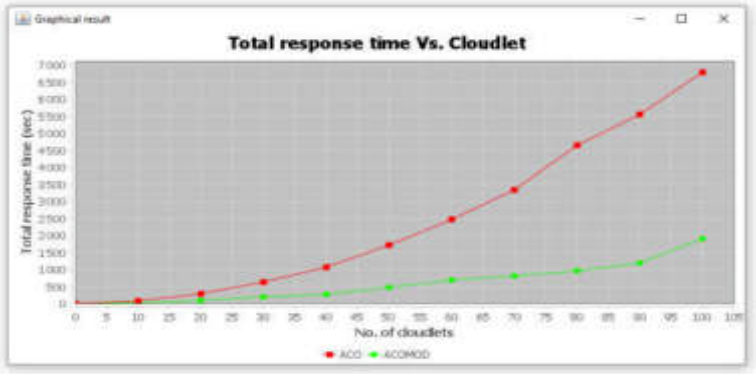


Fig. 5.1: The first total response time

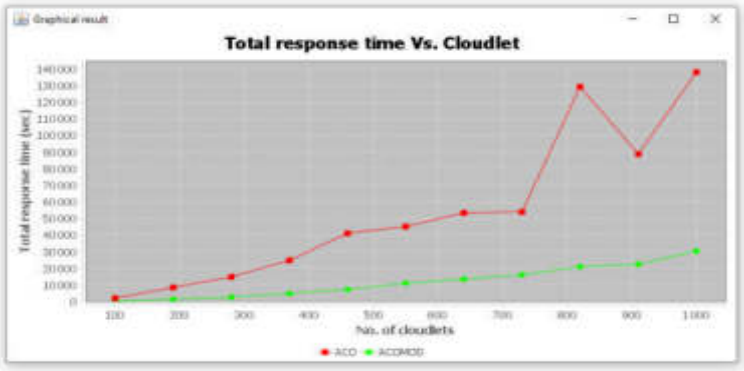


Fig. 5.2: The second total response time

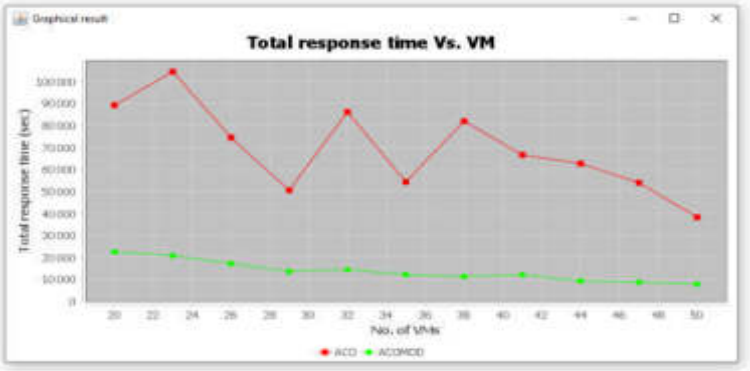


Fig. 5.3: The third total response time

5.2. Total execution cost. For this series of simulations, we calculated the total execution cost by varying the quantity of virtual machines (VM) and the effective of cloudlets. Figure 5.4 shows the simulation result for a number of cloudlets less than or equal to 100 with a number of virtual machines set to 10.

The result of the simulation for a number of cloudlets between 100 and 1000 on 50 virtual machines is

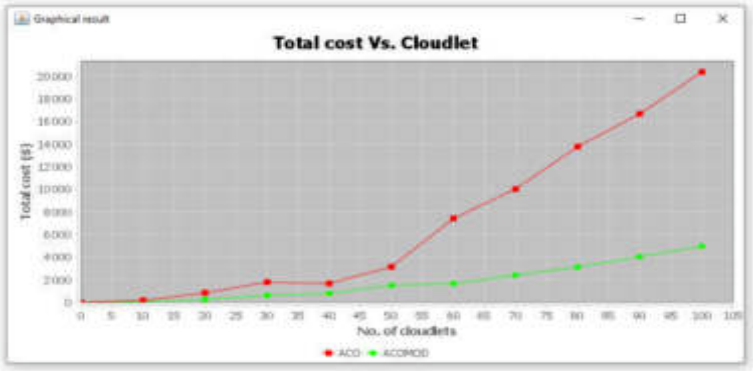


Fig. 5.4: The first total execution cost

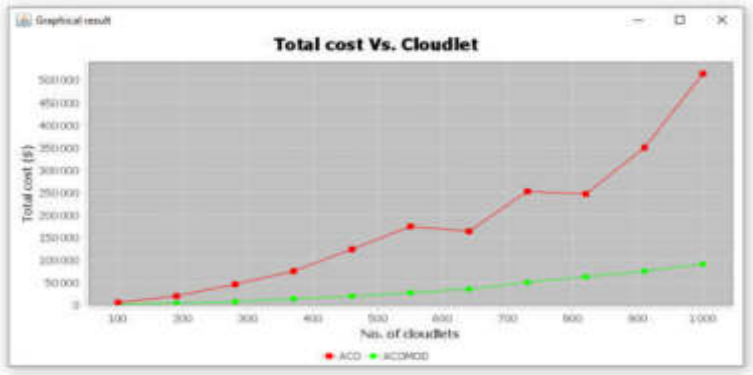


Fig. 5.5: The second total execution cost

Fig. 5.6: The third total execution cost

given in figure 5.5. For the simulations whose results are given in Figure 5.6, we varied the number of virtual machine (VM) (between 20 and 50) and we fix the number of cloudlets to 500.

In the graphs in (Figures 5.4, 5.5 and 5.6, we can see that the cost of executing queries has been reduced considerably, with a significant gain.

5.3. Average response time. For this simulation series, we calculated the average response time by increasing the quantity of virtual machines (VM) and the effective of cloudlets. Figure 5.7 shows the simulation result for a number of cloudlets between 1000 and 2000 with a number of virtual machines set to 100.

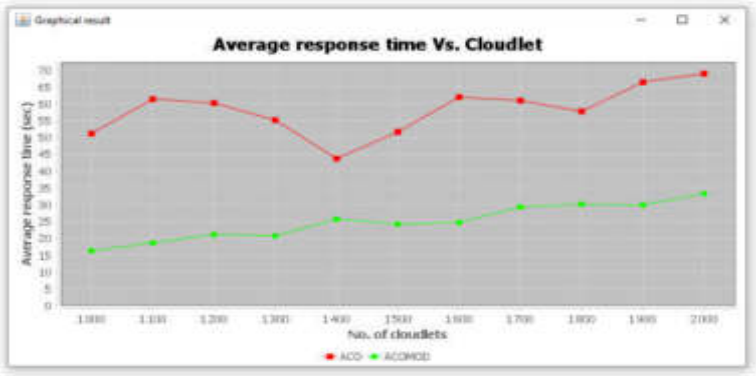


Fig. 5.7: The average response time

In the graphs in (Figures 5.7, we can see that the cost of executing queries has been reduced considerably, with a significant gain.

6. Conclusion. Cloud computing technology evoke an original model for the provision of computing resources. This technology has several advantages such as resource elasticity, pay-per-use payment model and others benefits.Cloud computing makes it possible to make resource allocations IT, but resources are not always sufficient. They often cannot meet user needs. Therefore, task scheduling and resource allocation mechanisms are needed to improve the optimization criteria. New strategies can employ some scheduling and allocation concepts to provide better task scheduling and resource allocation. A scheduling problem consists of organizing the execution of tasks over time, taking into account time constraints (deadlines, sequence constraints) and constraints relating to the availability of the required resources. The ACO algorithm, imitated from the act of foraging ants, which is a popular fuzzy optimization technique. The major principle of this act is the hidden dialogue between the ants by means of trails of chemical pheromones which allow them to find short routes between their home and their food germs. In this framework, the decisions made by all the ants are relevant and the experiences of all the ants are used at each iteration to build the new optimal solution. In our modest work, we have proposed a new modified variety of the ACO algorithm. We tested our algorithm under the CloudSim simulator. The approach reduces the total response time, the average response time, and the total execution cost. In summary, the simulation results obtained for the proposed scheduling strategy are satisfactory, very encouraging, and meet the objectives set. The improved ACO algorithm for task scheduling is not tested on real workflows. This point is very important to consider in future work. We also propose an improvement of the algorithm by hybridization with the gray wolf optimization algorithm which is an algorithm widely used recently in this type of work.

REFERENCES

[1] Ian Foster, Yong Zhao, Ioan Raicu, and Shiyong Lu, *Cloud Computing and Grid Computing 360-Degree Compared*, 2009.

[2] Marco Dorigo and Thomas Stützle, *The Ant Colony Optimization Metaheuristic: Algorithms, Applications, and Advances*, 2003.

[3] Liyun Zuo, Lei Shu, Shoubin Dong, Chunsheng Zhu and Takahirot Hara, *A Multi-Objective Optimization Scheduling Method Based on the Ant Colony Algorithm in Cloud Computing*, vol. 3, pp. 2687-2699, IEEE, December 2015.

[4] A. Arunarani and D.Manjula *Task scheduling techniques in cloud computing a literature survey*, Future Generation Computer Elsevier, vol. 91, pp. 407–415, 2019.

[5] S.K. Pande, S.K.Pande, and S. Das, *Task partitioning scheduling algorithms for heterogeneous multi-cloud environment*, Arab J. Sci. Eng. 43, pp. 913–933, 2018.

[6] R.P.Padhy, M.R.Patra and S.C.Satapathy, *Cloud computing: security issues and research challenges*, IRACST Int. J. Comput. Sci. Inf. Technol. Secur.,1(2), pp. 136–145, 2011.
[7] F. Yiqiu, X. Xia, and G. Junwei *Cloud computing task scheduling algorithm based on improved genetic algorithm*, IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, pp. 852–856, 2019.
[8] A. Gupta and R. Garg *Load balancing based task scheduling with ACO in cloud computing*, International Conference on Computer and Applications (ICCA), Doha, pp. 174–179, 2017.
[9] M. Gawali and S. Shinde, *Task scheduling and resource allocation in cloud is using a heuristic approach*, J. Cloud Comput. Adv. Syst. Appl.,7(4), pp. 1–16, 2018.
[10] Arvind Lal and C. Rama Krishna, *Critical Path-Based Ant Colony Optimization for Scientific Workflow Scheduling in Cloud Computing Under Deadline Constraint*, Ambient Communications and Computer Systems, Springer Singapore, pp. 447-461, 2018.
[11] Archana N. Jethava and Megha R. Desai, *Optimizing Multi Objective Based Dynamic Workflow Using ACO and Black Hole Algorithm in Cloud Computing*, 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp.1144-1147, 2019.
[12] M. Faisal and F. Albogamy, *Ant Colony Optimization Algorithm Enhancement for Better Performance*, IEEE World AI IoT Congress (AIIoT), pp. 0701-0710, 2023.
[13] C. Yin, Q. Fang, H. Li, Y. Peng,X. Xu and D. Tang, *An optimized resource scheduling algorithm based on GA and ACO algorithm in fog computing*, The Journal of Supercomputing, vol.80, pp. 4248–4285, 2024.
[14] A. Verma and S. Kaushal, *A hybrid multi-objective Particle Swarm Optimization for scientific workflow scheduling*, The Journal of Parallel Computing, vol.62, pp.1-19, 2017.
[15] L. Zimny, U. Schroeders, and O. Wilhelm, *Ant colony optimization for parallel test assembly*, The Journal of Behavior Research Methods, 2024.
[16] L. Bei, L. Wenlin, S. Xin and X. Xibin, *An improved ACO based service composition algorithm in multi-cloud networks*, The Journal of Cloud Computing, vol.13(17), 2024.
[17] H. Liu, A. Lee, W.Lee, P. Guo, *DAACO: adaptive dynamic quantity of ant ACO algorithm to solve the traveling salesman prob*, The Journal of Complex & Intelligent Systems, vol.9(4), 2023.
[18] A. J. Mohammed, K. I. Ghathwan and Y. Yusof, *Optimal robot path planning using enhanced particle swarm optimization algorithm*, Iraqi Journal of Science, pp. 178-184, 2020.
[19] C. Dudeja, *Fuzzy-based modified particle swarm optimization algorithm for shortest path problems*, Soft Computing, vol. 23, pp. 8321-8331, 2019.
[20] T. Chakraborty, S. Chakrabarti , and B. Hazra, *Study of β, ρ and Q_0 parameters for Shortest Path Estimation using Ant Colony Optimization*, 2019 IEEE Region 10 Symposium (TENSYMP), pp. 813-818, 2019.
[21] H. Alayed, F. Dahan, T. Alfakih, H. Mathkour, and M. Arafah, *Enhancement of Ant Colony Optimization for QoS-Aware Web Service Selection*, IEEE Access, vol. 7, pp. 97041–97051, 2019.
[22] J. Chen and J. Zhou, *An Improved Ant Colony Optimization for QoS-Aware Web Service Composition*, In: 2020 Eighth International Conference on Advanced Cloud and Big Data (CBD), pp 20–24, 2020.
[23] P. Shobeiri, M. Akbarian Rastaghi, S. Abrishami, and B. Shobiri. *PCP–ACO: a hybrid deadline-constrained workflow scheduling algorithm for cloud environment*. The Journal of Supercomputing, vol 80(6), pp. 7750-7780, 2024.
[24] R. Ghafari and N. Mansouri, *Improved Harris Hawks Optimizer with chaotic maps and opposition-based learning for task scheduling in cloud environment*, Cluster Computing, vol. 27(2), pp. 1421-1469, 2024.
[25] J. Kok Konjaang and Lina Xu, *Meta-heuristic Approaches for Effective Scheduling in Infrastructure as a Service Cloud: A Systematic Review*, Journal of Network and Systems Management, Vol. 29(2), 2021.

Edited by: Kavita Sharma
Special issue on: Recent Advance Secure Solutions for Network in Scalable Computing
Received: May 15, 2024
Accepted: Jul 23, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 540–551, DOI 10.12694/scpe.v26i2.3934

FORECAST OF TOBACCO RAW MATERIAL DEMAND BASED ON COMBINATION PREDICTION MODEL

BIN CHEN* JILAI ZHOU† HAIYING FANG‡ RENJIE XU§ AND WEIYI QU¶

Abstract. In order to improve the prediction accuracy of tobacco raw material demand, this paper presented a combined prediction model. Combined prediction model first used Holt-winters exponential smoothing method and SARIMA model to forecast the demand of cigarette raw materials respectively, and then used BP neural network to aggregate the results of these two predictions to get the final prediction result. Holt-winters exponential smoothing method, SARIMA model and combined prediction model were used to forecast the demand data of tobacco raw materials, respectively. For the prediction of the same material, the error of the combined prediction model were all less than the other two models. The prediction accuracy of combined prediction model was higher.

Key words: Demand prediction; Holt-winters exponential smoothing method; SARIMA model; BP neural network; Combined prediction model.

1. Introduction. In the practical production process of tobacco firms, in order to meet the demand of production and realize the continuity of production, firms often appear the phenomenon of tobacco raw material replenishment, which consumes a lot of manpower and material resources. Therefore, accurate prediction of tobacco raw material demand is helpful for enterprises to make more reasonable production plans, realize rational warehouse planning, reduce enterprise resource consumption and improve enterprise work efficiency.

For raw material demand forecast, the traditional forecast methods mainly include trend extrapolation, wavelet analysis, regression analysis, grey prediction and time series prediction. The trend extrapolation method is usually used for relatively simple functional models, such as linear functions [1]. Although wavelet analysis can analyze and study each time component, it is difficult to restore the original time series because each time component has no practical significance [2].

Regression analysis determines future demand by fitting curves of past data into regression equations. However, when dealing with samples of uncertain factors, its performance is not satisfactory [3]. Grey prediction method can deal with the samples of uncertain factors well and requires a relatively small sample size [4]. The time series method fully considers the periodicity, seasonality, randomness and other significant characteristics involved in the time factor in the prediction, which can carry out dynamic analysis on the data and judge the trend of the data change over time. Based on the above advantages, time series method has been applied to air environment prediction, financial field, special material demand analysis and other fields [5][6].

ARIMA model and Holt-Winters exponential smoothing method have been widely used. ARIMA focuses on using difference method to analyze rules in time series [7]. Compared with ARIMA, seasonal ARIMA (SARIMA) adds seasonal factors in the forecasting process, therefore it is more suitable for time series with periodicity [8].

Exponential smoothing is a time series prediction method of weighted average movement [9]. Holt-Winters exponential smoothing method differs from ordinary exponential smoothing in that it can analyze the horizontal, trend, and seasonal components of historical data simultaneously [10]. Through the analysis and observation of the historical data of tobacco raw materials, it is found that it has obvious seasonality, so SARIMA and Holt-Winters exponential smoothing method are suitable for its prediction model.

*Hongyun Honghe Tobacco (Group) Co., Ltd., China.
†Hongyun Honghe Tobacco (Group) Co., Ltd., China.
‡Hongyun Honghe Tobacco (Group) Co., Ltd., China.
§Hongyun Honghe Tobacco (Group) Co., Ltd., China.
¶Business School of Hohai University, Nanjing, China (Corresponding author, weiyi_hh2198@163.com).

In the 21st century, machine learning has been applied to the field of prediction by many scholars. Common machine learning methods are Random Forest (RF), BP Neural Network, artificial neural network (ANN) [11][12]. Among them, BP Neural Network has good performance in nonlinear regression and has been widely used [13]. On the other hand, combinatorial forecasting methods have also attracted the attention of many scholars [14][15].

According to previous studies, BP neural network, Holt-winters exponential smoothing method and SARIMA model have been widely used in the field of prediction, but few studies have combined the three and applied them to the tobacco field. A combined demand prediction model combining BP neural network, Holt-winters exponential smoothing method and SARIMA model was constructed to fit the historical data of tobacco raw materials and achieve accurate prediction of tobacco raw material demand.

2. Model Construction.

2.1. SARIMA Model. The SARIMA model evolved from the basic ARMA model. The ARMA model is suitable for solving the prediction problem with stationary series. In this case, seasonal difference needs to be added to the sequence, and a more complex SARIMA model is needed to realize the prediction of time series.

The SARIMA model can be expressed by each different parameter as SARIMA(p,d,q) $(P,D,Q)_S$, whose modeling formula is shown in equations (2.1)-(2.5) [16]:

$$\phi_p(L)\Phi_P(L^S)\Delta^d\Delta_S^D y_t = \theta_q(L)\Theta_Q(L^S)u_t \tag{2.1}$$
$$\phi_p(L) = (1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p) \tag{2.2}$$
$$\Phi_P(L^S) = (1-\Phi_1 L^S-\Phi_2 L^{2S}-\cdots-\Phi_p L^{PS}) \tag{2.3}$$
$$\theta_q(L) = (1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q) \tag{2.4}$$
$$\Theta_Q(L^S) = (1+\Theta_1 L^S+\Theta_2 L^{2S}+\cdots+\Theta_Q L^{QS}) \tag{2.5}$$

where p and P mean non-seasonal and seasonal autoregressive orders; ϕ_p and $\Phi_P(L^S)$ mean non-seasonal and seasonal autoregressive polynomials; d and D mean non-seasonal and seasonal difference orders; S means the seasonal cycle; Δ^d and Δ_S^D are difference and seasonal difference; y_t represents a time series; q and Q mean non-seasonal and seasonal moving average polynomials. $\theta_q(L)$ and $\Theta_Q(L^S)$ mean non-seasonal and seasonal moving average polynomials; u_t means the error.

This study adopted EViews software to construct SARIMA model, which could be divided into the following three steps:

Step 1: Stationarity test. In this paper, the stationarity of historical date series was determined by unit root test. If the series was non-stationary, it needs to make d-order difference; if the sequence had seasonality with period S, it needs to make D-order difference with step size S to make it into stationary sequence.

Step 2: Model selection. That is, the values of p, q, P, Q were certain. By observing the characteristics of the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) graph of the series after difference, the alternative combinations of each order of the model could be certain, and the optimal combinations were screened out by AIC, SC, HQC and other criteria.

Step 3: Model test. Each parameter of the model was determined in step 1 and step 2, and the residual sequence could be generated by running the model. In this paper, ADF (Augmented Dickey-Fuller) unit root test and Q test were used to test the residual error of SARIMA model.

Step 4: Model prediction. Made predictions according to the model selected in Step 3. By comparing the forecast data with the historical data, the prediction error was calculated. MAE, MAPE and RMSE were selected as the evaluation indexes. The three evaluation indexes showed that the lower the calculated value, the higher the prediction accuracy of the models. Each index formula is shown in

formula respectively:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{y}_i| \tag{2.6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{y}_i)^2} \tag{2.7}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{x_i - \hat{y}_i}{x_i}| * 100\% \tag{2.8}$$

Among them: x_i represents the historical data, $\hat{y}_i$ represents forecasts, n represents sample number.

2.2. Holt-Winters Exponential Smoothing Method. As a three-parameter smoothing method, exponential smoothing is mainly applied to time series with both linear and seasonal trends. The core idea of exponential smoothing method is to decompose time series into linear trend, seasonal trend and irregular change, then use exponential smoothing method to predict the above three trends or changes respectively, and finally establish addition or multiplication models for comprehensive prediction. The three parameters of Holt-winters exponential smoothing method include α, β and γ, which control the level, trend and season at time t, respectively. The value of each parameter ranges from 0 to 1. A larger value indicates a higher prediction ratio.

The addition model formula of Holt-winters exponential smoothing method is as follows:

$$u_t = \alpha(y_t - s_{t-T}) + (1-\alpha)(u_{t-1} + v_{t-1}) \tag{2.9}$$

$$v_t = \beta(u_t - u_{t-1}) + (1-\beta)v_{t-1} \tag{2.10}$$

$$s_t = \gamma(y_t - u_t) + (1-\gamma)s_{t-T} \tag{2.11}$$

The multiplication model formula of Holt-winters exponential smoothing method is as follows:

$$u_t = \alpha(\frac{y_t}{s_{t-T}}) + (1-\alpha)(u_{t-1} + v_{t-1}) \tag{2.12}$$

$$v_t = \beta(u_t - u_{t-1}) + (1-\beta)v_{t-1} \tag{2.13}$$

$$s_t = \gamma(\frac{y_t}{u_t}) + (1-\gamma)s_{t-T} \tag{2.14}$$

where, y_t is time series, u_t is linear trend, v_t represents the linear increasing rate of u_t, s_t is seasonal trend, T means seasonal cycle.

2.3. BP Neural Network. The structure of the BP neural network is shown in Figure 2.1. The original data is input layer, and then Data from the input layer pass to the hidden layer, hidden layer built-in transfer function used to process the input to the hidden layer of data, hidden layer will be processed data to the output layer, the output result z_1, z_2, $\cdots$, z_n. The BP neural network compares the output results with the original data and calculates the errors. If the errors meet the set requirements, the model can be determined. If the errors are greater than the set requirements, the errors will be reverse-transmitted to the hidden layer, and then the model parameters will be redetermined until the model errors meet the requirements.

There are three types of BP neural network data sets, namely training set, verification set and test set. In addition, in order to accelerate the training speed and improve the model performance and stability, BP neural network needs to normalize the original data. The formula is as follows:

$$S_t^{'} = \frac{S_t - min(S)}{max(S) - min(S)} \tag{2.15}$$

Among them: S means time sequence, S_t and $S_t^{'}$ means the value of the time series and normalized time series at time t, respectively. $max(S)$ and $min(S)$ mean the maximum and minimum value in the sequence.

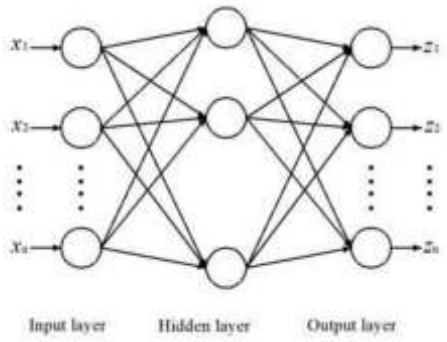


Fig. 2.1: Neural network structure diagram

2.4. Combinatorial Forecasting Model. Combinatorial forecasting techniques can be divided into two types: model grouping and result grouping. The model grouping method focuses on the analysis of the advantages and disadvantages of different models to put forward a combination method to realize the complementary advantages of the models. The common model grouping method includes the combination of qualitative prediction and quantitative prediction, linear prediction and nonlinear prediction, dynamic prediction and static prediction. The idea of result grouping method is to first obtain multiple prediction results by using different prediction models and methods, then appropriate methods are used to determine the weight of each model, and finally their weighted average is calculated and used as the predictive value. This paper chose the result grouping method and the expression of the weight model is as follows:

$$\hat{y}_t = w_{1t}y_{1t} + w_{2t}y_{2t} \tag{2.16}$$

In this formula: $\hat{y}_t$ is the output of BP neural network and represents prediction value of combination forecast model at time t. y_{1t} and y_{2t} are the input of BP neural network and represent the predicted values of the SARIMA model and Holt-winters exponential smoothing method at time t, respectively. w_{1t} and w_{2t} represent the respective weights of the predicted values of SARIMA model obtained by BP neural network and Holt-winters exponential smoothing method at time t.

The objective function of BP neural network is $L2$ norm loss function, and the formula is as follows:

$$L_2(\hat{y}, y) = \sum_{t=0}^{m} (y_t - \hat{y}_t)^2 \tag{2.17}$$

In this formula: y_t represents value of the time series at time t, $\hat{y}_t$ represents prediction value of combination forecast model at time t.

There were three steps to implement the combined prediction model:

Step 1: SARIMA model and Holt-winters exponential smoothing method were respectively used to forecast the past tobacco raw material demand data, and the respective prediction results were obtained.

Step 2: Selected the appropriate proportion and divided the predicted results obtained in Step 1 and the corresponding real values into training set, verification set and test set. And take the actual demand as the model test standard. Using training set to train the BP neural network, and verification set to debug the parameters.

Step 3: Taking the training set and verification set as the input after parameter debugging, running the BP neural network, and then output prediction result. Calculated the prediction error and evaluated the prediction effect. The evaluation indicators were MAE, RMSE and MAPE.

It is worth mentioning that when the time series has 0 or the seasonal variation of the series is not significant, the SARIMA model will no longer be applicable, thus affecting the prediction accuracy of the combined model.

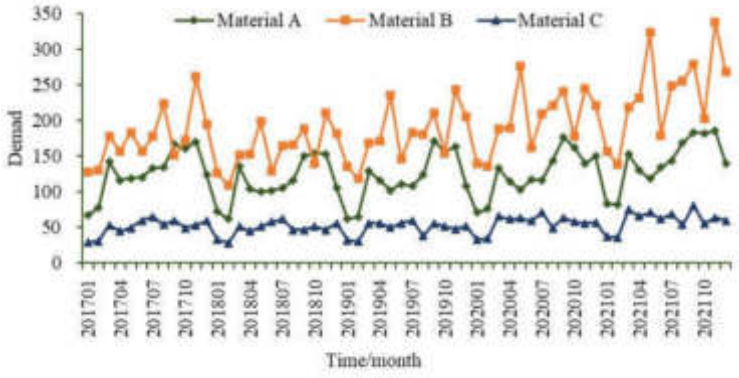


Fig. 3.1: Trend chart of historical data of materials demand

Table 3.1: Unit root text of DSX

		$t-Statistic$	$Prob$
ADF test statistic		-5.970426	0.0000
Test critical values	1%level	-3.615588	
	5%level	-2.941145	
	10%level	-2.609066	

3. Empirical Analysis.

3.1. Data Sources and Description. In this paper, the materials historical demand data of material A, B and C of Y Tobacco Company from January 2017 to December 2021 were selected as samples, in which material C was the main raw material of Y Tobacco Company's high-end products, and material A and B were the raw material of common products. Here were 60 periods of historical data for different materials. In order to verify the prediction accuracy of the combined prediction model, the data from July to December 2021 were taken as the test set of the combined model. The historical data of each material are shown in Figure 3.1. Among them, the demand for material A and material B decreased slightly in 2018, and the demand for raw materials showed an overall rising trend from 2019 to 2021. As high-end raw materials, the overall demand for material C was less than that for material A and material B. The overall data of the three kinds of material showed non-stationary time series with obvious seasonality.

Due to space limitations, this paper only shows the process of predicting raw material demand of material A by using models. The prediction results of each model of all materials will be shown in Section 3.4.

3.2. Prediction of SARIMA Model. The first-order difference of $S = 12$ was applied to the historical data series of raw material demand of material A from January 2017 to June 2021, that was, SARIMA$(p, 1, q)$ $(P, 1, Q)_{12}$. The sequence before the difference was called X and the sequence after the difference was called DSX.

First, unit root test was performed on DSX. As could be seen from Table 3.1, T-test statistic was -5.970426, which was less than -2.941145 at 5% test level, indicating that residuals pass ADF test and DSX was stationary time series.

ACF and PACF of DSX were observed, as shown in Figure 3.2. PACF of DSX was significantly non-0 when lagging 1 order, so $p = 1$ could be considered; ACF was significantly non-0 when lagging 1 order, so $q = 1$ could be considered; in addition, since seasonal difference with period 12 was carried out in the time series, $P = 1$

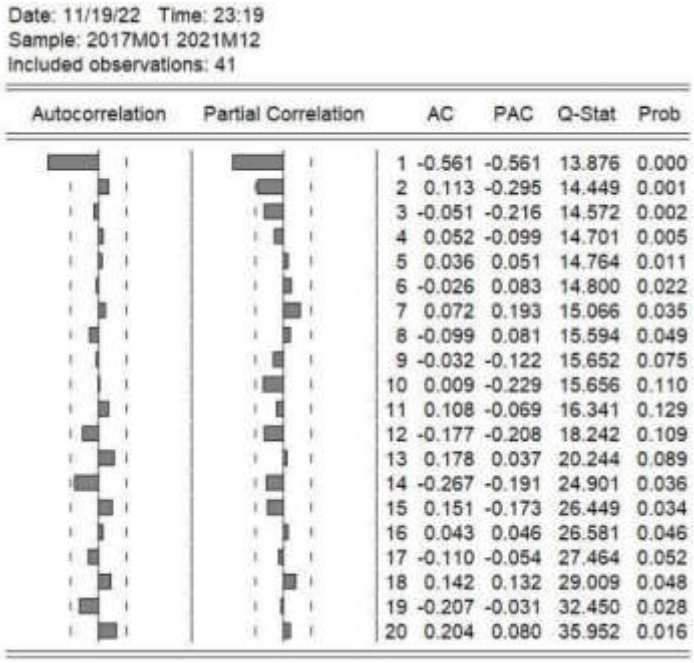


Fig. 3.2: ACF-PACF diagram of DSX

Table 3.2: Comparison results of SARIMA models

Model	Adjusted R-squared	AIC	SC	HQC
$SARIMA(1,1,1)(1,1,1)_{12}$	-	-	-	-
$SARIMA(1,1,1)(1,1,0)_{12}$	0.422709	7.864747	8.031925	7.925624
$SARIMA(1,1,1)(0,1,1)_{12}$	-	-	-	-
$SARIMA(1,1,0)(1,1,1)_{12}$	-	-	-	-
$SARIMA(0,1,1)(1,1,1)_{12}$	0.423823	7.882115	8.049293	7.942992
$SARIMA(1,1,1)(0,1,0)_{12}$	0.398515	7.856398	7.981781	7.902056
$SARIMA(0,1,0)(1,1,1)_{12}$	0.087883	8.316835	8.442218	8.362492
$SARIMA(1,1,0)(1,1,0)_{12}$	0.309010	7.994796	8.120179	8.040453
$SARIMA(0,1,1)(0,1,1)_{12}$	0.439722	7.833452	7.958836	7.879110
$SARIMA(0,1,1)(1,1,0)_{12}$	0.427561	7.840078	7.965461	7.885735
$SARIMA(1,1,0)(0,1,1)_{12}$	0.335584	7.980981	8.106364	8.026639
$SARIMA(1,1,0)(0,1,0)_{12}$	0.311724	7.959510	8.043099	7.989949
$SARIMA(0,1,1)(0,1,0)_{12}$	0.389070	7.847230	7.930819	7.877668
$SARIMA(0,1,0)(1,1,0)_{12}$	0.084441	8.278828	8.362416	8.309266
$SARIMA(0,1,0)(0,1,1)_{12}$	0.113267	8.268651	8.352240	8.299090

and $Q = 1$ could be considered.

Combined different values of P, p, Q, q into different SARIMA models, and calculate the Adjusted R-squared, AIC, SC and HQC values of each model. When the adjusted R-squared value was larger and the AIC, SC and HQC values were smaller, the prediction of the model was better. The comparison results of models could be seen in Table 3.2. Through comprehensive comparison and observation, $SARIMA(0,1,1)(0,1,1)_{12}$ performed best and selected as the model of material A.

The ADF test was carried out for the residual of $SARIMA(0,1,1)(0,1,1)_{12}$ model. The test results could

Table 3.3: ADF test of residual series

		$t-Statistic$	$Prob$
ADF test statistic		-5.354789	0.0001
Test critical values	1%level	-3.632900	
	5%level	-2.948404	
	10%level	-2.612874	

Table 3.4: Error of SARIMA$(0,1,1)(0,1,1)_{12}$ model

Time/month	Actual value	Prediction value	MAPE
2021.07	144	130.23	9.56%
2021.08	168	152.16	9.43%
2021.09	183	188.05	2.76%
2021.10	182	176.25	3.16%
2021.11	185	164.87	10.88%
2021.12	140	150.36	7.40%
MAE:11.82	RMSE:12.99	MAPE:7.20%	

be seen in Table 3.3. The T-test statistic was -5.354789, which was less than the -2.948404 at the test level of 5%, indicating that the residual has passed the ADF test. At the same time, the Q test with 20 order lag for the residual sequence was performed and $P = 0.294 > 0.05$ was obtained, indicating that the residual sequence passed the Q test and appeared as white noise sequence. If a sequence is a white noise sequence, it means that all information in the sequence has been extracted. Therefore, the SARIMA$(0,1,1)(0,1,1)_{12}$ model was valid and could be used to forecasting.

The SARIMA$(0,1,1)(0,1,1)_{12}$ model was used to fit and predicted the X series, the forecast data from July to December 2021 were used to compare with the real demand data, and calculated the prediction accuracy. Model errors were shown in Table 3.4, the MAPE was 7.20%, which was less than 10%. It was proved that the performance of the model was good, but it could also be seen that the stability of the model was poor for the prediction of different months.

3.3. Prediction by Holt-Winters Exponential Smoothing Method. EViews software was used for prediction, and the parameters in additive model and multiplicative model were the parameters with the highest fitting degree automatically calculated by the software. With the historical raw material demand data of A from January 2017 to December 2020 as the original data, the demand of material A from January 2021 to June 2021 was predicted. According to the size of the errors, judged which model of addition model or multiplication model had better prediction effect. Table 3.5 showed the prediction results of two models of Holt-winters. According to the table, compared with the multiplication model, most of the predicted values of the addition model were closer to the real values. But more specific calculations were needed, using errors to measure the performance of both models.

The evaluation indexes of the additive model and the multiplicative model were calculated and shown in Table 3.6. According to the evaluation results, each index of the addition model was superior to the multiplication model, and the addition model was selected as the prediction model of raw material demand for material A.

Taking the historical raw material demand data of A from January 2017 to June 2021 as the original data, the addition model was used to forecast the raw material demand from July 2021 to December 2021, and compared with the actual demand in the current month. The fitting error of the Holt-winters addition model could be seen in Table 3.7. The MAPE of addition model was 4.60% and less than 7.20% of SARIMA model, other indexes were also smaller. However, the maximal MAPE of different months was 10.60% and the

Table 3.5: Forecast result of Holt-winters models

Time/month	Actual value	model	
		Additive model	Multiplication model
2021.01	83	74.16	71.57
2021.02	82	76.41	73.58
2021.03	153	140.91	141.67
2021.04	130	118.66	118.19
2021.05	118	112.16	111.26
2021.06	134	118.66	118.05

Table 3.6: Fitting results of Holt-winters models

Model	Parameter			Fit evaluation		
	α	β	γ	MAE	$RMSE$	$MAPE$
Additive model	0.26	0	0	9.84	10.34	8.14%
Multiplication model	0.24	0	0	10.95	11.32	9.96%

Table 3.7: Fitting error of Holt-winters additive model

Time/month	Actual value	Prediction value	MAPE
2021.07	144	134.93	6.30%
2021.08	168	150.20	10.60%
2021.09	183	187.97	2.72%
2021.10	182	181.49	0.28%
2021.11	185	181.02	2.15%
2021.12	140	147.79	5.56%
MAE:9.13	RMSE:7.35	MAPE:4.60%	

minimum MAPE was 0.28%, the stability of the model was also not very well. In general, the fitting effect was better than that of SARIMA model.

3.4. Combination Model Prediction. The number of neurons in the BP neural network determines the complexity of the model, so as to adapt to the nonlinear mapping relationship of different complexity, but at the same time, the number of neurons also affects the computational complexity and training time. Therefore, the number of neurons of BP neural network needed to be confirmed. According to the empirical formula, the optimal number of neurons of hidden layers was between 2 and 8. Then, 35 groups of data from February 2018 to December 2020 were selected as the training set, the minimum error of the training target was set as 0.0001, the training times as 1000, and the learning rate as 0.01. Data from January to June 2021 were selected as the verification set. The number of each neuron was tested for 15 times, RMSE was used as the evaluation index, and the mean value of the 15 test results was taken. As can be seen from Figure 3.3, for the demand prediction of material A, the most appropriate number of hidden layer neurons was 2. As the number of neurons increased, the complexity of the model increased, and overfitting problems might occur, resulting in increased model errors.

After the number of neurons was determined, the combined prediction model could be constructed. The material A demand data from July 2021 to December 2021 were predicted. The prediction results and the values of each evaluation index were shown in Table 3.8. The MAPE of the combined prediction model decreased to

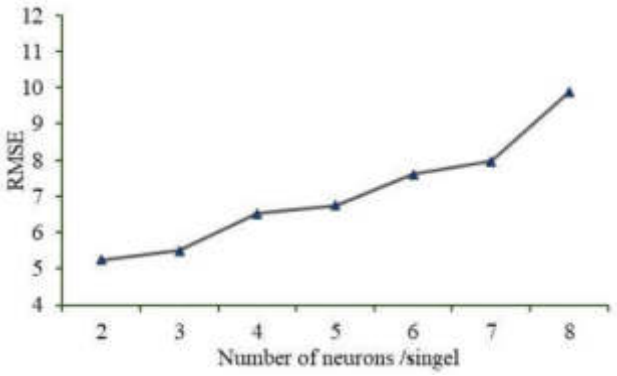


Fig. 3.3: Evaluation results of different neuron numbers

Table 3.8: Forecast results of combination forecasting model of material A

Time/month	Actual value	Prediction value	MAPE
2021.07	144	138.52	3.81%
2021.08	168	152.92	8.98%
2021.09	183	184.16	0.64%
2021.10	182	180.65	0.74%
2021.11	185	182.22	1.50%
2021.12	140	150.31	7.37%
MAE:7.90	RMSE:6.05	MAPE:3.84%	

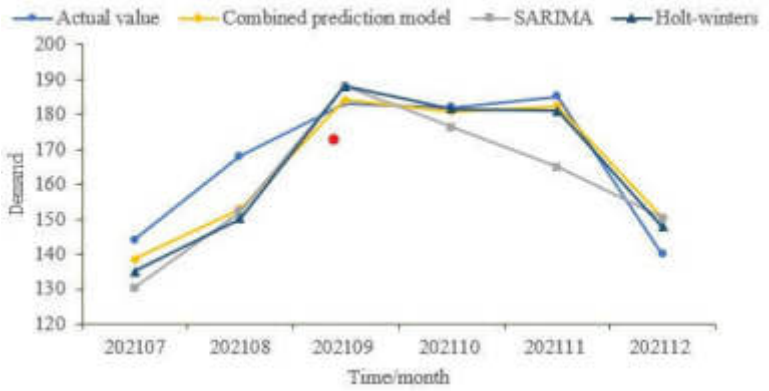


Fig. 3.4: Comparison of prediction results among three models of material A

3.84%.

The prediction results of the models for materials A, B and C were compared with the actual values of the current period to verify the validity of the combined model. The results were shown in Figures 3.4-3.6. By observing these figures, it could be known that the result of the combined prediction model was closer to the historical demand date, which represented it had good performance for prediction.

The accuracy indexes of different models of A, B and C were compared, and the results could be known in Table 3.9.

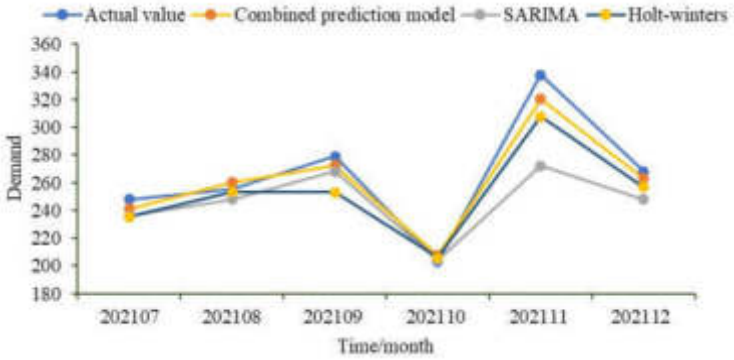


Fig. 3.5: Comparison of prediction results among three models of material B

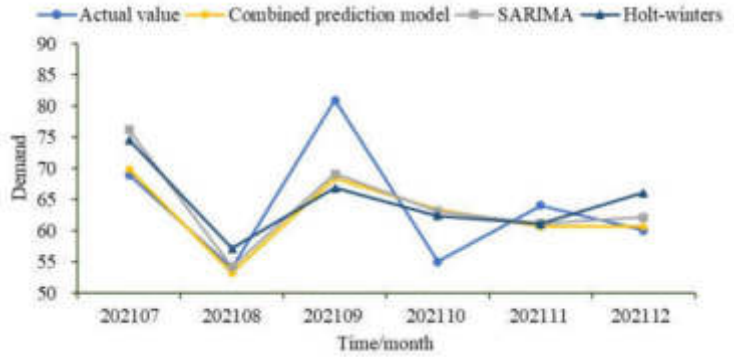


Fig. 3.6: Comparison of prediction results among three models of material C

Table 3.9: Comparison results of precision index among models of materials

Material	Model	MAE	RMSE	MAPE
Material A	Combined prediction model	6.03	7.90	3.84%
	SARIMA model	11.82	12.99	7.20%
	Holt-winters exponential smoothing method	7.35	9.13	4.60%
Material B	Combined prediction model	7.63	8.83	2.75%
	SARIMA model	19.47	29.04	6.49%
	Holt-winters exponential smoothing method	11.19	16.33	4.96%
Material C	Combined prediction model	4.44	6.33	6.64%
	SARIMA model	5.39	6.72	8.02%
	Holt-winters exponential smoothing method	6.53	7.54	9.90%

1. For different materials, the performance of three models of material C was worse than that of material A and B, which might be because the annual growth rate of overall demand for material C from 2019 to 2021 changes greatly. The growth rate of material A from 2019 to 2021 was 4.19%, 6.07% and 13.32%, respectively. Material B was 12.21%, 11.86% and 18.00%, while material C was 1.56%, 15.04% and 8.47%, respectively. Compared with material A and B, the demand growth rate of material C was

more volatile.

2. For different models, the MAPE of the three models was all less than 10%, which proved that the three models had better performance for the prediction of materials demand. Compared with the other two models, the prediction errors of the combined prediction model were reduced by 3.36% and 0.76% for material A, 3.74% and 2.21% for material B, and 1.38% and 3.26% for material C, respectively. It was proved that the overall fitting error of the combined prediction model was smaller for different order of magnitude of the material, and it had higher prediction accuracy.
3. For the prediction of material A, when the predicted values of SARIMA model and Holt-winters exponential smoothing method were directly averaged, the MAPE obtained was 4.60%, which was higher than 3.84% of the combined prediction model. It was proved that the weights of SARIMA model and Holt-winters exponential smoothing method were calculated by BP neural network, compared with direct average, the prediction accuracy can be improved effectively.

4. Conclusion. In this paper, a combination prediction model was proposed by combining BP neural network, Holt-winters exponential smoothing method and SARIMA model, and the effectiveness of the combination prediction model was analyzed by taking the historical data of tobacco raw material demand of Y enterprise as an example. The result grouping method and BP neural network were used to combine the predicted results of two models, and the combined predicted value was closer to the real value by calculating appropriate weight distribution.

1. Experimental results showed that the combined prediction model proposed in this paper had smaller prediction error and better fitting effect for different tobacco materials, and provided certain reference basis for cigarette factory to scientifically arrange production plan, inventory plan and cargo space allocation plan. In addition, the Holt-winters exponential smoothing method of BP neural network and the SARIMA model are often used in sales forecasting, financial forecasting and medical forecasting. Therefore, the combined forecasting model proposed in this paper may have strong applicability in these fields.
2. At the same time, it was found in the research process that the demand for cigarette raw materials in January and February was lower than that in other months due to the shutdown of cigarette factories during the Spring Festival, resulting in a large difference between the demand in January, February and other months. The Spring Festival and other holiday factors enhance the volatility of data. In addition, factors such as policy adjustments and market fluctuations will also have a great impact on the demand for cigarette raw materials, which may increase the prediction error. Therefore, in the subsequent research, factors such as holiday factors and policy adjustment factors could be introduced into the prediction model.

Acknowledgments. Thanks to the great support of the project team and Hohai University. Fundings from the Science and Technology Project (No. HYHH2020XX02) of Hongyun Honghe Tobacco (Group) Co., Ltd are gratefully acknowledged.

REFERENCES

[1] N. Y. Krakauer, M. D. Grossberg, I. Gladkova, and H. Aizenman, *Information content of seasonal forecasts in a changing climate*, Advances in Meteorology, vol. 2013, no. 1, Jan. 2013, pp. 1-12.

[2] S. Anil, K. Makoto, *Nexus between renewable energy certificates and electricity prices in India: evidence from wavelet coherence analysis*, Renewable Energy, Philadelphia, vol. 204, no. 0, Mar. 2023, pp. 836-847.

[3] K. Seungmi, K. Jaehwang, H. Ding, X. Xu, R. Chen, J. Guo and H. Fu, *Using multiple regression analysis to predict directionally solidified TiAl mechanical property*, Journal of Materials Science and Technology, vol. 104, no. 0, Mar. 2022, pp. 285-291.

[4] B. Men, Z. Wu, H. Liu, Z. Hu and Y. Li, *Improved grey prediction method for optimal allocation of water resources: a case study in Beijing in China*, Water Supply, vol. 19, no. 4, Jun. 2019, pp. 1044-1054.

[5] N. S. Muruganandam, U. Arumugam, *Dynamic ensemble multivariate time series forecasting model for PM2.5*, Computer Systems Science and Engineering, vol. 44, no. 2, Jun. 2022, pp. 979-989.

[6] L. Ana, H. P. Javier and M. Manuel, *A combined model based on recurrent neural networks and graph convolutional networks for financial time series forecasting*, Mathematics, vol. 11, no. 2, Jan. 2022, pp. 224-224.

[7] A. S. Ahmar, P. K. Singh, N. V. Thanh, N. V. Tinh and V. M. Hieu, *Prediction of BRIC stock price using ARIMA, SutteARIMA and Holt-Winters*, Computers, Materials and Continua, vol. 70, no. 1, Jan. 2022, pp. 525-534.

[8] O. Mahmod, I. Rachmah, S. A. Abubakar, Q. M. Bagus, and S. Rajalingam, *Model forecasting development for dengue fever incidence in Surabaya city using time series analysis*, Processes, vol. 10, no. 11, Nov. 2022, pp. 2454-2454.
[9] L. Rubio, A. J. Gutirrez-Rodriguez and M. G. Forero, *EBITDA index prediction using exponential smoothing and arima model*, Mathematics, vol. 9, no. 20, Oct. 2021, pp. 2538-2538.
[10] S. P. Kumar, P. A. Kumar, A. Sahil and K. Ravi, *Multiple forecasting approach: a prediction of CO2 emission from the paddy crop in India*, Environmental Science and Pollution Research, vol. 29, no. 17, Nov. 2021, pp. 25461-25472.
[11] P. Bajari, D. Nekipelov, S. P. Ryan, and M. Yang, *Machine learning methods for demand estimation*, The American Economic Review, vol. 105, no. 5, May 2015, pp. 481-485.
[12] H. Wumaier, J. Gao and J. Zhou, *Short-term forecasting method for dynamic traffic flow based on stochastic forest algorithm*, Journal of Intelligent and Fuzzy Systems, vol. 39, no. 2, Jun. 2020, pp. 1-13.
[13] F. O. John, S. Thokozani, *Path loss prediction in tropical regions using machine learning techniques: a case study*, Electronics, vol. 11, no. 17, Aug. 2022, pp. 2711-2711.
[14] C. Tarmanini, N. Sarma, C. Gezegin and O. Ozgonenel, *Short term load forecasting based on ARIMA and ANN approaches*, Energy Reports, vol. 9, no. S3, May 2023, pp. 550-557.
[15] S. Chatterjee, Y. Byun, *A synthetic data generation technique for enhancement of prediction accuracy of electric vehicles demand*, Sensors, vol. 23, no. 2, Jan. 2023, pp. 294-294.
[16] T. H. Noor, A. M. Almars, M. Alwateer, M. Almaliki, I. Gad and E. Atlam, *SARIMA: a seasonal autoregressive integrated moving average model for crime analysis in Saudi Arabia*, Electronics, vol. 11, no. 23, Nov. 2022, pp. 3986-3986.

Edited by: Jingsha He
Special issue on: Efficient Scalable Computing based on IoT and Cloud Computing
Received: May 14, 2024
Accepted: Aug 12, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 552–565, DOI 10.12694/scpe.v26i2.3916

ROBUST IDENTIFICATION ALGORITHM OF NETWORK COMMUNICATION SIGNALS VIA MACHINE LEARNING MODEL

PEIFENG SUN *AND GUANG HU †

Abstract. The efficiency of communication processing and control depends heavily on the recognition of network signals, however irregularities and mistakes frequently arise during the application process. In this work, we leverage machine learning models to automatically identify computer network communication signals, leveraging recent advancements in artificial intelligence technology. In the simulation, we employed a support vector machine (SVM) model, and we utilized parameter optimization to address the overlearning issue. The process of classifying modulation signals involves the extraction of feature parameters through the application of support vector machine and radial basis function neural network (RBFNN) models, respectively. Real-world network communication involves the observation and collection of signals from various viewpoints or feature spaces. These views provide a variety of detailed insights into the signal, and feature extraction is carried out for each view to produce the associated feature vectors. An extensive description of the signal can be generated by extracting the features from several viewpoints. Various viewpoints' feature vectors are combined and synthesized. The robustness of signal recognition can be increased and the bias and inaccuracy that could be generated by a single view can be minimized by combining the data from several perspectives. The support vector machine performs better than the radial basis function neural network, according to experimental findings. When the signal-to-noise ratio (SNR) is high, network communication signals function effectively. However, the latter (RBFNN) performs significantly worse in low SNR settings whilst the former (SVM) retains good accuracy. Therefore, when it comes to computer network communication signals, the support vector machine model is thought to be more reliable.

Key words: Classroom interaction, Data analysis, Compound foreign language, Business English teaching, Communication technology, Data fusion

1. Introduction. This work provides an enhanced approach based on the dual-attention module to compensate for the limitations of the conventional communication recognition mechanism. When using the conventional communication recognition mechanism, in order to accomplish the desired mobilization communication goal, the mobilization recognition instruction must be prepared ahead of time and linked to the appropriate execution system [1]. By computing the similarity weights between the constituents of the input signals, self-attentive mechanisms can recognize the global interdependence of the input signals and produce the output of a weighted sum. This process increases the accuracy of signal detection and aids in locating significant characteristics in the time-series data. The primary purpose of the channel attention mechanism is to draw attention to the relative relevance of the various input signal channels [2]. The channel attention mechanism increases the robustness of signal identification by enhancing the important channels and ignoring or attenuating the unimportant ones by computing the importance weights of each channel. The self-attention mechanism and the channel-attention mechanism work together in the dual-attention module to improve recognition performance in a complimentary manner. The model's ability to capture signal features more thoroughly and improve identification performance is made possible by the combination of the self-attention mechanism, which concentrates on global features, and the channel-attention mechanism, which concentrates on local importance. The input signal in the particular implementation first travels through the feature extraction layer, after which it is processed by the channel attention mechanism and the self-attention mechanism, in that order. Lastly, the processed features are combined and sent into the classifier so that it may be recognized. This model architecture can substantially decrease mistakes and flaws while enhancing the ability to recognize signals in complex situations. In light of the aforementioned background circumstances, it is therefore required to combine the real communication recognition demands in order to create a more solid and systematic recognition approach [3].

*Computer Department,Zhengzhou Preschool Education College,Zhengzhou,Henan 450099, China.(sunpf2009@126.com).
†Computer Department,Zhengzhou Preschool Education College,Zhengzhou,Henan 450099, China.

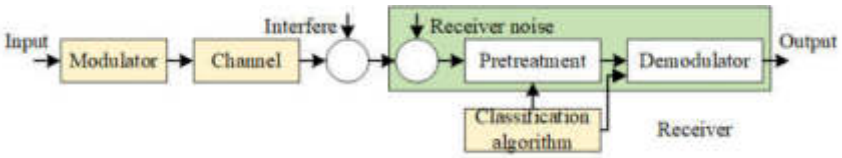


Fig. 1.1: Schematic Diagram of Modulation Identification.

Automatic identification methods for joint modulation via computer network communications are investigated and assessed in this work. It is necessary to process and modify the corresponding identification environment before developing the joint modulation automated identification technology for computer networks. Signal transmission and recognition are generally not too difficult while a computer network is operating, although this is only true in a single network setting [4]. The desired outcomes are frequently not obtained when there is a complicated environment or distinct recognition instructions. Capture identification is the primary mechanism used in traditional automatic identification systems. People are gradually becoming more aware of machine learning technology thanks to the invention of the automatic identification mode.

The two components of a modulation mode classifier's design are algorithm selection and signal preprocessing, as seen in Fig. 1.1. The first link is mainly for signal processing, including down conversion, noise suppression, parameter estimation and channel equalization. The next step is to select a proper classification algorithm, classify the signal after signal preprocessing, and then transmit the intercepted signal and the tag after signal classification to the demodulator to complete the demodulation of the intercepted signal.

Prior knowledge of the signal is typically difficult to come by in an electronic reconnaissance system. Starting from the spectrum, the third-party receiver will pick up the signal of interest and estimate its frequency hopping frequency, symbol rate, and other characteristics. To obtain the IF signal, the third-party receiver applies band-pass filtering and down conversion to the signal based on the characteristic parameters that were acquired in the preceding stage. Next, with a phase-locked loop, the carrier frequency and additional information of the IF signal are retrieved. After that, down conversion is used to retrieve the baseband signal. Then, in order to extract useful tactical information, modulation categorization and demodulation are done. The following study directions pertain to the modulation categorization of non-cooperative communication signals: (1) The different kinds of signal modulation patterns can be recognized using the classification method. Previous study has shown that the more modulation modes the algorithm can recognize, the better; however, this comes at the expense of a decrease in the system's overall recognition rate and an increase in its complexity. Consequently, creating a classification system with high identification performance and the ability to identify various modulation kinds will be a significant advancement.(2) A modulation classification method with a high identification rate in low SNR environments is created with the complicated ECM environment in mind. (3) The modulation classification algorithm's engineering realization. (4) The generalization and fitness of the algorithm. The two primary categories of modulation classification specific methods at the moment are maximum likelihood ratio detection method and statistical pattern identification method. The former establishes the threshold of the likelihood ratio result and thereafter ascertains the signal's modulation mode by calculating the likelihood function of the gathered signal. The latter is dependent on feature extraction.

Preprocessing is used to extract the feature parameters from the recorded signal. The retrieved parameters are then evaluated according to predefined decision criteria to get the pattern classification results. The classifier's tasks include summarizing the input feature vector into an appropriate category in line with a preset rule and completing the mapping of the feature space to the decision space, which yields the final recognition result. In the test phase, the classifier is fed the feature quantity of the test sample to determine the classification outcome; in the training phase, the classifier is trained using the feature quantity of the training sample until it fulfills the preset rules. Tree classifiers, neural network classifiers, support vector machines, and other classifiers based on machine learning algorithms are common classifiers. A hierarchical branch structure resembling a tree is used by the tree classifier. Each layer distinguishes the type of signals according to different parameters, and gradually completes the classification of multiple signals; Neural network

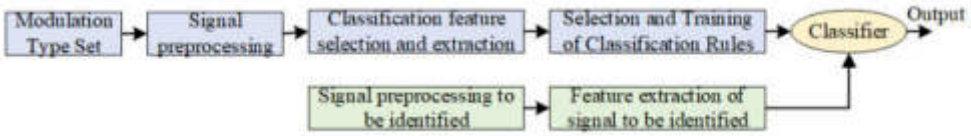


Fig. 2.1: Pattern recognition based on feature extraction.

classifier uses artificial neural network to train and test a variety of signals. Its disadvantage is that it is prone to local extremum, under learning and over learning problems. Support vector machine classifiers recognize and classify signals by constructing hyperplanes, and introduce kernel functions and relaxation variables to improve training error and generalization error. Therefore, it solves the problems of under fitting, over fitting and local extremum, and has good generalization. Nonetheless, the support vector machine classifier's performance will be impacted by its relaxation variables and kernel function, making the optimisation of these two parameters an unavoidable issue.

2. Related works.

2.1. Statistical pattern recognition. According to the concept of statistical models, modulation classification is essentially a pattern classification problem. The overall process is divided into two parts: (1) Training part. Given a set of alternative modulation types, select and extract the characteristic parameters of the training signal to form training samples for training the classifier until the classification output meets certain error requirements, or until an appropriate classifier decision threshold is generated; (2) In the performance test phase, feature parameters are extracted from test signals to form test samples for testing classifiers, and then the output prediction results are analyzed. Classifier, feature parameter extraction, and signal preprocessing are the main elements of the pattern recognition-based modulation classification strategy, as shown in Fig.2.1. A classifier is an important implementation tool that works in tandem with feature selection and extraction to produce a variety of pattern recognition methods. Feature selection and extraction are intermediate links that play a major role in determining the system's performance.

A double mapping from signal space to observation space and back to feature space is how feature extraction is understood. The purpose of the former mapping, which is a part of signal preprocessing, is to extract several signal characterisation parameters; The later mapping, which maps high-dimensional observation space to low-dimensional feature space and lowers computational cost, is the fundamental component of pattern recognition.

2.2. Characteristic parameters and classifier model of network communication signals. For the classification of modulated signals, the instantaneous envelope, phase and frequency included in the signal are the most straightforward characteristics. Among signals of different modulation types, the modulation information of MASK signal is included in the envelope, the modulation information of MPSK signal is included in the phase, and the modulation information of MFSK signal is included in the frequency [5]. Scholars are the pioneers of these researches, and have creatively put forward a lot of modulation signal recognition methods based on the instantaneous characteristic parameters of the signal: The instantaneous amplitude, frequency, phase, and other properties of the signal are extracted using the Hillbert transform. The simulation findings indicate that at 10 dB signal-to-noise ratio, the recognition rate for AM, FM, USB, LSB, VSB, and DSB signals is 91%; subsequently, these signals were classified using neural networks. According to the simulation results, at a signal-to-noise ratio of 10 dB, the recognition rate reached 98%. Signal modulation recognition also frequently makes use of the signal's spectrum, power spectrum, and other properties [6, 7]. Researchers recognise USB, LSB, FSK, and other signals ingeniously by using the spectrum information of signals. According to the simulation results, at 10 dB signal-to-noise ratio, the recognition rate surpasses 95% [8, 9]. By using the power spectrum characteristics of the signal, the scholars simulated the classification of MSK, FSK, FM, BPSK, QPSK, OQPSK and other signals without prior information, and the recognition rate of the simulation results reached more than 95% [10, 11]. By using spectral correlation features, scholars simulated and realized the classification of ASK, FSK, MSK, PSK and QPSK signals. In the simulation results, the recognition rate

reached 97% at 0 dB SNR. The characteristics based on cyclic spectrum analysis are also commonly used in the recognition and classification of digital signals. Cyclic stationarity is very helpful to distinguish between digital communication signals and noise, because digital communication signals have cyclostationarity, while noise is generally non-stationary.

Its spectral correlation function has zero amplitude when the cyclic frequency is not zero, and a large value when the cyclic frequency is zero. Digital communication signals are just the opposite. When the cyclic frequency is zero, the amplitude of the spectral correlation function is zero. Otherwise, it has amplitude. When modulation identification is used to extract parameters, this characteristic greatly helps to lessen the effect of noise on the findings [12, 13]. Scholars simulated the classification of FSK, BPSK, QPSK, MSK and other signals by using cyclic spectrum correlation features. In the simulation results, the recognition rate reached 95% at 10 dB signal-to-noise ratio [14]. Scholars used cyclic spectrum analysis to classify BPSK, QPSK and OQPSK signals. In the simulation results, the recognition rate reached 95% at 8 dB signal-to-noise ratio. Higher order cumulants can characterize the distribution of constellation, and are generally used to distinguish MASK and MPSK signals. The parameters extracted by high-order statistics are robust to phase shift and frequency shift, and can also suppress colored Gaussian noise [15]. Scholars have proposed a digital signal classification method based on cumulants, which can effectively classify MPSK, MQAM and MPAM signals [16]. Scholars use the cumulant based extraction method to identify MPSK signals within a class. The simulation results show that the recognition rate exceeds 95% at 10 dB SNR [17]. Scholars classify MQAM and MPSK signals based on mixed cumulants. In the simulation results, the recognition rate reaches 100% at 10 dB SNR. Some scholars directly extract cyclic statistics from communication signals for modulation classification [18]. Based on cyclic cumulants, scholars have realized effective classification of 4PSK and 16QAM signals. In the simulation results, the recognition rate is close to 100% at 0 dB SNR. Scholars use cyclic cumulants to extract features and recognize SQAM signals with frequency offset. In the simulation results, the recognition rate is close to 90% at 8 dB SNR. Wavelet transform is comparatively appropriate for digital signal processing operations and has good local detail mining ability in both time domain and frequency domain, which may be used to better extract the delicate and instantaneous properties of signals. The scholars adopted Haar wavelet transform to capture the characteristics of instantaneous frequency and phase jump of the signal, and realized the classification of MPSK signal and MFSK signal.

The recognition rate is more than 90% at 10 dB signal-to-noise ratio. Scholars use wavelet packets to decompose IF communication signals, calculate the mean square deviation of the decomposition coefficient vector of the fine part and the approximation part of each layer, use these mean square deviations to form characteristic parameters, and simulate the classification of MASK, MFSK and MPSK signals. According to the simulation results, with a signal to noise ratio of 10 dB, the recognition rate exceeds 95%. Scholars adopted the classification algorithm of discrete wavelet transform combined with neural network, and simulated the recognition and classification of FSK, PSK and QAM signals. In the simulation results, the recognition rate at 5 dB signal-to-noise ratio reached 97%. The wavelet transform-based feature parameters provide good time-frequency domain fine features and operate well in low SNR environments. The constellation diagram is another widely used method for analyzing digital modulation signals. A trustworthy characteristic parameter for distinguishing between MPSK and MQAM signals is a constellation diagram. It usually takes some prior signal understanding for researchers to recreate constellation information. Clustering is the most effective method for reconstructing constellation information. Researchers classified 16QAM, QPSK, and 8PSK signals by reconstructing constellation diagrams from signal symbols that were first introduced with noise using the fuzzy C-means clustering technique.

The simulation results show that the recognition rate exceeds 90% at 5 dB SNR, and exceeds 90% at 0 dB SNR for 8PSK and V.29 signals. However, constellation based methods need to know some prior information or excellent estimation parameters in advance. Therefore, when using constellation based feature extraction methods, excellent preprocessing process and an effective constellation reconstruction method for fuzzy parameters must be adopted. In addition to the features described above, there are other feature extraction algorithms. Scholars used Lempel Ziv complexity method to classify seven kinds of signals. In the simulation results, the recognition rate is 95% at 5 dB signal-to-noise ratio; BPSK, QPSK, CW, BFSK and QFSK signals are classified by information dimension and box dimension analysis. The simulation results show that the recognition rate

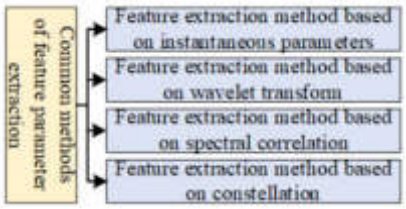


Fig. 3.1: Common Methods of Characteristic Parameters.

exceeds 98% at 5 dB SNR. Scholars used cloud mode parameter extraction method to classify V.29 and 8PSK signals. In the simulation results, the recognition rate reached 95% at 5 dB SNR.

Classifiers are very important; feature extraction and selection are even more important. After completing the mapping from the feature space to the decision space, the classifier's job is to summarize the input feature vector into a suitable category in accordance with a predetermined rule. In order to determine the classification outcome during the test phase, the feature quantity of the test sample is fed into the classifier. In the training phase, the classifier is trained using the feature quantity of the training sample until it meets the preset rules. Common classifiers include support vector machines, neural networks, trees, and other machine learning algorithm-based classifiers. The tree classifier uses a tree like hierarchical branch structure. Each layer distinguishes the type of signals according to different parameters, and gradually completes the classification of multiple signals; Neural network classifier uses artificial neural network to train and test a variety of signals. Its disadvantage is that it is prone to local extremum, under learning and over learning problems. Support vector machine classifiers recognize and classify signals by constructing hyperplanes, and introduce kernel functions and relaxation variables to improve training error and generalization error.

Therefore, it solves the problems of under fitting, over fitting and local extremum, and has good generalization. However, the kernel function and relaxation variables of support vector machine will affect the performance of the classifier, so the optimization of these two parameters is an inevitable problem for support vector machine classifier. Compared with the classification method based on maximum likelihood ratio, the classification method based on feature extraction has the following advantages: (1) The method is relatively simple, which is conducive to engineering implementation; (2) The complexity is small, and it is easy to classify online in real time; (3) The signal classification can be realized without too much prior knowledge. Its disadvantage is that this method will have a certain dependence on the number of training samples. Comprehensive analysis shows that the classification method based on feature extraction has a wider application prospect, especially in the field of electronic reconnaissance and electronic countermeasures. Therefore, the work carried out in this paper is based on pattern recognition methods.

3. Methodology.

3.1. Feature parameter extraction. The different feature extraction techniques for the modulation classification of network communication signals (Fig.3.1).

The first is a feature extraction technique based on immediate parameters. The easiest features to classify modulated signals based on are the signal's immediate envelope, phase, and frequency. The modulation information of the MASK signal is included in the envelope among signals with various modulation types. The MPSK signal's modulation information is contained in the phase, and the MFSK signal's modulation information is contained in the frequency. The second method of feature extraction is wavelet-based. Wavelet transform is very suitable for digital signal processing and may be used to better extract the subtle and instantaneous features of signals because of its strong local detail mining capabilities in both the time and frequency domains. The wavelet transform-based feature parameters provide good time-frequency domain fine features and can perform well in low SNR environments. A technique for extracting features that relies on spectral correlation is the third. Digital signals are frequently recognised and classified using features derived from cyclic spectrum analysis, and cyclostationarity is a valuable tool for differentiating digital communication signals from noise.

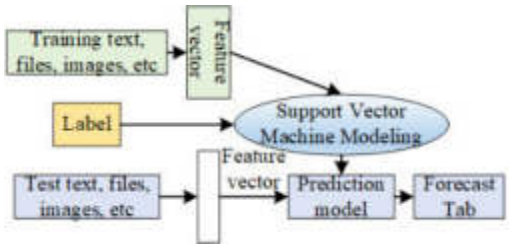


Fig. 3.2: Structural model of support vector machine.

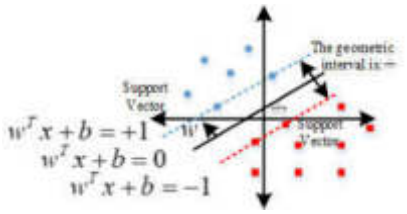


Fig. 3.3: Support Vector Machine Mathematical Model.

Because digital communication signals are cyclostationary, their spectral correlation function has zero amplitude when the cyclic frequency is not zero and a large value when the cyclic frequency is zero, while noise is usually non-stationary. Conversely, digital communication signals have amplitude when the cyclic frequency is non-zero and zero amplitude otherwise. This specific feature is very helpful in reducing the impact of noise on modulation recognition results. The fourth method extracts features using a constellation diagram. The constellation diagram is another widely used method for analyzing digital modulation signals. A trustworthy feature parameter for distinguishing between MPSK and MQAM signals is a constellation diagram. Usually, some prior information of signals is needed to reconstruct constellation information. Clustering is the most effective method to reconstruct constellation information. However, constellation based methods need to know some prior information or excellent estimation parameters in advance. Therefore, when using constellation based feature extraction methods, excellent preprocessing process and an effective constellation reconstruction method for fuzzy parameters must be adopted.

3.2. Classification of signals using support vector machines. Scholars proposed a unique concept in 1995: the Support Vector Machine (SVM). In order to achieve the best generalization ability, SVM uses limited sample information to search for the best compromise between the model's complexity and prediction ability, or, more specifically, between the model's ability to identify arbitrary samples and the accuracy of its training samples. The VC dimension theory and the structural risk minimization concept serve as the foundation for this strategy. Additionally, the SVM has shown good performance in addressing high-dimensional and nonlinear pattern problems due to the incorporation of the relaxation variable and kernel function. These benefits can be extended to other problems, such as function fitting. Its excellent generalisation ability makes it suitable for solving small sample problems (Fig.3.2).

In Fig.3.3, the central gap in the picture is supported by two classification planes, shown by dotted lines. The center hyperplane (black solid line) is the same distance from the two categorization planes. The two categorization planes will have a few "support" points. These "support" points are characterized as support vectors because they provide vivid support for the hyperplane; these support vectors make up a small portion of the sample.

In Fig.3.3, the mathematical definition of hyperplane is:

$$w^T x + b = 0 \tag{3.1}$$

The function interval is:

$$\gamma = y\left(w^T x + b\right) = yf(x) \tag{3.2}$$

The mathematical definition of geometric interval is:

$$\tilde{\gamma} = \frac{y\left(w^T x + b\right)}{||w||} = \frac{|f(x)|}{||w||} \tag{3.3}$$

where x is the eigenvector, y is the label value, and the function interval $y\left(w^T x + b\right) = yf(x)$ is $|f(x)|$ in essence, which is defined as the function interval for convenience; The geometric interval $\tilde{\gamma}$ is an intuitive distance from a point to a hyperplane, which has practical meaning.

3.3. Classification of signals using RBF neural networks. The bionics principle, which imitates brain neurons and learns from the outside environment to gather experience, is the basis of the artificial neural network concept. The experience is then stored in synapses that are connected to the neurons. The role of storing experience and the capacity to conduct computation are formed by artificial neural networks, which arrange basic processing units into a large-scale parallel distributed processing mechanism. Its large-scale parallel distributed architecture and capacity for learning generalization serve as examples of its potent computing capability. Properties and capacities of artificial neural networks include the following:

(1) Nonlinear: One crucial characteristic of artificial neurons, especially when modeling some nonlinear systems, is their ability to be classified as either linear or nonlinear.

(2) Input output mapping: Neural network is a popular algorithm with teacher learning or supervised learning. It uses labeled training samples to continuously update the synaptic weights of each artificial neural unit. By selecting a training sample from the training set each time and sending it to the network for training, the synaptic weight of the network is adjusted continuously according to the difference between the set statistical criteria and the actual response generated by the network operation, until the synaptic weight in the network system reaches a stable state that does not continue to change. Before the weight is stable, the network may receive the training samples that have been traversed throughout the period several times in a different order. Therefore, the neural network's foundation for realising the learning function from the training samples is the creation of input-output mapping.

(3) Adaptability: The artificial neural network will constantly change the synaptic weight value until it adapts to the external environment due to changes in the external environment. When the network is in a constantly changing external environment, the synaptic weight of the unit can change with the change of the environment. If neural network is used for pattern resolution, it can be developed into an adaptive pattern resolution method.

(4) Evidence response: The artificial neural network can select the pattern information in pattern resolution, and also provide the confidence information of the pattern. When there are patterns that cannot be judged, the confidence information obtained can be used to reject and screen out these patterns, thus improving the performance of the neural network (Fig.3.4). Neural unit is the basic unit of network operation and processing, which is composed of three basic links: synapse, adder and activation function:

$$u_k = \sum_{j=1}^{m} w_{ki} x_i \tag{3.4}$$

$$y_k = \Psi\left(u_k + b_k\right) \tag{3.5}$$

Fig.3.5 below depicts the neural network multilayer perceptron model.

3.4. Function of radial basis neural network architecture. Three layers must be designed in order to construct the Radial Basis Function (RBF) neural network depicted in Fig.3.6.

1. The input layer consists of m source nodes, where m is the input feature vector x's dimension.

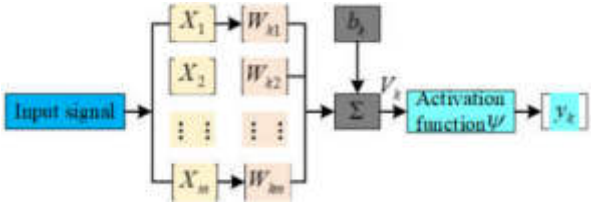


Fig. 3.4: Neuron model.

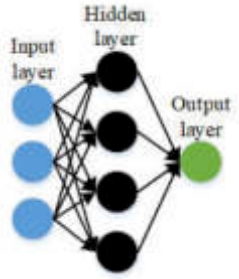


Fig. 3.5: Neural Network Model.

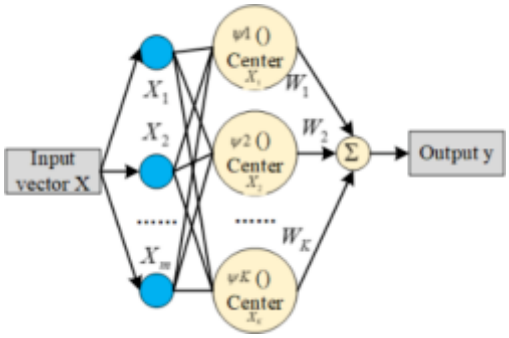


Fig. 3.6: RBF neural network model.

2. Hidden layer: Each of the K calculation units that make up this layer has the radial basis function:

$$\Psi_j(x) = \Psi\left(\|x - x_i\|\right), i = 1, 2, \ldots, K \tag{3.6}$$

The i source node x_i determines the radial basis function's centre. The radial basis function network's source node and hidden node are connected directly and without weight, in contrast to the multi-layer perceptron. The output layer is made up of just one adder unit. Since the Gaussian function is used as the radial basis function in this work, each hidden central layer calculation unit in Fig.3.6 is defined as:

$$\Psi_i(x) = \Psi\left(x - x_i\right) = \exp\left(-\frac{1}{2\sigma_i^2}\|x - x_i\|^2\right), i = 1, 2, \ldots, K \tag{3.7}$$

4. Experiments.

4.1. SVM network performance identification performance analysis. After manually adjusting the kernel parameter g and the penalty factor C, Fig.4.1 displays the simulation prediction performance; Fig.4.2

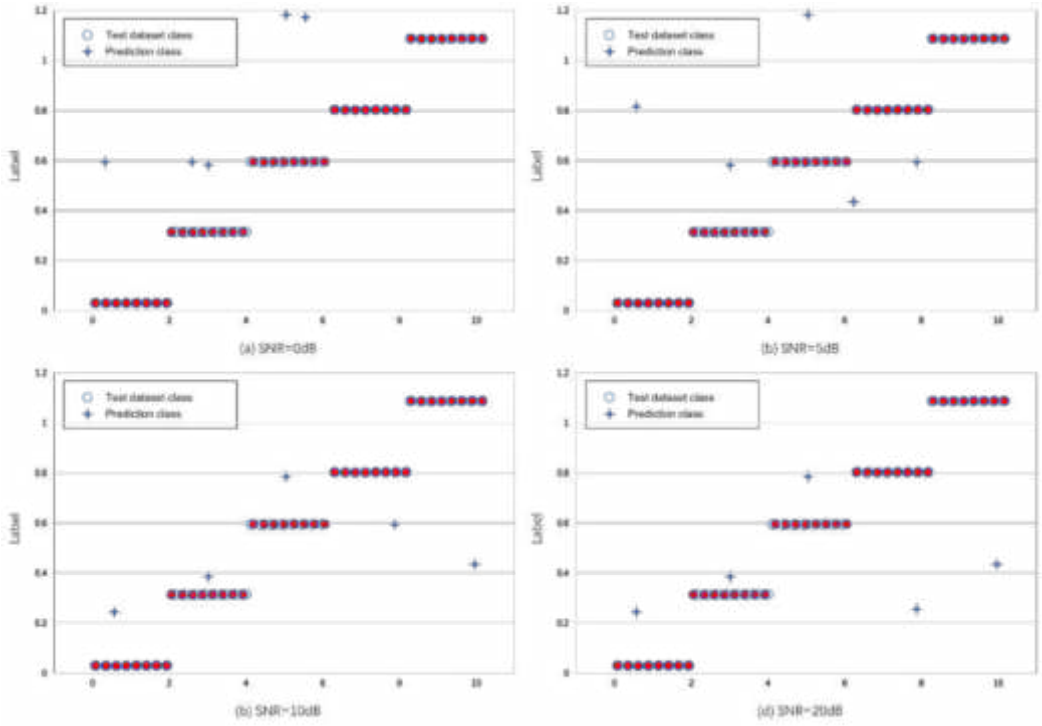


Fig. 4.1: SVM prediction accuracy when parameters are not optimized (SNR=0,5,10,20dB).

Table 4.1: Comparison between the results of manual parameter adjustment (bh) of c and g parameters and the results of genetic algorithm parameter optimization (GA algorithm).

SNR (dB)	0	5	10	15	20
C (bh)	40000	46583	46583	60000	50000
g (bh)	0.001	0.001	0.001	0.001	0.001
Accuracy (bh)	96.65%	97.76%	96.65%	96.65%	98.87%
C (ga)	76.6	93.2	78.1	22.3	22.1
g (ga)	0.054	0.0636	0.01	0.088	0.026
Accuracy (ga)	85.58%	90.00%	93.31%	91.13%	95.58%

shows the simulation prediction effect of C and g parameters after genetic algorithm optimization (when only the first four feature parameters are added).

The value of penalty factor C achieved by manual parameter adjustment is appalling when compared to the simulated performance before and after parameter optimization, and the identification rate essentially stays constant when the signal to noise ratio varies. In this case, the problem of hard spacing (overlearning) arises. The justification for this is that the minimum distance required to exist between every sample point and the categorization plane is known. The unfortunate outcome is that the model's capacity to generalize is reduced as a result of its ease of constraint by a small number of points. The parameters are healthier and the recognition rate is healthier with the trend of the signal to noise ratio after the genetic algorithm optimization, despite the fact that the recognition rate lowers. See Table4.1, Fig.4.3 and Fig.4.4.

When the aforementioned five characteristic parameters are applied, as the simulation results in Fig.4.4 demonstrate, the modulation recognition of nine signals 2ASK, 4ASK, 2FSK, 4FSK, 2PSK, 4PSK, QAM, AM and FM is realized based on support vector machine simulation, and the simulation recognition accuracy rate

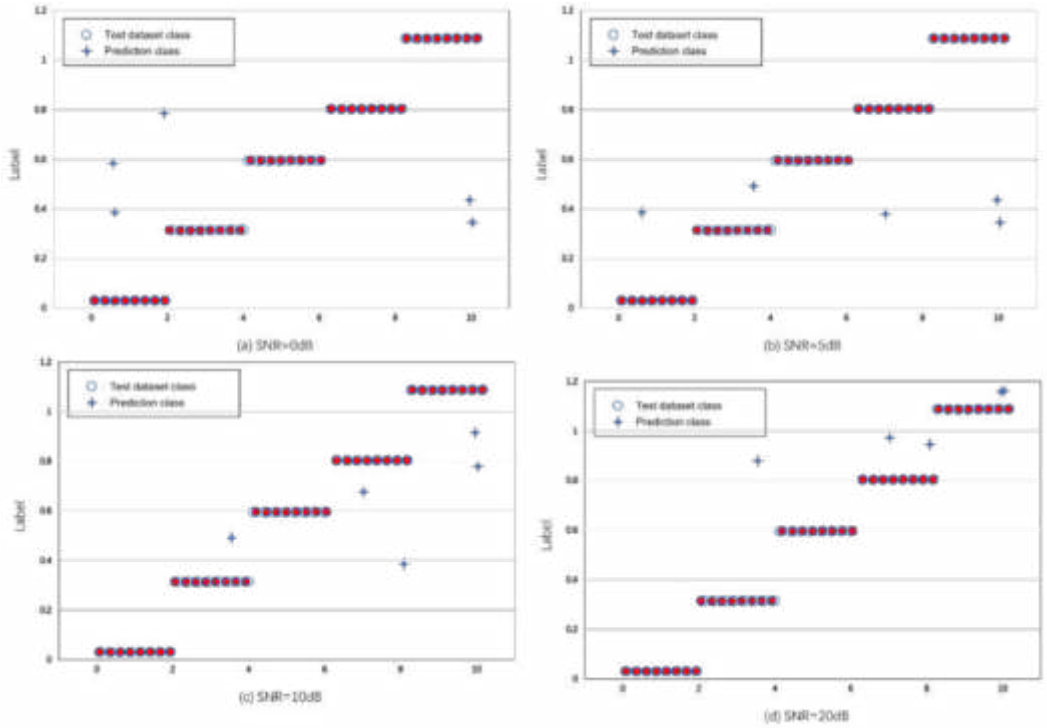


Fig. 4.2: SVM prediction accuracy after parameter ga optimization (snr=0,5,10,20dB).

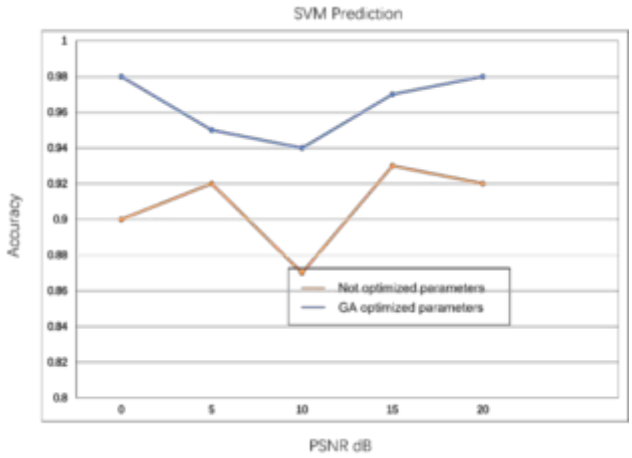


Fig. 4.3: Comparison of Prediction Results before and after Parameter Optimization.

exceeds 80% at - 10 dB signal-to-noise ratio; The simulation accuracy is close to 100% when the signal-to-noise ratio is equal to 0 dB. The simulation results validate the efficiency of the approach presented in this research. Simultaneously, it is discovered that the primary error arises during the identification of MFSK and 4PSK signals at low signal-to-noise ratios.

4.2. RBF network signal performance identification and analysis. According to the simulation results in Table4.2, Fig.4.5 and Fig.4.6, when the feature parameter group obtained in this paper is based on

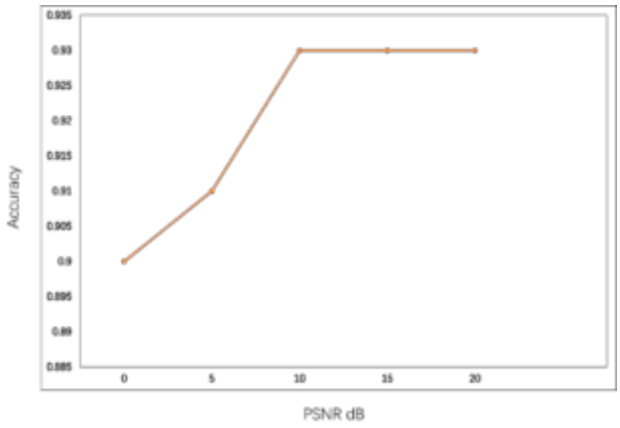


Fig. 4.4: Line Chart of SVM Prediction Results.

Table 4.2: Recognition rate and corresponding parameters of RBF neural network.

Signal-to-noise ratio	-10dB	-7dB	-5dB	-3dB	0dB	5dB	10dB	15dB	20dB
Basis function width	16	4	15	11	12	15	15	16	13
Number of hidden layer neurons	150	125	125	125	125	75	50	50	50
RBF	36.65%	54.42%	71.13%	75.54%	91.13%	100%	100%	100%	100%

Table 4.3: Comparison of Simulation Prediction Data Results of SVM Algorithm and RBF-NN Algorithm.

SNR/Accuracy	-10dB	-7dB	-5dB	-3dB	0dB	5dB	10dB	15dB	20dB
SVM	81.13%	87.76%	87.76%	94.46%	98.87%	100%	100%	100%	100%
RBF	36.65%	54.42%	71.13%	75.54%	91.13%	100%	100%	100%	100%

RBF neural network, it effectively realizes the recognition and classification of AM, FM, 2ASK, 4ASK, 2FSK, 4FSK, 2PSK, 4PSK and QAM signals, and the simulation accuracy reaches 100% at 5 dB SNR; However, the recognition performance is poor at low SNR, and the simulation accuracy is only 71.13% at - 5 dB SNR.

According to the simulation results in Table4.2, Fig.4.5 and Fig.4.6, when the feature parameter group obtained in this paper is based on RBF neural network, it effectively realizes the recognition and classification of AM, FM, 2ASK, 4ASK, 2FSK, 4FSK, 2PSK, 4PSK and QAM signals, and the simulation accuracy reaches 100% at 5 dB SNR; However, the recognition performance is poor at low SNR, and the simulation accuracy is only 71.13% at - 5 dB SNR.

4.3. Comprehensive comparative analysis of algorithms. In this paper, based on the MATLAB simulation platform, the support vector machine algorithm and the radial basis function neural network algorithm are used to classify the modulation signals according to the method of combining the obtained instantaneous parameters, cyclic spectrum analysis parameters and wavelet packet decomposition and reconstruction parameters. The simulation comparison between the two algorithms is as follows. See Table4.3 and Fig.4.7.

Based on the simulation results presented in Table4.3 and Fig.4.7, it is evident that the support vector machine outperforms the RBF neural network in this paper's modulation recognition problem, particularly when the signal to noise ratio is low and the machine still maintains a high accuracy rate. The support vector machine, for instance, can get an appreciable accuracy rate of more than 80% under the - 10 dB signal to noise ratio, whereas the radial basis function neural network rapidly degrades to only 36.65%, leading to significant performance deterioration. The simulation results of the two algorithms for this topic, in the author's opinion,

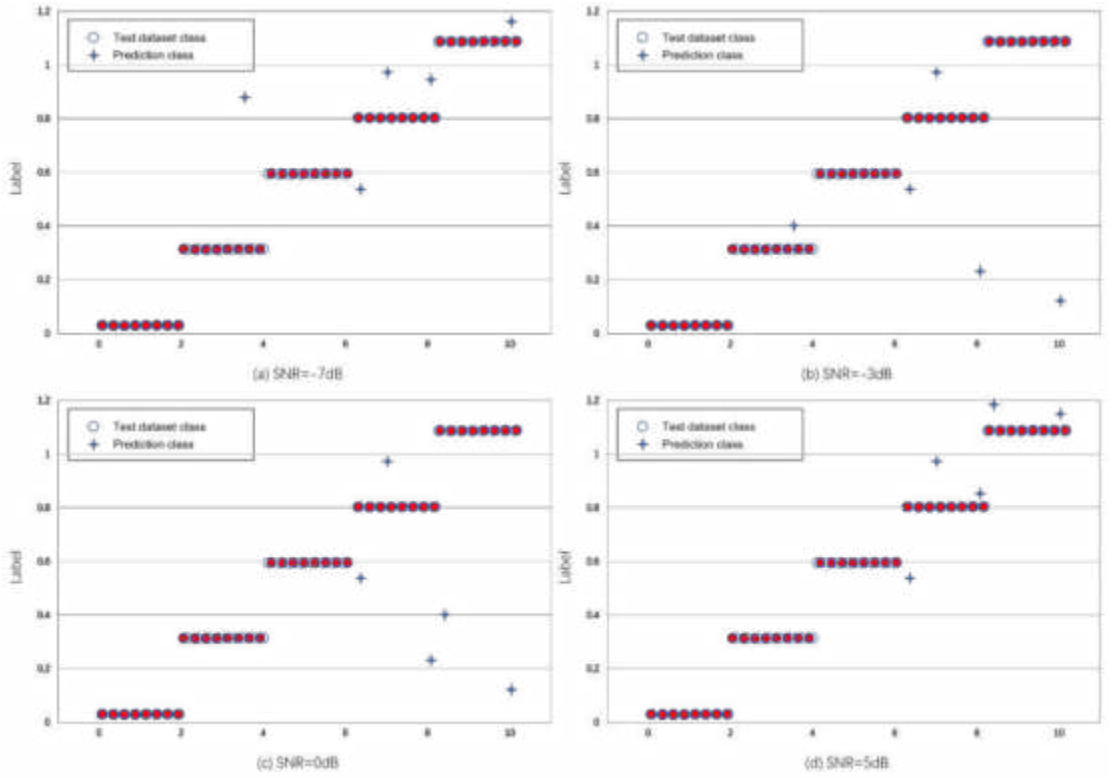


Fig. 4.5: RBF Neural Network Prediction Results.

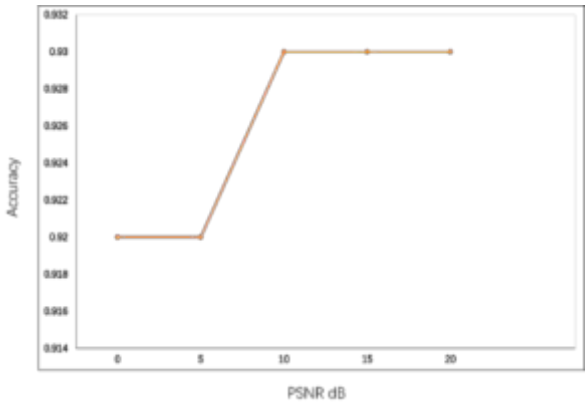


Fig. 4.6: RBF Neural Network Prediction Result Curve.

differ significantly. On the premise that the MATLAB program in this paper does not make mistakes, this difference can also be attributed to the difference in the principles of the two algorithms: RBF neural network is essentially an interpolation approximation idea, an extreme idea, which is too rigid and flexible, and its results are easily affected by the data model and the selected interpolation function; Support vector machine is essentially a compromise idea. The purpose of introducing relaxation variables and penalty factors is to fight for what should be fought for and give up what should be given up. Therefore, at high SNR, the classification

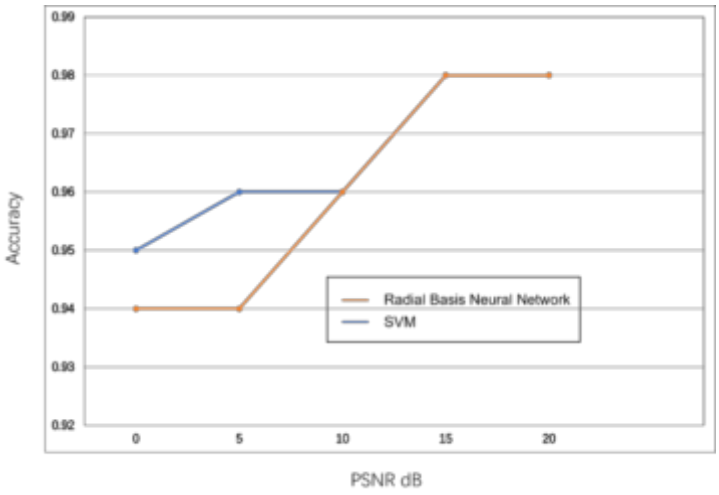


Fig. 4.7: Comparison of broken line trend predicted by SVM algorithm and RBF-NN algorithm.

features are good, and both have good performance; The classification characteristics deteriorate at low signal-to-noise ratios, and the radial basis function neural network is unable to give up a few outliers, making it little and small, while the support vector machine gives up by making the right trade-off in order to improve the overall situation.

5. Conclusion. Between signal detection and demodulation, a crucial technology in non-cooperative communication, is the process of modulation signal recognition. The relaxation variable and kernel function parameters in this research are provided based on the support vector machine's modulation recognition. After that, support vector machines are employed for signal classification, with the feature parameter group extracted in this study serving as the basis. In comparison to the same kind, the year-over-year recognition rate increases to 100% at a signal-to-noise ratio of 5 dB. Support vector machines perform better than radial basis function neural networks, especially when there is a low signal to noise ratio and the machine still maintains a high accuracy. Therefore, we verify that SVM in machine learning model can recognize network communication and has strong robustness.

Data Availability. The experimental data used to support the findings of this study are available from the corresponding author upon request.

REFERENCES

[1] Alsubari, S. N., Deshmukh, S. N., Alqarni, A. A., Alsharif, N., H., T. et al. *Data Analytics for the Identification of Fake Reviews Using Supervised Learning*. CMC-Computers, Materials & Continua, 70(2), (2022)3189–3204.
[2] Zhong, J. *Communication network array signal synchronous transmission method based on Gaussian fuzzy algorithm*. Wireless Networks, 28(5),(2022), 2289-2298.
[3] Verkhivker, G. M., & Di Paola, L. *Dynamic network modeling of allosteric interactions and communication pathways in the SARS-CoV-2 spike trimer mutants: Differential modulation of conformational landscapes and signal transmission via cascades of regulatory switches*. The Journal of Physical Chemistry B, 125(3), 850-873.
[4] Vézquez-Rodríguez, B., Liu, Z. Q., Hagmann, P., & Misic, B. *Signal propagation via cortical hierarchies*. Network Neuroscience, 4(4),(2020) 1072-1090.
[5] Zhao, C., Liu, X., Zhong, S., Shi, K., Liao, D., & Zhong, Q. *Secure consensus of multi-agent systems with redundant signal and communication interference via distributed dynamic event-triggered contro*l. ISA transactions,(2021) 112, 89-98.
[6] Zhou, Y., & Jiao, X. *Intelligent analysis system for signal processing tasks based on LSTM recurrent neural network algorithm*. Neural Computing and Applications, 34(15),(2022) 12257-12269.
[7] Toro, U. S., ElHalawany, B. M., Wong, A. B., Wang, L., & Wu, K. *Machine-learning-assisted signal detection in ambient backscatter communication networks*. IEEE Network, 35(6), 120-125.

[8] Rahman, M. L., Zhang, J. A., Huang, X., Guo, Y. J., & Lu, Z. *Joint communication and radar sensing in 5G mobile network by compressive sensing*. IET Communications, 14(22),(2020) 3977-3988.
[9] Sendak, M. P., Gao, M., Brajer, N., & Balu, S. *Presenting machine learning model information to clinical end users with model facts labels*. NPJ digital medicine, 3(1),(2020) 1-4.
[10] Allugunti, V. R. *A machine learning model for skin disease classification using convolution neural network*. International Journal of Computing, Programming and Database Management, 3(1),(2022) 141-147.
[11] Toğaçar, M., Ergen, B., Cömert, Z., & Özyurt, F. *A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models*. Irbm, 41(4), 212-222.
[12] C. Zhang, M. Li and D. Wu, *"Federated Multidomain Learning With Graph Ensemble Autoencoder GMM for Emotion Recognition,"* in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 7, pp. 7631-7641, July 2023, doi: 10.1109/TITS.2022.3203800.
[13] Tran, M. Q., Amer, M., Abdelaziz, A. Y., Dai, H. J., Liu, M. K., & Elsisi, M. *Robust fault recognition and correction scheme for induction motors using an effective IoT with deep learning approach*. Measurement, 207, 112398.
[14] Fan, J., Wu, L., Zhang, J., Dong, J., Wen, Z., & Zhang, Z.*Deep Learning-Aided Modulation Recognition for Non-Orthogonal Signals*. Sensors, 23(11),(2023) 5234.
[15] Awajan, A. *A novel deep learning-based intrusion detection system for IOT networks*. Computers, 12(2),(2023) 34.
[16] An, T. T., & Lee, B. M. *Robust Automatic Modulation Classification in Low Signal to Noise Ratio*. IEEE Access, 11, 7860-7872.
[17] Zhang, Y., Peng, Y., Sun, J., Gui, G., Lin, Y., & Mao, S. *GPU-Free Specific Emitter Identification Using Signal Feature Embedded Broad Learning*. IEEE Internet of Things Journal.
[18] Ali, Jehad, Rutvij H. Jhaveri, Mohannad Alswailim, and Byeong-hee Roh.*"ESCALB: An effective slave controller allocation-based load balancing scheme for multi-domain SDN-enabled-IoT networks."* Journal of King Saud University-Computer and Information Sciences 35, no. 6 (2023): 101566.

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/Emerging Wireless Networks
Received: May 13, 2024
Accepted: Aug 8, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 566–576, DOI 10.12694/scpe.v26i2.3917

MOBILE LEARNING AND RESOURCE SHARING MODE OF HIGHER EDUCATION BASED ON 5G MOBILE COMMUNICATION TECHNOLOGY

XINCHANG LI *AND YUXIN GUO †

Abstract. Currently, the development of mobile learning and resource sharing models for higher education, along with the application of 5G and other mobile communication technologies to education, hold significant scientific value. These models will help improve the speed and efficiency of retrieving learning resources for higher education while also organizing and managing learning resources more effectively. The complementarity and growth of mobile learning technology for traditional online learning technology are the major features that this research combines. This research builds a mobile learning and resource sharing mode for higher education based on 5G mobile communication technology, utilizing the Moodle online teaching platform. It also designs and implements a mobile learning model, as well as the system architecture, functional modules, learning mode, and learning process. Ultimately, the system model's primary functional modules are put into practice and put through testing. encouraging mobile learning to assist in a variety of ways in the teaching field. The test probability value P (siq)=0.00, which is significantly less than the significant level of 0.05, and the overall accuracy rate of 74.7% indicate that the AKAZE algorithm is utilized to optimize the model, according to the experimental results. As a result, we think that using a mobile learning instructional design mode will benefit students' academic achievement.

Key words: 5G mobile communication technology; Higher Education; Mobile learning; resource sharing

1. Introduction. The widespread application of mobile devices in education can personalize learning, enrich classroom content, and improve student performance [1]. Building an effective learning resource management platform is crucial for efficient mobile teaching [2]. However, current platforms face issues such as resource redundancy, low retrieval precision, outdated content, and simplistic storage formats [3, 4]. Additionally, traditional concepts and technologies hinder effective resource usage and sharing, impacting the effectiveness of education informatization.

With increasing international competition, lifelong learning has been embraced globally, forming a systematic approach to continuous education from childhood through old age, including school education and in-service training [5]. In China, rapid economic transformation necessitates continual skill improvement among workers, making both formal education and in-service training essential [6]. Mobile learning supports lifelong learning by promoting independent and collaborative learning, reducing resistance to formal learning models, and encouraging continuous engagement and self-improvement [7]. However, maximizing the benefits of mobile learning requires a robust foundational system for mobile learning and resource sharing. Teachers, burdened with heavy teaching loads, often lack the time and programming knowledge needed to develop custom learning systems [8, 9]. Existing commercially developed systems are often costly, inflexible, and unsuitable for open and independent learning management. Thus, constructing a mobile learning and resource sharing model based on 5G technology for higher education has significant practical value. It can free teachers from complex platform operations, allowing them to design mobile learning courses tailored to their teaching methods and experiences, thereby promoting mobile learning in education [10, 11].

2. Model of Mobile Learning and Resource Sharing in Higher Education. The teaching process is complex, and each link is crucial, so any factor in the teaching design process cannot be ignored. From the overall perspective, the teaching design model consists of four basic elements: teaching objects, teaching objectives, teaching strategies and teaching evaluation. Its simplified model is shown in Fig. 2.1.

In fact, the learning mode of mobile learning is "personal, seamless, spontaneous, anytime, anywhere"[12].

*Zhengzhou Preschool Education College; Zengzhou Henan 45000 China.(18638759389@163.com).
†Zhengzhou Preschool Education College; Zengzhou Henan 45000 China.

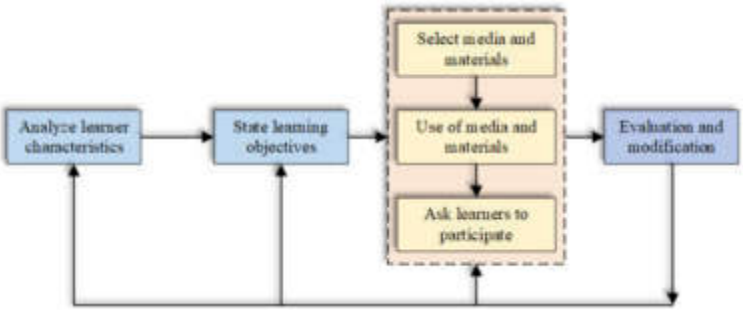


Fig. 2.1: Simple Design Model of ASSURE Mobile Teaching.

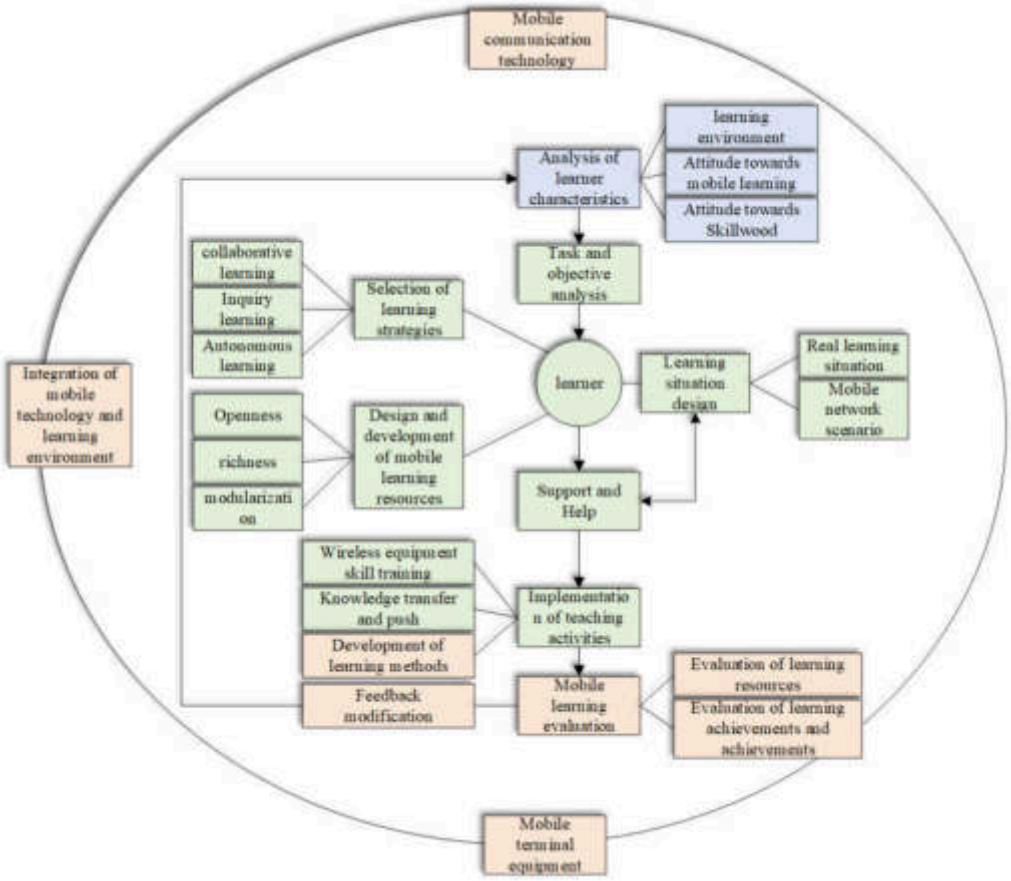


Fig. 2.2: Teaching Design Mode of Mobile Learning.

It focuses on interaction with external learning environment. It is applied to traditional education, online learning, enterprise training and other aspects. To better play a role in mobile learning, improve teaching efficiency and quality, and ultimately achieve good learning results, good teaching design is essential, as shown in Fig. 2.2.

In the process of mobile teaching, the process of resource sharing design and development is generally from front to back: preliminary analysis, resource design, resource development, resource implementation and

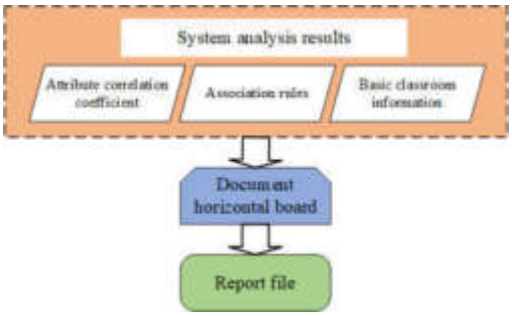


Fig. 2.3: Resource sharing design mode in mobile learning process.

resource evaluation, as shown in Fig. 2.3.

In mobile learning, time (t), space (S), content (C), technology (LE), mental factors (MM), and learning methods (M) all play crucial roles. Traditional learning time is often discontinuous, whereas mobile learning time is continuous. Traditional learning space is generally fixed, while mobile learning space is flexible and can include both physical and virtual environments. Mobile learning can structure courses to fit personal learning needs.

In this context, technology (LE) encompasses the network and technical devices enabling mobile learning. Psychological factors (MM) include learners' interests, motivation, and abilities. Learning methods (M) integrate all parameters related to the transmission of learning content, such as technology and pedagogy.Thus, the factors representing mobile learning under 5G communication technology can be summarized as follows:

$$Mlearn = f\{t, s, LE, c, IT, MM, m\} \tag{2.1}$$

where: t is the time (continuous in mobile learning), S is the space (unrestricted), C is the content (structured for personal learning), LE is the learning environment/technology, MM are the mental factors (interests, motivation, abilities), M are the learning methods.

From this, the representation factors of mobile learning under 5G communication technology can be formulated as:

$$c = f\{MM, soc, edu\} \tag{2.2}$$

where: MM are the mental factors (learners' interests, motivation, and background), Soc are the socially responsible factors, Edu is the educational relevance.

To further characterize the model flow of mobile learning and resource sharing, we use the anisotropy of image brightness diffusion to build the scale space, which needs to be solved by partial differential equations. The nonlinear diffusion equation can usually be expressed as:

$$\partial L/\partial t = div(c(x, y, t).\nabla L) \tag{2.3}$$

where L is the image brightness matrix, time t is the scale parameter, div represents the image divergence, (x, y, t) represents the conduction function, and ∇ represents the gradient calculation. The conduction function structure can be expressed as Eq. 2.4:

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|) \tag{2.4}$$

Then, we use Gaussian smoothing to process the gradient value of the post image L_σ . The form of function g can be expressed as Eq. 2.5-Eq. 2.6:

$$g_1 = \exp(-\frac{|\nabla L_\sigma|^2}{\lambda^2}) \tag{2.5}$$

Table 3.1: User Registration Information.

Serial No	Field Name	data type	length	Primary key	Is it empty	Field Description
1	name	varchar	10	Yes	Nonempty	Student Name
2	password	varchar	20	No	Nonempty	Login password
3	number	varchar	20	No	Nonempty	Student ID
4	phone	varchar	13	No	Can be empty	phone number

$$g_3 = \begin{cases} 1, |\nabla L_\sigma|^2 = 0 \\ 1 - \exp\left[-\frac{3.315}{\left(\frac{|\nabla L_\sigma|}{\lambda}\right)^8}\right], |\nabla L_\sigma|^2 > 0 \end{cases} \tag{2.6}$$

Among them, λ is a contrast factor used to control the diffusion degree, which can determine the integrity of edge information retained in the sampling process. The larger λ is, the less edge information retained in the mobile learning model.

The basic idea of additive operator splitting diffusion is to decompose a complex multidimensional problem into several simple one-dimensional problems, and then solve them separately to take the average value[13]. Compared with the traditional explicit solution, this method can adopt a larger step size for iteration, with faster convergence and higher stability. First, the original equation is discretized into an implicit difference scheme, as shown in Eq. 2.7:

$$\frac{L^{i+1} - L^i}{\tau} = \sum_{l=1}^{m} A_l \left(L^i\right) L^{i+1} \tag{2.7}$$

where τ is the iteration step size, and A_l is the matrix of the diffusion degree of the feature image on the scale l. The solution of this equation can be expressed as Eq. 2.8:

$$L^{i+1} = \left(I - \tau \sum\nolimits_{l=1}^{m} A_l \left(L^i\right)\right)^{-1} L_i \tag{2.8}$$

This manuscript takes Eq. 2.8 as the basic mathematical model to build a mobile learning and resource sharing model for higher education based on 5G mobile communication technology.

3. Methods.

3.1. Database structure design. MySQL database provides encryption connection with other multiple databases, and can also perform data batch processing. In this study, five data tables of mobile learning and resource sharing models are created, and the Chinese description of each field is given:

The first part is the user registration information table. As shown in Table3.1. This form is used to store student registration information, including student name, login password, student ID and mobile phone number. The student's name and login password are used for daily login verification. To protect student privacy, mobile phone number can be left blank.

The second part is the login information table, as shown in Table3.2. This table is used to record students' mobile learning application login. When students log in to the student client each time, they will automatically send the login time and student name to the database server.

The third part is the mobile learning progress record table, as shown in Table3.3. This table is used to record the learning progress of students at the mobile learning site, including student name, learning time, learning location, and learning duration. After the data information is stored, it is convenient for the teaching assistant to monitor the learning progress of each student in real time and supervise the learning externally.

The fourth part is the question information table, as shown in Table3.4. This table is used to record the information of students' questions, including their names, time and content.

Table 3.2: User Registration Information.

Serial No	Field Name	data type	length	Primary key	Is it empty	Field Description
1	login Time	varchar	20	no	Nonempty	login time
2	Login Name	varchar	10	Yes	Nonempty	Student Name
3	number	varchar	20	No	Nonempty	Student ID
4	phone	varchar	13	No	Can be empty	phone number

Table 3.3: Mobile Learning Progress Record.

Serial No	Field Name	data type	length	Primary key	Is it empty	Field Description
1	name	varchar	10	Yes	Nonempty	Student Name
2	time	varchar	20	No	Nonempty	Time of learning
3	title	varchar	100	No	Nonempty	Learning location
4	Study Time	int	10	No	Nonempty	Learning duration

Table 3.4: Mobile Learning Progress Record.

Serial No	Field Name	data type	length	Primary key	Is it empty	Field Description
1	name	varchar	10	Yes	Nonempty	Ask students' names
2	time	varchar	20	No	Nonempty	Question time
3	content	varchar	200	No	Nonempty Questions	
4	Study Time	int	10	No	Nonempty	Learning duration

Table 3.5: Reply to Questions.

Serial No	Field Name	data type	length	Primary key	Is it empty	Field Description
1	name	varchar	10	Yes	Nonempty	Ask students' names
2	Post time	varchar	20	No	Nonempty	Question time
3	time	varchar	20	No	Nonempty	Reply time
4	content	varchar	800	No	Nonempty	Reply content

The fifth part is the question reply information table, as shown in Table3.5. This table is used to record the information that the assistant teacher replies to the student's questions in the mobile learning situation. The assistant teacher makes targeted replies to the student's questions and stores them in the database server. The student client queries the database server's question reply to information table after each login. If the assistant teacher replies to the student's questions, download the reply to information.

Regarding data sources, the research object for this study was university mathematics class 1301 students. After a period of observation and consideration of the students' computer scores at the conclusion of the previous semester, the students were classified into two groups: Group A, which demonstrated strong computer acceptance ability and excellent computer performance, and Group B, which demonstrated moderate computer acceptance ability. To assess the simulation and development of the mobile learning model, however, and to guarantee the representativeness of the experimental results, 16 students from A and B, respectively, were chosen for the experimental group and the control group. Moreover, these students had not yet learned EXCEL2010 Chart Creation and Editing[14].

3.2. Model operation and optimization. SQLite, the underlying optimization algorithm of the model, is used in the application layer, application framework layer, system runtime, and Linux kernel layer of the mobile device operating system, as shown in Fig. 3.1. SQLite is integrated in the system runtime and runs on the Linux kernel layer.

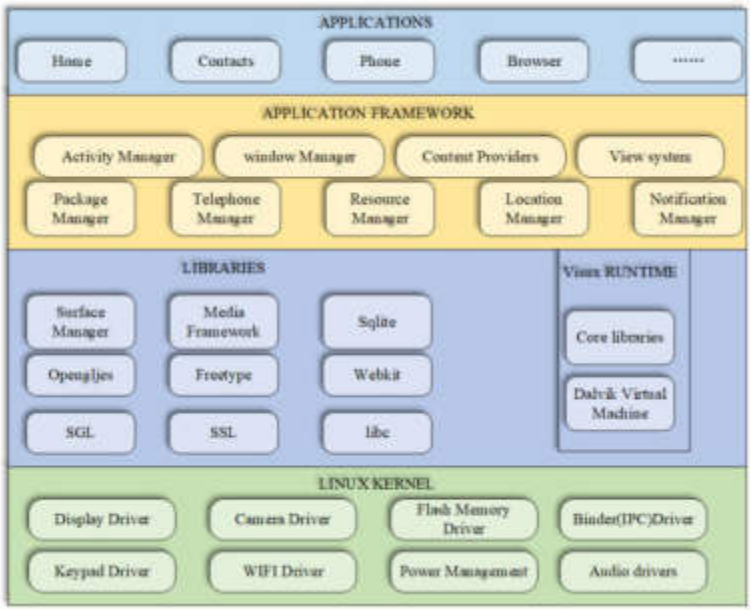


Fig. 3.1: Architecture of mobile device operating system.

Table 4.1: Reply to Questions.

-	Group	number	Average	standard deviation	Mean value of standard error
fraction	1	32	91.07	4.691	0.829
fraction	2	32	84.29	9.093	1.608

Since the data on the mobile learning mobile client needs not only to be uploaded to the database server but also stored in the local database, the tables existing on the database server also exist in the local database. The author will not repeat the structure of each table here. However, MySQL databases can create data tables through the Navi cat for MySQL visual interface without writing code. SQLite needs to use code to penetrate data tables.

4. Case study.

4.1. Empirical Results Output and Hypothesis Verification. Learners who choose mobile learning should complete the specified tasks within a specified class time (tasks must be submitted within the specified time), change the file name of the completed tasks to "student number name", and then submit the tasks to the server[15]. The author logs in to the server to download the task files submitted by all learners. After collecting the final learning task results of all learners, the author scores all the learners in turn according to the scoring rules and enters the learner's scores into the score sheet in turn, and finally collects the student scores of the experimental group and the control group. After the students in the experimental group finish their study, they should evaluate the recognition of the teaching design model constructed through the self-made questionnaire of this study. We compared the scores of the control group and the experimental group as samples and conducted independent sample T test through SPSS software. The test results are shown in Table4.1.

After students submit their homework, they use the mobile learning platform to send the teaching design model identification questionnaire shown in Table4.2 to the experimental object group, and the experimental object fills in the questionnaire and sends it back to the mobile learning platform. Collect, sort out and analyze the results fed back by the experimental subjects, as shown in Fig. 4.1: 68.75% of the experimental subjects are very interested in using WeChat to carry out "computer based" mobile learning activities; 56.26% of the

Table 4.2: Mobile Learning Model Recognition Questionnaire.

Title No	Sub item	strongly agree	agree with	uncertain	disagree	strongly disagree
1	I am very interested in using WeChat to carry out "Computer Foundation" mobile learning activities	-	-	-	-	-
2	Using WeChat to carry out "Computer Foundation" mobile learning activities has brought convenience to my study	-	-	-	-	-
3	I think the combination of mobile network situation and real situation in teaching design makes my computer learning easier, more convenient, and more effective	-	-	-	-	-
4	I think this teaching design mode has strengthened the communication between me and other learners	-	-	-	-	-
5	I think the activity design of teaching design strengthens the final learning effect	-	-	-	-	-

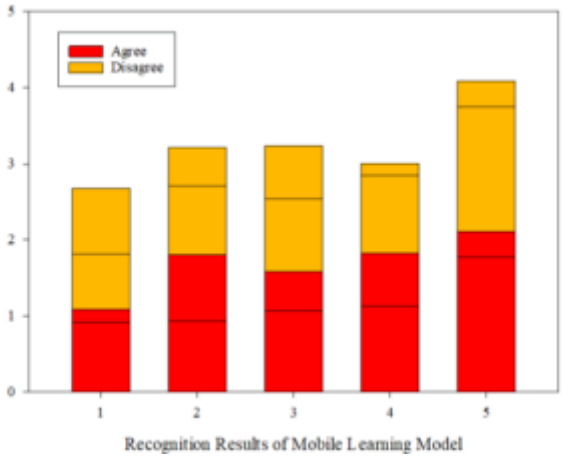


Fig. 4.1: Recognition Results of Mobile Learning Model.

subjects believed that mobile learning had brought convenience to learning; 72% of the subjects believed that instructional design made learning easier, more convenient and more effective; 85% of the subjects believed that the instructional design model strengthened the communication with other learners; 56.25% of the subjects believed that instructional design strengthened the final learning effect. The above data shows that the instructional design can effectively strengthen the communication between learners, make learning easier, more convenient, and effective, and strengthen the final learning effect.

4.2. Algorithm performance comparison. The AKAZE (Accelerated KAZE) algorithm is a powerful tool for identifying and describing features in images. It focuses on generating binary descriptors and multi-scale feature detection to optimize the image processing model.AKAZE finds feature spots in a picture by applying Nonlinear Diffusion Filtering. With the help of this filter, the image can be smoothed across many scales without losing important edge characteristics.

To better compare and analyze the performance difference between AKAZE optimization algorithm and other feature detection algorithms in the construction of mobile learning models, three algorithms are used to match the image features of models in the mobile learning and resource sharing experimental group and the control group. Mularczyk mobile learning database sets variables for factors affecting learning image features,

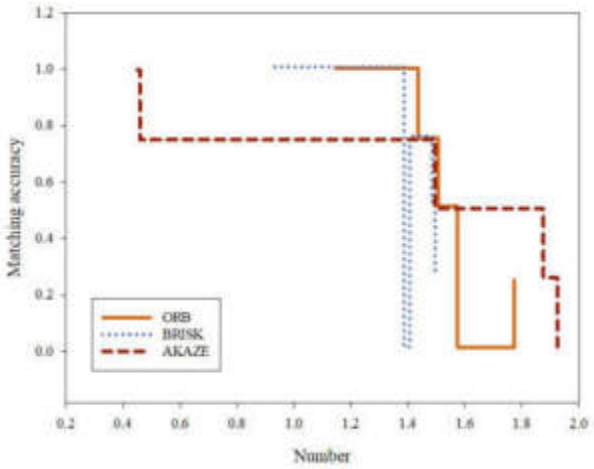


Fig. 4.2: Optimization Results of Mobile Learning Experimental Group.

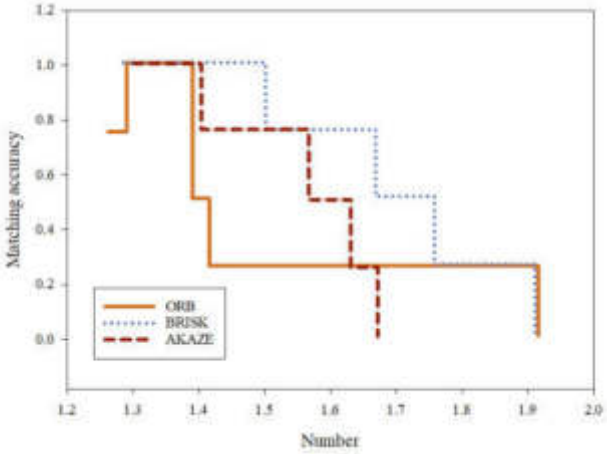


Fig. 4.3: Optimization results of mobile learning control group.

such as image blur, compression, illumination intensity, angle of view, rotation, and scaling. Five groups of images are selected, and each group of images takes 6 variable values for a certain influencing factor, keeping consistent with other factor variables. The first group, Boat learning group, is composed of six images with different rotation and scaling degrees; The second group, Leuven learning group, was composed of six images with different brightness; The third group, Ubc learning group, had six images with different compression levels; The fourth group, Bikes learning group, was composed of six images with different degrees of blur; The fifth group, Graf learning group, is composed of six images with different perspectives. The first image of each group of images is a preset image, and the other five images are matching images. The matching situation and preset situation are matched with three algorithms respectively, and the matching time and accuracy are compared. The learning situation matching process uses ratio ratio scheme for calculation and screening. Since ORB algorithm, BRISK algorithm descriptor and AKAZE algorithm M-LDB descriptor are all stored in binary form, hamming distance is used for rough matching during feature point matching, and possible feature point pairs are selected and filtered and eliminated by RANSAC algorithm, as shown in Fig. 4.2, Fig. 4.3, Fig. 4.4 and Fig. 4.5.

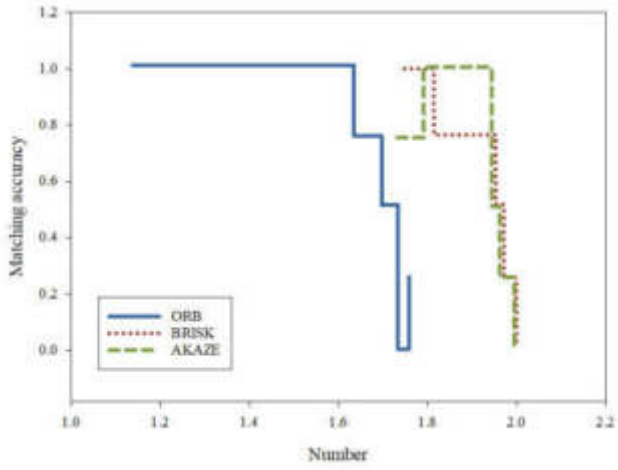


Fig. 4.4: Optimization results of mobile learning control group.

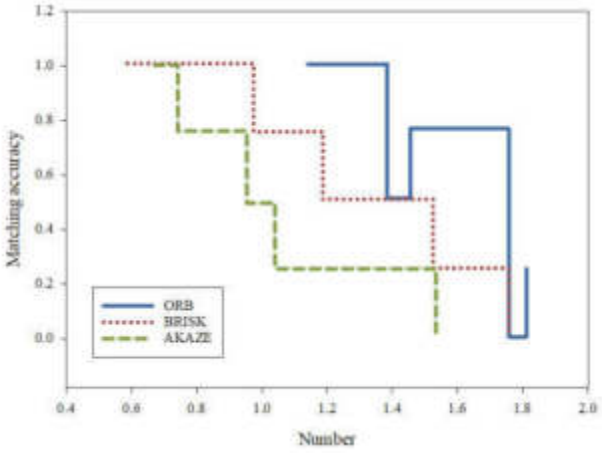


Fig. 4.5: Optimization results of mobile learning control group.

Fig. 4.2 shows the matching accuracy of the three algorithms under different rotation scaling angles. The average matching accuracy of ORB algorithm is 61.2043%, that of BRISK algorithm is 66.5227%, and that of AKAZE algorithm is 74.6991%. ORB algorithm can deal with small rotation and scaling of images well, but it cannot deal with large rotation and scaling of images. BRISK algorithm is stable, and AKAZE algorithm is the most stable.

Fig. 4.3 shows the performance analysis and comparison data of ORB, BRISK and AKAZE algorithms from the two dimensions of matching accuracy and matching time. The ordinate represents the matching accuracy, in percentage form, and the abscissa represents the matching image number in each group of images.

Fig. 4.4 shows the matching accuracy of the three algorithms under different learning durations. The learning durations of matched images decrease in turn. The average matching accuracy of ORB algorithm is 73.1271%, that of BRISK algorithm is 76.6142%, and that of AKAZE algorithm is 81.8454%. The three algorithms are similar to each other in that the image is affected by the change of learning time, but the matching accuracy of AKAZE algorithm is higher than that of the other two algorithms.

Fig. 4.5 shows the matching accuracy of the three algorithms under different fuzzy degrees. The average

matching accuracy of AKAZE algorithm is 76.6311%. Both ORB algorithm and BRISK algorithm cannot deal with the situation of blurred images very well. Especially in the matching process of the mobile learning experiment group, the correct rate of using ORB algorithm to match is only 29.2035%, while AKAZE algorithm has better relative performance.

This research combines three algorithms to match the model picture features of the experimental and control groups, allowing for a more thorough comparison and analysis of the performance difference between the AKAZE optimization algorithm and other feature detection algorithms in mobile learning model development.The factors that alter the properties of the taught images—such as picture blurring, compression, light intensity, viewing angle, rotation, and scaling—are configured in the Mularczyk Mobile Learning Database. Five sets of photos were chosen, and for each set of images, six variable values were taken for a specific influencing factor while maintaining consistency with the other factor variables.

5. Conclusion. Based on the development of 5G and other communication technologies, a model of mobile learning and resource sharing in higher education was created in this study. The model was then empirically verified and its algorithmic efficiency was examined. The test probability value P(siq) = 0.00, which is much smaller than the significant level of 0.05, and the overall accuracy rate of 74.7% are obtained after optimizing the model using the AKAZE algorithm, according to the experimental data. As a result, we think that using the m-learning and resource sharing instructional design model can help students achieve better academically. It is particularly crucial to clarify that the test probability value P (siq) = 0.00 denotes a very significant result, meaning that the likelihood of the obtained result in the statistical test is near to zero. This suggests that the model's performance was significantly improved by using the AKAZE algorithm for optimization, and it is highly unlikely that this gain was caused by chance. This results provides more proof of the mobile learning and resource sharing model's dependability and efficacy in raising student achievement. In conclusion, this work not only offers a theoretically sound model of resource sharing and mobile learning, but also conducts tests to validate the model's applicability and viability. Subsequent investigations may refine the model even more to enhance precision and investigate additional communication technology uses in mobile education.

Data Availability. The experimental data used to support the findings of this study are available from the corresponding author upon request.

REFERENCES

[1] Jingchun Zhou, Dehuan Zhang, Wenqi Ren, Zhang Weishi.*Auto Color Correction of Underwater Images Utilizing Depth Information*, vol. 19, pp. 1-5, 2022, IEEE Geoscience and Remote Sensing Letters. doi: 10.1109/LGRS.2022.3170702.

[2] Chen, S., Liang, Y. C., Sun, S., Kang, S., Cheng, W., & Peng, M. *Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed*. IEEE Wireless Communications, 27(2),(2020) 218-228.

[3] Lu, Y., Huang, X., Zhang, K., Maharjan, S., & Zhang, Y. *Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles*. IEEE Transactions on Vehicular Technology, 69(4),(2020) 4298-4311.

[4] Luo, X., Zhang, C., & Bai, L. *A fixed clustering protocol based on random relay strategy for EHWSN*. Digital Communications and Networks, 9(1),(2023) 90-100.

[5] Tariq, F., Khandaker, M. R., Wong, K. K., Imran, M. A., Bennis, M., & Debbah, M. *A speculative study on 6G*. IEEE Wireless Communications, 27(4),(2020) 118-125..

[6] Al-Rahmi, A. M., Al-Rahmi, W. M., Alturki, U., Aldraiweesh, A., Almutairy, S., & Al-Adwan, A. S. *Acceptance of mobile technologies and M-learning by university students: An empirical investigation in higher education*. Education and Information Technologies, 27(6),(2022) 7805-7826.

[7] Sun, X. *5G joint artificial intelligence technology in the innovation and reform of university English education*. Wireless Communications and Mobile Computing, 2021,(2021) 1-10.

[8] Nguyen, M. N., Tran, N. H., Tun, Y. K., Han, Z., & Hong, C. S. *Toward multiple federated learning services resource sharing in mobile edge networks*. IEEE Transactions on Mobile Computing, 22(1), 541-555.

[9] Yao, S., Li, D., Yohannes, A., & Song, H. *Exploration for network distance teaching and resource sharing system for higher education in epidemic situation of COVID-19*. Procedia Computer Science, 183,(2021) 807-813.

[10] Ahmad, W. S. H. M. W., Radzi, N. A. M., Samidi, F. S., Ismail, A., Abdullah, F., Jamaludin, M. Z., & Zakaria, M. *5G technology: Towards dynamic spectrum sharing using cognitive radio networks*. IEEE access, 8,(2020) 14460-14488.

[11] Ahmad, T. *Scenario based approach to re-imagining future of higher education which prepares students for the future of work*. Higher Education, Skills and Work-Based Learning, 10(1),(2020) 217-238.

[12] Alghayadh, F. Y., Ramesh, J. V. N., Quraishi, A., babu Dodda, S., Maruthi, S., Raparthi, M., ... & Farouk, A. *Ubiquitous learning models for 5G communication network utility maximization through utility-based service function chain deployment*. Computers in Human Behavior, 156, 108227.

[13] Siriwardhana, Y., Porambage, P., Liyanage, M., & Ylianttila, M. *A survey on mobile augmented reality with 5G mobile edge computing: Architectures, applications, and technical aspects*. IEEE Communications Surveys & Tutorials, 23(2), 1160-1192.

[14] Kaur, J., Khan, M. A., Iftikhar, M., Imran, M., & Haq, Q. E. U. *Machine learning techniques for 5G and beyond*. IEEE Access, 9,(2021) 23472-23488.

[15] Al-Maroof, R., Akour, I., Aljanada, R., Alfaisal, A., Alfaisal, R., Aburayya, A., & Salloum, S. *Acceptance determinants of 5G services*. International Journal of Data and Network Science, 5(4),(2021) 613-628.

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/Emerging Wireless Networks
Received: May 13, 2024
Accepted: Jul 17, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 577–586, DOI 10.12694/scpe.v26i2.3987

INTEGRATIVE DEVELOPMENT OF RESEARCH TRAVEL AND CULTURAL HERITAGE PROTECTION BASED ON 5G COMMUNICATION AND MOBILE BASE STATION FROM THE PERSPECTIVE OF GEOGRAPHY – TAKING LUSHAN WORLD GEOPARK AS AN EXAMPLE

YANYAN CHEN*

Abstract. Energy efficiency is one of the important issues that must be taken into consideration at the beginning of 5G communication design; in this study, the base station energy consumption model is selected for subsequent network energy efficiency assessment. With the increasingly strict requirements for base station antennas, the design of base station antenna units has been put forward with higher requirements. This thesis tracks the latest research advances in 5G communication requirements and key technologies from various research organizations at home and abroad. The world cultural heritage research trip integrates cultural heritage with heritage tourism and youth heritage education in the present environment of the merger of culture and tourism. The interpretation and presentation of legacy, as exemplified by Mount Lushan World Geopark, serves to communicate to the general public the significance of heritage. According to this study, heritage education is an activity that management agencies, research institutes, and local community members arrange either alone or in collaboration with other tourism and educational institutions with the goal of raising public awareness of heritage and educating people about heritage through research excursions. This document outlines this idea and lists the organizers, research objectives, range of activities, and learning materials for cultural heritage research excursions. The optimized MIMO-D2D system model is demonstrated to be effective in Lushan World Geopark by the experimental findings. This report suggests five strategies and actions for cutting-edge study abroad programs and the preservation of China's intangible cultural heritage.

Key words: 5G communication; Mobile base station; Research travel; Cultural heritage protection; MIMO-D2D system model optimization

1. Introduction. Intangible heritage is a precious historical heritage of mankind [1]-[2]. As a precious memory of traditional culture, intangible cultural heritage has special value for human existence and development [3]-[4]. Looking at both the inside and the outside, harmonious culture must depend on excellent traditional culture, which is the foundation and root of the construction of a harmonious society and a harmonious culture in China. World cultural heritage research and study tours help young people better understand the world cultural heritage. China is a treasure in the history of human culture and a reflection of the culture with national characteristics [5]-[6]. Social harmony is not only the goal that humans pursue together, but also the guarantee for sustainable development of human society [7]-[8].

Promoting young people's awareness of world heritage is the goal of both UNESCO's World Heritage Youth Education and the World Heritage Center's World Heritage Education Program, which was established in 1994. Taking Mount Lushan World Geopark as an example, research trips can be employed as a new form of world heritage teaching activities. One of the best methods to combine tourism and culture is through study travel, and a well-liked tourist destination is World Heritage [9]-[10]. Thus, the focus of this book is the practice of study travel, or the merging of tourism, education, and cultural heritage. It makes cultural heritage conservation, heritage tourism, and worldwide youth heritage education easier [11].

Examine the device's current condition to determine whether it is normal. Then, use parameter settings to suitably adjust the network status to enhance service performance and guarantee [13]. Control is based on monitoring, and monitoring is the basis of control. Consequently, WinForm-based mobile base station device management and management, together with other relevant technologies, can be used to achieve device management and detection [14]. Its main job is to collect operational parameters and status data from various network nodes and devices and deliver it to the mobile base station's device administrator in an understandable

*School of Fine Arts and Design, Heze University, Heze, China (hzch66666@163.com).

visual format. Simultaneously, the mobile base station operator or equipment administrator transmits control instructions to the network's devices to establish configuration parameters, carry out network monitoring and configuration tasks, and guarantee that the devices function as needed. Differential protection for distribution networks is one of them. The three 5G communication slice network application scenarios all have fairly consistent uRLLC. Large coverage eliminates the need for new line construction, which contrasts sharply with optical fiber's high laying cost [15].

Research on research travel focuses on geographical research, research policy analysis, regional research practice[16]. Whether it is cultural heritage or research travel, the academic research perspective is increasingly rich. Some scholars have proposed to carry out research travel activities for primary and secondary school students relying on regional cultural resources[17]. Taking the Lushan World Geopark as an example, some scholars have begun to discuss the application [18]. This study fully considers the multiple cultural heritage resources in the study area, as well as the supporting research tutors, transportation conditions, regional development level and other auxiliary factors to carry out research travel, so as to enrich the theoretical research perspectives and ideas of cultural heritage and research travel [19]-[20].

2. Comparing and analyzing many cases. We will compare a number of cases from different geographic or cultural contexts to improve the generalizability of the conclusions in this research. These examples include research on the Grand Canal's cultural legacy, Wuqiao's acrobatics, and Qingxian's martial arts. We can better examine the universality of fusing study tours with cultural heritage by looking at these situations.

2.1. Examining the Grand Canal's Cultural Legacy. The Grand Canal, which connected Beijing and Hangzhou and included a rich cultural legacy along its course, was a significant water conservation project in ancient China. Work together with regional administrations to create a guideline for educational excursions centered around Grand Canal culture. Programs for design studies that address the canal's engineering technology, history, and culture. Provide boats and additional modes of transportation, as well as qualified guides and interpreters. However, planning transportation is challenging due to the Grand Canal's geographic expanse. Study visits are divided into portions, and to increase productivity, contemporary transportation is employed.

2.2. Wuqiao Acrobatics Research. Wuqiao has a rich cultural history in acrobatics and is the home of Chinese acrobatics. Create regulations to safeguard and advance the culture of acrobatics. Create a curriculum that incorporates historical explanations of acrobatics, practice sessions, and performances. Provide performance spaces, practice equipment, and acrobatic study bases. But acrobatic instruction needs expert supervision, and safety concerns are common. To guarantee safety, employ qualified coaches and acrobats and closely oversee the training regimen.

2.3. Research on Martial Arts in Qing County. Qing County is rich in martial arts cultural materials and is considered one of the birthplaces of traditional Chinese martial arts. Work together with martial arts associations and the local government to create policies that will support martial arts culture. creating courses covering Wushu's history, fundamentals, and performance evaluation. establishing a camp for wushu research and training with the goal of offering qualified instructors and training supplies. Wushu training does, however, come with a danger of injury and demands a high level of physical condition. Provide medical staff and create a scientific training program that will guarantee a progressive increase in training intensity.

By comparing the aforementioned numerous cases, we are able to confirm the broad applicability of the study tours and cultural heritage combination as well as provide an overview of potential implementation challenges in various geographic and cultural contexts and their corresponding strategies for resolution. This will improve the findings' applicability and generalizability and serve as a useful guide for upcoming model replications in other areas. The research in this paper will be more broadly relevant by including these case studies from various geographic and cultural contexts. Additionally, readers will be provided with the problems that may arise in practice and their resolution solutions.

3. MIMO-D2D system model optimization. When there are MIMO system, the signal can be expressed as:

$$y_k = H_k v_k s_k + \sum_{i=1, i\neq k} H_k v_{S_i} s_i + \eta_k, k = 1, 2, \ldots, K \tag{3.1}$$

where y_k is the received signal vector, H_k is the channel gain matrix, V_k is the transmit signal vector, s_i is the noise vector.

Compared with ZF precoder, the algorithm has higher complexity. According to the minimum mean square error theory, the target user's precoding matrix can be obtained:

$$v^{opt} = \arg\min E\left[\|s - s\|_F^2\right] = \arg\min E\left[\left\|s - \frac{1}{\beta}(Hvs + \eta)\right\|_F^2\right] E\left[ss^H\right] = NT\sigma^2 \tag{3.2}$$

where v^{opt} is the precoding matrix, Hvs is the conjugate transpose of s, $NT\sigma^2$ is the noise variance, and F is the unit matrix.

The path loss can be described by mathematical expression:

$$P_r(d) = \frac{P_t G_t G_t \lambda^2}{(4\pi)^2 d^2 L} \tag{3.3}$$

ZF and MMSE precoding schemes have certain restrictions on the number of antennas at the receiver and transmitter. In contrast, the maximum SLNR based scheme has no limit on the number of antennas and is not limited by the application scenarios. Its basic idea is to maximize the ratio leaked to other users and channel noise power, so as to enhance the target signal while weakening the signal leaked to other users. The formula of signal-to-noise ratio can be expressed as:

$$\mathrm{SLNR}_k = \frac{\|H_k v_k\|_F^2}{\sum_{i=1, i=k}^{K} \|H_i v_k\|_F^2 + \sigma^2} \tag{3.4}$$

4. Methods.

4.1. Data selection. With Lushan Mountain accounting for 18. the total number of mountains in the region is 100.The Ministry of Culture and Tourism has named five batches of national intangible cultural heritage representative project inheritors, with 15 people in Lushan Mountain; Hebei Provincial People's Government has released 6 groups of provincial intangible cultural heritage lists, with Lushan Mountain accounting for 102 items. There are 117 individuals (alive) in Lushan Mountain among the five groups of representative inheritors of the province's intangible cultural assets that have been disclosed by the Department of Culture and Tourism. The Lushan area's intangible cultural heritage resources are abundant in variety and number, particularly when it comes to the categories that are part of provincial initiatives, as indicated by the statistical results presented in Table 4.1. The variety of intangible cultural heritage project types is highlighted by the seven study topics, which span more than five categories. Tangible space resources of cultural heritage research resources include cultural relics protection units, patriotic education bases, red tourism classic scenic locations, research practice bases, and A-level scenic spots. These tangible spaces already have the necessary infrastructure, which can meet the basic conditions for carrying out research travel, greatly reduce the difficulty of developing research travel, save human, material and financial resources.

It is worth mentioning that Lushan, as the city with the longest mileage in the cities that the Beijing Hangzhou Grand Canal flows through, has a high historical status and research value. Because it has certain tangible space, taking Lushan World Geopark as an example, it is classified as tangible space resources in this study, further enriching the types of tangible space resources in Lushan area. As Table 4.1 for their differences.

The buffer zone and protection barrier established by the World Cultural Heritage are not the extent of the research trip's operations. The division of administrative regions will not limit the scope of the world cultural heritage research trip's operations to the cultural region in which it is located. The concept and content of the activity, which is often linear and has a clear tourism route planning, define the extent of the World Cultural Heritage study trip. In essence, MIMO technology is a mathematical abstraction of a multi-antenna mobile communication system. Multipath effect has historically been a negative issue impacting system performance in communication systems. The difference is that in MIMO systems, multipath becomes a favorable factor. It can obtain multiple data stream gains between devices, making information transmission more accurate. The MIMO system can be shown in Fig. 4.1- taking the Lushan World Geopark as an example.

Table 4.1: Comparison of Short Range 5G Communication Technologies.

Parameter name	Wi-Fi Direct	Bluetooth 4.0	D2D
Standard	802.11	Bluetooth SIG	3GPP;LTE- Advanced
Frequency band	2.4GHZ	2.4 GHZ	Authorized frequency band
Frequency band	5 GHZ	2.4 GHZ	Authorized frequency band
Maximum data rate	250Mb/s	1 Mb/s	1Gb/s
Maximum transmission distance	200M	10-100m	10-1000m
Device Discovery	ID broadcast	Manual pairing	Base Station Collaboration
Device Discovery	Embedded Soft Access Point	Manual pairing	Base Station Collaboration
Application	File sharing	Peripheral equipment	File sharing; Local video; Public safety; Cell relay
Application	Device connection	Discover Connections	File sharing; Local video; Public safety; Cell relay
Application	Group competition	File transfer	File sharing; Local video; Public safety; Cell relay

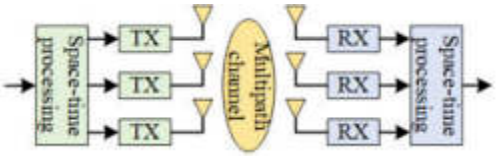


Fig. 4.1: Schematic Diagram of MIMO System.

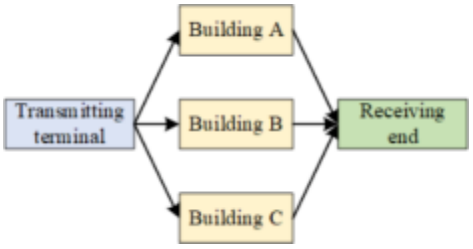


Fig. 4.2: Multipath Effect of Cultural Heritage Protection.

The wireless channel has the characteristics of multi-path effect, fading and Doppler effect. The result of multi-path effect is that the original signal is distorted or even wrong. The main reason is that the transmitter signal arrives at the receiver through different paths in the channel. Because the time of arrival at the receiver is different, the signal phases of each path overlap during signal synthesis. The process of multipath effect is shown in Fig. 4.2.

Understanding the world heritage and the need to safeguard it is one of the key goals of the World Cultural Heritage Study Tour. Recognize the nature, traits, importance, and background of the world; Recognize the world's knowledge, the veracity and historical information conveyed by cultural heritage, and the height at which it reflects as the primary subject. We can further study cultural heritage in the domains of art, history, society, and science by using actual data on the global cultural heritage. The learning content includes knowledge about art, archaeology, ancient architecture, religion, nature, and science. World cultural heritage research travel can meet the trainees' needs for experience and experience of culture and heritage, provide high-quality information for trainees, and meet their spiritual needs. In the end, students can obtain a valuable, satisfactory and pleasant experience. Fig. 4.3 shows the system model of MIMO-D2D. The transmitted signal of the cellular user through the multipath channel gain, and the D2D communication signal can also be more accurately transmitted to the

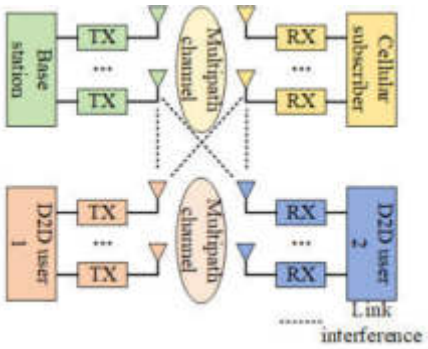


Fig. 4.3: MIMO-D2D system communication model optimization.

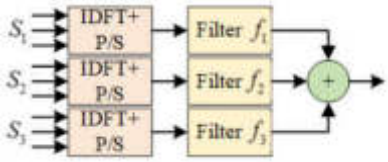


Fig. 4.4: UFMC System Model Optimization.

target D2D user.

4.2. Research assumptions. The current distribution network protection mostly adopts simple over-current and over-voltage protection methods, which do not need communication channels and cannot achieve sectional isolation, resulting in a large power failure range and difficult recovery after failure. Distribution network communication points are wide and highly dispersed. If optical fiber is used in a large area, the cost will increase dramatically. While wireless communication is fast deployed, widely covered and easy to be modified, 5G communication is fully feasible as a differential protection channel. Fig. 4.4 shows the entire bandwidth contains M subcarriers and is divided into B subcarriers, each subcarrier is composed of m subcarriers.

Intangible cultural heritage reflects a country's or a nation's self-identity and the extent to which it is recognized by the world. It is a key link to maintain a nation's or group's civilization. However, if we really want to turn various protection declarations and regulations into practical actions and make them have obvious effects, we can never easily achieve them. If we fail to achieve scientific, planned, up-to-date and sustainable protection, we may get twice the result with half the effort, hinder or even destroy these cultural treasures that we have inherited for several years. The simulation environment is completed in MATLAB, and the number of simulation experiments is 10000, as is shown in Table 4.2.

The main optimization points of balanced throughput in high throughput and low algorithm complexity, but they ignore the poor performance of large equipment connectivity in the 5G era, that is, the poor performance of system connectivity probability. The method of resource allocation algorithm based on power adjustment are given,as is shown in Fig. 4.5.

5. Simulation Result Analysis. The simulation scenario is that D2D users selectively reuse uplink resources of cellular users in a single LTE cell with a radius of 1km. The main simulation parameter settings are shown in Table 5.1. By comparing the proposed resource allocation (TPRA) method with the random resource allocation (RRA) method and the resource allocation (RCRA) method considering the D2D rate, the system throughput and power allocation under different algorithms are mainly compared.

Utilizing the cultural heritage resources of Mount Lushan area to create an atmosphere of study and travel, and to create a new business card of Mount Lushan in culture, education and tourism. Through the development

Table 4.2: Simulation Parameter Setting.

Parameter	Parameter value
Cell radius R/m	$10 < R < 500$
D2D pair distanceR_d/m	$10 < R_d < 25$
Number of simulations/time	10000
System bandwidth B/MHz	100
Carrier frequencyf_c /GHz	5.0
Noise power N/dBm	-112
Maximum transmission power of cellular users/dBm	26
Minimum SNR/dBm for cellular users	12
D2D user minimum signal-to-noise ratio/dBm	7
Path loss between any communication terminals	$22\log_{10}(d)$+44.4, 10<d<410;$40\log_{10}(d)$ +87.7, 410<d<500

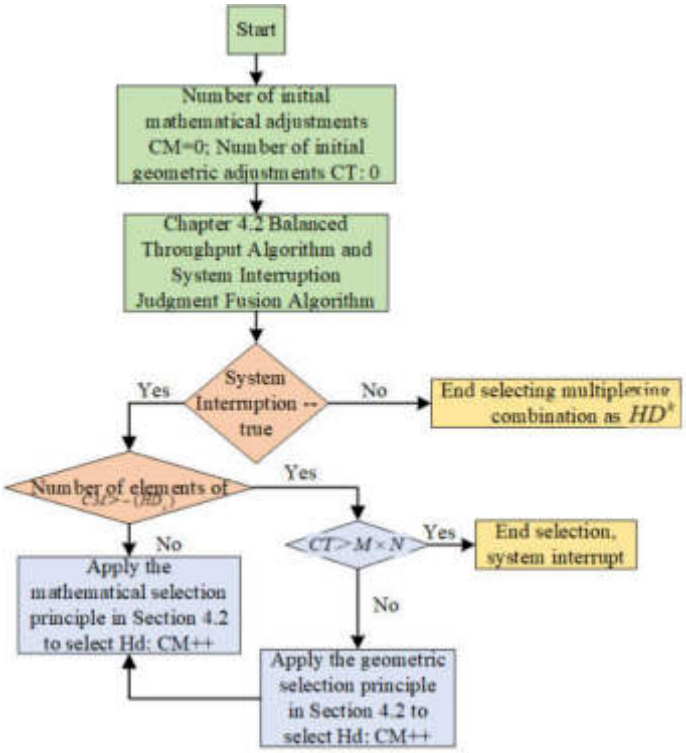


Fig. 4.5: Optimization flow chart of channel allocation algorithm based on 5G communication power adjustment management model and sub model.

and improvement of all kinds of study projects, improve the visibility of study, gradually expand the influence of Mount Lushan, the formation of Mount Lushan area unique study brand, and in the canal research, Wuqiao acrobatics research, Qingxian Wushu research and other outstanding areas of cultural heritage research IP, centered on the urban area of Mount Lushan. Ultimately, through the "tourism + education" form of study tours, the inheritance and development of values, while cultivating local primary and secondary school students' national sentiment, traditional culture and values education, and enhance cultural self-confidence - Lushan World Geopark as an example. Fig. 12 compares the CDF distribution of system throughput under the two algorithms.RRA is significantly lower than TPRA because the RRA algorithm fails to reasonably allocate the

Table 5.1: Simulation Parameter Setting of Mobile Base Station.

Parameter	Parameter value
Cell radius/km	1
Short distance communication distance/m	25 50
Bandwidth of one RB/kHz	180
RB quantity/piece	50
BTS transmission power/dBm	44
Cellular subscriber transmission power/dBm	25
Road loss/dB	$35.26+35\ \log_{10}(d)$
Number of D2D users	1 30
D2D maximum transmission power/dBm	25
Targetλ /dB	15
Gaussian white noise density/dB/Hz	-172
Iterations	32

resources of primary and secondary users, and the interference between primary and secondary users is not effectively controlled.

The TPRA (Throughput Prioritized Resource Allocation) algorithm aims at resource allocation by prioritizing system throughput. Although TPRA performs well in improving system throughput, its complexity is mainly due to the following aspects:

1. Computational complexity: TPRA must determine each user's priority inside each time slot and allocate resources according to these priorities. This calls for a lot of processing power, particularly when there are a lot of users.
2. Scheduling of Resources: For TPRA to guarantee that system throughput is optimal, resource allocation must be continuously monitored and adjusted. This puts a lot of pressure on network administration and control.

The following elements may have an impact on TPRA's real operational efficiency in large-scale networks:

1. Latency issue: In large-scale networks, TPRA latency may be considerable due to the requirement for intricate computations and real-time modifications, which can negatively impact user experience.
2. Hardware requirements: In order to guarantee that the algorithms can operate effectively, stronger hardware support is required due to the high computational demands of TPRA.
3. Scalability: While TPRA works well in small networks, as the number of users rises in large networks, its computational complexity and resource scheduling becomes more challenging, necessitating improvement to increase scalability.

Use the cultural heritage resources in Lushan area to create a research travel atmosphere, and create a new name card for Lushan in terms of culture, education, tourism, etc. Through the development and improvement of various research projects, we will improve the popularity of research, gradually expand the influence of Lushan , form a distinctive research brand in Lushan area, and create cultural heritage research IP in outstanding areas, such as Canal Research, acrobatics research in Wuqiao, martial arts research in Qingxian and other places with Lushan downtown as the center. Finally, we will inherit and carry forward the value in the form of "tourism + education" research travel, while cultivating the family and country feelings, traditional culture, values education of local primary and secondary students, and strengthening cultural self-confidence - take the Lushan World Geopark as an example. Fig. 5.1 compares the CDF distribution of system throughput under the two algorithms. RRA is significantly lower than that of TPRA. Because RRA algorithm does not allocate resources to primary and secondary users reasonably, the interference between primary and secondary users is not effectively controlled.

The overall path is to "issue standardized policies, formulate research plans", "focus on significant resources, customize research with characteristics", "focus on departmental linkage, improve research facilities", "integrate science and technology into tradition, and lead research through innovation". Then provide targeted research development path according to the above comprehensive potential index. When the signal-to-noise ratio in-

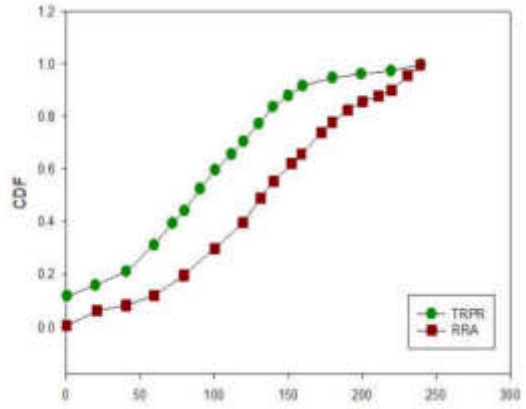


Fig. 5.1: System Throughput Comparison of Different Algorithms.

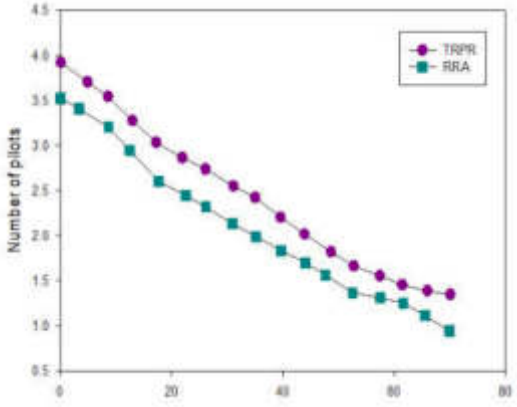


Fig. 5.2: Bit Error Rate of UFMC Constant SNR Compression Sensing Algorithm.

creases to a value that can be considered that the channel is distortion free and noise free, the reconstruction accuracy of the four algorithms is similar. At low SNR, RSAMP algorithm has the best performance, while OMP algorithm has the worst performance,as is shown in Fig. 5.2.

By prioritizing resource allocation to maximize system throughput, the TPRA algorithm efficiently enhances system performance. Regardless of the circumstance, the TPRA algorithm offers superior throughput, as evidenced by its higher and steady CDF curve.

Lower and unpredictable system throughput are the result of the RRA algorithm's inability to properly control inter-user interference because it distributes resources randomly. This suggests that in real-world applications, it is challenging to match the high throughput demand of the RRA algorithm.

Optimization based on the TPRA algorithm can be performed to further enhance system performance. This includes introducing an intelligent scheduling mechanism and a dynamic resource management strategy to increase the system's throughput and energy efficiency.

This thorough examination of Fig. 12 helps us to better understand the variations in system throughput under various algorithms as well as the direction of optimization, which serves as a crucial foundation for further study.

The higher the communication quality of high priority users. However, this algorithm will also have some impact on other low priority users. The more spectrum the mobile base station gets, the more obvious the

Table 5.2: Communication Rate of Low Priority Users under Different Spectrum Divisions.

Spectrum division (MHz)	$W_{MBS} = 2$, $W_{cf} = 18$	$W_{MBS} = 3$, $W_{cf} = 17$	$W_{MBS} = 4$, $W_{cf} = 16$	$W_{MBS} = 5$, $W_{cf} = 15$
Average rate of low priority users (Mbits/s)	24.1466	22.8047	21.4637	20.1222

impact will be, as shown in Table 5.2.

6. Conclusion. Higher technological demands are placed on base station antenna units due to the ever stricter regulations on base station antennas. This study traces the most recent developments in 5G communication requirements and essential technologies from both local and international research organizations in order to meet this background.

Research travel, especially when considering the integration of culture and tourism, integrates cultural heritage with heritage tourism and youth heritage education in terms of protecting cultural heritage. For instance, in the case of Mount Lushan World Geopark, the public is made aware of the significance of heritage through the interpretation and presentation of the site's legacy. The goal of heritage education and research tours is to increase public knowledge of heritage through the work of management agencies, research institutes, and members of the local community, either on their own or in conjunction with other tourism and education groups.

In general, research tours and studies on cultural heritage aid in the integration of tourism, education, and cultural heritage while providing young people with a deeper understanding of the world's cultural heritage. The study revealed that research trips were carried out with the assistance of local cultural resources, which enriched the theoretical research perspectives on research excursions and cultural heritage. Furthermore, the algorithm illustrates the benefits of system transmission performance under various SNR conditions by analyzing the UFMC constant SNR compression-aware algorithm. These benefits are particularly noteworthy in 5G and future wireless communication networks, as they offer high spectral efficiency and low bit error rate.

Data Availability. The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest. The authors declared that they have no conflicts of interest regarding this work.

Funding Statement. In 2022, the Shandong Provincial Undergraduate Teaching Reform Research Project "Intangible Cultural Heritage Ideology and Politics" Exploration of Teaching Practice of Collaborative Education in College Animation Majors M2022010.

REFERENCES

[1] Megeirhi, H. A., Woosnam, K. M., Ribeiro, M. A., Ramkissoon, H., & Denley, T. J. *Employing a value-belief-norm framework to gauge Carthage residents' intentions to support sustainable cultural heritage tourism.* Journal of Sustainable Tourism, 28(9),(2020) 1351-1370.

[2] Li, Y., Lau, C., & Su, P. *Heritage tourism stakeholder conflict: A case of a World Heritage Site in China.* Journal of Tourism and Cultural Change, 18(3),(2020) 267-287.

[3] Alazaizeh, M. M., Jamaliah, M. M., Mgonja, J. T., & Ababneh, A. *Tour guide performance and sustainable visitor behavior at cultural heritage sites.* Journal of Sustainable Tourism, 27(11),(2019) 1708-1724.

[4] Panzera, E., de Graaff, T., & de Groot, H. L. *European cultural heritage and tourism flows: The magnetic role of superstar World Heritage Sites.* Papers in Regional Science, 100(1),(2021) 101-122.

[5] Mariani, M. M., & Guizzardi, A. *Does designation as a UNESCO world heritage site influence tourist evaluation of a local destination.* Journal of Travel Research, 59(1), (2020) 22-36.

[6] Graziano, T., & Privitera, D. *Cultural heritage, tourist attractiveness and augmented reality: Insights from Italy.* Journal of Heritage Tourism, 15(6), (2020)666-679.

[7] Adie, B. A., Falk, M., & Savioli, M. *Overtourism as a perceived threat to cultural heritage in Europe.* Current Issues in Tourism, 23(14), (2020) 1737-1741.

[8] Gursoy, D., Akova, O., & Atsiz, O. *Understanding the heritage experience: a content analysis of online reviews of World Heritage Sites in Istanbul.* Journal of Tourism and Cultural Change, 20(3), (2022) 311-334.

[9] Bapiri, J., Esfandiar, K., & Seyfi, S. *A photo-elicitation study of the meanings of a cultural heritage site experience: A means-end chain approach.* Journal of Heritage Tourism, 16(1), (2021) 62-78.

[10] Müggenburg, H. *Beyond the limits of memory? The reliability of retrospective data in travel research.* Transportation research part A: policy and practice, 145, (2021) 302-318.
[11] Buckley, R. *Pandemic travel restrictions provide a test of net ecological effects of ecotourism and new research opportunities.* Journal of Travel Research, 60(7), (2021) 1612-1614.
[12] Caber, M., González-Rodríguez, M. R., Albayrak, T., & Simonetti, B. *Does perceived risk really matter in travel behaviour.* Journal of Vacation Marketing, 26(3), (2020) 334-353.
[13] Brown, A. E. *Who and where rideshares? Rideshare travel and use in Los Angeles.* Transportation Research Part A: Policy and Practice, 136, (2020) 120-134.
[14] Koushik, A. N., Manoj, M., & Nezamuddin, N. *Machine learning applications in activity-travel behaviour research: a review.* Transport reviews, 40(3), (2020) 288-311.
[15] L. Sun, J. Liang, C. Zhang, D. Wu and Y. Zhang, "Meta-Transfer Metric Learning for Time Series Classification in 6G-Supported Intelligent Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 3, pp. 2757-2767, March 2024, doi: 10.1109/TITS.2023.3250962.
[16] Chen, S., Law, R., Zhang, M., & Si, Y.*Mobile communications for tourism and hospitality: a review of historical evolution, present status, and future trends.* Electronics, 10(15), (2021) 1804.
[17] Siriwardhana, Y., De Alwis, C., Gür, G., Ylianttila, M., & Liyanage, M. *The fight against the COVID-19 pandemic with 5G technologies.* IEEE Engineering Management Review, 48(3), (2020) 72-84.
[18] Song, H., & Selim, G. *Smart heritage for urban sustainability: a review of current definitions and future developments.* Journal of Contemporary Urban Affairs, 6(2), (2022)175-192.
[19] Yücel, M., & Açikgöz, M. (2023). Optical communication infrastructure in new generation mobile networks. Fiber and Integrated Optics, 42(2), (2023) 53-92.
[20] Sheng, J., Cai, X., Li, Q., Wu, C., Ai, B., Wang, Y., ... & Yu, P. *Space-air-ground integrated network development and applications in high-speed railways: A survey.* IEEE Transactions on Intelligent Transportation Systems, 23(8), (2021) 10066-10085.

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/Emerging Wireless Networks
Received: May 20, 2024
Accepted: Aug 31, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 587–598, DOI 10.12694/scpe.v26i2.3873

OPTIMIZING ELECTRIC VEHICLE CHARGING INFRASTRUCTURE WITH EVGRIDNET BY INTERNET OF THINGS AND MACHINE LEARNING STRATEGIES

R. RAMANI *AND , A. NALINI †

Abstract. In the arena of renewable energy integration for electric vehicle (EV) infrastructure, it is a problem to efficiently use solar power with EV charging, considering grid constraints and user preferences. The current study proposes a new smart charging algorithm that utilizes IoT data for the dynamic optimization of EV charging patterns. A novel aspect of the study was a deep learning model, "EVGridNet", that reliably predicts solar energy output and grid prices. EVGridNet uses deep learning approaches to process accumulated data via IoT devices, facilitating fine-tuned adjustments to charging patterns through predictive analytics. This algorithm receives actual data on the generation of solar energy, the price of grid electricity, and other characteristics set by the user to optimize the usage of solar energy, limit the usage of grid electricity during the peak hours and meet all the needs of the user. The optimization process of the algorithm strategically manages energy sources, uses battery storage systems to exploit solar power effectively and uses grid electricity during low-cost periods, all within user preference parameters. The proposed system has the potential of reducing grid electric use by up to 25% for EV charging, and increasing the renewable energy electricity in EV charging to 40% as per simulation results. This will enable quick EV infrastructure scalability, low carbon emission, and energy independence. EVGridNet is an outstanding innovation in smart charging technology, which is a cost-effective and scalable solution to the renewable energy sector's primary challenge. One of the key aspects of the optimization process is the control of energy sources where battery storage systems allow for flexibility in using solar energy and grid electricity within certain pre-set thresholds.

Key words: EV, grid energy, renewable, IoT, CNN, GRU, dynamic, optimization, rate, infrastructure, data, solar energy.

1. Introduction. One of the most challenging aspects of the nascent electric vehicle (EV) infrastructure [1] is incorporating renewable sources of energy, such as solar energy [2], into EV charging points. The grid limitations during peak demand times and different user preferences increase the space of non-optimal use of renewable resources in traditional grid-dependent charging models, which is one of the main barriers to achieving more sustainable solutions.

The following study presents Smart EVGridNet to fill this research gap. It is proposed as a smart charging algorithm that employs IoT data [3] to optimize dynamic EV charging patterns. Solar energy generation, grid electricity prices, and user requirements are all taken into account by analyzing real-time data. This is done to enhance the use of solar power, create less reliance on the grid during peak periods, and accept user needs, hence filling the gap left by renewable energy integration as it is currently practiced. The novel EVGridNet model is based on deep learning, in which Convolutional Neural Networks (CNNs) [4] and Gated Recurrent Units (GRUs) are being used, among others coupled with Reinforcement Learning (RL) [5] and dynamic programming for sophisticated prediction of electricity prices and solar outputs. This prediction allows fine-tuning charging patterns in real-time, utilizing energy storage systems, and using the grid effectively when the charges are lowest. The system's ability to turn EV charging into an off-peak hour's activity and renewable energy supremacy is vital in achieving grid-independent and clean EV infrastructure. The impact of the EVGridNet program is not only energy savings; it tells the other story of the speedy expansion of EV infrastructure while lowering the level of carbon emissions and independence in energy. Smart charging [6] can keep up with the dynamics by adopting a holistic approach to decision-making under uncertainty, which allows it to navigate the complexity of the energy matrix and, therefore, becomes the ultimate solution to some of the renewable energy sector's challenges and the significant giant leap towards the establishment of smart charging infrastructure.

*Department of Electrical and Electronics Engineering, Dr. M.G.R. Educational and Research Institute, Tamilnadu, India. (Corresponding author, ramanikamal11@gmail.com)
†Department of Electrical and Electronics Engineering, Dr. M.G.R. Educational and Research Institute, Tamilnadu, India.

One of the most essential parts of the current research is the use of advanced neural networks for predictive control devices, which helps to improve the efficiency of solar power usage in real-time by controlling the charging patterns. In another way, the decrease in the distribution grid's energy consumption and the improvement of the renewable energy input to the charging system through EVGridNet establishes a basis for clean energy creation. The proposed model could be a significant tool for the transition of the EV charging ecosystem to being more environmentally friendly, cost-effective, and user-oriented. This model should provide support for the rapid development and scaling up of the EV infrastructure while minimizing the carbon emission situation in the process and reinforcing the independence of our local energy resources.Thus, the primary objectives of this study includes:

1. To improve EV charging infrastructure, EVGridNet was utilized, which helped accelerate the incorporation of solar energy and observably decreased non-renewable energy use.
2. To reduce the electricity grid load during peak demand times, the EVGridNet network will help decrease the load on the grid and optimize the charging patterns to take advantage of the off-peak periods and available solar power.
3. To develop a flexible charging infrastructure that could be configured to the inputs of real-time and user-related data.

2. Related Work. The work from [7] addressed EV charging patterns through Functional Data Analysis (FDA) analysis using five years of real-time data from 455 stations in the Kansas City of, Missouri district. The specific approach, the so-called "smooth functions" treatment of the data over time, is exceptional for its flexibility and effectiveness in monitoring daily, station-related usage by customers and energy consumption. The study recognizes the FDA's deficiency in obtaining time series from different periods, which is a significant problem in forming some analytical tools. This work is unique in the sense that the authors succeeded in applying the FDA course to evaluate EVs' charging behavior despite the difficulties involved, whilst the existing data may need to be more consistent. The study from [8] developed a model integrating long-term and short-term planning and decision-making processes to solve the problem of the most efficient placement of charging stations. Their innovative technique exploits a two-stage mixed integer linear programming model to achieve the optimal electricity flow costs, which can then incorporate renewable energy and V2G capabilities. To undermine the complexity of the model, they suggest a mixture of Apriori Progress Hedging And Sample Average Approximation (APH-SAA) algorithms, speeding up the algorithm and making it more efficient. Regardless of its positive side, the study acknowledges that the progressing hedging algorithm has limitations if capitalizing on large-scale problems. Therefore, the study shifts to the heuristic method for a smooth solution. Such research now provides a vital link between various operational factors and charging station planning design, which can be easily replicated for more cost-efficient and environmentally friendly infrastructure development.

Researchers in [9] proposed a solution for optimizing grid operations with growing EVs employing machine learning and the LSTM model. The method puts a great emphasis on the management of electric power demand. Moreover, it decreases losses and fluctuations and drops the economy tariffs, regardless of the charging technologies used. For managing complex and ever-changing loading patterns of the power grid and EVs, LSTM was chosen because of its ability to predict them accurately. The investigation proves that the impact of EV charging on the grid can be effectively managed through simulations; the benefits become apparent, especially during unforeseen load conditions. Notably, the project focuses on ML to improve grids in terms of efficiency and stability, which effectively minimizes challenges posed by the increasing EV acquisition. Conversely, vehicle-to-grid (V2G) technologies might not perform primary operations properly and can cause deep discharging of EVs, which leads to battery degradation and loss of customer satisfaction.

According to the study of [10], using machine learning, primarily deep LSTM, is proposing high-capacity management and routing tools for EV fleets. It measures the uncertainty of load data impact on the EV management system, which is a criterion in showing the LSTM's capacity to handle sudden data. On the other hand, though, this complexity in adopting deep learning approaches for managing the centralized system efficiently emerges as a potential limitation, mainly as more and more EVs come into play. Such growth encompasses larger data sets and affects businesses with parking facilities. As a result, these businesses require advanced technological solutions catering to the emerging EV market.

Researchers in [11] intended to optimize EV charging infrastructure using a genetic algorithm modified

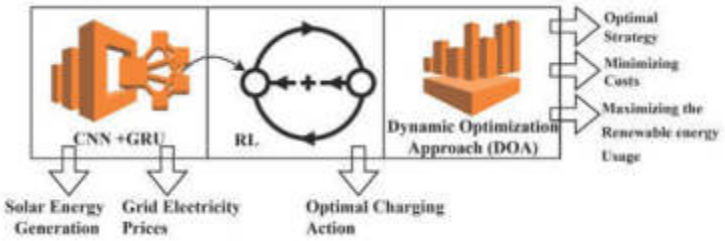


Fig. 3.1: EVGridNet Framework

by neural networks and deep learning architecture, namely data driven multi-objective optimization (DDMO). The main task of the associated framework is to improve these factors, such as the type of charge points, charging location, the amount of charging points and so on. Research consisting of simulation and optimization offers information about the best possible configurations, spatial arrangements and, importantly, the trade-offs between coverage costs and maintaining the target levels. On the one hand, the dependence on data accuracy and model quality for efficient solutions causes the processes to become more computationally complex and take more time. The proposed optimization method is comprehensive. Nevertheless, tuning the parameters is complex and requires efficient handling of computational problems.

The central theme of the study by [12] is a multi-agent-based simulation model for the strategic planning of EV charging facilities on the highways, which predicts user charging patterns and lowers the limitation of the traditional site selection methods by user charging behavioral modeling, and decision-making processes related to charging. The model framework is included in the open-source built-in multi-agent modeling tool MATSim, which simulates the behavior of agents making their travel decisions in a shared network which competes for scarce resources on the highway, like charging facilities. A narrow disadvantage is the belief that EV drivers do not charge at highway facilities while leaving out factors affecting their choice of nearby facilities, leading to significant challenges when simulated in real-world scenarios.

3. Methodology. The methodology integrates Convolutional Neural Networks (CNNs), Gated Recurrent Units (GRUs), and Reinforcement Learning (RL) with a dynamic programming approach to optimize electric vehicle (EV) charging patterns. Initially, it defines the state and action spaces based on time, energy levels in the Energy Storage System (ESS), and possible charging actions.

A policy with parameters α is initialized to guide decision-making. For each episode, the algorithm starts by observing the initial state, which includes the starting time and energy level. As it progresses through each time step, it uses the CNN-GRU framework to predict solar energy generation and grid electricity prices. Based on these predictions and the current state, the RL component determines the optimal charging action according to the policy, executes this action, observes the resulting reward and the next state, and then updates the policy parameters using the policy gradient method to maximize the expected cumulative reward. Simultaneously, the dynamic programming component systematically solves for the minimum cost or optimal strategy by defining a recurrence relation that considers both immediate costs and future outcomes, iteratively computing this for each state.

Finally, it traces back through these decisions to reconstruct the optimal charging strategy, effectively balancing the objectives of minimizing costs and maximizing the use of renewable energy within the operational constraints of the EV charging infrastructure.

This comprehensive approach leverages predictive analytics and decision-making under uncertainty to dynamically adjust and optimize EV charging patterns in a complex, ever-changing energy landscape. Figure 3.1 illustrates the overall framework of EVGridNet model.

3.1. Problem Definition. To formulate the problem of EV charging optimization through dynamic programming, the minimization of cost, the maximization of solar energy utilization, and the constraints are considered. Table 3.1 represents several computational notations for better computations.

Table 3.1: Computational Notations and Definitions

Notations	Descriptions
$T = \{t_1, t_2, \cdots, t_n\}$	Set of discrete time intervals
$S(t)$	Solar energy generation at t
$[G_p(t)]$	Electricity cost in the grid at t
C_{req}	Cumulative energy demand for charging within the horizon
$D_{iot}(t) = \begin{cases} 1, & online \\ 0, & offline \end{cases}$	IoT device status
$\xi(t)$	Energy stored at energy storage system (ESS), $\xi_{min} \leq \xi(t) \leq \xi_{max}$
$G(t)$	Energy utilization for charging (C) at grid

Objective Function. Considering the usage of $S(t)$ and $G(t)$ for charging, the primary objective is to minimize the $G_P(t)$ while meeting the user's C_{req} until the end of the optimization horizon is reached.

$$\min \mathrm{Z} = \sum\nolimits_{t \in T} [G_P(t) \bullet G(t)] \tag{3.1}$$

with three constraints:

$$\sum\nolimits_{t \in T} [G(t) + S(t) \geq C_{req}] \tag{3.2}$$

$$\xi_{min} \leq \xi(t) \leq \xi_{max} \tag{3.3}$$

$$C(t) \Leftarrow D_{iot}(t) = 1 \tag{3.4}$$

Equation 3.1, 3.2, 3.3 represents the energy requirement, ESS storage, and IoT device operational constraints, respectively.

The proposed EVGridNet comprises Convolutional Neural Networks (CNNs) for data representation and feature extraction, Gated Recurrent Units (GRUs) for sequence modeling [13], and Reinforcement Learning (RL) for decision-making and action selection. This model aims to leverage the strengths of each component: CNNs for spatial feature extraction (such as weather or temporal patterns), GRUs for capturing short-term trends (e.g., energy generation or market price), and RL for making control decisions (like an optimal charging pattern) which utilizes the dynamic programming algorithm based on specific constraints.

CNN Computations. For instance, processing data along with feature extraction for spatiotemporal data, such as weather conditions that impact solar power generation. The primary computation can be expressed as,

$$f_{cnn} = ReLU(I \times \omega_{cnn} + e_{cnn}) \tag{3.5}$$

From (3.5), we denote the input, ω and e represent the weight and biases of CNN, respectively with ReLU activation function.

GRU Computations. The comprehension of sequential patterns in solar energy production, grid electricity costs, and energy usage is gained by observation of this time series data. Consecutive readouts after a pass through the network of feature extractions, in combination with historical data on prices, energy generation, and demand, are indispensable to the input system. The operations of update gate (U_t), reset gate R_t), candidate ($\bar{h}_t$)and final activation (h_t) comprise the major processes of GRU in determining the patterns. Thus, the conventional process of GRU can be expressed as,

$$U_t = \mu([\mathrm{h}_{t-1}, i_t] \bullet \omega_U + e_U) \tag{3.6}$$

$$R_t{=}\mu\left([\mathrm{h}_{t-1}, i_t]\bullet\omega_R{+}e_R\right) \tag{3.7}$$

$$\widetilde{h_{tt}}{=}tanh\left([R_t{\times}i_t, \mathrm{h}_{t-1}]\bullet\omega{+}e\right) \tag{3.8}$$

$$\mathrm{h}_t{=}\left(U_t{\times}\widetilde{\mathrm{h}}_t\right)+\left[(1{-}U_t){\times}\bar{h}_{t-1}\right] \tag{3.9}$$

From (3.6) to (3.9), i_t and h_t denotes the input and hidden state at time t, and μ signifies the sigmoidal activation function.

RL Computations. At this phase, EVGridNet learns to acquire knowledge of the optimal charging strategy depending on the system's current state and the forecast made by the CNN-GRU framework. The system's present state comprises the efficiency of the ESS, the status of the IoT devices, the estimated solar output, and the price of grids for the given timeframe. Objectives are being achieved via a dynamic optimization approach (DOA) combined with RL decision-making strategies that can ensure the patterns are optimally maintained to minimize grid utilization and maximize solar usage. Thus, the core process of RL is expressed as,

$$\nabla_\alpha f_J\left(\alpha\right){=}E\left[\nabla_\alpha\log\rho_\alpha\left(A_t \mid S_t\right){\times}r\ \right] \tag{3.10}$$

From (3.10), ρ_α denotes the policy that was parameterized by α, A_t and S_t indicates the action and state at t, and r represents the reward function with respect to the f_j (objective function) of ρ.

The structured DOA includes four major steps: defining the state, formulating the recurrence relation, determining the base case, solving sub-problems, and reconstructing the optimal solution. Table 3.2 represents the entire computation process of EVGridNet with DOA.

The primary role of the computation procedure of EVGridNet with the DOA is to determine the best patterns for charging a given number of EVs by incorporating the forecasts of the amount of load required in the grid at a given time, reinforcement learning, and dynamic programming. This procedure, which involves the CNN-GRU framework to forecast the generation of solar energy and the price of grid electricity, and RL to identify the optimal charging actions, also involves learning and improving decision-making policies. It strives to reduce costs and enhance the use of renewable energy by determining the best actions for charging using the RL strategy with the best short-term and long-term results. Also, and most importantly, it helps in making sure that the ESS practices a limit of safety through prohibition of its capacity drop to a certain level as well as preventing it from being charged to excess. Therefore, this systematic approach helps manage and optimize the social needs of EVGridNet as an innovation by adapting and guiding the consumption of electricity for charging electric vehicles and ensuring that such usage is efficient for sustainable utilization of energy.

4. Materials and Tools Utilized. For experimental analysis of EVGridNet, Table 4.1 provides a set of hyperparameters along with their optimal values.

An Integrated Energy Management and Forecasting (IEMF) Dataset from [14] is utilized as the lifeblood of the study included in this research work, without which the accuracy of the proposed EVGridNet model cannot be realistically evaluated. The dataset is robust, fragmented, and ranging from Region 1 to Region 5, covering potential EV charging sites positioned along the Chennai-Bangalore highway in India. The specificity of this information - which can be broken down into solar generation, energy consumption, and pricing - is critical for this training and assessment function of EVGridNet. The dataset is dynamic and multi-layered, including the time resolution necessary to capture peak and off-peak fluctuations, spacious variation from varied locations along the highway, and all use cases that can vary in demand. Such a dataset does not only help the decision-making and calibration of the predictive analytics and optimization algorithms run by EVGridNet but also, the model developed from such a dataset operates in the right environment, constitutes a robust substratum for the practical applications, earning the model a chance to scale. EVGridNet utilized the train data and, from IoT devices, collected the attributes that are reflected in Table 4.2.

Overall, this dataset is highly useful for renewable energy management. Understanding the patterns of energy demand and the available supplies assists the operators in forecasting future energy needs. The addition of temperature and weather data also improves the model's ability to predict fluctuations in renewable energy reliability. This forecasting plays an important role in decision-making about energy storage, distribution, and transactions in the energy market.

Table 3.2: Computation Procedure of EVGridNet with DOA

Input: $S(t)$, $[G_P(t)]$, f_{cnn}, C_{req}, $D_{tot}(t)$, $\xi(t)$, $G(t)$, and ρ_α

Output: *optimal policy* $\rightarrow \rho_\alpha^*(A_t|S_t)$ *and the sequence of actions (charging) for each time interval according to the best-optimized policy:* $G(t_i \in T)$

Initialization

- Initialize state space S for a given t and $\xi \in ESS$.
- Initialize A based on potential $G(t)$ actions.
- Frame policy $\rho_\alpha(A_t|S_t)$ parameterized by α.

$\forall$[episodes $\neq$ convergence]

1: Observe initial S

$S_0 = (\xi_0, t_0)$

2: For $t \rightarrow t_0, t_0 + 1, \cdots, n-1$:

Predict $S(t)$and $[G_p(t)]$ *//using CNN-GRU outcome*

Compute A_t based on $\rho_\alpha(A|S_t)$

Perform $A_t$to observe r_t && S_{t+1}

Update $\xi \in ESS \leftrightarrow A_t$ && $S(t)$

Subject to:

$\xi + S(t) - G(t) \geq \xi_{min}$ *//ensures that ESS will not deplete below its minimum capacity*

$G(t) \leq \left[C_{req} - \sum_{t'}^{t}(S(t'))\right]^+$ *//considers the prevention of overcharging beyond* C_{req}

3: Policy Gradient Update

$\forall$steps of the episode

Compute: $\nabla_\alpha f_J(\alpha) = \mathbb{E}[\nabla_\alpha \log \rho_\alpha(A_t|S_t) \times r]$

Update αusing gradient ascent

$\alpha \leftarrow \alpha + \gamma J(\alpha)\nabla_\alpha$ *//*$\gamma \rightarrow$ *learning rate*

4: Dynamic Optimization

4.1 Defining the state

$\Psi(t, \xi)$ represents the minimum cost from t given ξ

4.2 Recurrence Relation

$\Psi(t, \xi) = min_{G(t)\geq 0}\{[G_P(t)] \cdot G(t) + \Psi(t+1, min\{s(t) + \xi - G(t), \xi_t\})\}$

4.3 Base case

$\Psi(k, \xi) = 0$; *//*$\forall \xi$ *at the final step* k.

4.4 Solving Sub-problems

Iteratively compute $\Psi(t, \xi)$ using 4.2.

4.5 Reconstruction of optimal solution

$G(t) = a[\xi, t]$

// 'a' denotes the auxiliary array of A to trace back the decisions

5. Performance Analysis. The realistic assessment for the EVGridNet model is carried out by selecting three state-of-the-art approaches - Dynamic Demand Management Optimization (DDMO), Apriori Progressive Hedging with Sample Average Approximation (APH-SAA) and Long Short-Term Memory (LSTM) networks - that are closely aligned with the main goals of the proposed approach. These approaches provide a solid basis for benchmarking by virtue of their alignment with the intrinsic targets of the electric vehicle charge scheduling optimization, enhancement of renewable energy use, and stabilization of grid interactions, thus supplying the entirety of context for assessing the indicated technological advancements and effectiveness of EVGridNet.

EVGridNet, the proposed EV charging infrastructure optimization model, is a smart technique combining deep learning models (CNNs and GRUs), decision-making technique (RL), and a optimization process (DOA). To evaluate the approach and achieve the work's objectives, the following four metrics and their mathematical computations are considered as a crucial factor of this research. The metrics are solar energy utilization rate (SEUR) [13], grid dependency reduction rate (GDRR) [12], cost savings (CS), and renewable energy utilization Rate (REUR).

SEUR. This metric refers to how much electric vehicle charging power is drawn from solar power. The best way to accomplish that part is to increase (or improve) this rate to wholly correspond with the main objective,

Table 4.1: Hyperparameter Specifications

Hyperparameter	Optimal Value
Learning Rate	0.01
Discount Factor	0.9
Epsilon	0.1
Number of GRU Layers	2
Number of GRU Units per Layer	64
Number of CNN Layers	3
Kernel Size in CNN	3x3
CNN Filters per Layer	32
Batch Size	32
Replay Buffer Size	10000
Policy Update Frequency	5
Training Episodes	1000
ESS Capacity Constraint	100 kWh
Minimum ESS Level	20 kWh

Table 4.2: List of Attributes Utilized in Dataset

List of Dataset Attribute
Timestamp
Solar Power Generation (kWh)
Grid Electricity Price (kWh)
EV Charging Demand (kW)
ESS State of Charge (SoC) (%)
Ambient Temperature (°C)
Weather Conditions
EV Charging Status
User Charging Preferences
Charging Station Utilization

which is to intensify the utilization of renewables for EV charging infrastructure. The primary computation of this metric can be expressed as,

$$SEUR = \sum_{t \in T} S(T) \bullet G(t)_{solar} / \sum_{t \in T} G(t)_{total} \tag{5.1}$$

From 5.1, $S(t)$ denotes available solar power at t, $G(t)_{solar}$ proportion of solar power utilized for charging at t, $G(t)_{total}$ indicates the overall charging power utilized at t.

Figure 5.1 shows that EVGridNet outperforms LSTM, DDMO, and APH-SAA in exploiting solar energy for EV charging under varying scenarios. In scenario 1, while monitoring modest solar availability and regular power demand for EVs, EVGridNet shows performance superiority, with the SEUR of 70% prevailing over the other approaches. Scenario 2 is an ideal situation with high irradiance solar availability and low vehicle demand; EVGridNet comes again on top, achieving a SEUR of 72%. The most underwhelming outcome is Scenario 3, which consists of a low solar availability level and a higher EV demand level, where EVGridNet still provides a SEUR of 75%. Situation 4 indicates that EVGridNet achieves a SEUR of 77% through its ability to apply high solar power efficiently during supply surges. EVGridNet does much better at high SEUR levels in all cases, demonstrating energy system robustness, especially when solar power is in short supply or demand is high. Its adaptability to a broader scope of scenarios is exhibited through changing circumstances. This suggests the smartness and efficiency of EV charging infrastructure management.

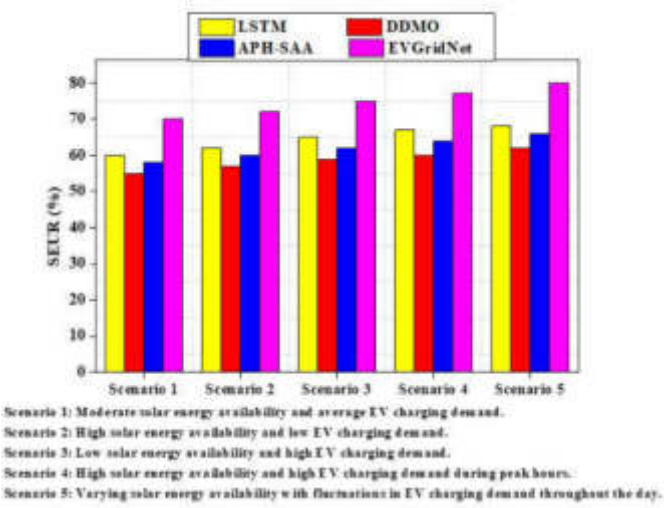


Fig. 5.1: Comparative Evaluation of SEUR across various Scenarios for different Approaches

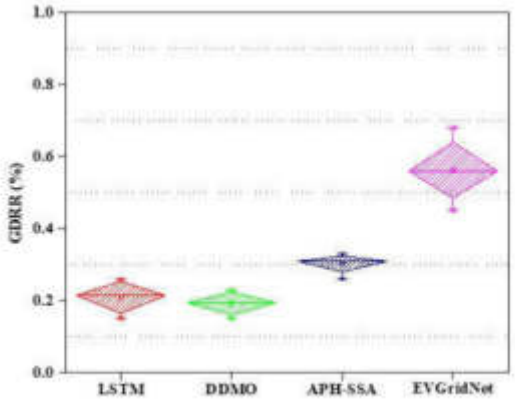


Fig. 5.2: Comparative Evaluation of GDRR across various Scenarios for different Approaches

GDRR. It computes the reduced loading of electricity for EV charging during peak demand periods to minimize the grid's stress duration and maintenance cost. The primary computation process of this metric can be expressed as,

$$GDRR{=}\Delta G_{grid,peak}/G_{grid,peak,baseline} \tag{5.2}$$

From (5.2), $\Delta G_{grid,peak}$ represents the reduction in grid source utilized at peak durations against baseline scenarios, $G_{grid,peak,baseline}$, which indicates the power utilization at peak durations without the induction of EVGridNet processes.

Figure 5.2 shows four EV charging optimization approaches assessed for GDRR metric throughout different scenarios. In Scenario 1, EVGridNet realizes a significant step ahead with the GDRR comprising only 0.6, implying that EVGridNet cuts the grid dependence by 60% during the peak times relative to the baseline scenario. In Scenario 2, EVGridNet narrows this gap further to 0.52, indicating a positive GDRR. Scenario 3 here reveals a relatively tight accordance, at 0.45 levels for EVGridNet and 0.26 levels for APH-SAA. In

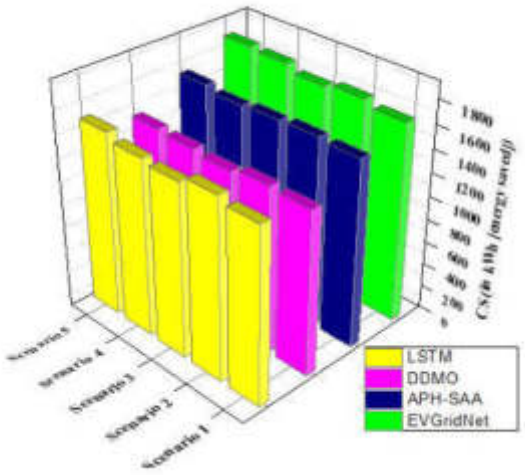


Fig. 5.3: Comparative Investigation of CS across various Scenarios for different Approaches

Table 5.1: Cost Saving Analysis of different Approaches across different Scenarios

Scenario	**CS (in kWh)**			
	LSTM	**DDMO**	**APH-SAA**	**EVGridNet**
Scenario 1	1500	1400	1600	1700
Scenario 2	1550	1420	1620	1750
Scenario 3	1490	1380	1610	1720
Scenario 4	1520	1410	1590	1780
Scenario 5	1580	1430	1630	1800

Scenario 4, the EVGridNet understudies demonstrate the most significant benefit to the network. Thus, it reaches a GDRR of 0.68, and the grid reduction enjoys a 68% decline. In comparison to LSTM (0.18), DDMO (0.23), and APH-SAA (0.33), this result shows that EVGridNet efficiently and effectively manages sources of energy. Overall, EVGridNet has the results of being the better approach in a consistent way, not only relieving grid stress but also optimizing the use of renewable energy for EV charging.

CS. CS ratios encapsulate the savings in charging costs actualized by the dynamic optimization of charging timetables, the application of low-cost grid electricity periods at the right time, and the complete utilization of solar power. The primary computation of CS can be expressed as,

$$CS = \left(\sum_{t\in T} G_P(t) \bullet G(t)_{baseline}\right) - \left(\sum_{t\in T} G_P(t) \bullet G(t)_{optimized}\right) \tag{5.3}$$

The $G(t)_{baseline}$ parameters from equation (5.3) states charging power utilized at t in a baseline scenario, and $G(t)_{optimized}$ denote the charging power utilized at t using optimized patterns.

Figure 5.3 and Table 5.1 depict 3D bar chart data and data table of CS metrics (in kWh), respectively, for EV charging optimization methods concerning various scenarios that construct a thought-provoking picture of electricity consumption reduction with the utilization of those methods. This is evident from the data, as EVGridNet takes the lead, starting at scenario 1 with 1700 kWh and ending with 1800 kWh by scenario 5. This reveals that energy efficiency and cost-effectiveness have, to a large extent, increased compared to the other techniques. It is worth noting that although LSTM provides excellent results, the savings are less, ranging from 1500 kWh to 1580 kWh over the exact scenarios. DDMO and APH-SAA came at the middle rates, with

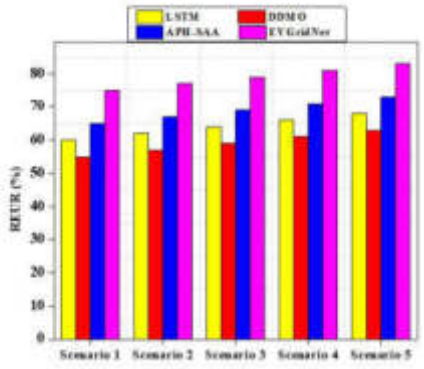


Fig. 5.4: Comparative Evaluation of REUR across various Scenarios for different Approaches

Table 5.2: REUR Analysis of different Approaches across different Scenarios

Scenarios	**REUR (%)**			
	LSTM	**DDMO**	**APH-SAA**	**EVGridNet**
Scenario 1	60	55	65	75
Scenario 2	62	57	67	77
Scenario 3	64	59	69	79
Scenario 4	66	61	71	81
Scenario 5	68	63	73	83

DDMO showing a range from 1400 to 1430 kWh and APH-SAA at a slightly better performance with 1600 to 1630 kWh.

Through Scenario 1 up to Scenario 5, the leap in the amount of savings by EVGridNet illustrates the proactive capacity of the system to effectively manage energy usage, with a high possibility of utilizing renewable sources and periods with low electric rates. The emerging insights in this section suggest that more sophisticated algorithms like EVGridNet make for more sustainable and economical charging facilities. EVGridNet's excellent performance is probably caused by the advanced machine learning techniques in this supervision net, which enables it to predict and adapt to the dynamic of energy generation and demand changes.

REUR. It is quantified as the proportion of renewable energy fed into EV charging, while the entire energy consumed for EV charging during a specific time is taken into consideration. The computation of REUR can be expressed as,

$$REUR{=}\Phi/\tau \tag{5.4}$$

From (5.4), τ denotes the overall energy consumed for EV charging, and Φ indicates the overall renewable energy consumed for EV charging.

Figure 5.4 and Table 5.2 display REUR percentage values for five possible scenarios obtained by applying these strategies for four different EV charging optimization approaches. Despite having a higher REUR percentage (75% to 83%) compared to the other scenarios throughout Scenarios 1 to 5, EVGridNet yields the highest RUE value at all times. Such performance shows a significant correlation, giving an elevated advantage to renewables among EV charging systems.

The results of the LSTM approach are moderate, achieving REUR rates as high as 68% without making any progress better than the previous value of 60%. In contrast, the performance of the DDMO approach is the poorest, with REUR values starting at 55% and only increasing to 63%. The general trend is that APH-SAA's performance is not as good as other neural networks (LSTM and DDMO) in real-life scenarios (accurate EV loads), but it is still better than EVGridNet (REUR varies from 65% to 73%). These outcomes demonstrate the

best EVGridNet efficiency, allowing the automatic scheduling of EV charging to raise the amounts of renewable energy that transforms into usable electricity. This fact that it is better compared to the performance of other systems suggests that it can adaptively manage the fluctuating energy sources in the electric system, thus causing efficient and sustainable charging service. The statistics of more than improved REUR percentage from EVGridNet confirm its possible application in a more extensive network, backing its scale and sustainability.

Thus, the usage of renewable energy can greatly increase due to the possibility of the EVGridNet model to optimize the patterns of EV charging by predicting solar energy production and grid prices accurately. It is not only beneficial in lowering the amount of carbon emission but also useful in shaping efficient transport modes. Through more efficient utilization of renewable energy sources, this particular model supports the design of a cleaner and environment-friendly future. At times, when many people return home from work, study, or engage themselves in other tasks, there is typically a significant demand for electricity. To mitigate the effects of this, EVGridNet charges electric cars for other parts of the day. This strategic approach can result in the improvement of the stability of grid operations, a reduction in energy costs and the probability of overloaded circuits, and, in some cases, a complete blackout. Thus, these replacements will drastically improve the grid load management process and, hence, contribute towards a more efficient power distribution system. The available charging times in most areas allow one to take advantage of cheaper night tariffs; thus, owners of EVs can save a lot of money. Such economic benefits may encourage more people to use electric technology cars as a method of transport, thus achieving a faster transition to sustainable transport. This is simple because reducing the cost of owning these cars has remained the most compelling reason for popularizing the use of electric cars in society. The ability of the EVGridNet model to change its structure allows it to be extended successfully to other regions and different conditions of the energy market. This is an important aspect given that the model shall be used in different settings to allow it to be used by many. However, when it comes to the application of smart grid technologies, two critical components should be considered: scalability and the ability to adapt. The blend of deep learning hierarchies [17] with reinforcement learning, along with the dynamic programming in the construction of EVGridNet, provides high technological standards for future advances in smart grid and renewable energy technologies. This approach can spur more advanced and updated research to come up with efficient energy systems as new ideas are discovered.

The effectiveness of the model within the proposed EVGridNet entirely depends on the real-time information obtained from the IoT devices as well as the weather and the energy market data. This means that the performance of the model can be impacted by inadequate or untimely data information, thus the need for correct and timely data information. The main issue to be noted is data dependency, which is a major problem that requires a solution for operations' proper performance. As regards the application of this research, the following are noted as follows:

Optimized EV Charging Infrastructure: Overall, the design of the EVGridNet model can improve the effectiveness and/or sustainability of charging stations for electric vehicles by managing the degree of reliance on the grid during peak hours and utilizing renewable energy resources.

Cost Savings for EV Owners: By shifting charging to off-peak times and utilizing cheaper electricity, the model can help EV owners save on energy costs.

Grid Load Management: The model described in this paper has the potential to revolutionize grid load management. Utility companies can use this model to accurately regulate loads, preventing overloads and reducing the risk of blackouts, thereby ensuring a more reliable and stable energy supply.

Scalability for Diverse Regions: Many such characteristics enable the implementation of the model across diverse regions, which have different energy market characteristics, thus encouraging green transportation solutions.

6. Conclusion and Future Work. The EVGridNet model demonstrates considerable potential in enhancing the efficiency and sustainability of EV infrastructure by integrating renewable energy sources. By addressing the inconsistencies between EV charging supply and demand, EVGridNet shifts charging activities to off-peak hours. This strategic approach not only reduces the reliance on the grid during peak times but also significantly increases the proportion of renewable energy used in the charging process. The simulation results indicate that this methodology can effectively reduce grid strain and optimize energy use, making it a robust solution for current energy challenges. Moreover, EVGridNet's capability to adapt to varying conditions

positions it as a foundational technology for developing more eco-friendly and adaptive plug-in vehicle charging systems, thereby promoting a greener and more sustainable transportation future. On the other hand, we are trying to deploy upper-level enhancement models by integrating more exact weather prediction models that can accurately anticipate solar generations. Furthermore, machine learning algorithms implementing adaptive user behavior patterns provide a more personalized charging schedule.

REFERENCES

[1] Wu, Y., Lu, Y., Zhu, Z., & Holguín-Veras, J. (2023). Optimizing Electric Vehicle Charging Infrastructure on Highways: A Multi-Agent-Based Planning Approach. Sustainability, 15(18), 13634. https://doi.org/10.3390/su151813634.

[2] Richardson, D.B., 2013. Electric vehicles and the electric grid: A review of modeling approaches, Impacts, and renewable energy integration. Renewable and Sustainable Energy Reviews, 19, pp.247-254.

[3] Solak, M., Faydasicok, O., & Arik, S. (2023). A general framework for robust stability analysis of neural networks with discrete time delays. Neural Networks, 162, 186-198.

[4] Hopkins, E., Potoglou, D., Orford, S., & Cipcigan, L. (2023). Can the equitable roll out of electric vehicle charging infrastructure be achieved. Renewable and Sustainable Energy Reviews, 182, 113398.

[5] Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. Computation, 11(3), 52.

[6] Shakya, A. K., Pillai, G., & Chakrabarty, S. (2023). Reinforcement learning algorithms: A brief survey. Expert Systems with Applications, 120495.

[7] Chen, C., Song, Y., Hu, X., & Guardiola, I. G. (2020). Analysis of Electric Vehicle Charging Behavior Patterns with Function Principal Component Analysis Approach. Journal of Advanced Transportation, 2020, 1–12. https://doi.org/10.1155/2020/8850654.

[8] Geetha, T. S., Amudha, V., & Chellaswamy, C. (2022). A Novel Dynamic Capacity Expansion Framework Includes Renewable Energy Sources for an Electric Vehicle Charging Station. International Transactions on Electrical Energy Systems, 2022, 1–25. https://doi.org/10.1155/2022/4813750.

[9] Shibl, M., Ismail, L., & Massoud, A. (2021). Electric Vehicles Charging Management Using Machine Learning Considering Fast Charging and Vehicle-to-Grid Operation. Energies, 14(19), 6199. https://doi.org/10.3390/en14196199.

[10] Mazhar, T., Asif, R. N., Malik, M. A., Nadeem, M. A., Haq, I., Iqbal, M., Kamran, M., & Ashraf, S. (2023). Electric Vehicle Charging System in the Smart Grid Using Different Machine Learning Methods. Sustainability, 15(3), 2603. https://doi.org/10.3390/su15032603.

[11] Farhadi, F., Wang, S., Palacin, R., & Blythe, P. (2023). Data-driven multi-objective optimization for electric vehicle charging infrastructure. IScience, 26(10), 107737. https://doi.org/10.1016/j.isci.2023.107737.

[12] Wu, Y., Lu, Y., Zhu, Z., & Holguín-Veras, J. (2023). Optimizing Electric Vehicle Charging Infrastructure on Highways: A Multi-Agent-Based Planning Approach. Sustainability, 15(18), 13634. https://doi.org/10.3390/su151813634.

[13] Barman, P., Dutta, L., Bordoloi, S., Kalita, A., Buragohain, P., Bharali, S., & Azzopardi, B. (2023). Renewable energy integration with electric vehicle technology: A review of the existing smart charging approaches. Renewable and Sustainable Energy Reviews, 183, 113518.

[14] Karthick Raghunath K M. (2023). Integrated Energy Management and Forecasting Dataset. IEEE Dataport. https://dx.doi.org/10.21227/zrgp-kh30.

[15] Kumbhar, A., Patil, N., Narule, M., Nadaf, S. M., & Basha, C. H. (2023). Reducing grid dependency and operating cost of micro grids with effective coordination of renewable and electric vehicle's storage. In Soft Computing for Problem Solving: Proceedings of the SocProS 2022 (pp. 639-653). Singapore: Springer Nature Singapore.

[16] Zheng, J., Zhu, J., & Xi, H. (2023). Short-term energy consumption prediction of electric vehicle charging station using attentional feature engineering and multi-sequence stacked Gated Recurrent Unit. Computers and Electrical Engineering, 108, 108694.

[17] Karthick Raghunath. K. M., K. S. Arvind, K. S. Suganya, and R. Vinothsaravanan, (2023). Smart Air Pollution Monitoring and Prediction System Using Deep Learning Enabled IoT Technology" for the book named "Artificial Intelligence for Future Intelligent Transportation", Chapter No. 3, ISBN 9781774913529, 2024.

Edited by: Dhilip Kumar V
Special issue on: Unleashing the power of Edge AI for Scalable Image and Video Processing
Received: May 8, 2024
Accepted: Aug 9, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 599–609, DOI 10.12694/scpe.v26i2.4077

DESIGN AND DEVELOPMENT OF MEMORY PIXEL ARCHITECTURE FOR SOBEL EDGE DETECTION

YELLAMRAJU SRI CHAKRAPANI *, NANDANAVANAM VENKATESWARA RAO †, AND MADDU KAMARAJU ‡

Abstract. Memory is a fundamental hardware structure used in various electronic and multi-media products. The design and development of large-size memory architectures have become increasingly complex due to the growing demands of high-resolution image processing applications. This research work aimes to design a specialized memory architecture module for Sobel edge detection, an essential technique in image processing. The proposed architecture (memory pixel size) consists of memory unit, comprising rows and columns, is determined by the image resolution. For effective Sobel edge detection, the image pixels must be stored in memory, and read operations are performed to access a 3x3 matrix of nine pixels. A critical consideration in developing this memory architecture is power dissipation, which is mitigated through the application of Clock Gating techniques. The proposed pixel memory architecture is implemented using MATLAB and Xilinx ISE software with Verilog HDL. The image pixel memory is developed using Block RAMs (BRAMs) and registers, and the 3x3 pixel matrix required for the Sobel edge detector is generated. Simulation, synthesis, and power analysis are conducted for the image pixel memory across source images with various resolutions, including 10x40, 10x20, 128x128, 320x240, and 512x512. The results indicate reduced power dissipation from 30% to 40% due to Clock Gating. This work demonstrates the effectiveness of the proposed memory architecture in reducing power consumption while maintaining performance. Future work aims to further enhance the performance of image pixel memory by decreasing the number of registers and improving pixel access times. Such module will provide a more efficient and scalable solution for high-resolution image processing applications.

Key words: Memory architecture, Image pixels, Power dissipation, Clock Gating

1. Introduction. An image is divided into small elements called pixels, each representing a different color. For example, an image with a resolution of m by n has m x n pixels. A 240 by 320 resolution image contains 76,800 pixels. The pixels are arranged in a matrix, where 'm' denotes the number of rows and 'n' denotes the number of columns. The pixel values, which range from 0 to 255, represent their intensities. In a grayscale image, a pixel intensity of 0 represents black, while an intensity of 255 represents white. Each pixel can be stored in an 8-bit register. The memory required to store the pixels is determined by the image resolution, with the address range corresponding to the number of rows. The data retrieved from a specified address represents the pixels in the columns [1].

Designing the hardware for image pixel memory is complex, especially when considering FPGA, due to the numerous registers and block memories required. Additionally, power dissipation is a crucial parameter to consider. This paper focuses on reducing power dissipation in the designed image pixel memory.

Few researchers have explored the development of memory architectures specifically for applications like Sobel edge detection systems. These architectures use input and output buffers, BRAM (Block Random Access Memory), registers, etc. Various methods, such as using external memory, registers with data reuse, BRAMs with data reuse, and combinations of registers and BRAMs with data reuse, are employed in designing image pixel memory for Sobel edge detection. Partitioning image frames into BRAMs utilizes all FPGA resources and reduces power dissipation . Techniques like partitioning, data reuse, loop pipelining, and merging are integrated into an optimized flow for behavioral synthesis . The clock gating technique, synchronized with the global clock, analyzes two clock gating methods.

*Department of ECE, Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India (srichakrapani@gmail.com).
†Department of ECE, Bapatla Engineering College, Bapatla, Andhra Pradesh, India (nvr68@gmail.com).,
‡Department of ECE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India (profmkr@gmail.com.).

A comparison of clock gating techniques implemented on FPGA and ASIC shows that FPGA implementation reduces dynamic power consumption by 50% to 80% compared to ASIC. Several low power dissipation techniques for System on Chip (SoC), such as clock and power gating, multi-voltage, and voltage scaling, are discussed. The RTL clock gating method with scan techniques is implemented in Synopsys Power Compiler to reduce power consumption and enhance Design For Test (DFT). Energy consumption from computations can be reduced using parallelism in FPGA memory architecture . The concept of gated clocking involves stopping the clock signal when idle, thereby saving power consumed by latches in this state.

In hardware, the Frame Buffer, Line Buffers, and Pixel Window are the most important parts for performing Sobel edge detection. The whole picture is kept in memory by the Frame Buffer, usually in the form of a 2D array with a pixel for each piece. In order to keep the current line and the two lines before it and after it in memory while the picture is being processed, line buffers are necessary. A 3x3 Sobel operator requires a minimum of three line buffers. By getting the correct neighborhood from the line buffers, the 3x3 Pixel Window may traverse the image and apply the Sobel operator to each pixel. Effective and efficient edge identification over the whole image is guaranteed by this organized technique [2].

1.1. Low Power Dissipation. Nowadays, electronic gadgets operate on batteries, which require more battery charge and life. To satisfy these requirements, power dissipation is an important factor in developing an Integrated Circuit (IC) in these electronic gadgets using CMOS VLSI technology. If low power dissipation is achieved, battery consumption will also be reduced. There are three sources of power dissipation, i.e leakage (static) power, short-circuit power, and dynamic (switching) power [8]. Whenever the system-circuit is in idle mode, then the power dissipated is static, which depends on the leakage current that occurs due to the flow of minority carriers in the sub-threshold region. The short circuit power dissipation arises whenever both NMOS and PMOS transistors are in active or saturation regions, i.e the power supply is directly connected to the ground. The dynamic power dissipation occurs whenever the system circuit signals change, which depends on switching activity per node (β), Switched Capacitance (C), Frequency (switching events per second, F), and Supply Voltage (VDD).

The power reduction methods are suggested at various abstraction levels of CMOS VLSI design flow such as system level, algorithm level, logic or gate level, transistor level, transistor level, etc. The power optimization methods in system level are low frequency clocks, off-chip components like ROM, RAM integration, etc. In algorithm level, minimizing the no. of operations, conditions and loop iterations will reduce the power dissipation. Parallel, pipelining, arithmetic architectures are used to minimize power dissipation in architecture level.

To optimize the power in logic or gate level, switching activity reduction, clock and bus loading optimization are proposed. In transistor level, the methods such transistor sizing, multi- threshold voltages etc. are employed for low power. If the methods like device scaling, optimization in placement and routing are applied for device level, then low power dissipation in obtained. In this paper, gate or logic level optimization of power dissipation is proposed by using gated clock technique for the implementation of image pixel memory for sobel edge detection system An efficient architecture for image pixel memory usually entails arranging pixel data such that it can be accessed and processed quickly. A 3x3 neighbourhood of pixels must be readily accessible for Sobel edge detection to work. This necessitates giving serious thought to the storage and access of pixels.

2. Literature survey. There has been a lot of research on memory architecture design and optimization for image processing applications, especially Sobel edge detection. This literature review summarises significant findings in optimizing FPGA memory designs, memory allocation, power minimization, and low-power designs using clock gating techniques. Optimal FPGA memory architectures for Sobel edge detection were the primary focus of Harald Devos and Dirk Stroobandt [1]. In this work proved that optimized memory architectures are crucial to making FPGA-based edge detection algorithms work better. This research shows how difficult it is to create FPGA-based solutions while balancing memory size, access speed, and computational efficiency. Deepayam Bhowmik, Robert Stewart, Greg Michaelson, and Andrew Wallace [2] investigated solutions for optimized memory allocation and power minimization for FPGA-based image processing. Their work, tackled the pressing issue of reducing power consumption and memory utilisation in high-resolution image processing jobs. Optimal memory allocation was highlighted as a critical component in their approaches to drastically reduce power consumption without sacrificing performance. In [3], Peng Zhang, Xu Cheng, and Jason Cong presented an integrated and automated memory optimization method for FPGA behavioral synthesis. This

Table 2.1: Comparison of Various Techniques for Sobel Edge Detection Architecture

Details	Technique	Research Gap	Limitations
Joshi, R et.al 2020 [11]	Novel bit-sliced near-memory computing	Complex Design	High power consumption
Osman, Z.E.M et al. 2010 [12]	Optimized processor	High Latency	Critical architecture
Singh, S et.al 2014 [13]	Analysis of hardware architectures	Deep analysis	Future demands
Kumar, S.et.al 2013 [14]	Segmentation with edge detection	FPGA prototype	Latency
Saidani, T et.al 2024 [15]	Model-based design method	Model edge with IoT	Power consumption
Pudi, D et.al 2023 [16]	DRRA and DiMArch architectures	Cannot support IoT models	Operational efficiency is less
Chang, Q et.al 2023 [17]	Multi-directional kernel	Operations possible with GPU processor	High power consumption
Yamini, V et.al 2024 [18]	SoC design with edge	Critical architecture	Less performance
Khalil, A. S et.al 2023 [19]	Enhanced system on chip	Less computational module	High latency
Orthy, M., et.al 2023 [20]	FPGA-based image	Faster performance Sobel edge model	High power consumption

methodology is optimizing memory utilization during behavioral synthesis of FPGA designs . By automating the optimization process, the suggested flow hoped to increase memory efficiency while decreasing design time. Low-power sequential circuit design using clock gating methods was investigated by M. Pedram and Xunwei Wu [4]. By selectively blocking the clock signal to inactive circuit portions, gives Fundamental Theory and Applications—showed how clock gating may drastically cut power usage. To design memory structures that use less power, this method is essential. A new and comparative assessment of clock gating in FPGAs was given by J. Raivainen and A. Mammela [5]. The work was concentrated on how clock gating may be used to reduce power consumption in FLPGA architectures. The study shed light on the advantages and methods of clock gating in systems that rely on field-programmable gate arrays (FPGAs).

Low Power Methodology Manual For System-on-Chip Design, written by David Flynn, Robert Aitken, Alan Gibbons, and Kaijian Shi [6], is an all-inclusive reference on low power design approaches, such as clock gating. An excellent resource for engineers and designers, this document offers practical solutions for attaining low power consumption in system-on-chip (SoC) systems.

Power reduction by RTL clock gating was covered in Mark Biegel's [7] presentation at SNUG, San Jose. The goal of Biegel's research was to find ways to drastically reduce power consumption by using clock gating at the register-transfer level (RTL). By incorporating power-saving strategies early on in the design process, designers may create more efficient goods in the end.

Methods for reducing digital CMOS circuit power consumption were investigated by R. W. Brodersen [8]. His work laid the groundwork for low-power design in digital circuits, including techniques like clock gating.

The effect of memory design and parallelism on FPGA communication energy was studied by A. Dehon and D. Lakata [9]. Their research showed that FPGA designs might minimize communication energy via the use of parallel processing and optimized memory structures. The significance of memory design to FPGA systems' total energy efficiency is highlighted by this study. To reduce power consumption in sequential circuits, methods for synthesizing gated clocks were proposed by P. Siegel and G. D. Micheli [10]. Their work, which was Tested and contributed to the field of low-power electronics by providing realistic approaches for incorporating clock gating in sequential circuit designs.

Table 2.1 clealry explains about various sobel edge detection models and its limitations.

3. Image Pixel Memory Architecture for Sobel Edge Detection. The proposed image pixel memory architecture for generating sobel edge detection pixel matrix is designed and developed for m x n image resolution, which contains m rows and n columns, as shown in Fig 3.1. The address is used to access the rows

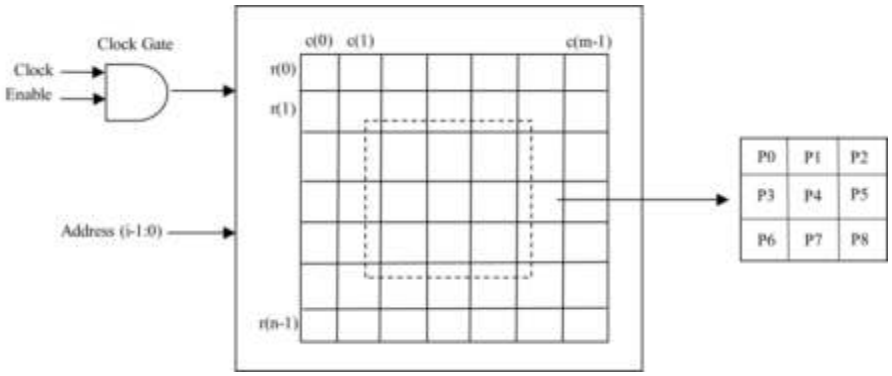


Fig. 3.1: Image pixel memory system and Pixel matrix for Sobel edge detection

from r(0) to r(n-1). Three consecutive rows are accessed to generate P0 to P8 pixels, representing the input matrix for computing sobel edge detection, and each pixel in the memory location comprises 8 bits. The 3x3 pixel matrix will be continuously generated from the image pixel memory from all the successive three rows and all columns i.e. first three rows r(0) to r(2) and columns c(0) to c(m-1), after that rows r(1) to r(3) and columns c(0) to (m-1), in the last rows r(n-3) to r(n-1) and columns c(0) to (m-1). The power dissipated during the implementation of image pixel memory can be minimized by employing clock gating [21]. In this technique, an enable signal is used for controlling the clock signal, i.e., whenever the clock signal is unnecessary, the enable signal is '0', and when it is required, the enable signal is '1'. So, the power consumed by the clock signal is reduced whenever idle or not required [22].

3.1. Image Dataset for experiment. In this section, dataset availability and its testing process have been explained. Many different picture libraries and image-processing tools may be found on SourceForge, an active site for open-source projects [23]. Anyone looking for open-source image libraries may utilize SourceForge's search tool to locate them. Just type in terms like "image library" or "image processing." Finding these projects is also more accessible by perusing the "Software" section and sorting by applicable categories. The OpenCV, ImageMagick, and GIMP open-source image libraries are some of the most prominent ones on SourceForge [24]. OpenCV's many picture and video analysis features have earned it a reputation as a powerful tool in computer vision and image processing. ImageMagick offers a powerful set of tools and libraries for working with bitmap pictures, whether in creation, editing, or conversion. While the main use of GIMP is to edit images, it also has scripts and plugins that may be utilized for a variety of image processing jobs [25].

3.2. Implementation of proposed Architecture. The entire process of implementing image pixel memory and matrix generation is shown in the Fig 3.2. First a colour source image is taken and then it is converted to grey-scale image, which is the source image in this process. The source images with different resolutions are taken, and the pixels are extracted as hexadecimal values by using MATLAB software, which are given as an input text file to Xilinx ISE Software.

The first thing that happens while processing images is reading them from memory. Then, a typical method for drawing attention to borders and transitions, Sobel edge detection, is employed to locate picture edges. Data extraction from random access memory (RAM) is the following step, after which pertinent picture information is retrieved. If you're working with picture pixel data in a line buffer, you may need to convert the values into hexadecimal format. The extraction of a 3x3 matrix from the picture data is an essential step because it is often utilised in many image processing methods. As a last step, a text file is generated from the processed picture data, containing the outcomes of all these actions. To keep everything running smoothly and at the correct time, it may be necessary to use a gated clock signal with clock enable outputs at various points in this process.

In this software, image pixel memory and matrix generation for the Sobel edge detection are developed using Verilog HDL as shown in Algorithm 3. The image pixels memory is generated according to the rows

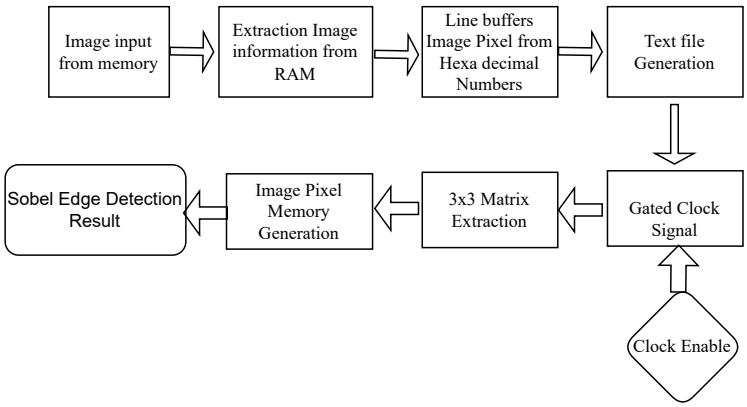


Fig. 3.2: Implementation process of image pixel memory and matrix generation for Sobel edge detection

and columns of image resolution. The 3x3 pixel matrix is formed using accessing the address of corresponding three rows and three columns. Clock gating technique is applied in the processing of image pixel memory and pixel matrix generation for dissipating less power. The clock gating signal is generated by performing AND operation with the clock and enabling signals.

Algorithm 3 Image Pixel Memory and Matrix Generation for Sobel Edge Detection

Require: Image width W, height H, and pixel data file in hexadecimal format
Ensure: 3x3 pixel matrices for Sobel edge detection
Initialize Parameters:
Set image dimensions $W \times H$
Allocate RAM memory sufficient to store the entire image based on its resolution
Load Pixel Data:
Read the pixels from the hexadecimal text file into RAM memory
Generate Gated Clock Signal:
Perform an AND operation between the clock and enable signals to obtain a gated clock signal
Process Image for Sobel Edge Detection:
while gated clock signal is positive edge **do**
    Step 1: Copy to Line Buffers
    Read pixel values from three consecutive rows in RAM into three line buffers
    Step 2: Extract 8-bit Pixel Matrix
    Extract 8-bit pixel values from the line buffers to form a 3x3 pixel matrix
    Ensure the correct 3x3 neighborhood by shifting the contents of the line buffers
    Step 3: Output Pixel Matrix
    Output the 3x3 pixel matrix for Sobel edge detection
    Step 4: Iterate Through Rows
    Move the 3x3 window across the entire image by iterating through all the rows
    Ensure each pixel is processed for edge detection
end while

Sobel edge detection's architectural design is shown in the block diagram that was supplied. One of the main goals is to use a 3x3 matrix and a threshold to find out whether each pixel is an edge by calculating its gradient magnitude.

Beginning with the 3x3 pixel matrix, every pixel, denoted as P1 through P9, has a width of 8 bits. The

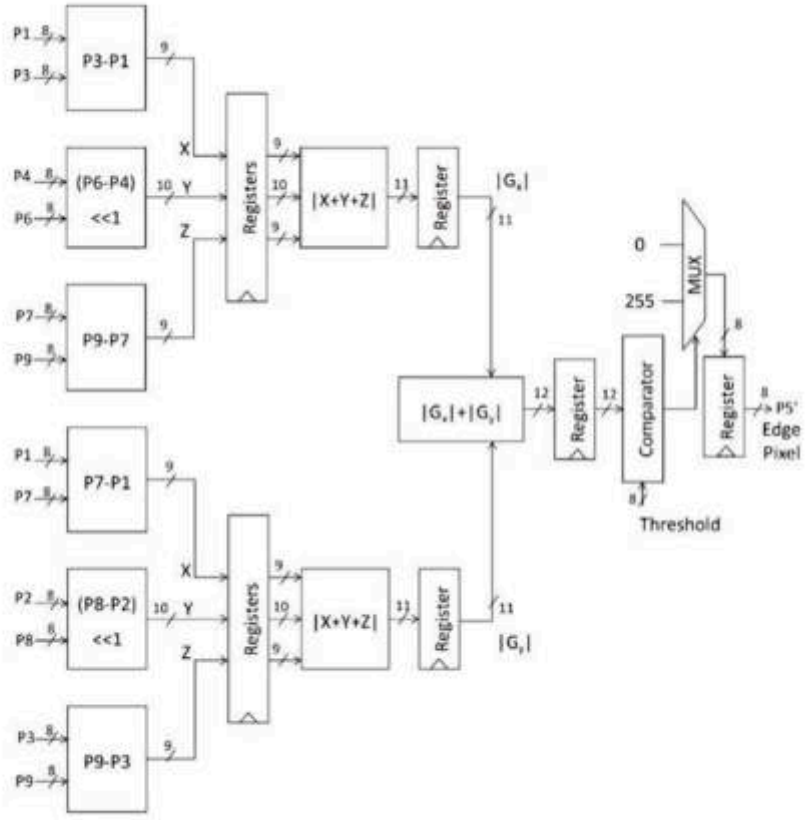


Fig. 3.3: Design Architecture for Sobel Edge Detection Module

horizontal and vertical gradients are computed by performing subtraction operations between certain pairs of pixels. To illustrate, the P3-P1 block takes the value of pixel P3 and subtracts it from pixel P1, producing an output with 9 bits. Other subtractor blocks, such as P6-P4, P9-P7, P7-P1, P8-P2, and P9-P3, carry out similar operations.

With the addition of the shift left operation ($\ll 1$), specific pixel values are effectively multiplied by 2 to highlight their role in the gradient computation, in addition to the subtraction operations. Next, adders labelled |X+Y+Z| are used to total the outcomes of these operations. After each round of subtractions and shifts, these adders take the absolute value of the total, checking that the gradient values are positive.

Combining the outputs of the required subtractor and shift blocks allows one to determine the horizontal gradient (Gx) and the vertical gradient (Gy). The absolute values of these gradients are briefly stored in registers in order to synchronise the processing phases. After that, we get a 12-bit value—the total gradient magnitude – by adding the absolute values of the horizontal and vertical gradients, |Gx| and |Gy|, respectively.

A comparator block is used to compare the magnitude of this gradient to a preset threshold. A pixel's inclusion or exclusion from an edge is determined by the threshold, an 8-bit integer whether the gradient magnitude is greater than the threshold determines the comparator's signal output.

The output of the comparator is then used by a multiplexer (MUX) to determine the final value of the edge pixels. The MUX will output 255 to indicate the existence of an edge if the gradient magnitude is greater than the threshold. Without an edge, the MUX will return a value of 0 if the gradient magnitude is less than the threshold. The value of the edge-detected pixel is represented by the final output, which is labeled as P5'.

(a) (b)

Fig. 4.1: a) Source image of resolution 128x128 b) Image pixels as hexadecimal values in text file

Table 4.1: Error Analysis on various test input images

Input Image	Description	Total Pixels	Error (in Pixels)	Error (%)
Image001	zero	1024	0	0
Image002	one	1024	4	0.16
Image003	two	1024	7	0.45
Image004	three	1024	3	0.21
Image005	four	1024	8	0.72
Image006	five	1024	6	0.58
Image007	six	1024	7	0.56
Image008	seven	1024	5	0.51
Image009	eight	1024	6	0.59
Image010	nine	1024	3	0.79

Finally, this design effectively executes Sobel edge detection on the input pixel data, producing a binary signal of edge existence according to the calculated gradients and the given threshold.

4. Results and Discussion. In this section, a brief discussion of Sobel edge detection for image pixel memory design of testing outcomes has been discussed. The source image of resolution 128x128, shown in Fig. 4.1a, is taken for generating image pixel memory and a 3x3 pixel matrix. This source image is given as input to MATLAB software. Then, image pixels are generated in a text file as hexadecimal values, as shown in Fig. 4.1b. The text file, which consists of image pixels in hexadecimal values, is given as input to Xilinx ISE software. Then, the pixels matrix generates image pixel memory using Verilog HDL. The implemented image pixel memory is made up of BRAMs and registers. The simulation process for the image pixel memory and pixels 'P0' to 'P8', is shown in Fig. 4.1.

Images 001 through 010 make up the test inputs in the table 4.1, which displays an error analysis for each. The analysis checks for pixel accuracy problems in each picture, which has a total of 1024 pixels. Image001, standing for the number zero, has a 0% mistake rate according to the statistics. On the other hand, Image002, which stands for the number one, has four mistakes—a rate of 0.16%. There are seven mistakes in Image003, which represents the number two, resulting in an error rate of 0.45%. Image004 (the third digit) has an error rate of 0.21%, with a total of 3 errors. Image005, which stands for the number four, has 8 mistakes, giving it an error rate of 0.72%. With a total of six mistakes, the error rate for Image006 (the fifth digit) is 0.58%. The error rate for Image007 (digit six) is 0.56%, with seven mistakes. The 0.51% error rate results from 5 mistakes in Image008, which represents the number 7. The error rate for Image009 (digit eight) is 0.59%, with six mistakes. At 0.79%, Image010 (the ninth picture) has the greatest error rate of all the photos due to its three mistakes. By measuring and comparing the error rates across many test pictures, this research sheds light

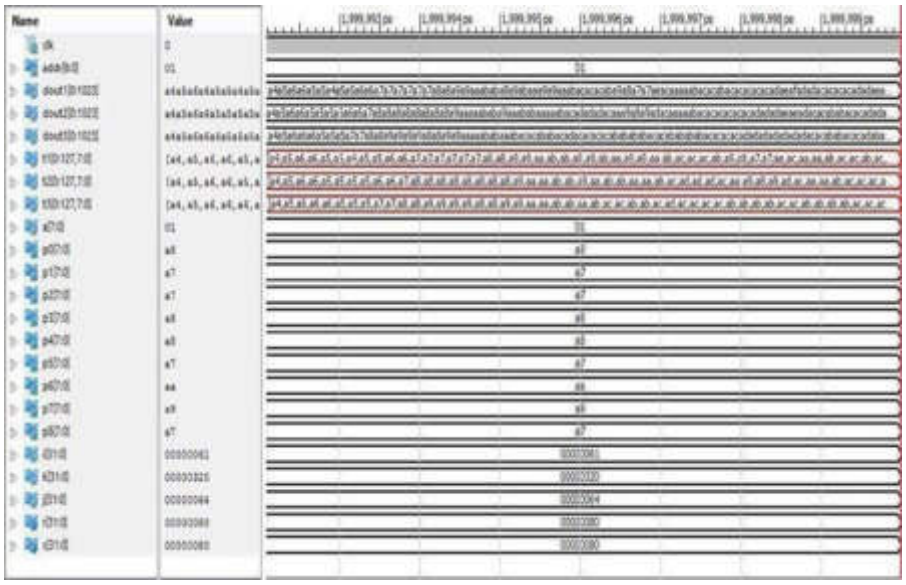


Fig. 4.2: Simulation waveforms for the image pixel memory and output pixels P0 to P8

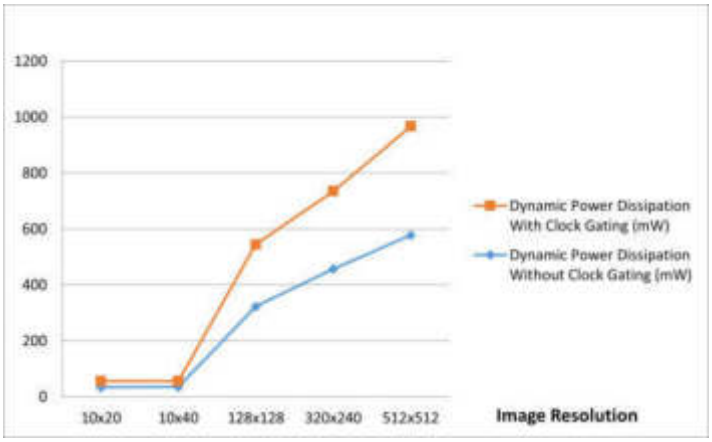


Fig. 4.3: Power dissipation of image pixel memory and 3x3 pixel matrix generation for different image resolutions

on the precision of image processing.

The parameters ‘r’ and ‘c represents the rows and columns of the image pixel memory and ‘addr’ is the address register which is used to access the rows. The registers ‘dout1’, ‘dout2’, ‘dout3’ represents the combined pixels data stored in three consecutive rows respectively and ‘t1’, ‘t2’, ‘t3’ are the registers that represent the individual pixels values in 8-bit format

The synthesis and power analysis process were performed after simulation. The power analysis of image pixel memory for 3x3 pixel matrix generation, is performed using XPower Analyzer of Xilinx ISE for different source image resolutions with and without clock gating, whose power dissipation values are shown in Fig. 4.2. With reference to Fig. 4.3 and by comparing source images with various resolutions with and without clock gating, the approximate dynamic power dissipation reduces by 31% to 40%. As the resolution of the source image increases the power dissipation reduction decreases. For example, the source image with resolution 10x40 has power reduction of 40%, whereas for source image with resolution 128x128 has power reduction of 31%.

Table 4.2: Comparison of Enhanced Edge Detection Models

Enhanced Edge Detection Model	Power Consumption	Latency Improvement	Performance
SoC Sobel Edge [19]	0.2	0.2	0.15
Multi-directional Sobel Operator Kernel [17]	0.22	0.25	0.13
ML Model-based Design Method [15]	0.25	0.26	0.2
Proposed Method	0.4	0.3	0.26

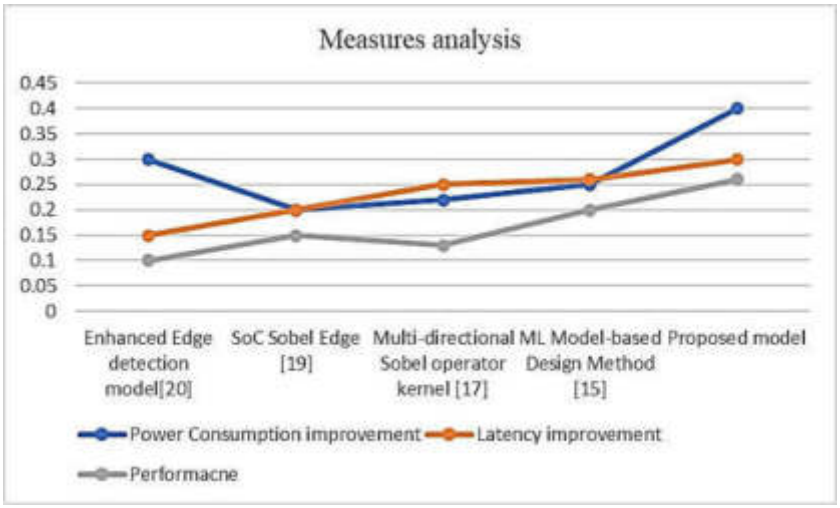


Fig. 4.4: Performance analysis

Table 4.2 and Figure 4.4 compare enhanced edge detection models based on three criteria: power consumption, latency improvement, and performance. The models compared include the SoC Sobel Edge method, the Multi-directional Sobel Operator Kernel, the ML Model-based Design Method, and the Proposed Method.

The SoC Sobel Edge model consumes 0.2 units of power, improves latency by 0.2 units, and has a performance rating of 0.15. The Multi-directional Sobel Operator Kernel consumes slightly more power at 0.22 units, achieves a latency improvement of 0.25 units, and has a performance rating of 0.13. The ML Model-based Design Method shows higher power consumption at 0.25 units, with a latency improvement of 0.26 units and a performance rating of 0.2. The Proposed Method has the highest power consumption at 0.4 units, but it also offers the most significant latency improvement of 0.3 units and the best performance rating of 0.26.

Edges in an image are critical features that mark regions where there is a sudden change in pixel intensity. In a grayscale image, which varies from black (representing the lowest intensity) to white (the highest intensity), edges appear where the intensity changes abruptly. To detect these changes, we analyze the slope of the intensity curve, which is effectively the first derivative of the intensity function.

The process of detecting edges involves a mathematical operation known as convolution. This operation applies a small matrix, or kernel, to the image to estimate derivatives. One of the most common edge detection techniques is the Sobel Operator, which utilizes two distinct kernels—one for detecting vertical edges and another for horizontal edges.

In vertical edge detection, the Sobel operator uses a 3x3 kernel to compute the gradient in the x-direction, which highlights horizontal edges by convolving this kernel with the image. Similarly, for detecting horizontal edges, another 3x3 kernel is used to calculate the gradient in the y-direction, which reveals vertical edges. By combining the gradients from both directions, we can determine the overall gradient magnitude at each pixel. This combined gradient helps identify edges where the magnitude exceeds a specified threshold.

Sobel Edge Detection is widely applied in various image processing tasks. It is instrumental in object

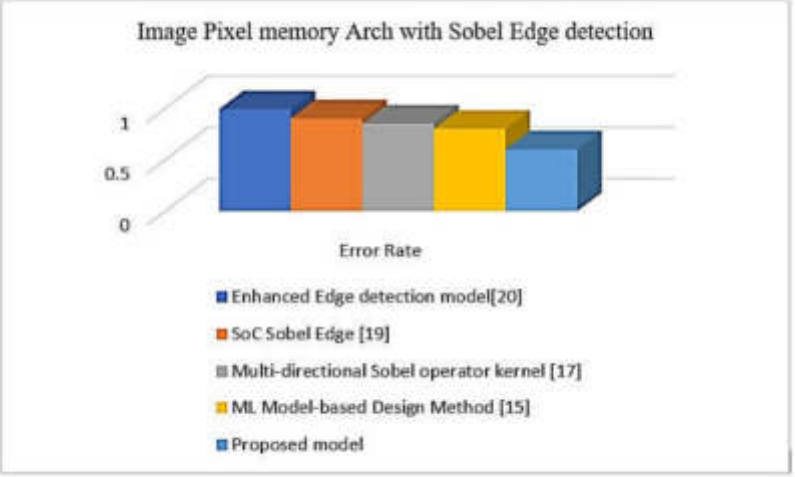


Fig. 4.5: Performance analysis

recognition, where it helps to identify the boundaries of objects within an image. It is also used in image segmentation to separate regions of interest and in pattern recognition to detect specific shapes or features. These practical applications highlight the importance of edge detection in analyzing and interpreting visual information shown in figure 4.5.

5. Conclusion. In this paper, the development and analysis of image pixel memory are thoroughly examined through the use of Block RAMs (BRAMs) and registers. The approach involves generating a 3x3 pixel matrix for a Sobel edge detector, utilizing MATLAB and Xilinx ISE software. The process encompasses various stages, including simulation, synthesis, and power analysis of the image pixel memory system. The analysis highlights a significant aspect of the study—power dissipation. The power consumption of the developed image pixel memory is evaluated across different source image resolutions, including 10x40, 10x20, 128x128, 320x240, and 512x512. The results indicate that the use of clock gating techniques effectively reduces power dissipation by approximately 30% to 40%. Clock gating, which selectively disables the clock signal to inactive parts of the circuit, proves to be a crucial method for enhancing the energy efficiency of the system. Looking forward, there is potential for further improvement in the performance of image pixel memory. Future work can focus on optimizing the system by reducing the number of registers and minimizing the pixel access time. Such improvements would not only enhance the efficiency but also contribute to more responsive and high-performance image processing applications. The ongoing development and optimization of image pixel memory systems are essential for advancing edge detection technologies and other image processing techniques.

REFERENCES

[1] Harald Devos, Dirk Stroobandt, *Optimizing the FPGA Memory Design for a Sobel Edge Detector. International Conference on Engineering of Reconfigurable Systems & Algorithms (ERSA). 2009.*

[2] Deepayam Bhowmik, Robert Stewart Greg Micharlson, Andraw Wallace, *Optimized Memory Allocation and power Minimization for FPGA based Image processing. Journal of Imaging. 2019, vol. 5, iss. 1, DOI: 10.3390/jimaging5010007*

[3] Peng Zhang, Xu Cheng, Jason Cong, *An Integrated and Automated Memory Optimization Flow for FPGA Behavioural Synthesis. In: 17th Asia and South Pacific Design Automation Conference. Sydney, NSW, Australia: IEEE, 2012, pp. 257-262. DOI: 10.1109/ASPDAC.2012.6164955*

[4] M. Pedram, Xunwei Wu, *Clock-gating and its application to low power design of sequential circuits. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications. 2000, vol. 47, iss. 3, pp. 415-420. DOI: 10.1109/81.841927.*

[5] J Raivainen, A Mammela, *Clock Gating in FPGA's: A Novel and Comparative Evaluation. In: Ninth Euromicro Conference on Digital System Design, Cavtat, Croatia: IEEE, 2006, pp. 584-590, DOI: 10.1109/DSD.2006.32.*

[6] David Flynn, Robert Aitken, Alan Gibbons, Kajian Shi, *Low Power Methodology Manual For System-on-Chip Design. Springer, 2007. ISBN 978-0-387-71818-7.*
[7] Mark Biegel, *Power Reduction through RTL Clock Gating. SNUG, San Jose, 2000, pp.1-11.*
[8] R. W. Brodersen, *Minimizing power consumption in digital CMOS circuits In: Proceedings of the IEEE, vol. 83, no. 4, pp. 498-523, 1995, DOI: 10.1109/5.371964.*
[9] D Lakata, A Dehon, *Impact of Parallelism and Memory Architecture on FPGA Communication Energy. ACM Transactions on Reconfigurable Technology and Systems. 2016. vol. 9, iss. 4, pp. 1-23. DOI:10.1145/2857057*
[10] P. Siegel, G. D. Micheli, *Saving power by synthesizing gated clocks for sequential circuits. IEEE Design & Test.1994. pp 32-41.*
[11] R. Joshi, M. A. Zaman, and S. Katkoori, *Novel bit-sliced near-memory computing based VLSI architecture for fast Sobel edge detection in IoT edge devices. In 2020 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS) (pp. 291-296). IEEE, December 2020.*
[12] Z. E. M. Osman, F. A. Hussin, and N. B. Z. Ali, *Hardware implementation of an optimized processor architecture for SOBEL image edge detection operator. In 2010 International Conference on Intelligent and Advanced Systems (pp. 1-4). IEEE, June 2010.*
[13] S. Singh, S. Saurav, R. Saini, A. K. Saini, C. Shekhar, and A. Vohra, *Comprehensive review and comparative analysis of hardware architectures for Sobel edge detector. International Scholarly Research Notices, 2014(1), 857912, 2014.*
[14] S. Kumar, and P. Pandey, *FPGA implementation of image segmentation by using edge detection based on Sobel edge operator. International Journal of Research in Engineering and Technology, 2(10), 198-203, 2013.*
[15] T. Saidani, R. Ghodhbani, M. Ben Ammar, M. Kouki, M. H. Algarni, Y. Said, ... and E. H. Abd-Elkawy, *Design and Implementation of a Real-Time Image Processing System Based on Sobel Edge Detection using Model-based Design Methods. International Journal of Advanced Computer Science & Applications, 15(3), 2024.*
[16] D. Pudi, R. Ryansh, V. Goudu, S. Boppu, and A. Hemani, *Implementation of Sobel Edge Detection on DRRA and DiMArch Architectures. In 2023 26th Euromicro Conference on Digital System Design (DSD) (pp. 16-23). IEEE, September 2023.*
[17] Q. Chang, X. Li, Y. Li, and J. Miyazaki, *Multi-directional Sobel operator kernel on GPUs. Journal of Parallel and Distributed Computing, 177, 160-170, 2023.*
[18] V. Yamini, S. A. Hussain, G. Chandra Sekhar, P. Avinash Kumar, P. Lehitha, B. Sree Venkata Teja, ... and P. K. Sanki, *An SoC System for Real-Time Edge Detection. Journal of Electronic Materials, 1-8, 2024.*
[19] A. S. Khalil, M. Shalaby, and E. Hegazi, *An Enhanced System on Chip-Based Sobel Edge Detector. In 2023 International Telecommunications Conference (ITC-Egypt) (pp. 179-183). IEEE, July 2023.*
[20] M. Orthy, S. M. R. Islam, F. Rashid, and M. A. Hasan, *Implementation of Image Enhancement and Edge Detection Algorithm on Diabetic Retinopathy (DR) Image Using FPGA. IET Circuits, Devices & Systems, 2023(1), 8820773, 2023.*
[21] K. M. Rao, P. S. Kumar, T. V. Reddy, D. Nilima, K. Saikumar, and A. M. Khlaif, *"Ultra Low Power High Speed DFT Implementation For ASIC SoC,"* in 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), IEEE, April 2024, pp. 1-6.
[22] R. Revathi, R. Vatambeti, K. Saikumar, M. A. Alkhafaji, U. R. Khairy, and S. Noori, *"An advanced online mobile charge calculation using artificial intelligence,"* in AIP Conference Proceedings, vol. 2845, no. 1, AIP Publishing, September 2023.
[23] K. Mannepalli, K. B. Raju, J. Sirisha, K. Saikumar, and K. S. Reddy, *"LOW complex OFDM channel design using underwater-acoustic-communication using machine learning techniques,"* in 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, December 2021, pp. 1505-1513.
[24] D. S. Kumar, C. S. Kumar, S. Ragamayi, P. S. Kumar, K. Saikumar, and S. H. Ahammad, *"A test architecture design for SoCs using atam method,"* International Journal of Electrical and Computer Engineering, vol. 10, no. 1, pp. 719, 2020.
[25] N. Soumya, K. S. Kumar, K. R. Rao, S. Rooban, P. S. Kumar, and G. N. S. Kumar, *"4-bit multiplier design using cmos gates in electric VLSI,"* International Journal of Recent Technology and Engineering, vol. 8, no. 2, pp. 1172-1177, 2019.

Edited by: Dhilip Kumar V
Special issue on: Unleashing the power of Edge AI for Scalable Image and Video Processing
Received: May 31, 2024
Accepted: Aug 24, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 610–618, DOI 10.12694/scpe.v26i2.3888

A COMPUTER SYSTEM OPERATION AND MAINTENANCE INTERACTION PLATFORM BASED ON ARTIFICIAL INTELLIGENCE

WENQIU WU*

Abstract. To address the challenge of inefficient data management in computer operation and maintenance across various scales, the author suggests an artificial intelligence-driven platform for computer system operation and maintenance interaction. A computer operation and maintenance management approach has been devised, leveraging artificial intelligence technology. This method entails the segmentation of the computer operation and maintenance structure to efficiently manage signals transmitted by computers, ensuring swift signal transmission. With the integration of artificial intelligence, the approach achieves superior space utilization and data throughput. By comparing past computer operation and maintenance management methods. The experimental results show that the maximum space saving percentage can reach 75%, and the minimum is also higher than 40%. The throughput is the highest at various data scales, with an average throughput of 704.67 MB/s. The higher the average throughput value, the higher the management efficiency. Conclusion: Computer operation and maintenance management methods based on artificial intelligence technology have better management effects, can better save space, and improve management efficiency.

Key words: Artificial intelligence technology, Operation and maintenance management, Network intelligence

1. Introduction. As computer networks grow larger and data volumes increase, traditional network management and maintenance approaches fall short of meeting the demands of modern network environments. To tackle this issue, there's been a widespread adoption of artificial intelligence technology in computer network management, ushering in innovative solutions to address the evolving needs of networks[1]. Artificial intelligence can discover network faults and abnormal behaviors by learning and analyzing large amounts of network data, and provide precise network management and optimization strategies. Furthermore, artificial intelligence contributes to enhancing the security and reliability of networks while boosting the quality and efficiency of network services. The pivotal role of computer network technology in the advancement of modern society is underscored by its significant development and positive impact[2,3]. While bringing many conveniences, there are many security risks in the operation of computer networks, especially malicious attacks that can cause serious losses and pose a great threat to computer network security.

Artificial intelligence is a product of the healthy development of computer science. Introducing artificial intelligence technology into computer network operation and maintenance can improve the running speed of computers, help them achieve high-speed operation, and serve users to the maximum extent possible. Artificial intelligence technology encompasses a burgeoning scientific and technological domain dedicated to exploring and advancing theories, methods, technologies, and application systems aimed at replicating, augmenting, and broadening human intelligence. It endeavors to emulate the information processing capabilities of human consciousness and cognition. Hence, judicious utilization of artificial intelligence technology holds promise for more effectively serving users[4-6]. With the advancement of technology, artificial intelligence technology is constantly improving. Adding artificial intelligence technology to computer network operation and maintenance can not only improve data security, but also pose certain threats. Continuous in-depth research on artificial intelligence technology can effectively promote the progress of artificial intelligence, but also promote the progress of computer network technology and deepen network intelligence [7]. The design of a computer resource security monitoring and management system can monitor and manage computer network resources in real time, quickly locate and solve faults that occur, timely prevent security risks, and ensure the security of computer network resources.

*Information Engineering College, Jilin Engineering Vocational College, Siping, 136100, China (Corresponding author, WenqiuWu@163.com)

2. Literature Review. Artificial Intelligence (AI) is a technological science that utilizes computer technology to simulate and implement human intelligence. It simulates human thinking and intelligent behavior, enabling machines to have certain automation and intelligence capabilities, enabling them to reason, learn, judge, and make decisions like humans [8,9]. The core of artificial intelligence technology is machine learning, which enables machines to learn and extract patterns from large amounts of data by constructing and training models, thereby continuously optimizing and improving their performance. CHENSiyu et al. designed how to manage railway security systems more efficiently and comprehensively, and built an intelligent security operation and maintenance platform. This platform uses "CMDB" as the database for integrated operation and maintenance management of "supervision and control", achieving a unified security equipment asset configuration management library, integrated sharing of security information, standardization of security workflow management, real-time status monitoring of security equipment operation, environmental monitoring, intelligent identification of prohibited items, fault prediction and other functions. It improves the overall intelligence level of security equipment management, is conducive to detecting safety risks in station operation, improving security work efficiency, and ensuring the smooth and efficient operation of security work [10]. Li, W. et al. have pioneered the development of a distributed system merging mixed reality (MR) and Internet of Things (IoT). Their methodology involves two key steps: firstly, generating a digital model by integrating design blueprints with real-world environments, then constructing a Unified Perception Network (UPN) MR system using a game engine and the OpenXR platform. Secondly, they establish an IoT cloud platform, leveraging API collections and cloud services, to seamlessly interface with the MR system. Sensor data communication with MR devices is facilitated via the Socket method, while a Kalman algorithm-based data filtering model is employed to enable information exchange between on-site workers and backend managers[11]. ZONGPan et al. conducted a comprehensive and detailed analysis of the application of integrated operation and maintenance systems in airport luggage systems, and analyzed the advantages and design requirements of operational systems to improve the operational speed of airport luggage systems and the level of aviation information construction [12].

Entering the information society, human beings have increasingly high requirements for data processing. In order to meet the processing needs of massive data such as scientific and engineering calculations, supercomputers are widely used in various industries. Computer network operation and security management are very important and necessary, and should be given high attention. Computer information management should be done seriously, and risk management efforts should be strengthened. In response to potential security risks in computer networks, efforts should be made to highlight the important role of computer network operation, maintenance, and management. Computer information technology should be applied effectively to prevent common computer network security issues. At the same time, the computer network information security management system should be continuously optimized to achieve better computer network operation, maintenance, and security management.

3. Method.

3.1. Selection of Computer Operation and Maintenance Data Collector. In order to ensure that the system can obtain the most primitive data information, each component collects raw data from the system's hardware or software devices and stores it in the local database of the system. According to the operational needs of the system, the author selected the SCS2458-KF168-290 model data acquisition and transmission instrument as the data collector for the network operation and maintenance management system. This model of data collector has many interfaces, not only supporting communication and transmission functions such as GPES/4G/5G/Ethernet, but also supporting the unique IC165-5900 communication protocol in cloud computing environments. To bolster the online monitoring and early warning capabilities of the network operation and maintenance management system, a data collector is deployed to interface with diverse sensors within the network's operational environment. This collector gathers data from these sensors and subsequently uploads it to the network operation and maintenance server for analysis and processing[13]. Figure 3.1 is a schematic diagram of the connection of the network operation and maintenance data collector.

3.2. Selection of Computer Operation and Maintenance Data Memory. Considering the need for massive data storage, the author chose JF165-1650 model data storage as the storage hardware for massive data, with an output power of 48KHz and a normal operating voltage of 22V. This memory integrates 6

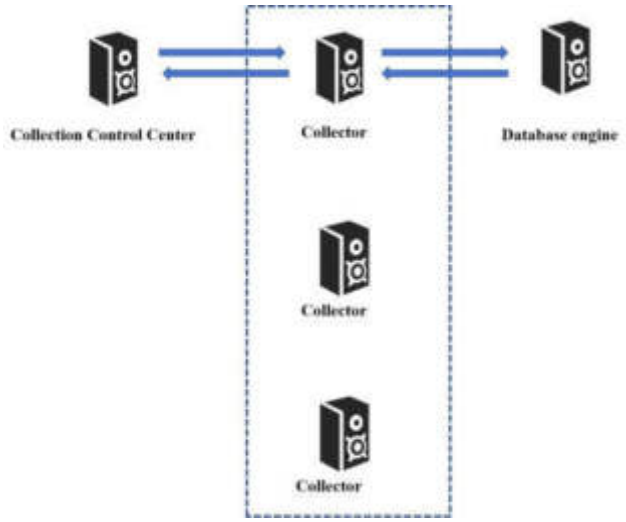


Fig. 3.1: Connection diagram of computer operation and maintenance data collector

labeled controller domain network bus interfaces, which can achieve real-time storage of computer operation and maintenance data by connecting other hardware structures to the controller domain network bus. At the same time, in order to ensure the stability of network operation and maintenance data storage and transmission, 6 analog data transmission paths and 2 universal input/output ports are reserved to meet the needs of data input and output.

3.3. Design of Computer Operation and Maintenance Data Storage Scheme. The basic function of computer operation and maintenance is to store massive amounts of data. Combining the above hardware conditions, the author designs a computer operation and maintenance data storage technology scheme based on digital twin technology [14]. Among them, the first type of data refers to data that describes the data structure. Generally, there is a data structure first, and then the relevant data information content is filled in. The second type of data is inconvenient to use, usually filled data and data without data structures, such as audio, video, images, etc. The third type of data contains implicit or irregular data, such as sound, graphic files, etc. The author uses MongEG as the storage database for the computer operation and maintenance management system, which can complete the storage of the three different types of data structures mentioned above. At the same time, MongEG supports a wide range of languages, and its syntax is closer to object-oriented query languages. Therefore, it can achieve most single table query functions, making it easier to quickly find the data resources that need to be queried during operation and maintenance.

3.4. Develop intelligent operation and maintenance cycles for network centers. Assuming that the maintenance cycle of the system throughout its entire lifespan is X, the interval between each operation and maintenance cycle should be Ti, where i is the number of operation and maintenance cycles, with values of i=1,2,3,..., n. In the previous cycle, when the information reliability of the network center exceeds the preset reliability threshold L, the system needs to automatically recognize and complete the maintenance of the network center; When the reliability threshold L is reached for the Nth time, it indicates that there is a problem with the network center at this time, and it is necessary to perform replacement operation and maintenance on the relevant modules; If the network center experiences failure during the operation and maintenance cycle, a minimum maintenance approach should be adopted to restore some of its problematic modules, in order to minimize operation and maintenance costs while ensuring the reliability of the network center operation [15]. Based on the above analysis, the formula for calculating the network operation and maintenance cycle is:

$$exp(\sum_{T_i} h_i(t)) = K(t) \tag{3.1}$$

In the formula, T_i represents the time interval of network operation and maintenance, represents the loss rate during the i-1st and i-th operation and maintenance cycles of the network center, and K (t) represents the running time of the network center. By setting the operation and maintenance cycle, the intelligent operation and maintenance of the network center are completed according to the big data based network operation and maintenance data storage technology scheme mentioned above, thereby ensuring the security of massive data in the cloud computing environment.

3.5. Managing Computer Signals. After dividing the computer operation and maintenance management structure, it is necessary to predict the transmission destination of computer signals, because predicting the transmission process and destination of computer signals in advance, planning the transmission path in advance, can save transmission time and improve computer transmission efficiency. In the actual work process, in order to ensure the effectiveness of computer operation, the setting of information transmission points will be minimized as much as possible. Although there are few information transmission points that cannot fully integrate various functions in the computer at once, if the signal receiving point moves, it will reduce the possibility of signal interruption. Therefore, in order to ensure the effectiveness of the computer during use, it is necessary to predict the direction of the transmitted network signals. The linear discrete state equation and observation equation of computer signals are represented as follows.

$$\begin{cases} A_k = f[A_{k-1}, k-1] + \omega[A_{k-1}, k-1] \cdot B_{k-1} \\ C_k = c[A_{k-1}, k] + D_k \end{cases} \tag{3.2}$$

In equation 3.2, A_k denotes the matrix characterizing the communication network's discrete state, while C_k represents the observation function associated with the communication network's discrete state, B_{k-1} represents the network noise matrix, and D_k represents the noise observed by the communication network in a discrete state. Communication network signals are mostly linear during transmission. Therefore, in the calculation process, the communication network signals need to be linearized to obtain the prediction equation for the direction of the communication network signals, as follows:

$$A_{k,k-1} = f[\widetilde{A}_{k-1}, k-1] \tag{3.3}$$

Equation 3.3 enables the calculation of the transmission dynamics of communication network signals, facilitating the determination of signal destinations. Moreover, it enhances the reliability of destination prediction. Once the calculation aligns with the actual conditions of the communication network, it enables effective management of ongoing signal transmissions within the network.

3.6. Calculation of operation and maintenance management indicators based on artificial intelligence technology. Utilizing equations 3.2 and 3.3, the signal transmission path can be determined, and the prediction error arising during the transmission process can be computed. This function can be expressed as follows:

$$S = E[\widetilde{A}_k, \widetilde{A}_k^T] \tag{3.4}$$

In equation 3.4, $\widetilde{A}_k$ represents the possible error value of signal k during transmission, and the effectiveness of the transmission process can be determined based on the results.

In practical management work, data from multiple inspection points can be transmitted to the data center, and then analyzed and statistically analyzed through the data center, and various problems that arise in the computer can be handled. Afterwards, if similar problems occur again, they can be quickly addressed based on past experience to ensure the transmission efficiency of the computer.

In addition, in computer operation and maintenance management, identifying faults that occur during the communication process is also a major challenge. Previous communication methods cannot quickly locate communication faults, which may lead to communication network interruptions and cause huge losses [16].

For small networks, signal transformation is often used to determine the fault point. However, for some large-scale communication networks with long-distance transmission, using this method to locate the fault point poses certain difficulties. Hence, to enhance the efficiency and precision of fault detection and expedite troubleshooting

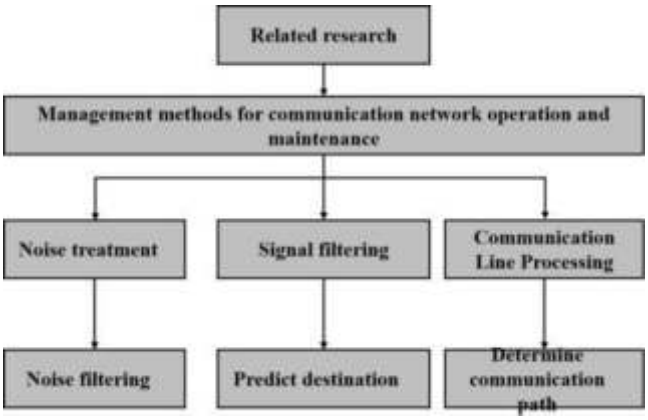


Fig. 3.2: Division of Communication Network Operation and Maintenance Management Process

during computer operation, the integration of artificial intelligence technology becomes imperative. Leveraging artificial intelligence, the author can pinpoint network faults, assess the performance of communication network operation and maintenance management methods, and compute indicator values to evaluate their efficacy. The detailed calculation process unfolds as follows:

$$P = \frac{\sum_{k=1}^{n} S_k}{\sqrt[n]{S_K}} \tag{3.5}$$

Equation 3.5 encapsulates the iteration count of communication network signal transmission denoted by 'n', while 'P' signifies the specific numerical value of the operation and maintenance management method indicator. Through computation of these numerical values, the rationality of the operation and maintenance management method can be assessed, thereby validating its efficacy.

3.7. Refine the communication network operation and maintenance management structure. In order to better carry out communication network operation and maintenance management work, the process of communication network operation and maintenance management can be refined. For various problems that may arise during the process, the author proposes targeted suggestions and solutions to make the communication network operate better. The specific division of the communication network operation and maintenance management structure is shown in Figure 3.2.

Drawing from an analysis of historical trends in communication network user growth and usage patterns, and in alignment with practical requirements, a streamlined operation and maintenance management process for the communication network is devised [17]. In practical implementation, the signal transmitted through the communication network undergoes noise processing as a preliminary step, as depicted in Figure 3.3.

Using wavelet denoising method to process signals, avoiding the transmission speed of signals not being fast enough due to the presence of noise during subsequent transmission, and preventing users from receiving information quickly. After dealing with signal noise, filter the transmitted signal to reduce unnecessary signal interference with communication quality. In addition, for noisy signals transmitted in, certain corrections should be made to ensure the correctness of signal transmission. In addition, after noise processing and signal filtering, it is necessary to process the path where the signal is about to be transmitted. By analyzing and diagnosing the communication line, the subsequent communication path can be determined, which can make the signal transmission process more accurate and rapid [18].

By calculating the numerical values of operation and maintenance management, it is possible to verify whether the operation and maintenance management method is reasonable. In summary, the design content is integrated to ensure that each part is interrelated and independent, and applied to the calculation of communication network operation and maintenance management method indicators. The communication network operation and maintenance management method is designed, as shown in Figure 3.4.

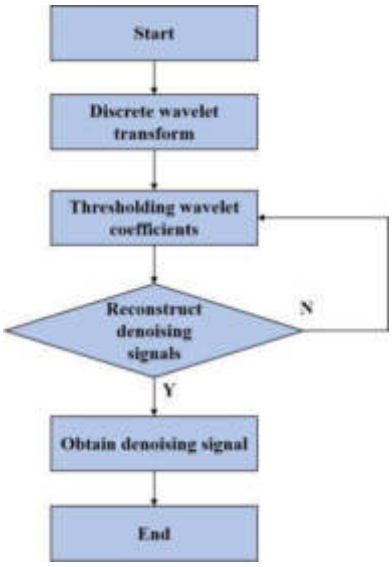


Fig. 3.3: Communication Network Signal Noise Processing Process

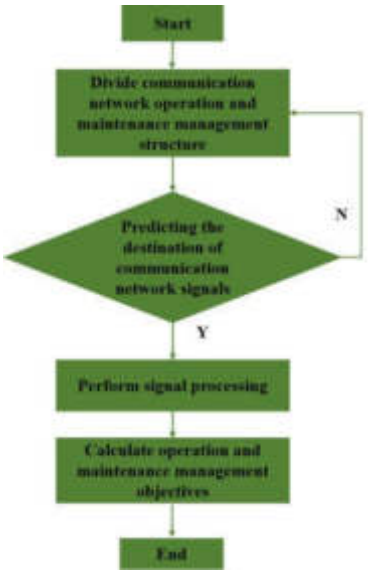


Fig. 3.4: Communication Network Operation and Maintenance Management Flowchart

3.8. Experimental Preparation. In order to objectively analyze the effectiveness of computer operation and maintenance management methods based on artificial intelligence technology, the author conducted simulation experiments. The testing environment is built on a Windows system, with MySQL as the database and Java as the main language used for backend development. The Angular framework is mainly used for frontend development, as shown in Figure 3.5. The dashed box represents the application, where users interact with the application and transmit data through the server.

4. Results and Discussion. Table 4.1 provides a comparative analysis of the efficacy of four operation and maintenance management methods employing signal acquisition and reception equipment alongside high-

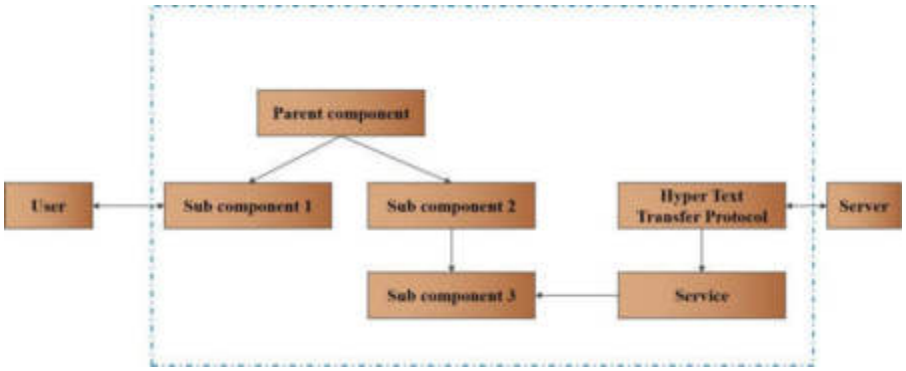


Fig. 3.5: Angular framework structure

Table 4.1: Space savings percentage of four management methods under different data scales/unit:%

Data scale	Method 1	Method 2	Method 3	Method 4
1000	45.1	40.5	40.3	39.1
2000	48.4	43.2	40.5	38.2
3000	47.5	45.4	43.2	40.5
4000	50.5	45.2	40.4	38.5
5000	60.2	50.2	45.7	42.4
6000	75.3	60.3	55.4	49.7

Table 4.2: Throughput of Four Management Methods at Different Data Scales Unit: MB/s

Data scale	Method 1	Method 2	Method 3	Method 4
1000	801	705	681	651
2000	751	685	674	643
3000	651	635	602	597
4000	749	723	705	684
5000	601	651	634	603
6000	683	661	652	635

precision time measuring instruments. Method 1, devised by the author, utilizes artificial intelligence technology for computer operation and maintenance management. Method 2 employs big data for management purposes, while Method 3 relies on neural networks. Finally, Method 4 represents the conventional approach to computer operation and maintenance management.

Comparative analysis shows that Method 1 can save up to 75% of space, while Method 2 can save up to 60% of space, Method 3 can save up to 55% of space, and Method 4 can save up to 49% of space, therefore, compared to other management methods, Method 1 has a higher percentage of space savings and better compression effect[19]. Note that, in practical applications, the computer operation and maintenance management method based on artificial intelligence technology designed by the author has a higher percentage of space savings, lower data transmission costs, better space savings, and better management effects.

In computer operation and maintenance management, the size of incoming signal data directly impacts the overall effectiveness of the management method. To assess the efficacy of operation and maintenance management methods across various data scales, four management approaches are compared based on data throughput. The throughput results are detailed in Table 4.2.

According to Table 4.2, Method 1 has the highest throughput for each data scale, with an average through-

put of 704.67 MB/s. The higher the average throughput value, the higher the management efficiency. Although Method 2 has similar throughput to Method 1 for different data scales, there is still a certain gap, with an average throughput of 673.33 MB/s. Method 3 has relatively less data throughput compared to the first two methods, with an average throughput of 655.5 MB/s. Method 4 has the lowest average throughput, which is 633 MB/s. In summary, the computer operation and maintenance management method based on artificial intelligence technology designed by the author has higher data throughput, greater space savings, higher management efficiency, and better management effects at different data scales[20]. Through the above two experiments, it can be seen that in practical applications, compared with traditional operation and maintenance management methods, the operation and maintenance management method based on artificial intelligence technology designed by the author can better save space, transmit data faster, improve management efficiency, and achieve better management results at different data scales.

5. Conclusion. The author proposes a computer system operation and maintenance interaction platform based on artificial intelligence. Artificial intelligence technology has strong data analysis capabilities and can analyze various states and behaviors of computers. Applying it to computer operation and maintenance management will bring huge changes to computer operation and maintenance management. The computer operation and maintenance management method based on artificial intelligence technology designed by the author can effectively improve network management efficiency and ensure network reliability. Although there are some shortcomings, with the continuous improvement of artificial intelligence technology in China, these shortcomings will gradually be made up for.

REFERENCES

[1] Yu-dianOUYANG, KunXIE, Gao-gangXIE, & Ji-gangWEN. (2022). A data recovery algorithm for large-scale network measurements: association learning based tensor completion. Acta Electronica Sinica, 50(07), 1653-1663.

[2] Tsuzuki, R. (2022). Development of automation and artificial intelligence technology for welding and inspection process in aircraft industry. Welding in the World, 66(1), 105-116.

[3] Zhao, L., Zhu, D., Shafik, W., Matinkhah, S. M., Ahmad, Z., & Sharif, L., et al. (2022). Artificial intelligence analysis in cyber domain: a review:. International Journal of Distributed Sensor Networks, 18(4), 121-131.

[4] Yang, H. (2023). Research on the development trend of electronic communication technology based on intelligent network. Modern Electronic Technology, 7(2), 26-30.

[5] Noordt, C. V., & Misuraca, G. (2022). Exploratory insights on artificial intelligence for government in europe:. Social Science Computer Review, 40(2), 426-444.

[6] Nouraldeen, R. M. (2023). The impact of technology readiness and use perceptions on students' adoption of artificial intelligence: the moderating role of gender. Development and Learning in Organizations: An International Journal, 37(3), 7-10.

[7] Qi, K. (2022). A research on the application of virtual network technology in computer network security. Electronic Research and Applications, 6(4), 1-6.

[8] Jarrahi, M. H., Kenyon, S., Brown, A., Donahue, C., & Wicher, C. (2023). Artificial intelligence: a strategy to harness its power through organizational learning. Journal of Business Strategy, 44(3), 126-135.

[9] Nabi, G. (2023). Machine learning, artificial intelligence, and digitalisation of healthcare: convergence of science and technology:. Scottish Medical Journal, 68(2), 37-38.

[10] CHENSiyu. (2022). Research and discussion on big data operation and maintenance platform of railway passenger security inspection system. Engineering technology of foreign science and technology periodical database (abstract edition), 5(2), 93-109.

[11] Li, W., Ye, Z., Wang, Y., Yang, H., Yang, S., & Gong, Z., et al. (2023). Development of a distributed mr-iot method for operations and maintenance of underground pipeline network. Tunnelling and Underground Space Technology, 133, 104935-.

[12] ZONGPan, & LIUYuan. (2022). Application of integrated operation and maintenance system in airport baggage system. Engineering technology of foreign science and technology periodical database (abstract edition), 10(1), 1-14.

[13] Elyan, E., Vuttipittayamongkol, P., Johnston, P., Martin, K., Mcpherson, K., & Carlos Francisco Moreno-García, et al. (2022). Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artificial Intelligence Surgery, 2(1), 24-45.

[14] Do, K. T., Gip, H., Guchait, P., Wang, C. Y., & Baaklini, E. S. (2023). Empathetic creativity for frontline employees in the age of service robots: conceptualization and scale development. Journal of Service Management, 34(3), 433-466.

[15] LuZHANG, JiapengLIU, & DongmeiTIAN. (2022). Application of stacking-bagging-vote multi-source information fusion model for financial early warning. Journal of Computer Applications, 42(01), 280-286.

[16] Eduardo Pérez. (2022). A simulation-driven online scheduling algorithm for the maintenance and operation of wind farm systems. Simulation: Journal of the Society for Computer Simulation, 98(1), 47-61.

[17] Wang Chuan, Y. H. (2023). Identification and traceability system for coil weld based on manufacturing operation management. Manufacturing Automation, 45(1), 125-128.
[18] Liang, X., Chen, P., Chen, C., Che, W., Yang, Y., & Tan, Z., et al. (2022). Comprehensive risk assessments and anesthetic management for children with osteogenesis imperfecta: a retrospective review of 252 orthopedic procedures over 5years. Paediatric anaesthesia, 32(7), 851-861.
[19] Gao, Y. G. (2022). Audit optimization based on computer vision technology. Computer and Communication (English), 10(10), 50-58.
[20] Wang, J., & Zhao, Y. (2022). Study on the matching scheme between early warning information dissemination demands and communication network dissemination channels. Telecommunications Science, 38(5), 104-113.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 10, 2024
Accepted: Jun 25, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 619–629, DOI 10.12694/scpe.v26i2.3892

RESEARCH ON INTENTION RECOGNITION METHODS BASED ON DEEP LEARNING

QIANG LI*, FENG ZHAO†, PAN GAO‡, HUANHUAN LI§, AND LINFENG YE¶

Abstract. In order to improve the accuracy of intelligent speech interaction robots, the author proposes a deep learning based intention recognition method for research. By introducing the GloveBibGRU-Self attention classification prediction model, an intention recognition function module is constructed, and the ROS distributed architecture is adopted to integrate the system functional modules, achieving intelligent voice interaction between humans and machines. The simulation results show that the speech intention recognition using the proposed method has higher accuracy. Compared with the intention recognition methods based on DCNN model, CNN-LSTM model, and GRU Self attention model constructed unidirectionally, the recognition accuracy is higher than 8. 02%, 4. 06%, and 2. 13%, respectively, and has better recognition effect, In terms of feature extraction, the training time of BiGRU is shortened by four times compared to traditional extraction methods based on BiLSTM models, resulting in higher training efficiency. According to the experimental findings, the speech interaction system developed utilizing the suggested intention recognition method maintains a high level of accuracy and efficiency in understanding user English speech commands. With an average accuracy rate of 89.72% and recognition times consistently below 0.35 seconds, it is evident that the proposed method is applicable for real-world speech interactions. The intent recognition method based on Glove2BiGRU-Self attention can be applied to English speech interaction in intelligent speech robots.

Key words: Artificial intelligence, Intention recognition, Predictive classification, Deep learning, English voice interaction

1. Introduction. The intersection of human-machine collaboration has emerged as a prominent research focus within the intelligent robotics domain, finding widespread applications across entertainment services, advanced manufacturing, military operations, medical rehabilitation, and beyond. As artificial intelligence technology advances rapidly, it has catalyzed significant transformations in societal dynamics. The seamless integration of AI and robotics technology, aimed at enhancing human-machine synergy, has evolved into a pivotal trend shaping the trajectory of robotics development [1].

Experts and scholars in the field of robotics believe that human-machine collaboration can be seen as a necessary attribute of new industrial robots, and the research and development of collaborative robot technology is the focus of robotics technology. In order to reduce collaboration risks in human-machine collaboration, collaborative robots often use flexible driving mechanisms to improve the performance of contact based human-machine interaction, and the robot body also uses lightweight components. However, although passive flexible structures have certain flexibility and can ensure a certain level of human-machine cooperation safety, they still have a certain degree of stiffness, and robots cannot adjust their own flexibility. In many fields such as entertainment services, advanced manufacturing, military, medical rehabilitation, etc., there is a very high demand for flexibility in contact based human-machine interaction. Only passive flexibility is difficult to meet complex human-machine cooperation tasks [2]. If robots can autonomously adjust flexibility according to task requirements and interaction environments, the performance of human-machine collaboration will be greatly improved. This requires robots to have a certain level of cognitive ability towards the external environment, among which identifying the movement intentions of collaborators is extremely important. We hope that robots have a certain ability to recognize the movement intentions of collaborators. According to task requirements, robots should not only passively follow human movements, but also actively approach human movement intentions, improving the smoothness and comfort of human-computer interaction [3]. Although human-machine

*State Grid Information & Telecommunication Group Co.,Ltd.,Beijing 102211,China (Corresponding author, QiangLi58@c163.com)
†State Grid Information & Telecommunication Group Co.,Ltd.,Beijing 102211,China (FengZhao83@126.com)
‡State Grid Information & Telecommunication Group Co.,Ltd.,Beijing 102211,China (PanGao137@163.com)
§State Grid Information & Telecommunication Group Co.,Ltd.,Beijing 102211,China (HuanhuanLi7@126.com)
¶State Grid Information & Telecommunication Group Co.,Ltd.,Beijing 102211,China (LinfengYe7@163.com)

collaboration technology has received increasing attention from scholars and has achieved certain research results, problems such as difficult recognition of human motion intentions, poor flexibility of robot motion, and low efficiency of human-machine collaboration still exist.

2. Literature Review. The specific content of interaction intention recognition research varies for robots in different fields. For industrial robots, interaction intention recognition focuses on inferring the tasks that people will be performing, such as recognizing the intention of others to grab a certain item or operate the machine. For humanoid robots such as Sophia and Nadine, in order to improve the quality of human-computer interaction, it is necessary for robots to recognize human social intentions. The author focuses on how to enable humanoid robots to recognize human social intentions, namely interactive intentions [4]. According to different feature fusion methods, human-computer interaction intention recognition methods can be divided into rule-based methods and data-driven methods. Lou, H. et al. introduced a novel hybrid model named SACL, which combines self-attention mechanism, convolutional neural network (CNN), and long short-term memory network (LSTM). They began by collecting a dataset of common passenger travel issues in civil aviation airports using web crawlers. After preprocessing the data, they generated a word vector matrix. The model then utilizes a serial structure of CNN and attention mechanism to capture local information of the problem from the word vector matrix. Simultaneously, LSTM captures the global structural information of the text. By concatenating these two feature vectors, the model obtains a comprehensive representation of the text, which is fed into a fully connected neural network and softmax layer for text classification. The SACL model was trained using the Gradient Descent Method (GDM) and compared its performance with four other models. Results indicate that the proposed SACL model effectively identifies passengers' intentions in asking questions, which holds significant implications for enhancing the accuracy and efficiency of answer extraction in civil aviation airport passenger quality assurance systems [5]. Pan, Y. et al. introduced a novel robot teaching system that relies on detecting robot contact states and recognizing human motion intentions. This innovative system can accurately identify the contact status of the robot's hand joints and extract motion intention cues from surface electromyographic signals of the human body to guide the robot's movements. Furthermore, a dedicated module for robot motion mode selection has been developed, enabling the control of the robot's motion along a single axis, in a linear trajectory, or for repositioning purposes. Experimental findings demonstrate that the system effectively facilitates online robot teaching across three distinct motion modes[6]. Wang et al. introduced a deep learning approach to assess the condition of substation switchgear. This method leverages image data captured by the robot's optical camera, employing deep learning algorithms for analysis and detection. Initially, the method curates a dataset comprising images of substation switchgear for model training. Subsequently, utilizing the Yolov3 object detection network, an automatic recognition model is constructed to evaluate the status of substation switchgear. Experimental outcomes indicate that this method achieves an impressive accuracy rate of 9% in automatically identifying the condition of substation switchgear [7].

On the basis of the above research, the author proposes an intention recognition method based on the Glove2BiGRU-Self attention classification prediction model. The model is introduced to construct an intention recognition function module, and through the integration of different functional modules, intelligent human-machine speech interaction of robots is achieved.

3. Research Method.

3.1. Intention recognition. Intent recognition is the process in speech interaction systems where a robot determines the user's intention based on user instructions and provides timely feedback during communication between the user and the machine[8]. It typically includes steps such as preprocessing, text vectorization, feature extraction, and feature classification. The general process of intent recognition is shown in Figure 3.1:

As shown in the figure, intention recognition in robot intelligent speech interaction is a process of constructing a model for classification prediction based on manually selected speech data, feature processing, and then constructing a model for classification prediction. It belongs to a type of classification prediction model. Designers typically use machine learning and deep learning methods to train models. Compared to machine learning, deep learning takes less time and produces better learning results. Therefore, deep learning is adopted as the author's training method.

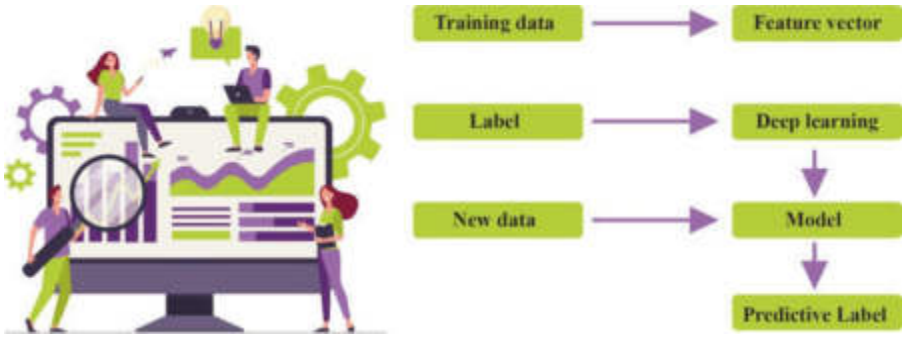


Fig. 3.1: Consciousness recognition processing flow

3.2. Glove_BiGRU_Self-attention model. Traditional intention recognition methods based on template matching or artificial feature sets have high costs and low scalability. In order to enhance the model's understanding performance of user intentions, a user intention classification method based on Glove2BiGRU-Self attention is proposed to help the model learn speech features more fully and quickly. The commonly used deep learning algorithm for consciousness classification model is LSTM, which can solve the long order dependency problem of RNN. However, in computation, due to the design of many variables and parameters, it consumes a considerable amount of time and resources [9]. Therefore, using GRU instead of LSTM, GRU is a variant of LSTM, which reduces one gate compared to LSTM, has a simpler structure, and still has good learning performance, which can significantly save training time and improve model learning efficiency. The user awareness recognition classification model based on Glove-BiGRU-Self attention mainly includes five layers, namely input layer, Glove layer, BIGRU layer, Self attention layer, and Softmax layer.

In the input layer, stuttering segmentation is used to process the user instructions converted into text. For words that are not in the dictionary, an HMM model based on Chinese character word formation ability is used for segmentation. Then, a stuttering segmentation machine is used for segmentation. Finally, all segmentation results are input into the Glove layer for semantic feature representation; Using Glove tool to convert the segmented words in the input layer into word vectors, and utilizing the semantic properties between word vectors to solve the word vectors; In the BiGRU layer, feature extraction is performed on the transformed word vectors. In order to make the information contained in each node more complete, bidirectional GRU is used to model the speech data in both forward and backward directions, making the output of this layer more global; In order to improve the efficiency of feature extraction, Self attention is introduced to calculate different weights in the sample sequence features of the BiGRU model, automatically learning more important semantic features; Input sentence vectors containing word level feature weights into the Softmax layer, and use the Softmax classifier for text classification to obtain the final intent category labels, achieving intent recognition of user instructions [10]. The specific calculations for each step of the GloveBibGRU-Self attention classification model are as follows:

By using a stutterer to segment instruction statements, Glove tool converts each word into a corresponding word vector. Using the Euclidean distance formula to calculate the semantic similarity between two word vectors, the Euclidean distance formula is as follows 3.1:

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.1}$$

Construct a co-occurrence matrix X based on the number of times word i and adjacent word j appear together in the context window, X_{ij}, as matrix elements. According to the decay function Decay=1/d, it can be inferred that the distance between two words and the weight of the semantic relationship between words in the total count are inversely proportional. The farther the distance, the smaller the weight [11]. Construct an approximate relationship between word vectors and co-occurrence matrices, and solve word vectors w_i^T and w_j

based on their relationship.

$$w_i^T w_j + b_i + b_j = log(X_{ij}) \tag{3.2}$$

Among them, b_i and b_j are the bias terms of w_i^T and w_j.

Input the obtained word vector into the BiGRU model, set the model update gate and reset gate to z_i and r_i, respectively, and select the input information based on the reset gate. The value calculation of the reset gate at time t is shown in equation 3.3:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{3.3}$$

Among them, x_t is the input at time t; W_r represents resetting door permissions; h_{t-1} represents the hidden layer output at time t-1, σ is the Sigmad activation function.

The value of the update gate determines which new information can be retained in the cellular state. Firstly, the update gate forgets and updates the previous and added information, then, the tanh layer creates a new candidate value ht based on the reset gate value, and finally combines the above two parts to achieve the update between the new and old meta cells ht and h_{t-1}. The specific steps are shown in equations 3.4 to 3.6:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{3.4}$$

$$h_t = tanh(W_h x_t + r_t * U_h h_{t-1}) \tag{3.5}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h_t \tag{3.6}$$

Among them, the weight of the update gate is W_z, z_t is the value of the update gate at time t, and ht-1 is the output of the hidden layer at time t-1; X_t is input at time t.

A unidirectional GRU typically outputs the current state in order from front to back. But the input and output of a moment are often related to the state of both before and after the moment. Therefore, in order to make the node information more complete, a bidirectional GRU is used to establish the association between the node and the time before and after. Using GRU to model from front to back and from back to front respectively, the outputs $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ of the forward GRU and the backward GRU are calculated as equations 3.7 to 3.8:

$$\overrightarrow{h_t} = \overrightarrow{GRU}(x_i, \overrightarrow{h_{t-1}}) \tag{3.7}$$

$$\overleftarrow{h_t} = \overleftarrow{GRU}(x_i, \overleftarrow{h_{t-1}}) \tag{3.8}$$

The output of BiGRU is determined by two unidirectional GRUs, and the weighted sum is used to calculate the output ht of BiGRU at time t, as shown in equation 3.9:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{3.9}$$

If the frame rate of the speech sample is T, the final output H of the BiGRU layer is as shown in equation 3.10

$$H = [h_1, h_2, h_3, \cdots, h_T] \tag{3.10}$$

Due to the coherence of human language, in intention recognition tasks, nodes need to combine contextual information in order to fully express themselves. The comprehensive context information content BiGRU makes the extracted features of the model at this layer more globally representative, which is more conducive to the subsequent classification and recognition of the model [12].

In order to enable the model to quickly filter out valuable key information from a large amount of speech data and capture more direct semantic relationships, the fully connected output of the BiGRU layer is connected

to the input of the Self attention layer. The speech data is encoded based on the weight calculated by the Self attention layer. By calculating the different weights of each word, the model can focus on more important lexical segments when recognizing a sentence. The use of Self attention as the attention function results in a simpler and more efficient Self attention structure with lower computational costs compared to Attention. Using self attention to associate any two frames in speech through computation can make feature usage more efficient. Assuming the model weight is w_w and the bias is b_w, a multi-layer perception mechanism is used to obtain the hidden representation (u_{it}) of h_{it}, as shown in equation 3.11:

$$u_{it} = tanh(w_w h_{it} + b_w) \tag{3.11}$$

Calculate the similarity between the vector u_{it} and its context vector u_w to measure the importance of the word. Use the softmax function to calculate the normalized weight a_{it}, the a_{it} is related to the input state and uw at each moment, calculated as equation 3.12:

$$a_{it} = \frac{exp(u_{it}^T y_w)}{\sum_t exp(u_{it}^T u_w)} \tag{3.12}$$

u_w can be seen as a semantic representation of input, which is randomly initialized and learned together with the input during the training process. Finally, calculate the sentence vector based on the normalized weight a_{it}, as shown in equation 3.13:

$$S = \sum_{i=1}^{n} a_{it} h_{it} \tag{3.13}$$

The output sentence vector already contains various word level weight information of the input state. Build an intention recognition classifier in the Softmax layer, input the sentence vector output from the previous layer into the classifier for text classification. If the weight of the Softmax classifier is w_f and the bias coefficient is bf, then the probability y of the input x belonging to a certain intention category can be calculated as equation 3.14:

$$y = softmax(w_f s + b_f) \tag{3.14}$$

Among them, $y \in R^k$, k represents the total number of user intent categories.

Using a minimum loss function to update and optimize the parameters in the model, the optimization function is defined as equation 3.15:

$$L(y, p) = \sum_{t=1}^{N} \sum_{j=1}^{C} y_{ij} log(p_{ij}) \tag{3.15}$$

Among them, N and C represent the total number of intent recognition categories in the dataset. a_{it} is the intention category label of the i-th sample in the j-th category [13,14].

3.3. Overall design of intelligent English voice interaction system. The intelligent English voice interaction system refers to a human-computer interaction system that answers user questions or executes related actions based on user English voice commands. Based on understanding user instructions, the system utilizes databases or networks to crawl relevant information and answer user questions. The author adopts the ROS distributed architecture to integrate the functions of the robot intelligent voice system, which includes four main functional modules: voice wake-up, speech recognition, intention recognition, and speech synthesis. The specific architecture is shown in Figure 3.2.

From the framework diagram, it can be seen that the ROS Master is the core of the entire system, and the four modules communicate with the ROS Master in both directions for unified control. The specific workflow of the entire system is as follows: When the user issues a voice command, the microphone is used to collect the voice signal of the command and determine whether it belongs to the device wake-up word. When the instruction

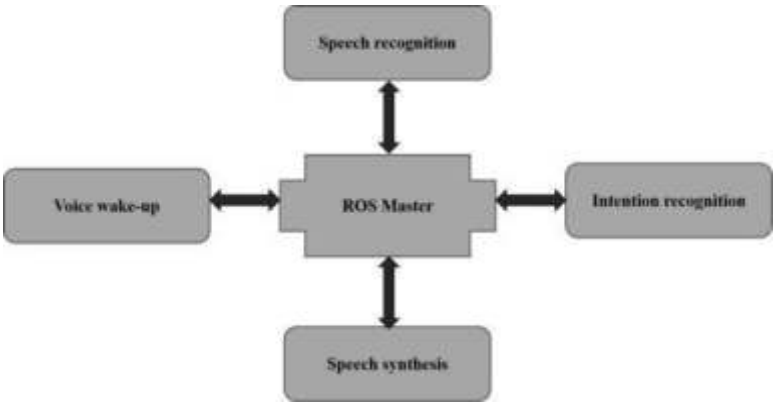


Fig. 3.2: Communication framework diagram of intelligent voice interaction system functional modules

content contains wake-up words, the device enters the working state. When the user raises a question or issues control instructions, the device will match the data information of the local intelligent question answering library through AIML when in offline state, and perform offline command word recognition. If the recognized command word is an input question, the device will search for answers in the question answering library, and the voice service of the voice cloud platform will synthesize the voice service, convert the text information into speech, and then play it; If the recognized command word is a control command, the device will execute the required action based on the command content, achieving offline voice interaction of the device. When the device is in network mode, the search range of voice command recognition will expand to the entire voice cloud platform, which will analyze the user's intentions. If the identified content involves third-party applications, the system will translate it into structured information, send it to relevant applications, process it, and return the processing results to the voice cloud platform in a structured form. If the returned information is judged as a control command, move according to the content of the command; If the returned message is judged as a question, the voice cloud platform will send the returned text reply back to the device, and the voice synthesis service of the voice cloud platform will synthesize the voice and play the result [15].

4. Result analysis.

4.1. Experimental Environment Construction. The experiment is based on ROS integrated intelligent voice interaction robot's various functional modules. ROS is a universal software framework for robots, which combines individually designed functional modules in a loosely coupled manner using a distributed structure. It is often used in the development and application of multi node and multi task robots. Due to the need for ROS to run on Linux, Ubuntu16. 04 in the Linux operating system and the corresponding ROS version Kinetic were chosen as the experimental testing platform, and they were installed and configured separately on the Linux system.

4.2. Data Sources and Preprocessing. The experiment selected the Frames English conversation dataset for intention recognition and training of the BiGRU Self attention model. Frames is a complex artificial dataset for natural language understanding research launched by a deep learning company, Maluuba, in recent years. It is mainly used in areas such as machine reading comprehension, intelligent question answering, and text mining. It contains 19 986 question and answer pairs, mostly based on text recording. 6429 text records were selected as the total sample for the experiment, and the total sample was annotated using five categories: Encyclopedia, chat, Q&A, search, and news. The number of samples in each category was 1089, 1971, 1431, 1018, and 920, respectively. Divide the total sample into three parts: Training set, validation set, and test set according to 7:2:1 [16].

4.3. Evaluation indicators. In this experiment, recognition accuracy was selected as the evaluation index for evaluating the intention recognition effect of classification algorithms, and the calculation method is

Table 4.1: Classification recognition results of models on different categories of file texts (unit:%)

Model	Encyclopedia	Chatting type	Q&A category	Search class	News	Average
DCNN	85.17	77.21	78.42	82.30	81.6	80.85
CNN-LSTM	86.32	82.03	82.35	86.84	86.51	84.84
GRU-Self-attention	88.08	86.35	83.03	87.66	88.70	86.74
BiGRU-Self-attention	89.64	89.03	86.31	88.60	90.81	88.88

shown in equation 4.1:

$$Accurary = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.1}$$

Among them, TP represents the number of times the model identifies correct results as "correct"; FP represents the number of times the model identifies correct results as "errors". TN represents the number of times the model identifies incorrect results as "correct"; FN represents the number of times the model identifies incorrect results as "errors".

4.4. Parameter settings. This experiment sets the Epoch value of the BiGRU Self attention model to 10 and the Attention_size to 128. Using Glove for word vector initialization, with a word vector dimension of 200; Using BiGRU for feature extraction with 100 hidden units; Using adaptive Adam algorithm for training optimization, the learning rate and learning rate decay rate are set to 0.001 and 0.0001.

4.5. Analysis of experimental results.

4.5.1. Verification of intent recognition classification model. In order to verify the intention recognition performance of the BiGRUSelf attention model used in intelligent speech interaction robots, experiments were conducted to compare the classification accuracy of classification methods used in other speech recognition studies such as DCNN, CNN-LSTM, GRU Self attention, etc. on the dataset. The recognition results of each model on different categories of file texts are shown in Table 4.1.

As shown in the table, BiGRU Self attention consistently has the highest recognition accuracy on various categories of file texts, with an average recognition accuracy of 88.99%, better recognition and classification performance [17]. Compared to traditional recognition methods such as DCNN and CNN-LSTM, the recognition results of BiGRU Self attention have been significantly improved, with significant improvement effects; Compared to GRU Self attention using only unidirectional GRU for feature extraction, BiGRU Self attention using bidirectional GRU contains more information, has a more comprehensive understanding of the context, and further improves recognition accuracy by an average of about 2. 03%, resulting in superior model performance.

Feature extraction is an important step in intent recognition. In order to verify the improvement of the BiGRU Self attention model in terms of training speed, the same 1000 data points were selected from the dataset to train BiGRU and BiLSTM feature extraction from commonly used intention recognition classification models. The training time for both is shown in Figure 4.1.

As shown in the Figure, under the same training data conditions, BiGRU has a faster training speed, a training time shortened by about 4 times compared to BiLSTM, and a higher training efficiency. It can be inferred that compared to intention recognition models based on Glove2BiLSTM-Self attention, Glove2BiGRU-Self attention should also have faster training speed and more obvious advantages.

4.5.2. Verification of Speech Interaction System Based on Intent Recognition. In order to verify the effectiveness of the author's design intention recognition method in the intelligent voice interaction system of robots, 500 user interaction statements of different categories were selected from the dataset with 100 entries per category, and intention recognition tests were conducted on intelligent voice interaction robots equipped with intention recognition modules[18]. The test is divided into three groups, and the recognition results of the robot are counted according to different categories. The average of the three tests is taken as the final value of the test. The measurement indicators for the effectiveness of robot intent recognition include recognition accuracy and recognition time. The recognition accuracy of the three tests is shown in Table 4.2.

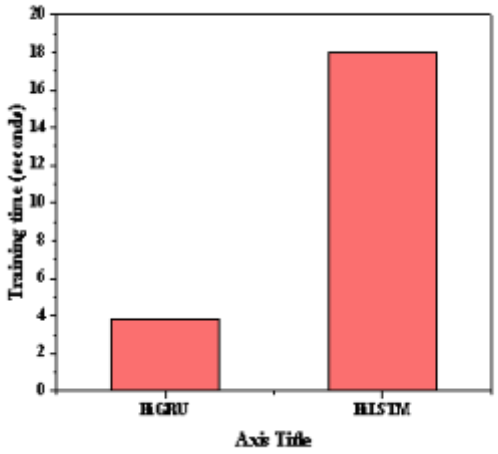


Fig. 4.1: Comparison of model training speed (Unit: seconds)

Table 4.2: Display of Test Results for Intelligent Voice Interaction Robot Intent Recognition Module (Unit:%)

Category	Test 1	Test 2	Test 3	average
Encyclopedia	91.55	93.15	88.15	91.05
Chatting type	88.14	86.51	93.17	89.23
Q&A category	86.57	89.84	88.11	88.17
Search class	91.52	86.54	85.01	87.65
News	91. 53	93.16	91.50	92.06

From Table 4.2, it can be seen that the robot's intention recognition accuracy for five types of interaction statements in three tests is above 88%, and the recognition results are stable and high. The average recognition accuracy is similar to the test results of Glove2BiGRU-Self attention, proving that the intention recognition model proposed by the author can be successfully applied to the intelligent speech interaction system of the robot.

On the basis of the above intent recognition results, remove the parts that were identified incorrectly, and then conduct a recognition time test on the parts that were identified correctly. Define the end time of user command issuance as the initial time, and record the completion time of robot intent recognition one by one. The number of tests is also divided into three, and the data categories involved in the test include five categories. The average time of three tests for each category of data is used as the test result. The time required for robots to complete correct intent recognition under different types of data is shown in Figure 4.2.

From Figure 4.2, it can be seen that the speech interaction robot loaded with Glove2BiGRU-Self attention not only has a high recognition accuracy but also has a fast recognition speed for the intention recognition of five categories of user interaction statements, with recognition time all within 0.34 seconds. Glove2BiGRU-Self attention still has high performance on robot humans, ensuring that the robot can accurately and quickly interact with user speech intelligently.

4.5.3. Practical application verification of the proposed method.

(1) Speech instruction recognition. In order to verify the implementation of an intelligent speech interaction system based on intent recognition in robots, the operation of the intent recognition module in the speech interaction system was tested using the voice command of "The weather in Yibin today" as an example. After the voice recognition node recognizes the voice command "The weather in Yibintoday" sent by the user and

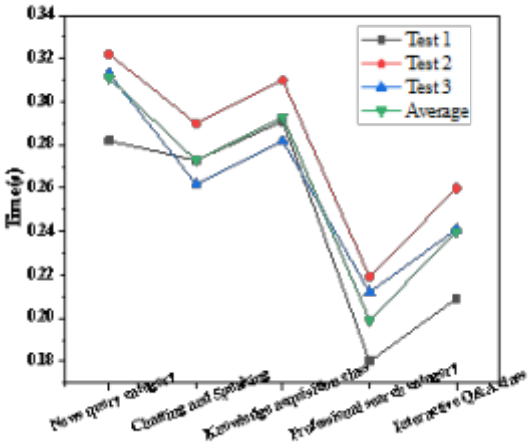


Fig. 4.2: The time required for intelligent voice robots to correctly complete intent recognition (Unit: seconds)

Table 4.3: Intelligent Voice Interaction Results of Robots under Five Categories of Communication Conversations

Category	Interrogative sentence	Answer	Result determination
Chatting type	What's the scenery like in Yibin	I heard Yibin is a very nice place	Meet expectations
Q&A category	It's bargain time	March 29, 2023 at 5:10 p.m	Meet expectations
News	What is the score of the Warriors today	109∼120	Meet expectations

publishes it as text, it automatically subscribes to the command text, searches for answers through the Internet, and publishes the search results as text of "Yibin: Wednesday, March 29, cloud to light rain southwest wind, the lowest temperature 12 degrees", and then performs text to speech conversion by the voice synthesis node, and finally plays them through robot voice. According to the test results, it can be concluded that intelligent voice robots can accurately recognize the intentions of user voice commands. In order to further verify the recognition performance of robots in human-machine voice interaction for user intentions, a multi category topic dialogue and communication were conducted with robots through voice input. In the five categories of dialogue communication, each representative is selected, and the results of robot intelligent voice interaction are shown in Table 4.3.

From the Table 4.3, it can be seen that the robot's response content is consistent with the user's voice input intention, which proves that the robot can perform intention recognition through an intelligent voice interaction system equipped with the Glove2BiGRU-Self attention intention recognition module, and search and output results based on the user's instruction intention.

(2) Mobile instruction recognition. In order to further verify the application effect of the design intent recognition method in intelligent voice interactive robots, the robot is controlled for movement through voice commands. Custom dialogue construction is carried out in the intent recognition module, and the constructed mobile command information includes the following content:

Instruction 1:"move_go_forward"
Instruction 2:"move_rotate_left"
Instruction 3:"move_rotate_right"
Instruction 4:"move_go_back"

The program matches keywords based on the text output of voice commands. When the flag is 0, it

indicates the inclusion of any keyword, indicating that the program matches successfully. The robot will move based on the successfully matched instructions. According to the command "forward left right backward", the little turtle made corresponding movements, proving that the intelligent voice interaction system built using the proposed Glove2BiGRU-Selfattention intention recognition method can control the robot's actions by matching the command content, and the movement results meet the requirements of voice commands.

5. Conclusion. In summary, the designed intelligent speech interaction robot based on intention recognition is constructed by introducing the Glove2BiGRU-Self attention classification prediction model, completing the construction of the system intention recognition function module, and integrating various functional modules of the system using ROS distributed architecture to achieve intelligent speech interaction between users and robots. The results show that compared to the intent recognition methods based on DCNN model, CNN-LSTM model, and unidirectional GRU Self attention model, the proposed intent recognition method based on GloveBibGRU-Self attention model significantly improves the recognition accuracy, reaching 88.88%. In addition to accuracy, the recognition efficiency of the model has also improved. In terms of feature extraction, compared to traditional BiLSTM based feature extraction methods, the BiGRU feature extraction method reduces training time by 4 times and has higher training efficiency. When applied to actual robot voice interaction systems, it still has high performance, achieving rapid interaction between human-machine speech. Through research, using GRU instead of traditional LSTM for speech feature classification simplifies the model structure, improves model training efficiency, and through the bidirectional modeling of GRU, the feature nodes contain richer information, resulting in higher generalization of extracted features. This enables the model to learn features more effectively, improve classification accuracy, and enable robots to accurately and quickly respond to user speech instructions. However, there are still shortcomings in this study. With the increase of training data, the classification and recognition time of the model also increases, and the recognition accuracy correspondingly decreases. Further research and improvement are needed to ensure that the model still performs well in more diverse and flexible speech instructions in reality.

REFERENCES

[1] Xing, Y., Lv, C., Wang, H., Cao, D., & Velenis, E. (2020). An ensemble deep learning approach for driver lane change intention inference. Transportation Research Part C: Emerging Technologies, 115, 102615.
[2] Wu, Z., Chen, Y., Zhao, B., Kang, X., & Ding, Y. (2021). Review of weed detection methods based on computer vision. Sensors, 21(11), 3647.
[3] Pérez-Hernández, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H., & Herrera, F. (2020). Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance. Knowledge-Based Systems, 194, 105590.
[4] Xiong, D., Zhang, D., Zhao, X., & Zhao, Y. (2021). Deep learning for EMG-based human-machine interaction: A review. IEEE/CAA Journal of Automatica Sinica, 8(3), 512-533.
[5] Lou, H., Zhao, H., & Deng, W. . (2022). Research on civil aviation passenger question intention recognition based on text classification method of self-attention and deep neural network, 22(7), 4337-4347.
[6] Li, W., Shi, P., & Yu, H. (2021). Gesture recognition using surface electromyography and deep learning for prostheses hand: State-of-the-art, challenges, and future. Frontiers in neuroscience, 15, 621885.
[7] Wang, L., Kou, Q., Zeng, Q., Ji, Z., Zhou, L., & Zhou, S. . (2022). Substation switching device identification method based on deep learning. 2022 4th International Conference on Data-driven Optimization of Complex Systems (DOCS), 1-6.
[8] Li, G., Liu, F., Sharma, A., Khalaf, O. I., Alotaibi, Y., Alsufyani, A., & Alghamdi, S. (2021). Research on the natural language recognition method based on cluster analysis using neural network. Mathematical Problems in Engineering, 2021, 1-13.
[9] Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., ... & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. Information Fusion, 80, 241-265.
[10] Zang, H., Cheng, L., Ding, T., Cheung, K. W., Wei, Z., & Sun, G. (2020). Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. International Journal of Electrical Power & Energy Systems, 118, 105790.
[11] Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked, 20, 100372.
[12] Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. Plant Methods, 17, 1-18.
[13] Wan, S., Qi, L., Xu, X., Tong, C., & Gu, Z. (2020). Deep learning models for real-time human activity recognition with smartphones. Mobile Networks and Applications, 25(2), 743-755.

[14] Guo, L., Lu, Z., & Yao, L. (2021). Human-machine interaction sensing technology based on hand gesture recognition: A review. IEEE Transactions on Human-Machine Systems, 51(4), 300-309.
[15] Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences, 11(9), 4164.
[16] Teng, F., Song, Y., & Guo, X. (2021). Attention-tcn-bigru: an air target combat intention recognition model. Mathematics, 9(19), 2412.
[17] Guo, L., Lu, Z., & Yao, L. (2021). Human-machine interaction sensing technology based on hand gesture recognition: A review. IEEE Transactions on Human-Machine Systems, 51(4), 300-309.
[18] Buerkle, A., Eaton, W., Lohse, N., Bamber, T., & Ferreira, P. (2021). EEG based arm movement intention recognition towards enhanced safety in symbiotic Human-Robot Collaboration. Robotics and Computer-Integrated Manufacturing, 70, 102137.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 10, 2024
Accepted: Jun 25, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 630–640, DOI 10.12694/scpe.v26i2.3893

POWER DATA ANALYSIS AND PRIVACY PROTECTION BASED ON FEDERATED LEARNING

YAJIE LI* SHUTING CHEN† XINMIAO HU ‡ SEN XU § AND MAO FAN¶

Abstract. In order for active distribution network operators to carry out power business such as load forecasting without meter reading rights, the author proposes a research on power data analysis and privacy protection based on federated learning. The author proposes a federated learning load forecasting framework for industry user data protection by selecting weather and time factors as the correlation factors of load. On this basis, the author constructed an industry user dataset and established a load forecasting model based on Long Short Term Time Series Network (LSTNet). At the same time, the FedML framework was used to establish a sub industry load forecasting framework based on federated learning. The results indicate that: The accuracy of the industry specific load forecasting framework based on federated learning proposed by the author is less than 9 p.u., and the theoretical maximum value of SMAPE is 210 p.u., indicating that this method has universality and universality and can be applied in different industries. The training method of this scheme is parallel, although it increases the interaction time by 1 minute, the interaction time accounts for a smaller proportion compared to the training time, and the time consumption is interaction time (1 minute)+single training time (94 minutes). Conclusion: The method can enable users in the same industry to conduct federated training without sharing load data, and support active distribution network operators in related business operations while protecting user electricity privacy. It has better predictive performance, fewer model numbers, and shorter time consumption.

Key words: Long short-term time series network, Load forecasting, Federated learning, FedML framework, Privacy protection

1. Introduction. With the increasing number of distributed power generation resources such as renewable energy and the extensive integration of various intelligent terminal devices such as smart homes, a large amount of data flow will be generated between power grid enterprises and power users, as well as between electrical equipment and control centers. The smart grid has generated an unprecedented amount of raw information that can accurately assess situational awareness, improve the intelligence, efficiency, and sustainability of multiple industrial systems [1]. Researchers generally believe that the true value of smart grids lies not in the physical interconnected devices themselves, but in the vast amount of crude and unrefined information they contain, as well as how to efficiently, quickly, and meaningfully process this information. Therefore, in recent years, the analysis and processing of data in smart grids have received widespread attention, among which data privacy protection and anomaly detection have always been hot and difficult research topics [2]. On the one hand, due to the collection, transmission, and processing of massive data, frequent communication among various participants in the smart grid has led to increasingly serious data privacy issues, such as sensitive data directly exposing user privacy information [3]. Therefore, privacy protection technology is needed to process smart grid data to ensure the privacy and security of data in the smart grid and promote the practical development of smart grid applications.

On the other hand, the problem of false data injection, which has not been effectively solved in traditional power grids, has become more serious in the information-based and digital smart grid environment. More smart meters and other devices provide malicious users with more opportunities to modify power data. Due to

*State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (Corresponding author, YajieLi9@163.com)
†State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (ShutingChen3@126.com)
‡State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (XinmiaoHu8@163.com)
§Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China. State Grid Xinjiang Electric Power Co., Ltd, Urumqi, Xinjiang, 832000, China (SenXu7@126.com)
¶Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China. State Grid Xinjiang Electric Power Co., Ltd, Urumqi, Xinjiang, 832000, China (MaoFan738@163.com)

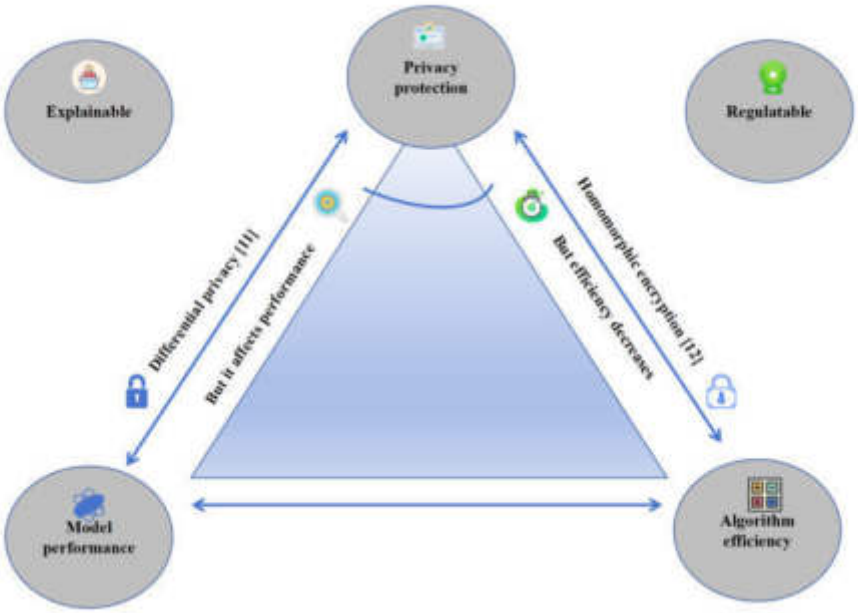


Fig. 1.1: Power Data Analysis for Federated Learning

the fact that the power supply of smart grids is generally provided based on real-time electricity consumption data [4]. If real-time electricity consumption data is not accurate, resulting in a mismatch between electricity supply and actual electricity consumption, it will not only disrupt the normal power supply order, bring huge economic losses to power companies and legitimate users, but also cause damage to the stability of the power grid. Therefore, how to effectively detect abnormal data in smart grids is an important research direction at present. As shown in Figure 1.1:

2. Literature Review. For anomaly detection of electricity consumption data in smart grids, researchers are committed to improving detection models to obtain more efficient and accurate model performance, such as from ordinary neural network models to convolutional neural network models, and then to XGBoost (Xtreme Gradient Boosting), hybrid models, etc. [5]. Truong, N uses the XGBoost model to detect normal and abnormal users, which is based on multidimensional smart meter data such as electricity consumption data, longitude and latitude data, and communication protocol data of instruments. The final accuracy is as high as 91% [6]. Ma, C. proposed a hybrid neural network model that combines convolutional and fully connected methods, specifically combining convolutional computation as a wide component and fully connected as a deep component to detect abnormal users, resulting in an AUC of up to 78% [7].

Li, Z attempted to use feedforward neural networks (FFN) to implement privacy protected abnormal electricity consumption data detection based on inner product function encryption, which is one of the function encryption methods. In addition, the scheme also achieves dynamic billing and load monitoring, without the need to learn fine-grained power consumption readings [8]. However, this scheme currently does not support the combination with convolutional neural networks. In addition, function encryption can only provide one ciphertext for a general function, which has the drawback of not being able to provide any number of keys for multiple general functions. This limits its ability to adapt to smart grid frameworks that are currently developing towards distributed frameworks.

Federated learning is a distributed machine learning environment, whose core concept is to model data without moving, make data available but not visible, and achieve collaborative joint modeling of data without leaving the local client [9]. Federated learning solves the problem of ”data silos” caused by data fragmentation

and can fully utilize participant data, at the same time, the data does not leave the domain, protecting the data security of the participating parties and reducing the possibility of leakage at the original data level.

Liu, X. first proposed a user level differential privacy federated learning framework, which provides users with different privacy protections by calculating the upper bound of the sensitivity of any user to protect all data of any user. The purpose of differential privacy protection is to hide the contribution of individual users in model training, balancing privacy loss and model performance. Users who trust the server will use user level differential privacy. This way, users can safely join federated learning [10]. Mothukuri, V. proposed a differential privacy federated learning protection method based on Gaussian mechanism, which can more accurately obtain privacy loss during model training [11]. Wang, Y. proposed a new framework NbAFL, a staged differential privacy federated learning model, which calculates gradients for each client, adjusts its variance, adds noise that follows a Gaussian distribution to prevent parameter information leakage, and meets the requirements of global DP. While protecting user privacy, the accuracy and convergence speed of the model are guaranteed. In addition, this method compares the convergence performance of the model under different pruning thresholds through pre training and selects the best pruning threshold [12]. Wang, R. proposed a federated learning protection method based on localized differential privacy mechanism, which mainly adds noise disturbance to user local data and passes it to the central server, improving model accuracy and reducing performance loss [13].

The author designed a federated learning load forecasting model for active distribution networks to protect industry user reading data. Industry user datasets and corresponding data preprocessing schemes were constructed according to industrial industry categories. Based on long and short time series networks (LSTNet), a customer local load forecasting model was established. The model parameters were aggregated using the Federated Average (FedAvg) algorithm, and a sub industry power load forecasting framework based on federated learning was established using the FedML framework. The industry load forecasting model was obtained while protecting user data privacy. The analysis results of the example show that this scheme has better predictive performance, fewer models, and shorter time consumption, and can obtain accurate predictive models without mastering user data.

3. Research Methods.

3.1. Load forecasting based on LSTNet.

3.1.1. Factors affecting load. The load curve can be decomposed into regular components, uncertain components, and noise components; The regular component is the periodic variation of load in the time dimension; The uncertain component is the non cyclical changes in load caused by factors such as weather and economy; The noise component refers to the impact of other factors that cannot be physically explained on the load [14]. Common weather factors include temperature, humidity, precipitation, and air pressure, which have a significant impact on short-term load changes; The time factor is reflected in the combined effect of short-term and long-term repetition patterns, mainly including the influence of workdays, holidays, and seasons.

3.1.2. Construction of Industry User Datasets. The author selects weather, load, and time as the characteristic types of input data for the load forecasting model. Weather includes temperature, relative humidity, precipitation, wind speed, and pressure to reflect the impact of weather factors on load; Time data includes holiday information, workday information, year, month, day, hour, and week information, used to reflect the periodicity of load, among them, holiday information is determined based on the statutory holidays in China in 2019 and 2020, and workday information includes workdays generated during the week due to compensatory leave. These constructed industry datasets only need to be retained locally by users and do not need to be provided to active distribution network operators, effectively protecting user data privacy.

3.1.3. Load forecasting model based on LSTNet. The author uses LSTNet as the load forecasting model, which consists of three parts: convolutional layer, loop layer, loop jump layer, fully connected layer, and autoregressive layer. LSTNet can use convolutional layers to extract short-term patterns and dependencies between variables from the input multivariate time series, and then use loop layers and loop jump layers to capture long-term and longer-term dependencies between variables [15]. The input of the model is a multivariate

time series

$$X = \begin{bmatrix} x_0(1) & x_1(1) & \cdots & x_n(1) \\ x_0(2) & x_1(2) & \cdots & x_n(2) \\ \cdots & \cdots & & \cdots \\ x_0(m) & x_1(m) & \cdots & x_n(m) \end{bmatrix}$$

where n represents the number of features of the multivariate time series, where there are a total of 27 features including load, weather, and time; M represents the length of the input time series.

The weight coefficients corresponding to the convolutional layer are W_C and b_C, the weight coefficients corresponding to the cyclic layer are W_R and b_R, the weight coefficients corresponding to the cyclic jump layer are and , the bias of the fully connected layer is b_D, the weight coefficients corresponding to the autoregressive layer are W_{AR} and b_{AR}, respectively. Record the weight coefficients $W = \{W_C, W_R, W_S, W_{AR}\}$ and $b = \{b_C, b_R, b_S, b_D, b_{AR}\}$ bias of each layer in LSTNet as equation 3.1:

$$\omega \in \{W, b\} \tag{3.1}$$

In the formula, ω represents the weight coefficients and biases of each layer of neurons, i.e. the model parameters of LSTNet.

3.2. Industry specific load forecasting framework based on federated learning.

3.2.1. Overview of Federated Learning. Federated learning is a machine learning method based on distributed datasets, first proposed by Google's McMahan. Federated learning includes two processes: Model training and model inference. During the model training process, multiple clients train locally with their own datasets, and the server collects model parameters to update the global model. Each client continues to train based on this update, ultimately obtaining a model that can be shared by multiple parties [16]. Model inference is the application of trained federated learning models to new data. Federated learning is divided into horizontal federated learning, vertical federated learning, and federated transfer learning based on the overlap between feature space and sample space of data from different clients [17]. Among them, horizontal federated learning is suitable for situations where the client has overlapping feature spaces but different sample spaces. The applicable conditions for horizontal federated learning are shown in equation 3.2.

$$\begin{cases} X_i = X_j \\ I_i \neq I_j \end{cases} \quad \forall D_i, D_j, i \neq j \tag{3.2}$$

In the formula, D_i and D_j represent the datasets of client i and client j, respectively; X_i and X_j represent the feature spaces of client i and client j, respectively;I_i and I_j are the sample spaces for client i and client j, respectively.

In the author's application scenario, different power end-users in the same industry each have their own dataset, with different samples in the dataset, but the corresponding features of the samples are similar and the same (load shape, weather information, time information). If future users refuse to share their meter reading data from a privacy protection perspective, active distribution network operators can use the author's proposed federated learning algorithm framework for industry user load forecasting, users only need to train their own load forecasting model locally using their own load data, and share the load forecasting model parameters with active distribution network operators instead of specific user reading load data. Active distribution network operators use the FedAvg algorithm to aggregate these model parameters, and use the FedML framework to establish a sub industry power load forecasting framework based on federated learning. Under the premise of protecting user data privacy, they obtain the industry's load forecasting model.

3.2.2. Federated training process based on FedAvg algorithm. Federated learning is based on the FedAvg algorithm to obtain a global model. The FedAvg algorithm is a horizontal federated optimization algorithm proposed by McMahan, which runs gradient descent algorithm in parallel on multiple user side clients. In each round of interaction, the central server collects parameters for aggregation and sends them back to the client for further training.

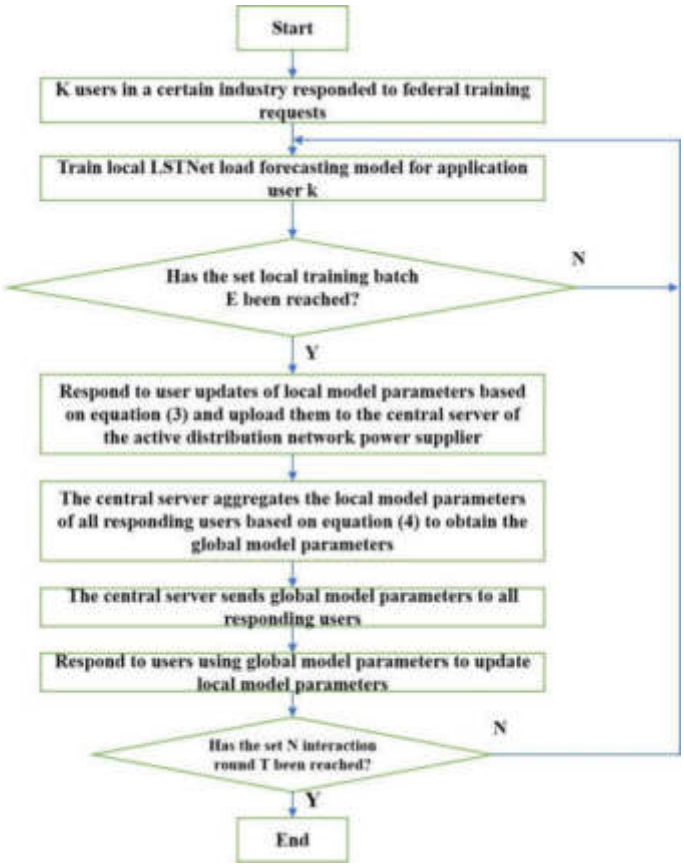


Fig. 3.1: Federated training flowchart based on FedAvg algorithm

The users participating in federated learning are responsive users, and the model parameters are the weight coefficients or biases of neurons in each layer of LSTNet. The active distribution network operator/power supply company sets up a central server to aggregate local model parameters uploaded by response users. Response users use local datasets for local training based on LSTNet load forecasting models, and do not share datasets with each other. After the local training reaches the set training batch, the corresponding user uploads the local model parameters to the central server. After collecting the local model parameters uploaded by all responding users participating in federated training, the central server generates global model parameters based on the FedAvg algorithm and sends them back to each responding user. Respond to users updating their local model parameters with global model parameters and continue training, repeating this process until the set interaction round is reached. The specific process is shown in Figure 3.2.

Firstly, the central server located in the active distribution network operator sends the initialized model parameters to ω_0 to all responding users. After the user receives the model parameters, the LSTNet model is trained locally based on the gradient descent algorithm. In the t-th round of interaction, the application user k first performs local training, and the model parameters $\omega_{k,t}$ after training are shown in equation 3.3.

$$\omega_{k,t} = \omega_{k,t-1} - \eta \nabla f(\omega_{k,t-1}) \tag{3.3}$$

In the formula, $\omega_{k,t}$ represents the weight coefficients or biases of neurons in each layer of LSTNet in response to user k's completion of local training in the t-th round of interaction; η is the learning rate; ∇f represents gradient descent.

Then, each responding user will input the trained model parameters from equation 3.3 ωk. Send to the

central server located in the active distribution network operator/power supply company. The central server collects model parameters updated by all responding users ωk. After t, aggregate the parameters according to equation 3.4 to generate global parameters [18].

$$\omega_t = \sum_{k=1}^{K} \frac{n_k}{n} \omega_{k,t} \tag{3.4}$$

In the formula, n_k represents the number of samples in response to user k; N is the total number of samples for all responding users; K is the total number of responsive users. The central server sends the global parameter ω_t to all responding users, who update the local parameters based on the global parameters and continue training until all interaction rounds are completed, that is

$$\omega_{k,t} = \omega_t \tag{3.5}$$

3.2.3. Industry specific load forecasting framework based on federated learning. The main body of the industry specific load forecasting framework based on federated learning includes active distribution network operators/power supply companies, central servers, users, and electricity sales companies. The entire process is divided into the following 7 steps:

Step 1: The active distribution network operator/power supply company sends a training request to the corresponding users in a certain target industry.

Step 2: Considering that users need to complete training locally, they can decide whether to participate in federated training. The participating users are response users who preprocess their own dataset locally and input it into the LSTNet load forecasting model for local training. After completing one round of training, the neuron parameters of each layer of LSTNet are passed to the central server [19].

Step 3: The central server aggregates local model parameters passed by response users based on the FedAvg algorithm to generate global model parameters, and passes the global model parameters to the response users. After multiple interactions, the central server obtains the industry global model.

Step 4: The central server transfers the industry global model to the active distribution network operator/power supply company.

Step 5: The active distribution network operator/power supply company distributes corresponding rewards based on the contribution level of the responding users to the overall industry model.

Step 6: The electricity sales company submits a model demand application to the active distribution network operator/power supply company based on the required industry.

Step 7: Actively return the required industry global model to the distribution network operator/power supply company and obtain profits.

3.3. Example analysis. The author verifies the effectiveness of the proposed scheme through numerical examples. The experimental environment used is described as follows: CPU is 3.70 GHz Intel Core i7-8700K, GPU is NVIDIA TITAN XP 12 GB, memory is 30 GB, operating system is Ubuntu 18.04 LTS, Python version is 1.7.2, CUDA version is 11.3, Python version is 3.8, federated learning is implemented based on FedML library, federated optimization algorithm is FedAvg algorithm, and single machine deployment mode is adopted [20]. The parameter settings for the LSTNet model used for load forecasting are shown in Table 3.1.

3.3.1. Dataset Description. The dataset selected for the calculation is a 730 day multivariate feature dataset classified by industry in a certain city from January 1, 2021 to December 30, 2022, with a time granularity of 48 points per day. Randomly select three responsive users from three industries: pharmaceutical manufacturing, food manufacturing, and rubber and plastic products in the constructed dataset. Each responsive user dataset is divided into training, validation, and testing sets in a ratio of 8:2:2 (corresponding to 584, 74, and 74 d, respectively).

3.3.2. Evaluation indicators. In order to measure the predictive performance of the model, it is necessary to use evaluation indicators to calculate the prediction error of the model. Common evaluation indicators include mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE) [21]. The author selected RMSE and SMAPE (their

Table 3.1: LSTNet model parameters

Parameter Name	Parameter values
The number of hidden units in convolutional layers	130
Convolutional layer convolution kernel width	7
The length of convolutional kernels in convolutional layers	28
Number of hidden units in the loop layer	129
Loop - skip layer hidden unit count	129
Loop skip layer skip hidden unit count	50
Learning rate	0.002
Loss rate	0.3
Number of samples in a batch	65
Training batch	210 times

Table 4.1: Federated Learning Training Scenario Settings

Scene	Interaction round/time	Local training batch/time	Is it divided by industry
1	15	20	yes
2	15	25	yes
3	20	15	yes
4	25	15	yes
5	25	15	no

values are λ RMSE and λ SMAPE, as an evaluation indicator for short-term load forecasting results, is expressed in equations 3.6 and 3.7, respectively.

$$\lambda_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{3.6}$$

$$\lambda_{SMAPE}\& = \frac{200}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{|\hat{y}_i| + |y_i|} \tag{3.7}$$

In the formula: N represents the total number of samples; $\hat{y}_i$ and y_i are the predicted and actual values at time i, respectively. RMSE amplifies the difference between larger and smaller errors using the square term, making it more sensitive to data with larger prediction deviations. A smaller RMSE indicates better model prediction accuracy. SMAPE is a revised version of MAPE, which solves the problem of MAPE being unable to calculate when the actual value is 0 and MAPE punishing negative errors more than positive errors. The range of SMAPE values is [0210], and a smaller SMAPE indicates better predictive performance.

4. Result analysis.

4.1. Analysis of Federated Learning Optimization Results. This section sets up federated learning training scenarios based on Table 4.1. Among them, the interaction round refers to the number of times the central server in each industry updates global parameters to each responding user, while the local training batch refers to the number of local iterations for each responding user. The load forecasting model based on industry electricity consumption characteristics is trained on an industry basis. This section sets up three industries, namely pharmaceutical manufacturing, food manufacturing, rubber and plastic products. Each industry uses three datasets that respond to users. The FedAvg algorithm is used to capture the common electricity consumption characteristics of a single industry and obtain the final industry load forecasting model [22].

According to Tables 4.2 and 4.3, it can be seen that Scenario 4 has the best predictive performance. Comparing scenarios 1 to 5, it can be seen that under the same total number of training iterations, the

Table 4.2: Comparison of RMSE for Predictive Accuracy of Federated Learning in Different Scenarios

Scene	RMSE/kW		
	pharmaceutical industry	Food manufacturing industry	Rubber and plastic products industry
1	131.57	79.56	162.58
2	130.48	80.79	166.43
3	129.78	80.02	167.36
4	127.12	76.55	160.78
5	131.79	84.37	178.34

Table 4.3: Comparison of SMAPE prediction accuracy for federated learning in different scenarios

Scene	SMAPE/p.u.		
	Pharmaceutical industry	Food manufacturing industry	Rubber and plastic products industry
1	3.58	5.68	4.53
2	3.57	5.85	4.52
3	3.55	5.72	4.65
4	3.47	5.45	4.38
5	3.58	5.99	4.89

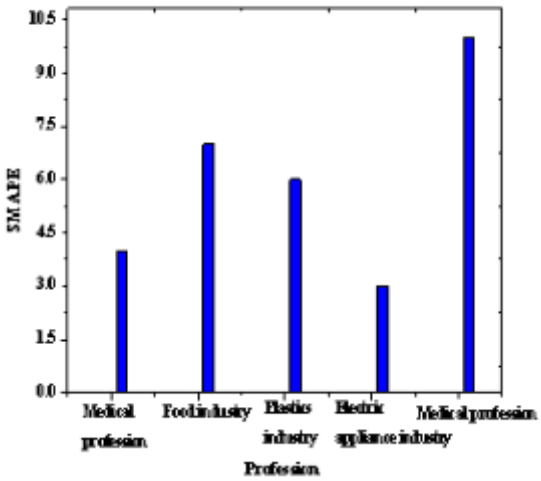


Fig. 4.1: Comparison of SMAPE prediction accuracy in different industries

interaction round has a greater impact on the performance of load forecasting than the local training batch. Comparing Scenario 4 and Scenario 5, it can be seen that the scenario based on different industries has better predictive performance than the scenario based on no industry, as it captures the electricity consumption characteristics of the industry.

Apply scenario 4 to more industries to test the generalization and adaptability of this method in different industries. The prediction accuracy SMAPE under different industries is shown in Figure 4.1.

From Figure 4.1, it can be seen that the author's proposed federated learning based industry specific load forecasting framework has a prediction accuracy SMAPE value of less than 9 p.u. in different industries,

Table 4.4: Comparison of RMSE prediction accuracy between this scheme and scheme 1

Programme	RMSE/kW		
	pharmaceutical industry	Food manufacturing industry	Rubber and plastic products industry
Option 1	126.59	79.56	158.83
Author's Proposal (Scenario 4)	127.11	76.55	160.74

Table 4.5: Comparison of RMSE prediction accuracy between this scheme and scheme 1

Programme	SMAPE/p.u.		
	pharmaceutical industry	Food manufacturing industry	Rubber and plastic products industry
Option 1	3.45	5.68	4.38
Author's Proposal (Scenario 4)	3.45	5.45	4.37

while the theoretical maximum value of SMAPE is 210 p.u., indicating that the method has universality and universality and can be applied in different industries.

4.2. Industry specific prediction plan. In industry specific prediction schemes that do not consider privacy protection, all users are required to upload raw table reading data. The smart meter uploads the user's local meter reading data to the server of the active distribution network operator/power supply company through a gateway, and accumulates it by industry. The active distribution network operator/power supply company establishes an industry load forecasting model based on the accumulated data of all users [23]. This section takes Scenario 4 in Table 4.1 as an example to compare and analyze it with industry specific prediction schemes that do not consider privacy protection. For the convenience of expression, scheme 1 in the following text refers to industry specific prediction schemes that do not consider privacy protection.

Tables 4.4 and 4.5 show the comparison of prediction accuracy between the proposed federated learning scheme and scheme 1 by the author. It can be seen that in predicting the cumulative load of industry users, the performance of this scheme is similar to that of scheme 1, indicating that the performance of the proposed privacy protection based federated learning prediction model for power load in different industries is similar to that of traditional independent prediction schemes without considering privacy protection. However, this scheme does not require uploading original meter reading data to active distribution network operators/power supply companies, effectively protecting user privacy.

4.3. Independent prediction scheme for individual users. In a user independent prediction scheme that does not consider privacy protection, users need to upload raw meter reading data, and the active distribution network operator/power supply company establishes different prediction models for each user's meter reading data. This section takes scenario 4 in Table 4.1 as an example to compare and analyze it with the sub user independent prediction scheme that does not consider privacy protection [24]. For the convenience of expression, scheme 2 in the following text refers to a sub user independent prediction scheme that does not consider privacy protection.

The accuracy of this scheme is slightly inferior to scheme 2, but it considers privacy protection and only requires the establishment of three industry prediction models, which has advantages in computing speed and memory usage.

4.4. Comprehensive comparative analysis. Table 4.6 provides a comprehensive comparison between this scheme and the scheme that does not consider privacy protection.

In terms of predictive performance, this scheme is similar to scheme 1 but slightly inferior to scheme 2. In terms of the number of models, this scheme is the same as scheme 1, both representing the number of industry users. However, compared to scheme 2 (where the number of models and users is the same), this scheme has a relatively fewer number of models, and its advantage is more pronounced as the number of users participating

Table 4.6: Comprehensive comparison between this scheme and traditional schemes

Programme	Do you consider privacy protection	SMAPE/p.u. pharmaceutical industry	Food manufacturing industry	Rubber and plastic products industry	Number of models /piece	Transfer data volume /MB	time consuming /min
Option 1	no	3.46	5.68	4.38	3	65.06	278
Option 2	no	3.56	5.48	4.32	8	65.07	838
Author's proposal	yes	3.48	5.45	4.37	4	1070	94

in training increases. In terms of data transmission, traditional solutions only require the transmission of user read table data, which is the number of users (8) × the number of features (27) × the number of days (730) × the number of load points within a day (49) × the size of floating-point numbers (8). In this scheme, each responding user needs to transmit local model parameters to the central server during each interaction round, with a data transmission amount of users (8) × the number of model parameters x the number of interaction rounds (20) × the size of floating-point numbers (8). The number of model parameters is very large, with only the convolutional layer, loop layer, and loop skip layer having 129 × 129 × 49 parameters . Due to the consideration of privacy protection, this scheme generates a large amount of communication data, and the transmission cost is higher compared to traditional schemes. In terms of time consumption, the average time for a single training session is 94 minutes. Scheme 1 requires the establishment of a model for each industry dataset, so the time consumption is the number of industries (3) × the time for a single training session (94 minutes); Option 2 requires the establishment of a model for each user's dataset, so the time required is the number of users (9) multiplied by the single training time (94 minutes); The training method of this scheme is parallel, although it increases the interaction time by 1 minute, the interaction time accounts for a smaller proportion compared to the training time, and the time consumption is interaction time (1 minute)+single training time (94 minutes). In summary, although this scheme increases the amount and cost of data transmission, it ensures the privacy of table reading data, maintains better predictive performance, fewer models, and shorter time consumption.

5. Conclusion. The author proposes a sub industry power load forecasting framework based on federated learning from the perspective of data protection for meter reading. Firstly, select weather, economic, and time factors as the main influencing factors of load to construct a dataset; Then, establish a load forecasting model based on LSTNet; Finally, a sub industry load forecasting framework based on federated learning was established using the FedAvg algorithm and the FedML framework. Case analysis shows that active distribution network operators can obtain predictive models without knowing user data. Although it increases the amount of transmitted data, it has better predictive performance, fewer models, and shorter time consumption.

Federated learning has high requirements for communication, and in situations where the required global model size is large, network bandwidth limitations and the number of working nodes may exacerbate the communication bottleneck of federated learning. Subsequently, feature filtering can be performed on weather features to determine the optimal weather features to reduce irrelevant meteorological interference, reduce the amount of data transmitted by the model, improve prediction accuracy, and enhance the generalization ability of the prediction model. At the same time, further consideration will be given to quantifying and certifying the contribution of users to the model. Rewards will be distributed based on the contribution of responsive users to the prediction model. Economic measures will be used to encourage more users to actively participate in federated learning, resulting in economic benefits for users. Active distribution network operators will have a more accurate understanding of industry load forecasting models, forming a virtuous cycle.

REFERENCES

[1] Wei, K., Li, J., Ding, M., Ma, C., Su, H., Zhang, B., & Poor, H. V. (2021). User-level privacy-preserving federated learning: Analysis and performance optimization. IEEE Transactions on Mobile Computing, 21(9), 3388-3401.

[2] Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., ... & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE transactions on information forensics and security, 15, 3454-3469.
[3] Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., ... & He, B. (2021). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3347-3366.
[4] Li, Y., Wang, R., Li, Y., Zhang, M., & Long, C. (2023). Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach. Applied Energy, 329, 120291.
[5] Liu, D., & Simeone, O. (2020). Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control. IEEE Journal on Selected Areas in Communications, 39(1), 170-185.
[6] Truong, N., Sun, K., Wang, S., Guitton, F., & Guo, Y. (2021). Privacy preservation in federated learning: An insightful survey from the GDPR perspective. Computers & Security, 110, 102402.
[7] Ma, C., Li, J., Ding, M., Yang, H. H., Shu, F., Quek, T. Q., & Poor, H. V. (2020). On safeguarding privacy and security in the framework of federated learning. IEEE network, 34(4), 242-248.
[8] Li, Z., Sharma, V., & Mohanty, S. P. (2020). Preserving data privacy via federated learning: Challenges and solutions. IEEE Consumer Electronics Magazine, 9(3), 8-16.
[9] Jia, B., Zhang, X., Liu, J., Zhang, Y., Huang, K., & Liang, Y. (2021). Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT. IEEE Transactions on Industrial Informatics, 18(6), 4049-4058.
[10] Liu, X., Li, H., Xu, G., Chen, Z., Huang, X., & Lu, R. (2021). Privacy-enhanced federated learning against poisoning adversaries. IEEE Transactions on Information Forensics and Security, 16, 4574-4588.
[11] Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619-640.
[12] Wang, Y., Bennani, I. L., Liu, X., Sun, M., & Zhou, Y. (2021). Electricity consumer characteristics identification: A federated learning approach. IEEE Transactions on Smart Grid, 12(4), 3637-3647.
[13] Wang, R., Lai, J., Zhang, Z., Li, X., Vijayakumar, P., & Karuppiah, M. (2022). Privacy-preserving federated learning for internet of medical things under edge computing. IEEE journal of biomedical and health informatics, 27(2), 854-865.
[14] Badr, M. M., Mahmoud, M. M., Fang, Y., Abdulaal, M., Aljohani, A. J., Alasmary, W., & Ibrahem, M. I. (2023). Privacy-preserving and communication-efficient energy prediction scheme based on federated learning for smart grids. IEEE Internet of Things Journal, 10(9), 7719-7736.
[15] Fang, C., Guo, Y., Wang, N., & Ju, A. (2020). Highly efficient federated learning with strong privacy preservation in cloud computing. Computers & Security, 96, 101889.
[16] Song, M., Wang, Z., Zhang, Z., Song, Y., Wang, Q., Ren, J., & Qi, H. (2020). Analyzing user-level privacy attack against federated learning. IEEE Journal on Selected Areas in Communications, 38(10), 2430-2444.
[17] Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Federated learning for data privacy preservation in vehicular cyber-physical systems. IEEE Network, 34(3), 50-56.
[18] Zhou, C., Fu, A., Yu, S., Yang, W., Wang, H., & Zhang, Y. (2020). Privacy-preserving federated learning in fog computing. IEEE Internet of Things Journal, 7(11), 10782-10793.
[19] Lakhan, A., Mohammed, M. A., Nedoma, J., Martinek, R., Tiwari, P., Vidyarthi, A., ... & Wang, W. (2022). Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare. IEEE journal of biomedical and health informatics, 27(2), 664-672.
[20] Yin, L., Feng, J., Xun, H., Sun, Z., & Cheng, X. (2021). A privacy-preserving federated learning for multiparty data sharing in social IoTs. IEEE Transactions on Network Science and Engineering, 8(3), 2706-2718.
[21] Wen, M., Xie, R., Lu, K., Wang, L., & Zhang, K. (2021). Feddetect: A novel privacy-preserving federated learning framework for energy theft detection in smart grid. IEEE Internet of Things Journal, 9(8), 6069-6080.
[22] Thapa, C., Chamikara, M. A. P., & Camtepe, S. A. (2021). Advancements of federated learning towards privacy preservation: from federated learning to split learning. Federated Learning Systems: Towards Next-Generation AI, 79-109.
[23] Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. IEEE journal of biomedical and health informatics, 27(2), 778-789.
[24] Singh, S., Rathore, S., Alfarraj, O., Tolba, A., & Yoon, B. (2022). A framework for privacy-preservation of IoT healthcare data using Federated Learning and blockchain technology. Future Generation Computer Systems, 129, 380-388.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 10, 2024
Accepted: Jun 25, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 641–650, DOI 10.12694/scpe.v26i2.3894

DESIGN OF SECURITY AND PRIVACY MODELS FOR OPTIMIZING THE SELECTION OF CLOUD SERVICE PROVIDERS IN CLOUD COMPUTING ENVIRONMENTS

FAN YANG*, FUQIANG TIAN†, HONGYU WU‡, JUN MOU§, SHILEI DONG¶, AND MAONAN LIN‖

Abstract. In order to solve the problem of difficult access to personalized and high-quality services for cloud users, the author proposes a security and privacy model for optimizing the selection of cloud service providers in the cloud computing environment. This model first divides user nodes into three types based on historical transactions between nodes: familial nodes, unfamiliar nodes, and ordinary nodes; Secondly, in order to protect the privacy information feedback from nodes, a trust evaluation agent is introduced as the subject of trust evaluation, and a trust value evaluation method based on user type is designed; Finally, considering the dynamic nature of trust, a new trust update mechanism based on service quality is proposed by combining transaction time and transaction amount. The experimental results show that compared with the AARep model and PeerTrust model, this model not only has advantages in scenarios with a lower proportion of malicious nodes, but also improves interaction success rates by 12% and 18%, respectively, in harsh scenarios where the proportion of malicious nodes exceeds 72%. This model overcomes the low success rate of interaction between user nodes and service nodes in cloud environments and has strong resistance to malicious behavior.

Key words: Trust model, Personalized cloud services, User type, Privacy protection, Service quality, Update mechanism

1. Introduction. Cloud computing represents the convergence of various computing and networking technologies like distributed computing, parallel computing, and virtualization, delivered over the internet to governments, businesses, and individuals. It offers three primary service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). These models enable users to access IT resources and services remotely, facilitating tasks such as storage, computing power, and software applications without the need for on-site infrastructure [1].

Cloud computing is an emerging network computing model. Compared with traditional information technology, it has overwhelming advantages such as saving IT investment, fast and simple deployment, on-demand resource allocation, low usage costs, powerful computing power, and unlimited storage capacity. It has been strongly advocated and promoted by governments and enterprises around the world, bringing about significant changes in the computing and business fields. In order to quickly seize the high ground in the field of cloud computing and enhance enterprise competitiveness, global IT giants such as Amazon, Microsoft, Google, Alibaba, etc. are actively developing cloud computing platforms and have successively launched their own cloud computing products. The currently recognized cloud computing platforms include Amazon Simple Storage Service Platform (S3) and Elastic Computing Platform (EC2), Google's App Engine, IBM's Blue Cloud, Microsoft's Azure Cloud Platform, and Alibaba's Alibaba Cloud [2]. With the strong promotion of the government, the domestic cloud computing industry chain has gradually formed, and innovative achievements such as virtualization technology, distributed computing technology, big data processing and mining, and artificial intelligence have begun to be applied. Cloud applications in industries such as transportation and automotive cloud, logistics cloud, medical cloud, and financial e-commerce cloud are emerging and developing [3,4].

With the rapid development and popularization of cloud computing, more and more individual and enterprise users are considering outsourcing their private data to cloud service providers to enjoy affordable data

*Sichuan Zhongdian Oixing Information Technology Co., Ltd, Chengdu, Sichuan, 610000, China (Corresponding author, FanYang9@ 126.com)
†Sichuan Zhongdian Oixing Information Technology Co., Ltd, Chengdu, Sichuan, 610000, China (FuqiangTian5@163.com)
‡Sichuan Zhongdian Oixing Information Technology Co., Ltd, Chengdu, Sichuan, 610000, China (HongyuWu6@126.com)
§Sichuan Zhongdian Oixing Information Technology Co., Ltd, Chengdu, Sichuan, 610000, China (JunMou5@163.com)
¶Sichuan Zhongdian Oixing Information Technology Co., Ltd, Chengdu, Sichuan, 610000, China (ShileiDong9@126.com)
‖Sichuan Zhongdian Oixing Information Technology Co., Ltd, Chengdu, Sichuan, 610000, China (MaonanLin@163.com)

storage and computing services. However, the emergence and development of any new thing is a double-edged sword, and cloud computing is no exception. Compared to traditional information technology, it has enormous technological advantages and commercial potential, but also brings many new problems and challenges [5-6]. Among them, cloud security issues bear the brunt and have become the main factor restricting the further development and application promotion of cloud computing. According to a 2009 cloud computing survey report by Gartner, over 70% of businesses do not adopt cloud computing primarily due to concerns about data security and privacy. In a research report by Gartner and IDC, data security and privacy issues have been consistently listed as the top challenges in cloud computing technology [7,8]. This is because once user data is outsourced to a remote cloud service provider, the data will be detached from the direct physical control of the data user, and user data stored in the cloud will face a dual threat from the cloud service provider and external malicious attackers.

2. Literature Review. Cloud computing services are highly valued by companies for their myriad advantages. Nevertheless, safeguarding data privacy remains a paramount concern for users, as existing laws in this domain exhibit numerous inconsistencies that necessitate refinement. Alkhasawneh, A. et al. explored the legal framework governing privacy concerns in cloud computing, highlighting deficiencies and advocating for additional provisions to enhance consumer experiences, bolster service quality, and fortify personal data protection. The paper concluded with a set of recommendations directed towards both government entities and private companies, aiming to augment the accountability of cloud computing service providers in safeguarding personal data from privacy breaches[9]. Haipeng, S. et al. introduced a novel cross-domain identity authentication protocol with a focus on privacy protection. This protocol offers several key advantages, including a self-authentication key generation algorithm which allows mobile terminals to generate their own public/private key pairs, thus eliminating security vulnerabilities associated with third-party key distribution and custody. Additionally, cross-domain identity authentication is facilitated through blockchain technology, enabling the calculation of alliance keys between edge servers. The protocol ensures simplicity and efficiency in the cross-domain authentication process. Moreover, it ensures the revocability of identity authentication, ensuring that once a mobile terminal is logged out or exits the system, its legitimate identity immediately becomes invalid to uphold the security of system resources. The protocol's security has been demonstrated and its effectiveness verified under the assumptions of the discrete logarithm problem and the Diffie-Hellman problem[10]. Wang et al. introduced an innovative privacy protection approach grounded in K-anonymity principles. This method not only safeguards the query and location privacy of cloud users but also reconciles the tension between privacy protection and service quality. Through simulation experiments, the efficacy of this approach has been validated, affirming its ability to uphold user privacy while maintaining service standards[11]. YiDING et al. investigated a trusted privacy service computation model tailored for common application scenarios of convolutional neural networks. They delved into data and model computation techniques that leverage homomorphic encryption to safeguard data privacy. Additionally, they devised a method for service certificate storage and rights allocation computation utilizing blockchain and smart contract technology, ensuring the transparency, reliability, and traceability of service computation. This research also pioneered a novel cloud environment resource and data service model, facilitating the seamless integration of resources and fostering a sharing economy among resource providers, model owners, and users. Finally, they conducted experimental analyses to scrutinize the efficacy of the privacy protection methods integrated into the model [12].

The author proposes a Trust Model based on User Types and Privacy Preservation for the Personalized Cloud Services (P3Trust) for personalized cloud services. This model first transforms the subject of trust evaluation from selfish user nodes to objective and fair trust evaluation agents. The trust evaluation process is transparent to user nodes and service nodes, and user nodes will not be able to obtain sensitive historical information. It also improves the security of entity privacy information such as identity, interests, and evaluations on the transmission channel, effectively suppressing malicious behaviors such as collusion fraud and malicious recommendations, making the results of trust evaluation more realistic; Secondly, an efficient trust value evaluation method based on user types is proposed. This method divides users into three types based on historical transactions between user nodes and service nodes: familial users, unfamiliar users, and ordinary users. Different trust value calculation methods are used for different user types, which not only improves the efficiency of trust evaluation, but also reasonably solves the initialization problem of trust relationships between new user

Table 3.1: Trust Relationship Table

User	Historical comprehensive trust value	Last transaction interest	Last transaction review	Last transaction request time
u_1	T_{1j}	Q_{1j}	$E(Q_{1j})$	$(t_{11}^{(1)}, \cdots, t_{1n}^{(1)})$
u_2	T_{2j}	Q_{2j}	$E(Q_{2j})$	$(t_{21}^{(1)}, \cdots, t_{2n}^{(1)})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_k	T_{kj}	Q_{kj}	$E(Q_{kj})$	$(t_{k1}^{(1)}, \cdots, t_{kn}^{(1)})$

nodes and service nodes; Finally, a trust update mechanism based on service quality is proposed, taking into account the transaction time and amount between user nodes and service nodes. The experimental findings indicate that the trust evaluation outcomes of P3Trust accurately, impartially, and authentically portray the trust dynamics between user nodes and service nodes. Moreover, these results underscore the robustness of P3Trust even in challenging environments, highlighting its reliability and effectiveness.

3. Research Methods.

3.1. Trust Relationship. Trust relationship fundamentally relies on the subjective endorsement of one entity by another, encompassing subjective, ambiguous, and uncertain traits that resist precise measurement. Currently, trust lacks a standardized definition, with trust value serving as the sole quantitative metric for assessing the trust dynamics between user nodes and service nodes.

3.2. Definition and representation of models.

Definition 1: User node. The individuals who make service requests in the system and the processes representing them are represented by U, which represents the set of all user nodes. $u_i(u_i \in U)$represents the i-th user node.

Definition 2: Service nodes. The individuals and representative processes that provide services to user nodes in the system, represented by O as the set of all service nodes, and $o_j(o_j \in O)$as the j-th service node.

Definition 3: Trust evaluation agents. Individuals deployed by cloud computing authorities who can fairly and objectively evaluate the trust level of U in O on behalf of user nodes, represented by A as the trust evaluation agent in the cloud environment.

Definition 4: Personalized services. Services that reflect the U feature and are of interest, or services that reflect the O feature and are good at, are represented by Q_{ij} as the personalized service vector for the last request of o_j by u_i, and $E(Q_{ij})$ as the evaluation vector for the personalized service Q_{ij} by ui[13,14].

Definition 5: Trust Relationship Tables (TRTs). The trust information table saved by the trust evaluation agent records the historical interaction information between U and O, and its structure is shown in Table 3.1.

Definition 6: Historical related direct trust. The current direct evaluation of U on O based on the direct interaction experience between U and O history is quantitatively represented by historical related direct trust values.

Definition 7: Historical related indirect trust. The current indirect evaluation of U on O based on the similarity between U and the reference node is quantitatively represented by historical related indirect trust values [15].

Definition 8: Comprehensive trust related to history. The overall evaluation of U on O is obtained by combining the direct evaluation of U on O and the indirect evaluation of U on O, and is quantitatively represented by historical related comprehensive trust values.

3.3. Trust Model Design. The design concept of a trust model for personalized cloud services based on user type and privacy protection is that the trust evaluation agent is the main body of trust evaluation, and the trust evaluation agent adopts corresponding trust value evaluation methods based on user type to dynamically obtain the degree of trust of user nodes in service nodes. The goal of this model is to enhance the accuracy of trust assessment while protecting the privacy of user nodes, and help them obtain high-quality services to improve resource utilization. Based on the historical transactions between user nodes and service nodes, the

Table 3.2: Classification of User Nodes and Their Trust Value Evaluation Methods

Customer type	Is there a historical transaction record	$\geqslant T_b$	$\leqslant t_b$	Service Node O
Family nodes	Yes	Yes	Yes	$O = \{o \mid T_o^{(n)} = \max_{j=1}^{n}(T_{oj}^{(n)})\}$
Unknown node	No			$O = \{o \mid T_o^{(n)} = \max_{j=1}^{n}(R_{(N\_user,o_j)}^{(o\_user)})\}$
Ordinary node	Yes	No	Yes	$O = \{o \mid T_o^{(n)} = \max_{j=1}^{n}(T \oplus R_{(o_j)})\}$
	Yes	Yes	No	
	Yes	No	No	

trust evaluation agent defines user nodes as familial nodes, unfamiliar nodes, and ordinary nodes, and the trust value evaluation method for each type of node is different, as shown in Table 3.2.

Definition 9: Family nodes. User nodes in the system who have had historical transactions with service nodes, and whose historical comprehensive trust value is greater than the minimum trust value Tb acceptable to the user node, and whose transaction time interval does not exceed the time threshold tb. The trust evaluation method between family nodes and service nodes is shown in section 3.3.1.

Define 10: Unfamiliar nodes. The user node that requests service for the first time has no historical transactions with the service node, that is, a new user node. The trust evaluation method between unfamiliar nodes and service nodes is shown in section 3.3.2.

Definition 11: Ordinary nodes. All user nodes in the system, except for familial and unfamiliar nodes. The trust value evaluation method between ordinary nodes and service nodes is shown in section 3.3.3.

3.3.1. Trust evaluation of family nodes on service nodes. If the user node is a family node and its transaction interval is within the time threshold range of ($|t_n^{(1)} - t_{(n-1)}^{(1)}| \leqslant t_b$) , it indicates that the user node has requested a transaction again in a short period of time. Generally, the user node's interests (that is personalized needs) will not change significantly; Secondly, when the historical transaction trust value of a user node is not less than its minimum acceptable trust value ($T_{so_j}^{(n-1)} \geqslant T_b$), it indicates that the transaction record has high reference value. Therefore, the trust evaluation agent selects the appropriate service node for the family node among the service nodes it has traded with. Trust has a time-dependent characteristic, which is reflected in the fact that family nodes are more willing to trust recent transaction records. Therefore, the author introduces a time decay function and defines it as follows 3.1:

$$S^{(n)} = \frac{t_n^{(1)} - t_{n-1}^{(1)}}{\sum_{j=1}^{k}(t_j^{(1)} - t_{j-1}^{(1)})} \tag{3.1}$$

where $t_i^{(1)}$ represents the i-th transaction request time; The larger the $S^{(n)}$, the longer the time interval between the last transaction between the family node and the service node and the current transaction.

Using $\lim_{x \to -\infty} e^x = 0$, define the time decay factor $\Delta t^{(n)} = e^{-S(n)}$ to measure the freshness of transactions.

Based on the comprehensive time related characteristics, the trust evaluation agent calculates the historical related direct trust value $\widehat{T}_{so_j}^{(n)}$ based on historical transaction records, which is defined as the following equation 3.2:

$$\widehat{T}_{so_j}^{(n)} = T_{so_j}^{(n-1)} * \Delta t^{(n-1)} \tag{3.2}$$

Among them: $\widehat{T}_{so_j}^{(n-1)}$ represents the comprehensive trust value of the previous transaction, $\Delta t^{(n-1)}$ is the time decay factor, and $\widehat{T}_{so_j}^{(n-1)}$represents the reference value of historical transactions for this transaction. According to equation 3.2, the trust evaluation agent calculates the historical related direct trust values between the family node and the service node that has been traded, as shown in Table 3.2. The maximum $\widehat{T}_o^{(n)}$ is selected as the basis for this transaction of the family node. In theory, this trust value evaluation method can improve the efficiency and accuracy of trust value evaluation.

3.3.2. Trust evaluation of service nodes by unfamiliar nodes . One of the hotspots in trust research is how to initialize the trust relationship between unfamiliar nodes (new users) and service nodes in a cloud environment. In social networks, there is a saying that "different paths do not conspire", which is reflected in the following two aspects in cloud environments:

1) Strange nodes are more willing to refer to the historical transaction records of nodes with similar interests;

2) Strange nodes are more willing to refer to the historical transaction records of nodes with similar evaluations. Indirect trust relationship is the basis for trust evaluation agents to help unfamiliar nodes determine service nodes by referring to the historical transactions of other nodes when they have insufficient understanding of the service node.

The author proposes an indirect trust value evaluation method based on interest similarity and evaluation similarity to initialize the trust relationship between unfamiliar nodes and service nodes.

1) Interest similarity. In social networks, when unfamiliar nodes request personalized services, they are more willing to refer to the historical transaction records of nodes with similar interests. Therefore, the similarity of interests between unfamiliar nodes and reference nodes is an important influencing factor in indirect trust evaluation. Interest similarity refers to the degree of similarity between personalized demand vectors of any two nodes. Let $Q(N\_user) = (n\_q_1, n\_q_2, \cdots, n\_q_n), Q(O\_user) = (o\_q_1, o\_q_2, \cdots, o\_q_n)$ be the personalized demand vectors for the unfamiliar node N_user and the reference node O_user, and use the cosine similarity between Q(N_user) and Q(O_user) to represent the interest similarity between N_user and Q_user (3):

$$P\_Sim^{(Q)}_{(O\_user)}(O\_user) = \frac{Q(N\_user) \cdot Q(O\_user)}{|Q(N\_user)| * |Q(O\_user)|} \tag{3.3}$$

2) Evaluate similarity. Similarly, when a stranger node requests personalized services, in addition to considering nodes with similar interests, they are also willing to refer to the historical transaction records of nodes with similar evaluations. Therefore, the evaluation similarity between the stranger node and the reference node is another important influencing factor in indirect trust evaluation.

Evaluation similarity refers to the degree of consistency between any two nodes in evaluating the same behavior. Let $S' = \{S'_1, S'_2, \cdots, S'_n\}$ be the public service provided by the trust evaluation agent, and the evaluation vectors for N_user,O_user.and S' after interaction are $E^{(s')}(Q(N\_user)) = (e(n\_q_1), e(n\_q_2), \cdots, e(n\_q_n))$, $E^{(s')}(Q(O\_user)) = (e(o\_q_1), e(o\_q_2), \cdots, e(o\_q_n))$.

Trust relationships originate from social networks, which have subjective characteristics, uncertainty, and fuzziness. The essence of the above characteristics of trust relationships is their gray nature. Therefore, the gray correlation coefficient between $E^{(S')}(Q(N\_user))$ and $E^{(S')}(Q(O\_user))$ is used to define the evaluation similarity between N_user and Q_user 3.4.

$$E\_Sim^{(Q)}_{(N\_user)}(O\_use) = \frac{\Delta_{min} + \rho\Delta_{max}}{\Delta + \rho\Delta_{max}} \tag{3.4}$$

Among them: ρ for the resolution coefficient, it is usually taken as 0.5; Δ_{Min}, Δ_{Max}, Δ is the minimum difference, maximum value, and absolute difference between the two poles of $E^{(S')}(Q(N\_user))$ and $E^{(S')}(Q(O\_user))$

3) The synthesis of historical related indirect trust values. In order to initialize the trust relationship between unfamiliar nodes and service nodes, an indirect trust value evaluation method is proposed. The trust evaluation agent combines the interest similarity $P\_Sim^{(Q)}_{(O\_user)}(N\_user)$ and evaluation similarity $E\_Sim^{(Q)}_{(O\_user)}(N\_user)$ of N_user and O_user, as well as the historical comprehensive trust value $T^{(n-1)}_{(O\_user)}$ of O_user. According to equation 3.5, the historical related indirect trust value $\widehat{R}^{(O_1 user_j)}_{(N\_user,O_i)}$ of O_user and each service node is calculated, as shown in equation 3.5:

$$\widehat{R}^{(O_1 user_j)}_{(N\_user,O_i)} = P_{Sim^{(Q)}}{}_{(O\_user)}(N\_user) * T^{(n-1)}_{(O_{user})} * E_{Sim^{(Q)}}{}_{(O\_user)}(N\_user) * \Delta t^{(n-1)} \tag{3.5}$$

As shown in Table 3.2, the trust evaluation agent selects the $T^{(n)}_O$ with the highest historical indirect trust value as the basis for the initial transaction of the unfamiliar node, and feeds back the service node

information to the user node. The trust evaluation method proposed by the author for unfamiliar nodes and service nodes provides a solution to the problem of initializing the trust relationship of new user nodes, with high accuracy [16].

3.3.3. Trust evaluation between ordinary nodes and service nodes . For ordinary nodes, they are not unfamiliar nodes, but their historical transaction records do not meet the requirements of family nodes. Therefore, they not only have their own historical transaction records, but also refer to the historical transaction records of other nodes to evaluate trust relationships. Therefore, trust evaluation agents cannot only evaluate the trust relationship between them and service nodes based on $\widehat{T}_{(o_j)}^{(n)}$ or $\widehat{R}_{(user,o_j)}^{(other\_users)}$. On the basis of the previous text, the author proposes a historical related comprehensive trust value evaluation method to evaluate the trust relationship between ordinary nodes and service nodes, and defines the historical related comprehensive trust value according to equation 3.6.

$$\widehat{T} \oplus \widehat{R}_{(o_j)} = \alpha \widehat{T}_{(o_j)}^{(n)} + (1-\alpha)\widehat{R}_{(user,o_j)}^{(other\_users)} \tag{3.6}$$

Among them: α is a historical related direct trust factor; $\widehat{T}_{(o_j)}^{(n)}$ is the historical related direct trust value; $\widehat{R}_{(user,o_j)}^{(other\_users)}$ is the indirect trust value related to history. Finally, as shown in Table 3.2, the trust evaluation agent selects the largest $\widehat{T}_{O}^{(n)}$ as the basis for this transaction of the ordinary node.

3.4. Trust Update Mechanism Based on Service Quality. Given the dynamic nature of trust in cloud computing environments, the trust relationship between any two nodes is not static. Therefore, after the transaction between the user node and the service node is completed, it is necessary to update the trust relationship between the user node and the service node in a timely manner [17]. The author proposes a new Quality of Service based Trust Updating Mechanism (QoS UM). The steps of QoS UM are: After the user node completes the transaction with the service node, the user node pays the fee to the service node and provides feedback on the service evaluation. The trust value update mechanism is activated, and the trust evaluation agent updates the TRT based on the user node's satisfaction with the transaction. Firstly, the evaluation of the service by the user node is the most important reference factor for trust updates. Therefore, the trust evaluation agent defines the satisfaction $N^{(Q)}$ of the user node with this transaction based on the Q and E(Q) of the user node according to equation 3.7, as follows:

$$N^{(Q)} = Q * E(Q)^T \tag{3.7}$$

Secondly, the transaction volume between user nodes and service nodes can reflect the quality of service provided by the service node. Therefore, the transaction volume between user nodes and service nodes is one of the important factors for trust updates. The transaction volume related factor $M^{(n)}$ is defined according to equation 3.8:

$$M^{(n)} = \frac{(M_n)^{\omega}}{\sum_{j=1}^{p}(M_j)^{\omega}} \tag{3.8}$$

Among them: ω represents the adjustment factor for transaction volume, taking values based on actual circumstances; M_j represents the j-th transaction amount between the user node and the service node.

Finally, user nodes are more willing to trade with service nodes that can cooperate stably in the long term. If the transaction time between user nodes and service nodes is longer, it reflects that the service node can provide high-quality services. Therefore, the transaction time between user nodes and service nodes is another important reference factor for trust updates. The transaction time related factor $I^{(n)}$ is defined according to equation 3.9:

$$I^{(n)} = \frac{(t_n^{(2)} - t_n^{(1)})^k}{\sum_{j=1}^{n}(t_j^{(2)} - t_j^{(1)})^k} \tag{3.9}$$

Among them: κ Represents the adjustment factor for trading time, taking values based on actual circumstances; $t_i^{(1)}$ represents the request time for the i-th transaction; $t_i^{(2)}$ represents the end time of the i-th transaction.

Table 4.1: Experimental Parameter Settings

Parameter	Parameter values	Parameter	Parameter values
Number of service nodes	100	α	0.5
Number of user nodes	50	T_b	0.3
Types of interests	3	τ	0.5
ψ	0.5	t_b	0.003
ω	2	κ	2

The trust evaluation agent obtains the service quality NMI_{so_j} of this transaction by integrating the satisfaction $N^{(Q)}$, transaction volume related factor $M^{(n)}$, and transaction time related factor $I^{(n)}$ of this transaction, which is defined as the following equation 3.10:

$$NMI_{so_j} = N^{(Q)} * M^{(n)} * I^{(n)} \tag{3.10}$$

The trust evaluation agent integrates historical transaction records and the service quality of this transaction, and updates the trust relationship according to equation 3.11. The following equation 3.11:

$$T_{so_j}^{(n)} = \psi * \widehat{T}_{so_j}^{(n)} + (1 - \psi) * NMI_{(so_j)} \tag{3.11}$$

Among them: ψ represents the weight of historical transaction records, taking values based on actual circumstances.

4. Result analysis. This section verifies the effectiveness and correctness of P3Trust through simulation experiments. As a reference, both the PeerTrust model and AARep model were implemented simultaneously [18].

4.1. Experimental Environment. The simulation experiments were conducted on a hardware setup consisting of an Intel Core2 Quad processor clocked at 2.83 GHz with 2.00 GB of RAM. The software environment utilized for the simulation experiments comprised Windows XP operating system and the Matlab 7.1 simulation platform. The experimental scenario is file download service, personalized service attribute vector: Download service=[file quality response time download speed], user interest vector: Interest vector=[0.5 0.2 0.3]. The evaluation index is the transaction success rate between user nodes and service nodes η. define it as the following equation 4.1.

$$\eta = N_{(T_{so} \geqslant \tau)} / \sum N \tag{4.1}$$

Among them: τ is the minimum acceptable comprehensive trust value; $N_{(T_{SO} \geqslant \tau)}$ represents a trust value that exceeds the minimum acceptable comprehensive trust value τ total number of times; $\sum N$ represents the total number of transactions.

In real situations, there are many types of service nodes, and this experiment sets the following two types of service nodes:

1. General node: This type of service node can provide high-quality services based on the corresponding services registered in the trust evaluation center.
2. Malicious nodes: The services provided by these service nodes do not match the service information registered in the trust evaluation center [19].

The experimental parameter settings are shown in Table 4.1.

This experiment sets the following three types of user node transaction situations:

1. 40% of transactions are related to family nodes;
2. 20% of transactions are first-time transactions, that is transactions with unfamiliar nodes;
3. 40% of transactions are ordinary node transactions.

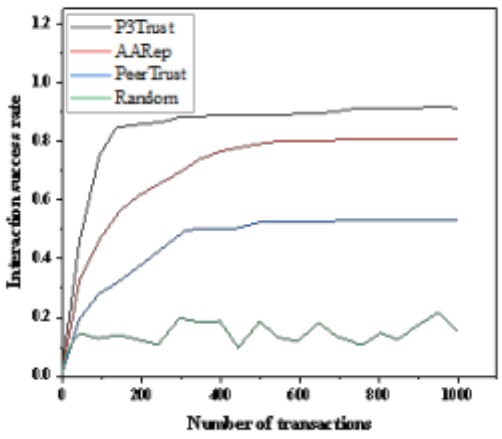


Fig. 4.1: Comparison of interaction success rates under 30% malicious nodes

During each simulation cycle, user nodes make personalized requests to the trust evaluation agent to download resources they have never owned. The trust evaluation agent first searches for TRT and determines the type of user node; Then, corresponding trust evaluation methods are used to evaluate the trust relationship between user nodes and service nodes, and suitable service nodes are selected for user nodes; Finally, after the transaction is completed, the user node evaluates the service and feeds it back to the trust evaluation agent. The trust evaluation agent updates the trust relationship based on QoS UM and saves it in the TRT.

4.2. Experimental Results.

4.2.1. Efficiency and accuracy of P3Trust trust assessment . This section examines the efficiency and accuracy of trust assessment in P3Trust, and compares it with the PeerTrust model, AARep model, and random selection model [20]. In scenarios where the proportion of malicious service nodes is 30% and 70%, the curves of the interaction success rate obtained through P3Trust model, PeerTrust model, AARep model, and Random model with respect to the number of transactions are shown in Figure 4.1 and Figure 4.2, respectively.

As shown in Figure 4.1, in the interaction scenario of 30% malicious service nodes, except for the Random model which is irregular, the other three models can all lead to an increase in interaction success rate. When trading around 600 times, the interaction success rate of the PeerTrust model can remain stable at around 0.5, the interaction success rate of the AARep model can remain stable at 0.7-0.8, and the P3Trust model can remain stable at around 0.9 after trading around 200 times. This is because P3Trust provides a reasonable trust value initialization method. In the harsh scenario of 70% malicious service nodes, the P3Trust model still has a higher interaction success rate than the other three models, as shown in Figure 2. Therefore, P3Trust can improve the accuracy of trust assessment while improving the efficiency of trust assessment [21].

4.2.2. Robustness of P^3 Trust. This section examines the ability of P3Trust to withstand harsh environments and analyzes its differences in resistance to harsh environments compared to the PeerTrust model, AARep model, and Random model. The curves of the success rate of interaction with the proportion of malicious nodes after 2000 transactions for P3Trust model, PeerTrust model, AARep model, and Random model are shown in Figure 4.3.

From Figure 4.3, it can be seen that as the proportion of malicious nodes increases, the interaction success rate of the other three trust models decreases continuously, except for the random model which is irregular. P^3 Trust can achieve an interaction success rate of around 0.8 in a good environment with a malicious node proportion of about 20%, and in a harsh environment with a malicious node proportion of about 72%, P^3 Trust maintains an average transaction success rate that is 12% and 18% higher than AARep and PeerTrust [22].

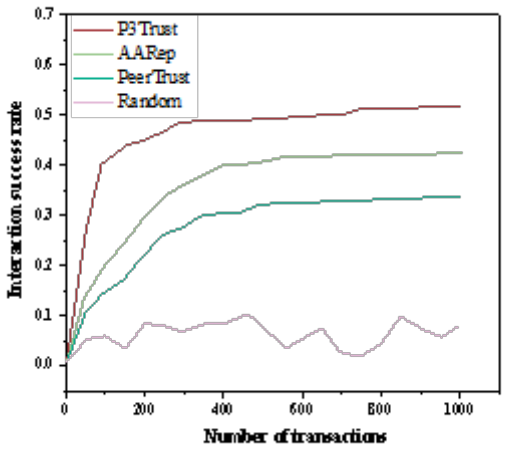


Fig. 4.2: Comparison of interaction success rates under 70% malicious nodes

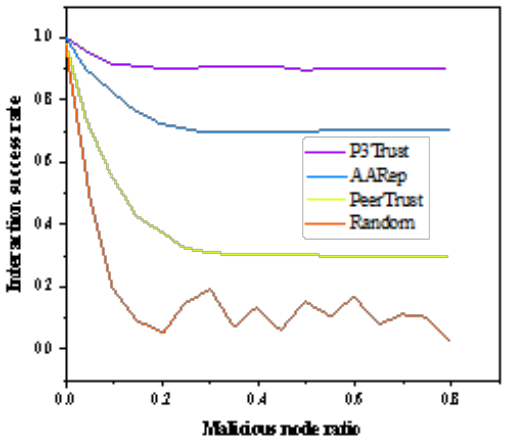


Fig. 4.3: Comparison of interaction success rates under different malicious nodes

Respectively, the reason is that the subject of trust evaluation in the P3Trust model is a fair and objective trust evaluation agent, and an efficient trust value evaluation method has been proposed. Therefore, the P3Trust model has strong resistance to harsh environments.

5. Conclusion. Through the discussion of existing trust models in cloud environments, the author proposes a trust model for personalized cloud services based on user types and privacy protection - P3Trust. The model first classifies user nodes based on historical transactions between nodes; Secondly, in order to improve the efficiency and accuracy of trust evaluation, a trust evaluation agent was introduced as the main body of trust evaluation to protect the privacy information feedback from nodes, and a trust value evaluation method based on user type was proposed; In addition, in order to reflect the dynamism of trust and improve the integrity of trust models, fully considering the transaction time and transaction amount of cloud services, the

author proposes a trust update mechanism based on service quality. The simulation results show that compared with existing trust models, P3Trust has significant improvements in the effectiveness and robustness of curbing malicious behavior, and improves the accuracy of trust evaluation.

REFERENCES

[1] Tricomi, G., Merlino, G., Panarello, A., & Puliafito, A. (2020). Optimal selection techniques for Cloud service providers. IEEE Access, 8, 203591-203618.

[2] Mohamed, A. M., & Abdelsalam, H. M. (2020). A multicriteria optimization model for cloud service provider selection in multicloud environments. Software: Practice and Experience, 50(6), 925-947.

[3] Asghari, P., Rahmani, A. M., & Javadi, H. H. S. (2022). Privacy-aware cloud service composition based on QoS optimization in Internet of Things. Journal of Ambient Intelligence and Humanized Computing, 13(11), 5295-5320.

[4] Rahimi, M., Jafari Navimipour, N., Hosseinzadeh, M., Moattar, M. H., & Darwesh, A. (2022). Toward the efficient service selection approaches in cloud computing. Kybernetes, 51(4), 1388-1412.

[5] Gupta, I., Gupta, R., Singh, A. K., & Buyya, R. (2020). MLPAM: A machine learning and probabilistic analysis based model for preserving security and privacy in cloud environment. IEEE Systems Journal, 15(3), 4248-4259.

[6] Kumar, R. R., Kumari, B., & Kumar, C. (2021). CCS-OSSR: a framework based on hybrid MCDM for optimal service selection and ranking of cloud computing services. Cluster Computing, 24(2), 867-883.

[7] Reddy, M. I., Rao, P. V., Kumar, T. S., & K, S. R. (2024). Encryption with access policy and cloud data selection for secure and energy-efficient cloud computing. Multimedia Tools and Applications, 83(6), 15649-15675.

[8] Ermakova, T., Fabian, B., Kornacka, M., Thiebes, S., & Sunyaev, A. (2020). Security and privacy requirements for cloud computing in healthcare: Elicitation and prioritization from a patient perspective. ACM Transactions on Management Information Systems (TMIS), 11(2), 1-29.

[9] Alkhasawneh, A. , & Khasawneh, F. . (2023). Legal issues of consumer privacy protection in the cloud computing environment: analytic study in gdpr, and usa legislations. Int. J. Cloud Comput., 12, 40-62.

[10] Haipeng, S. , Yu'An, T. , Congwu, L. I. , Lei, L. , Qikun, Z. , & Jingjing, H. U. . (2022). An edge-cloud collaborative cross-domain identity-based authentication protocol with privacy protection. Chinese Journal of Electronics, 2021, 1-11.

[11] Wang, T. , Xu, L. , Zhang, M. , Zhang, H. , & Zhang, G. . (2022). A new privacy protection approach based on k-anonymity for location-based cloud services. Journal of Circuits, Systems and Computers, 31(05), 2343-2358.

[12] YiDING, WeiSHEN, Hai-shengLI, Qiong-huiZHONG, Ming-yuTIAN, & JieLI. (2022). Blockchain trusted privacy service computing model for cnn. Acta Electronica Sinica, 50(06), 1399-1409.

[13] Nezafat Tabalvandani, M. A., Hosseini Shirvani, M., & Motameni, H. (2024). Reliability-aware web service composition with cost minimization perspective: a multi-objective particle swarm optimization model in multi-cloud scenarios. Soft Computing, 28(6), 5173-5196.

[14] Wu, Q., He, K., & Chen, X. (2020). Personalized federated learning for intelligent IoT applications: A cloud-edge based framework. IEEE Open Journal of the Computer Society, 1, 35-44.

[15] Wu, Q., Chen, X., Zhou, Z., & Zhang, J. (2020). Fedhome: Cloud-edge based personalized federated learning for in-home health monitoring. IEEE Transactions on Mobile Computing, 21(8), 2818-2832.

[16] Kashevnik, A., Lashkov, I., Ponomarev, A., Teslya, N., & Gurtov, A. (2020). Cloud-based driver monitoring system using a smartphone. IEEE Sensors Journal, 20(12), 6701-6715.

[17] Gill, S. H., Razzaq, M. A., Ahmad, M., Almansour, F. M., Haq, I. U., Jhanjhi, N. Z., ... & Masud, M. (2022). Security and privacy aspects of cloud computing: a smart campus case study. Intelligent Automation & Soft Computing, 31(1), 117-128.

[18] Gill, S. H., Razzaq, M. A., Ahmad, M., Almansour, F. M., Haq, I. U., Jhanjhi, N. Z., ... & Masud, M. (2022). Security and privacy aspects of cloud computing: a smart campus case study. Intelligent Automation & Soft Computing, 31(1), 117-128.

[19] Murthy, C. V. B., Shri, M. L., Kadry, S., & Lim, S. (2020). Blockchain based cloud computing: Architecture and research challenges. IEEE access, 8, 205190-205205.

[20] Wang, J., Yan, Z., Wang, H., Li, T., & Pedrycz, W. (2022). A survey on trust models in heterogeneous networks. IEEE Communications Surveys & Tutorials, 24(4), 2127-2162.

[21] Papenmeier, A., Kern, D., Englebienne, G., & Seifert, C. (2022). It's complicated: The relationship between user trust, model accuracy and explanations in ai. ACM Transactions on Computer-Human Interaction (TOCHI), 29(4), 1-33.

[22] Ramu, N., Pandi, V., Lazarus, J. D., & Radhakrishnan, S. (2020). A novel trust model for secure group communication in distributed computing. Journal of Organizational and End User Computing (JOEUC), 32(3), 1-14.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: Nay 10, 2024
Accepted: Jun 25, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 651–662, DOI 10.12694/scpe.v26i2.3929

ANALYSING THE CLASSIFICATION OF ARTISTIC STYLES OF PAINTING IN ART TEACHING FROM THE PERSPECTIVE OF EMOTIONAL SEMANTICS

YAN SONG *

Abstract. With the advancement of Internet information technology, a great number of art paintings have appeared on key small network platforms. These art paintings include not only a huge quantity of representational information, but also a large amount of semantic information; yet, there is currently a dearth of more systematic research on sentiment semantic analysis in painting art works. To provide basic support for study into the sentiment semantics of art paintings, we present a machine learning-based classification algorithm for painting art in art education from the standpoint of sentiment semantic analysis. To begin, we build machine learning models of different painting styles. Once machine learning is realized, we convert the color space into Lab color space and use the weighting function and the color values of the a and b channels to obtain the image's color entropy; to obtain the chunking entropy, we use the art image's chunking machine learning and the mean of the chunking machine learning; and to obtain the contour Entropy, we use the Contourlet transform to extract the image's contour information. With significant novelty and practical application value, this study presents a new direction for the study and practice of emotional level adaptive interaction in intelligent learning environments.

Key words: Sentiment semantic analysis, Machine learning, Deep learning, Painting classification

1. Introduction. Traditional culture is the spiritual source of a nation, and painting is one of the forms of expression of traditional cultural heritage[1]. Mankind's knowledge and understanding of the world, as well as his perception and emotion of real life, are expressed through the form of painting. From ancient times to the present, human civilization has progressed continuously and accumulated a huge amount of painting resources, such as Eastern landscape paintings, Western oil paintings, Dunhuang frescoes [2], and so on. These painting images are not only an effective carrier for passing down human civilization, but also an important force for human development and progress in the course of history[3]. In the current era of rapid social growth, smart mobile gadgets have been increasingly integrated into people's daily lives, and at the same time, digital painting images are gradually integrated into people's lives, and can be enjoyed anytime and anywhere through the Internet and electronic devices, which also makes it more convenient for researchers to explore them at a deeper level [4].

While research on computer cognitive science and artificial intelligence has reached a significant degree, it is still early in the development of machine cognitive aspects of emotion perception and expression [5]. Emotion is an essential element of human functioning and plays a crucial part in many human activities, including cognition, logic, planning, creation, and communication, as demonstrated by research in neurophysiology, brain science, and other fields [6]. The American study [7] is the first to propose allowing a computer to have emotions. The author made the argument that an intelligent machine can only be one that has emotions. According to study [8], since then, emotional computing research has drawn interest from nations all over the world in the field of information science. Affective computing aims to enable computers to perceive, comprehend, and communicate a wide range of emotional traits, among other things, with the ultimate goal of enabling computers to freely express and interact with one another just like people. The method of affective computing states that the gathering, analyzing, and modeling of emotional data, as well as the interpretation of emotional signals, are the principal areas of research [9]. The primary area of study in emotion computing is image emotion semantic recognition, and emotion computing will play a major role in experiments and research on artificial intelligence in the future [10].

The amount of image resources is currently increasing quickly, especially in the constantly updating Internet

*Zhengzhou Railway Vocational & Technical College,Zhengzhou,Henan, 450000, China (10497@zzrvtc.edu.cn).

Fig. 1.1: Example of style painting.

network, necessitating the efficient organization, administration, and application of these enormous image information resources. The majority of semantic-based research on the emotional semantics of images uses images from a variety of sources, such as landscape photographs, paintings, and life photographs in a variety of perplexing and complex categories. There is still a dearth of dedicated image resources in the field of image semantic research. The establishment of a comprehensive and applicable image material library for the analysis of the emotional semantics of images is a very necessary basic work. While the difficulty of experimental research may be lessened by the arbitrary selection of image resources, the accuracy of the experimental results for the study of the emotional semantics of images will be diminished.

Further systematic emotion semantic research on painting art works is now lacking. The thesis develops a semantic annotation system of painting emotion from the perspective of user cognition and based on the hierarchical semantic model of paintings, and offers an easy-to-use research platform for more scholars in the study of image emotion. The goals of this system are to establish a semantic dataset of Yunnan art paintings that has been annotated with standardised descriptions, and to provide basic support for the research on emotion semantics of art paintings.

2. Related work. Japan from the 1990s conducted research on perceptual engineering (kansei engineering), which is the combination of human emotions and engineering to design and manufacture goods [11]. In terms of market applications, the German multi-model shopping assistant EMBASSI, which mainly takes the psychology of shoppers and the demand for the goods environment in shopping malls as its research goal, establishes a convenient Internet e-commerce system [12], the multi-functional perception machine successfully developed and applied by the Harbin Institute of Technology [13], the research on robots with emotions based on artificial intelligence proposed by Tsinghua University, which is capable of controlling its own architecture [14], the research conducted by Beijing Jiaotong University to integrate multifunctional perceptual machines and

emotional computing with each other [15], the research conducted by the University of Science and Technology of China on interactive perceptual image retrieval technology based on image content [16], and the research conducted by Zhejiang University to create a virtual character and emotional system in E-Teatrix [17].

In contrast, there is still a lot of room for growth and little research being done in the subject of picture emotional semantics, both domestically and internationally. Low-level visual characteristics like color, texture, and edge outlines can be directly used to extract emotional semantic information from an image, which will then naturally reflect the image's rich emotional content. As a result, low-level visual features and machine learning techniques (e.g., SVM, Adaboost, neural networks, fuzzy clustering, etc.) are used in the great majority of current image-based emotion semantic studies for emotion annotation and recognition as well as emotion-based image categorization and retrieval. A fuzzy similarity based emotion classification method for color photos was proposed by study [18], which used color as a key characteristic. Study [19] used an adaptive fuzzy system and neural network based approach to study and compare two emotion models based on color templates. An emotion prediction system developed by Study [20] uses visual characteristics to automatically identify particular emotions in textile imagery. According to the link between color and emotion, study [21] suggests an emotion retrieval model based on the semantic description of visual colors.

More methodical studies on the semantics of emotion for pictorial artworks are still lacking. Most of the work that has already been done only focuses on a small number of specific issues related to emotion computation (such as semantic-based painting retrieval, painting art style classification, etc.), or compares and identifies a particular painter or painters, a particular type of painting style, etc. A multi-resolution Hidden Markov (MHMM) technique was presented in study [22] to classify traditional Chinese black-and-white ink drawings. An approach based on SVM classifiers and low-level characteristics was reported in study [23] to divide traditional Chinese paintings into two style categories: writing and brushwork. Research [24] examined the brushstroke features of Van Gogh's paintings in an effort to pinpoint his original creations. Study [25] used an RBF neural network to automatically classify historical western artworks according to both local and global visual characteristics. Study [26]: Using clustering and vector quantization (VQ) based on MPEG-7 descriptors, high-level visual features are extracted to classify and retrieve semantic information from artworks of 18 Western European artists from 6 categories representing various stylistic periods. In order to conduct scientific research on the scientific understanding of visual art, the study [27] uses the Curvelet transform, information theory technology, and Sparse Code coding on the paintings of six painters: Xu Beihong, a representative of Chinese ink painting, and Van Gogh, a representative of Western painting. These techniques are used to extract the digital features of each school of painting and to summarize the statistical characteristics and the correspondence between painting styles. In a comparative study on the influence that artists have on one another, Research [28] first created a two-level painting classification model and then used knowledge discovery techniques to analyze the influence that artists have on one another. These works do not address sentiment analysis, even if they target paintings and concentrate on the connection between low-level characteristics and high-level semantics.

2.1. Emotion in color. It is widely recognized that color has the power to awaken emotions, and the symbolic nature of color can trigger associations, so that color is associated with certain emotions and feelings, and it can be said that color has emotionality. Although the emotions evoked by color can vary somewhat because of different cultural backgrounds, personal experiences and psychological factors, there are still many common feelings in the psychology of color under factors such as human physiological structure and the physical characteristics of color. In art, architecture, design and painting, the use of color is used to make the picture "pleasant, powerful, melancholic" and other emotions, which is a common daily experience [12].

In the study of image sentiment, [13]and others in Japan used a color vector composed of the average of the RGB components of the L row and column colors to establish a mapping between them to sentiment words, while [14] divided the image into 32×32 sub-blocks, calculated the average color intensity values of the image sub-blocks, and they composed a vector as the image color features. In these studies, basically, a simple extraction approach is taken for the color features and not much consideration is given to the connection between features and sentiment.

The approach of [15] is pioneering, and based on Itten's theory about the semantics expressed by colors and lines in art paintings, he integrates the lower-level image features with several rules, logical and relational operations into features with certain semantic descriptive ability, i.e., higher-level image expressive features,

and then uses the expressive features for the derivation of image emotional semantics, so that the extracted features have a stronger ability to express emotion capability.

In general, color features are rarely used in current image retrieval systems to describe from the perspective of emotion, and emotion-based image color feature extraction must fully consider human feelings and psychology to construct appropriate color features, which is a further research trend for emotion-based image feature extraction.

2.2. Emotion in shapes and contours. Shapes also have their own aesthetic expression value, and certain shape features in images can stimulate people to produce perceptual awareness. For example, lines have rich expressive forms; vertical lines are clear and solemn, symbolizing dignity and eternity; horizontal lines are wide and still, indicating silence and stability; diagonal lines are vivid and energetic, with a sense of movement; circles and curves are soft, elegant, and rhythmic, in various shapes and forms.

Different shapes also convey different visual effects, and people often assign different shapes to different thoughts and emotions. For example, geometric shapes have a simple, simple, clear mechanical and indifferent sense; shape gives stability and solemnity; S shape gives change and liveliness; C shape causes centripetal flow, shape gives proper concentration; O shape gives a sense of roundness and relaxation; V shape produces instability; shape gives seriousness and silence; organic shape has a sense of liveliness and "human feelings and ignorance ".

Commonly used shape and contour features are boundary-based and region-based methods, and methods with better description are Fourier descriptors and invariant moment methods, which are invariant to translation, rotation, and expansion of the shape . In recent years there are some new methods such as finite element matching methods and wavelet transform description methods.

Emotion-based shape feature extraction should be relatively simple, its purpose is different from general shape feature extraction, and it does not require to identify the object type from the shape, so the requirements for shape are not strict, only need to judge the general line shape (straight line, curve, fold) or the basic area shape belongs to geometric shape (shape or V shape, etc.) or organic shape (O shape, S shape or natural shape, etc.). However, there are not many studies linking shape features with image emotion, and it is worthy of further in-depth exploration.

2.3. Emotion in texture. All images have surface texture, which is also called texture in design, and the material is different, the structure of the surface is different, and it gives different feelings. Smooth to give a sense of delicacy, soft to give a sense of warmth, rough to give a sense of old, hard to give a sense of strength, can produce different visual psychological effects, and people's emotions are closely linked.

Texture is an important and difficult to describe feature of an image, and there is still no accepted precise definition. An early typical representation of texture features is the symbiotic matrix approach based on traditional mathematical models. The so-called grayscale co-occurrence matrix $M(\triangleleft x, \triangleleft y)$ is mathematically represented as the joint frequency distribution of the simultaneous occurrence of two grayscale pairs in the image at positions $(\triangleleft x, \triangleleft y)$ apart, so that it can reflect the spatial dependence of the grayscale level texture. The drawback of the grayscale co-occurrence matrix method is that some of the texture properties it obtains (e.g., entropy) would have no corresponding visual content.

3. Methods. The goal of the picture feature extraction process, as seen in Fig. 3.1, is to recover objective underlying features like colour, texture, shape, etc. from the image, and the features are extracted mainly using computer vision and digital image processing techniques to obtain objective visual content features directly from the image, and the algorithms they correspond to are becoming increasingly mature. In the process of studying the emotional semantics of images, the feature extraction part of images is very heavy, and it is necessary to construct or select psychological and physiological models in combination with human physiological and psychological characteristics to find features, and to describe them using appropriate ways in order to play a positive role in the extraction of emotional semantics.

First, the edges were filtered and closed by morphological manipulation. Finally, the top six emotional semantic sub-images of the formed painting images in terms of density are used as the input features needed for classification. The flowchart of the emotion-semantic feature extraction method proposed in this chapter is shown in Fig. 3.2.

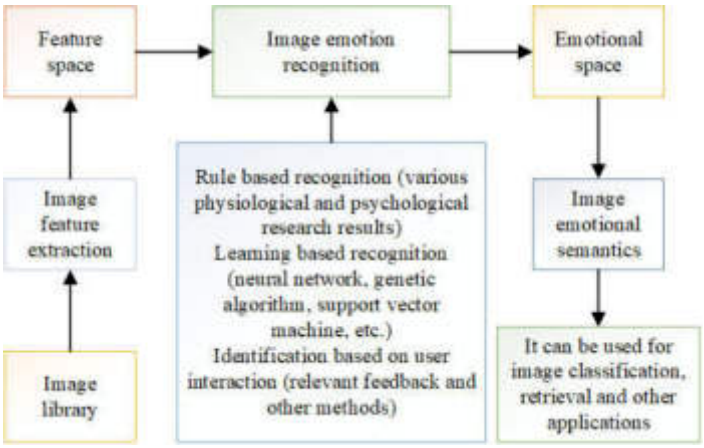


Fig. 3.1: Basic framework of emotion semantic extraction.

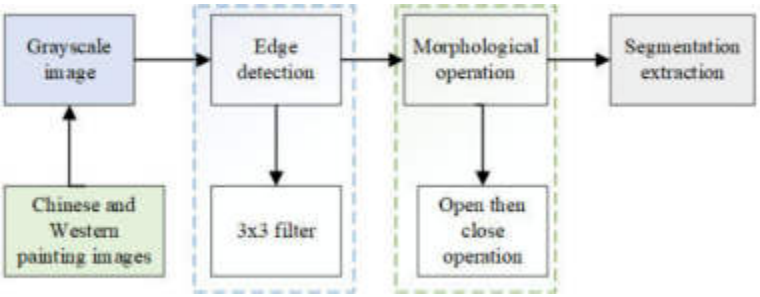


Fig. 3.2: Flowchart of extracting sentiment semantic features.

3.1. Methods of extracting sentiment semantic features in this paper.

3.1.1. Sobel operator edge detection method. Edge detection is one of the common methods for processing image edges. It is independent of the image content and acts mainly on the edges. Thus, in our model, we choose the edge as the first place if it contains the number of pixels between two thresholds (200 and 900).

Among them, the Sobel operator is an algorithm for edge detection based on the first-order gradient operation, which uses the convolution kernels G_X and G_Y to convolve in the X and Y directions of the image, respectively, in the process of extracting sentiment semantic features. The results computed from the convolution in X and Y directions are algebraically weighted and summed to obtain the edge results for the whole image. The convolution kernel G_X, G_Y is shown in Eq. 3.1:

$$G_X = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad G_Y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \tag{3.1}$$

The Sobel operator matrix has the following form:

$$G_x = f_x(x,y) = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \mathrm{G}_y = f_y(x,y) = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{3.2}$$

The edge operation method, which is essentially a gradient change in the grey value of a pixel point, is carried out by the drawing image, and the edge location of a certain region is where the gradient change is

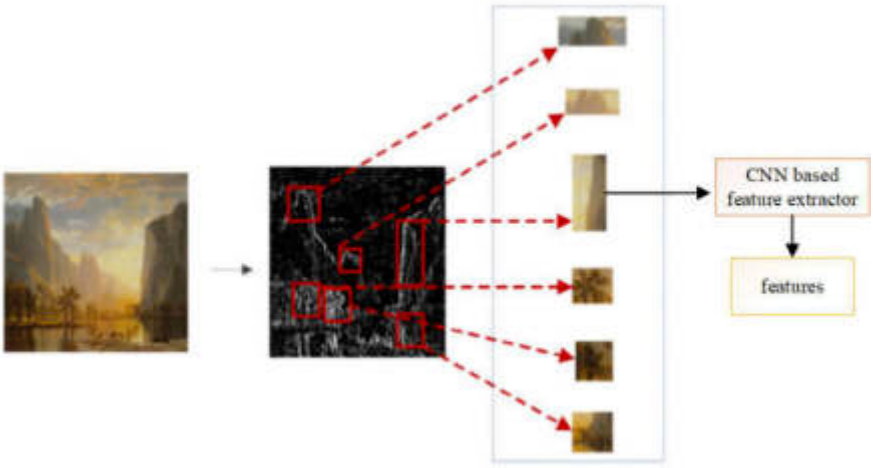


Fig. 3.3: Selection of emotional semantic sub-images.

located. The combined grey value represents the portion of the grey value that has changed in this instance. The following formula determines the point's grey value size.

$$G = \sqrt{G_x^2 + G_y^2} \tag{3.3}$$

In the process of extracting the emotional semantic features, the painting images are firstly converted into grayscale images, and the edge detection Sobel operator is used to perform the operation through a 3*3 filter.

3.1.2. Morphological operation. We have to select representative sentiment semantic features as the input for classification, because too sparse sentiment semantics will interfere with the classification results. The top six emotional semantic features with density extracted according to the above algorithm are used as the most representative emotional semantics of this painting image, and we call it selectable emotional semantics. As shown in Fig. 3.3, the process of extracting selectable sentiment semantics is shown.

Fig. 3.4 shows the block diagram of the system implementation.

3.2. Classification model based on sentiment semantic features. With the expansion of data set, CNN has been widely used in various fields of image processing, such as target recognition, classification and tracking. Therefore, based on the above mapping data set, we establish a convolution neural network classification model based on emotional semantic features.

3.2.1. CNN. This study uses morphological manipulation, segmentation, and extraction of the image after edge detection, and then uses the 64x64 image size as the CNN's input for learning and training. Fig. 3.5 depicts the architecture, and the structure is as follows:

1. Create six 60 × 60 mappings in the C1 layer in order to extract certain edge features from the original image pixels.
2. Each mapping in the S1 layer is lowered using a subsampling rate of 2, which aids in the extraction of important edge characteristics while lowering the model's parameters.
3. The second convolution layer in the C2 layer creates 12 mappings with a size of 26 x 26 in order to identify basic shape features from the edge features.
4. By creating fully connected layers, 2028 dimensional vectors are created in one layer of the S2 layer. In the output layer, 1014-dimensional features are finally obtained.

3.2.2. SVM. The essential idea of SVM is to transform the nonlinear problem in classification into a linear problem and finally get the optimal solution. SVMs maximize the distance between two types of images in a dataset and the hyperplane. Its diagram is shown in Fig. 3.6.

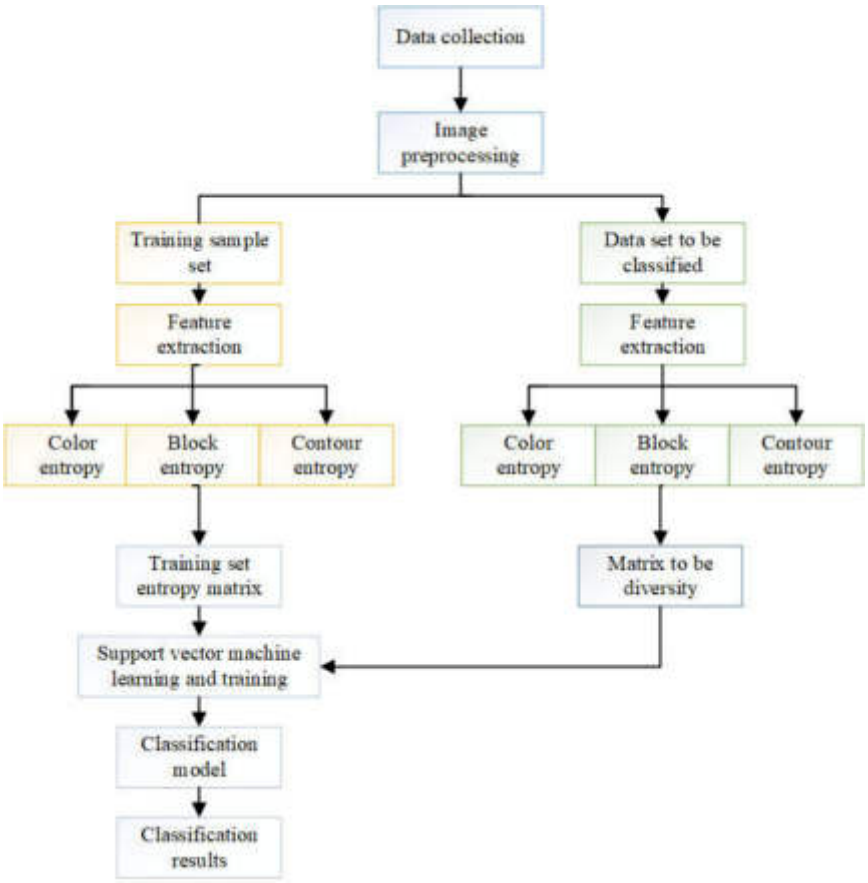


Fig. 3.4: Diagram of system implementation.

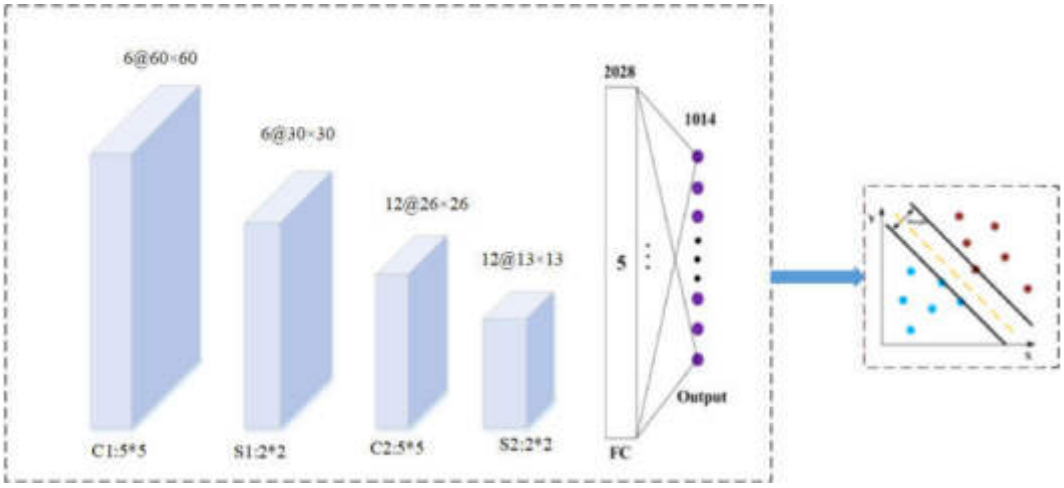


Fig. 3.5: CNN model.

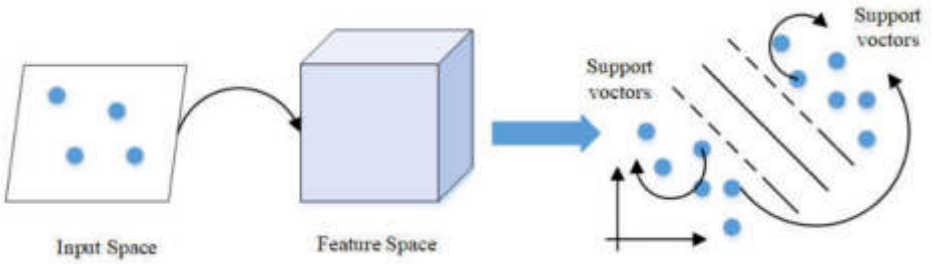


Fig. 3.6: SVM model.

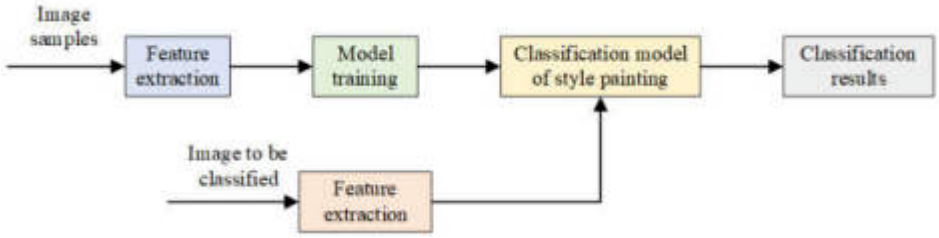


Fig. 4.1: Flow chart of the experiment.

For the test painting images, it is assumed that their affective semantic features are described as x, which is classified as a positive sample if $f(x) = 1$ and a negative sample if the opposite is true. In the algorithm of SVM for classification, we used generic values as parameters. The final decision of whether the test image belongs to good painting image or bad painting image is made by SVM.

4. Experiments. The experimental flow is shown in Fig. 4.1.

Step 1: To extract colour, block, and contour entropies, respectively, 100 paintings are randomly exhibited as simulation samples from seven genres, such as painting and sketching.

Step 2: The entropy is combined with [mxn] (m to show the training samples, and N to provide the sample attributes), and the associated tag set is established.

Step 3: Utilising the characteristic matrix as the training kit, choose the radial basis function kernel function, train the training kit repeatedly using libsvm, and obtain pattern classification for art instruction.

Step 4: Seven image styles such as patterns and patterns are selected as test samples, and the recognizable vectors of the images are obtained according to steps 1 and 2. The recognizable vectors are input into step 3 after the classifier is well trained for recognition, and the accuracy and results of the recognition are obtained.

Step 5: To retrieve the sorting results, set any label on the image to be recognised and enter it into the style to draw the classifier.

4.1. Accuracy and completeness. All experiments in this paper were conducted in Windows 7, 32-bit operating system, based on 2.3 GHz AMD Turion II CPU, 4 G memory, and programmed in Matlab2016. In this paper, experiments are conducted for seven different painting arts in art teaching: cartooning, drawing, oil painting, watercolor painting, branding, ink painting, and mural painting, and the proposed method is mainly validated by accuracy and completeness. One hundred paintings of each style were chosen at random from the remaining dataset to create the test set, while the other hundred paintings of each style comprised the experimental training set. Fig. 4.2 displays the outcomes.

Table 4.1 shows the accuracy of the classification. The findings of the experiment indicate that the categorization accuracy of oil, mural, and watercolor paintings is quite high, whereas that of cartoon and ink paintings

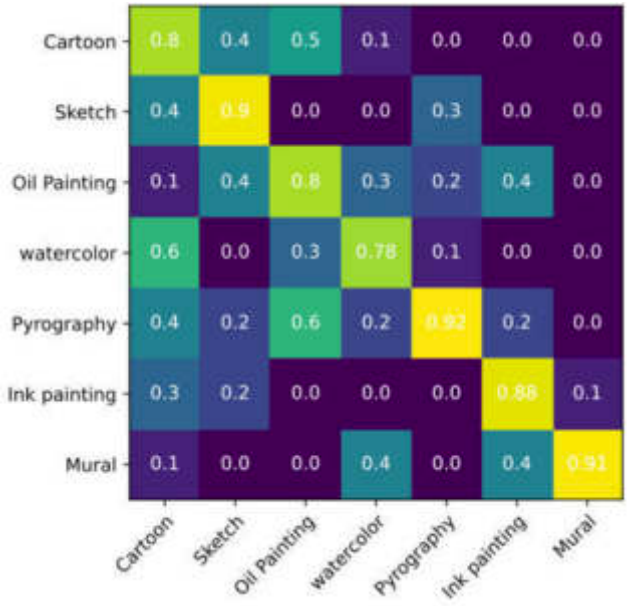


Fig. 4.2: The accuracy of different paintings.

Table 4.1: Rate of classification accuracy.

Painting animal species	Quantity of results for classification	Quantity of accurate categorizations	Accurate classification (%)
Cartoon	48	40	83.32
Sketch	50	44	88.00
Oil Painting	50	48	92.14
Painting in water colors	48	45	93.63
Pyrography	51	45	88.47
Chinese brush painting	55	44	79.62
Mural	46	46	95.75

is relatively poor. 309 of the 350 test photos that were selected had an average classification accuracy of 88.28%, meaning that they were accurately classified. Because cartoons and oil paintings have a particular style that is simpler to recognize than other types, the recognition accuracy is good. It is very difficult to identify and classify style paintings.

Table 4.2 shows the completion rates. Because branding and oil painting have such high recognition accuracy, their completion rates are comparatively high, often exceeding 90%, and they are not easily misidentified as other paintings, so the completion rates are high. The search rate of cartoons is relatively low, only 80%, because the recognition accuracy of cartoons is low, and they are easily classified as other types of paintings, so the search rate is low. The categorization accuracy has an impact on the search completion rate, and increasing the accuracy rate helps to increase the search completion rate.

4.2. Comparative analysis. The system finally achieves the classification of 5 types of paintings with different art styles by setting multiple classes of SVM classifiers and classifying them layer by layer, and each class of SVM classifiers adopts the specified image features to achieve their respective classification functions (see Fig. 4.3).

Fig. 4.4 shows the multiclass binomial tree classification system in the literature[7], which uses a 3-layer

Table 4.2: Classification rate.

Painting animal species	Quantity of results for classification	Quantity of accurate categorizations	Accurate classification (%)
Cartoon	40	10	82
Sketch	45	5	86
Oil Painting	46	4	95
Painting in water colors	44	5	87
Pyrography	45	5	91
Chinese brush painting	44	8	88
Mural	45	4	92

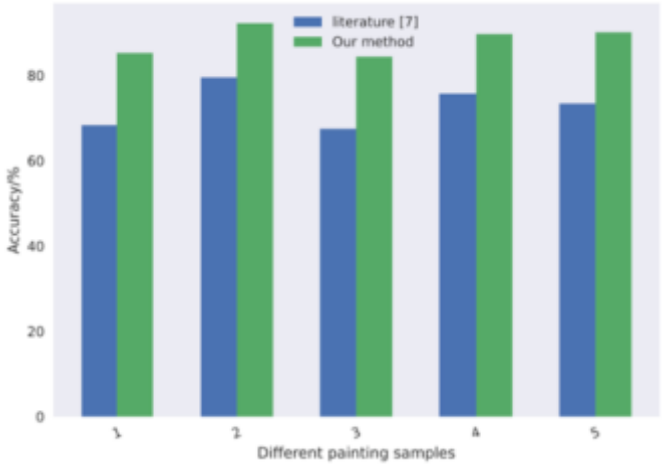


Fig. 4.3: The accuracy of painting samples.

4-class SVM classifier to progressively achieve classification of 5 classes of art style paintings, and each classifier uses different features to achieve classification of the specified style paintings from top to bottom layer by layer, and its overall classification accuracy is 85.56%. The classification accuracy of the algorithm in this paper for art painting styles is 88.28%, and the accuracy is improved. In addition, both the literature[7] and this paper use SVM classifier for classification, including 2 processes of training and classification. The literature [12] uses a hierarchical approach to perform at least three time-consuming SVM calculations on samples. The SVM is then used to train and classify the data set once, which speeds up the computational process when there are a lot of samples, improving the efficiency of the classification calculation.

We examined the teachers' level of satisfaction with the system after verifying the algorithm's accuracy. The particular outcomes are displayed in Fig. 4.5. As can be shown, the great majority of educators support the use of the algorithm, with 44% of educators between the ages of 35 and 39 being the largest group.

5. Conclusion. In this work, we suggest a painting-related machine learning-based categorization method for art education. The algorithm takes seven kinds of representative painting art in art teaching: cartoons, drawings, oil paintings, watercolor paintings, branding paintings, ink paintings and murals as objects, and uses neural networks and machine learning algorithms to train, test and classify different painting art in art teaching. The establishment of image semantic emotion model and accurate image semantic description standard model is a very challenging topic, in which the mapping relationship between the layers of image semantics, the normalisation of image emotion semantic feature description and the integration of theories from other related disciplines are important research directions, and the next research work includes the following:

1. Wearable devices based on psychological experiments can be added to the construction of the emotional

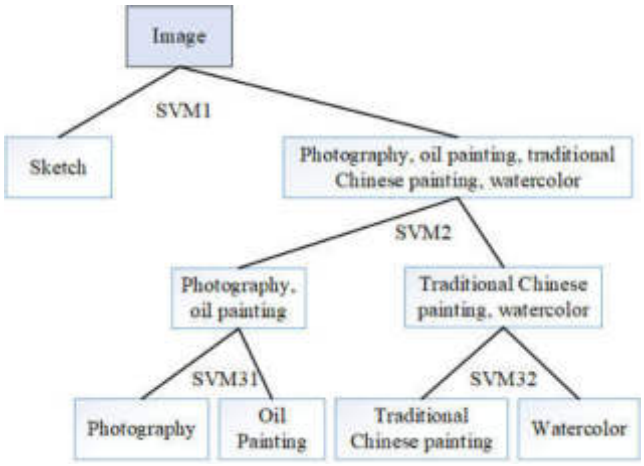


Fig. 4.4: SVM multiclass binomial tree classification for different art style classification.

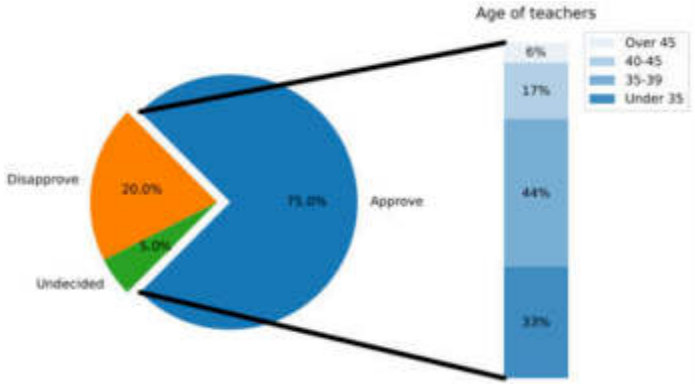


Fig. 4.5: Teachers' satisfaction with the method.

space model, such as human heart rate, visual attention area and brain waves and other physiological data.

2. Based on the visual human-computer interaction annotation to obtain the emotional semantics of the paintings, and find a better solution to solve the problem of semantic gap between the bottom visual features of the image to the higher level emotional semantics.

Data Availability. The experimental data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments. The author thanks the anonymous authors whose work largely constitutes this sample file.

REFERENCES

[1] Chen, C. L., Huang, Q. Y., Zhou, M., Huang, D. C., Liu, L. C., & Deng, Y. Y. *Quantified emotion analysis based on design principles of color feature recognition in pictures.* Multimedia Tools and Applications,(2023) 1-25.

[2] Kong, F. *Application of artificial intelligence in modern art teaching.* International Journal of Emerging Technologies in Learning (iJET), 15(13),(2020) 238-251.

[3] Zhong, S. H., Huang, X., & Xiao, Z. *Fine-art painting classification via two-channel dual path networks.* International Journal of Machine Learning and Cybernetics, 11(1),(2020)137-152.

[4] Di Wu, Yin Lei, Maoen He, Chunjiong Zhang, Li Ji, *"Deep Reinforcement Learning-Based Path Control and Optimization for Unmanned Ships", Wireless Communications and Mobile Computing, vol.* 2022, Article ID 7135043, 8 pages, 2022. https://doi.org/10.1155/2022/7135043.

[5] Santos, I., Castro, L., Rodriguez-Fernandez, N., Torrente-Patino, A., & Carballal, A. *Artificial neural networks and deep learning in the visual arts: A review.* Neural Computing and Applications, 33(1), (2021)121-157.

[6] Palanisamy, S.; Thangaraju, B.; Khalaf, O.I.; Alotaibi, Y.; Alghamdi, S.; Alassery, F. *A Novel Approach of Design and Analysis of a Hexagonal Fractal Antenna Array (HFAA) for Next-Generation Wireless Communication.* Energies 2021, 14, 6204. https://doi.org/10.3390/en14196204.

[7] Alsubari, S. N., Deshmukh, S. N., Alqarni, A. A., Alsharif, N., H., T. *Data Analytics for the Identification of Fake Reviews Using Supervised Learning.* CMC-Computers, Materials & Continua, 70(2),(2022) 3189–3204.

[8] Al-Mekhlafi, Fahd A., Reem A. Alajmi, Zainab Almusawi, Fahd Mohammed Abd Al GAlil, Pawandeep Kaur, Muhammad Al-Wadaan, and Mohammed S. Al-Khalifa. *"A study of insect succession of forensic importance: Dipteran flies (diptera) in two different habitats of small rodents in Riyadh City, Saudi Arabia."* Journal of King Saud University-Science 32, no. 7 (2020): 3111-3118.

[9] Ganczarek, J., Pietras, K., Stolińska, A., & Szubielska, M. *s and Semantic Violations Affect Eye Movements When Viewing Contemporary Paintings.* Frontiers in Human Neuroscience,(2022) 16, 808330.

[10] Zhang, J., Duan, Y., & Gu, X. *Research on emotion analysis of Chinese literati painting images based on deep learning.* Frontiers in Psychology, (2021)12, 723325.

[11] Beermann, U., Hosoya, G., Schindler, I., Scherer, K. R., Eid, M., Wagner, V., & Menninghaus, W. *Dimensions and clusters of aesthetic emotions: A semantic profile analysis.* Frontiers in psychology, (2021)12, 667173.

[12] Jingchun Zhou, Qian Liu, Qiuping Jiang, Wenqi Ren, Kin-Man Lam, Weishi Zhang*. *Underwater image restoration via adaptive dark pixel prior and color correction.* International Journal of Computer Vision, 2023. DOI :10.1007/s11263-023-01853-3.

[13] Guo, H., Liang, X., & Yu, Y. *Application of big data technology and visual neural network in emotional expression analysis of oil painting theme creation in public environment.* Journal of environmental and public health, 2022(1), 7364473.

[14] Xiu, D., He, L., Killikelly, C., & Maercker, A. *Prolonged grief disorder and positive affect improved by Chinese brush painting group in bereaved parents: A pilot study.* Journal of social work in end-of-life & palliative care, 16(2), 116-132.

[15] Jiang, W., Wang, X., Ren, J., Li, S., Sun, M., Wang, Z., & Jin, J. S. *MTFFNet: a Multi-task Feature Fusion Framework for Chinese Painting Classification.* Cognitive Computation, 13(5),(2021) 1287-1296.

[16] Tian, W. *Emotional information transmission of color in image oil painting.* Journal of Intelligent Systems,(2022) 31(1), 428-439.

[17] Li, J., Chen, D., Yu, N., Zhao, Z., & Lv, Z. *Emotion recognition of Chinese paintings at the thirteenth national exhibition of fines arts in China based on advanced affective computing.* Frontiers in Psychology,(2021) 12, 741665.

[18] Bian, J., & Shen, X. *Sentiment analysis of Chinese paintings based on lightweight convolutional neural network.* Wireless Communications and Mobile Computing, 2021(1), 6097295.

[19] Zhu, Y., Zhu, Y., Ge, N., Gao, W., & Zhang, W.*Visual Emotion Analysis via Affective Semantic Concept Discovery.* Scientific Programming, 2022(1), 6975490.

[20] Duan, Y., Zhang, J., & Gu, X. *A novel paradigm to design personalized derived images of art paintings using an intelligent emotional analysis model.* Frontiers in psychology, 12, 713545.

[21] Wang, D. *Research on the art value and application of art creation based on the emotion analysis of art.* Wireless Communications and Mobile Computing, 2022(1), 2435361.

[22] Feng, Y., Chen, J., Huang, K., Wong, J. K., Ye, H., Zhang, W., ... & Chen, W. *iPoet: interactive painting poetry creation with visual multimodal analysis.* Journal of Visualization,(2022) 1-15.

[23] Zhao, W., Zhou, D., Qiu, X., & Jiang, W.*How to represent paintings: A painting classification using artistic comments.* Sensors, (2021)21(6), 1940.

[24] Klettner, S. *Affective communication of map symbols: A semantic differential analysis.* ISPRS international journal of geo-information,(2020) 9(5), 289.

[25] Urquhart, L., & Wodehouse, A. *The emotive and semantic content of pattern: an introductory analysis.* The Design Journal,(2021) 24(1), 115-135.

[26] Mullennix, J. W., Hedzik, A., Wolfe, A., Amann, L., Breshears, B., & Ticjak, N. *Affective context and the interpretation of facial expressions in portrait paintings.* Empirical Studies of the Arts,(2022) 40(2), 174-191.

[27] Yan, J., Wang, W., & Yu, C. *Affective word embedding in affective explanation generation for fine art paintings.* Pattern Recognition Letters, (2022)161, 24-29.

[28] Démuth, A., Démuthová, S., & Keçeli, Y. *A Semantic analysis of the concept of Beauty (Güzellik) in Turkish language: Mapping the semantic domains.* Frontiers in Communication, (2022)7, 797316.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 14, 2024
Accepted: Jun 26, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 663–672, DOI 10.12694/scpe.v26i2.3957

QUALITY ANALYSIS AND PREDICTION METHOD OF SMART ENERGY METER BASED ON DATA FUSION

SIWEI WANG[*] JI XIAO[†] YINGYING CHENG[‡] YU SU[§] AND WENLI CHEN[¶]

Abstract. In order to study the quality analysis method of key links in smart energy meters, the author proposes a data fusion based quality analysis and prediction method for smart energy meters. This method is based on the relevant data of key links in the electric energy meter, and selects the data of the electric energy meter in research and development design, material procurement, production and manufacturing, acceptance testing, installation and operation, dismantling and scrapping as the sample data for model construction. The XGBoost algorithm classification method is used to establish an intelligent electric energy meter quality analysis model. Taking the dismantled electricity meter data of a certain power company as an example, this paper conducts modeling analysis and prediction of various quality issues of smart electricity meters, and conducts on-site verification. Based on the verification results, the model is continuously optimized. The results indicate that: The model was optimized using cross validation and grid search methods, and the final model achieved an accuracy rate of 0.74 and a recall rate of 0.82 on the validation set. This method can meet the actual needs of power grid business and objectively reflect the quality situation of key links in smart energy meters.

Key words: Energy meter failure rate, Time series, Fault characteristics, XGBoost algorithm, multiple linear regression

1. Introduction. With the development of human living standards and society, more and more power electronic components are being applied to the power system. With the integration of national photovoltaic poverty alleviation projects, the proportion of distributed photovoltaic grid connection is increasing. These devices have relatively superior performance, but they have caused an impact on the power quality, making the problem of power quality in the low-voltage platform area increasingly severe. How to comprehensively monitor and evaluate the power quality of low-voltage substation areas has become one of the hot topics for power supply enterprises and researchers [1].

Excellent power quality is an important guarantee for the safe and economic operation of the power grid. Power quality issues not only cause losses to electricity consuming enterprises and customers, but also seriously affect the power supply service indicators of power supply enterprises. Even serious power quality problems will impact the brand image of power supply enterprises [2]. If there is a voltage quality issue, excessive voltage can cause damage to equipment such as transformers, energy meters, and electrical appliances used by customers; Low voltage can bring huge obstacles to social production and human life, and serious low voltage problems may cause machines to malfunction and cause economic losses [3]. For example, the three-phase imbalance problem in the low-voltage substation area can slightly reduce the efficiency of low-voltage lines and distribution transformers, but in severe cases, it may cause serious consequences such as wire overload burning, switch burning, and even single-phase burning of distribution transformers [4]. How to carry out power quality monitoring in low-voltage substations is the primary issue in analyzing power quality issues. The power quality monitoring device can provide real-time data monitoring for power supply enterprise staff, and assist them in recording and analyzing the basic situation of power quality in low-voltage substation areas. The monitored operational data can also be used to analyze the problems of electricity customers and provide effective solutions for grassroots grid managers [5].

[*]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (Corresponding author, SiweiWang5@163.com)
[†]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (JiXiao13@126.com)
[‡]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (YingyingCheng65@163.com)
[§]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (YuSu32@126.com)
[¶]State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing, 400023, China. (WenliChen59@163.com)

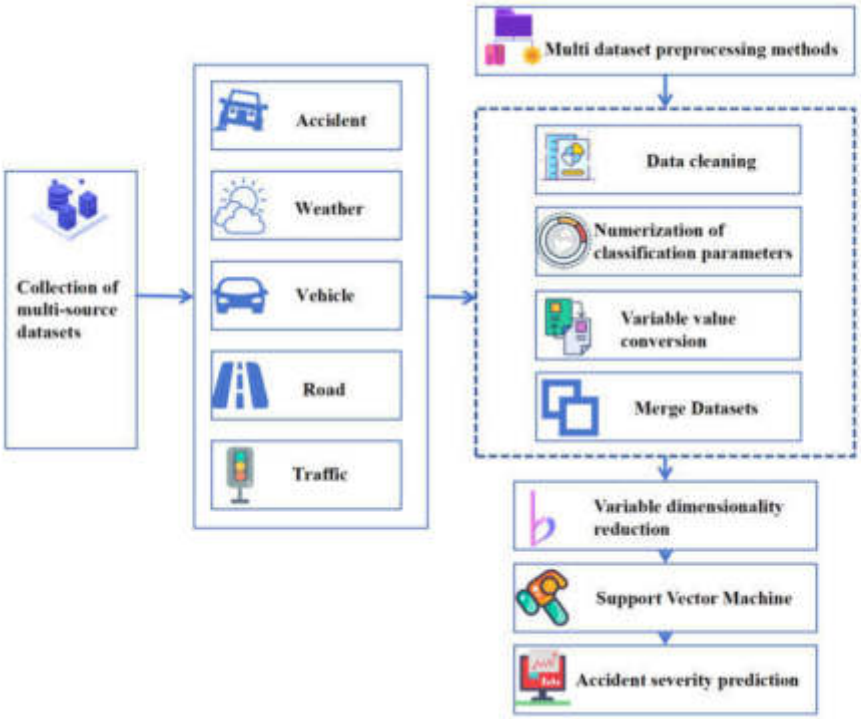


Fig. 1.1: Quality Analysis of Intelligent Energy Meters Based on Data Fusion

The power quality monitoring device can also record the type and geographical location of faults in power supply equipment, which helps to carry out power supply repair services and improve the efficiency of restoring power supply. With the deployment of digital transformation strategies for power supply enterprises, more and more perception devices are being applied to power supply, and a plethora of new technologies (big data, cloud platforms) are gradually being applied to various fields of power supply enterprises. Like the right wing front flag of Horqin, the power supply service resource scheduling and control system covers the entire area, and the perception ability of power grid equipment has been greatly improved [6]. The promotion and construction of business systems such as electricity information collection, online power grid, and power supply service resource scheduling and control provide strong data support for comprehensive monitoring of the operation status of the power grid. According to the author's statistics, all provincial companies of State Grid Corporation of China have established data service platforms and massive data platforms, but progress in data value mining, data analysis applications, data cleaning and integration is relatively slow [7] (Figure 1.1).

2. Literature Review. The price of power quality monitoring devices is expensive, and considering their cost, it is not possible to configure them in large quantities in the distribution network. Therefore, the optimization goal of minimizing power quality monitoring points has always been a research hotspot for power quality monitoring schemes. The goals of power quality monitoring are different, and the methods of configuring power quality monitoring points are also different, making it difficult to form a unified standard. Himeur, Y. proposed a monitoring device configuration scheme that takes into account the severity of voltage sag in substations, taking into account the number and observability of substation monitoring, and taking into account the number of monitoring points and the observability of voltage sag at each node of the entire network as constraints [8]; Chen, Y. proposed an equipment configuration optimization scheme for monitoring the entire network voltage using the fault point method, taking into account the number of monitoring points and the observability of voltage dips at each node, in response to voltage dips caused by line short circuits [9]; Nakutis, Ž. proposed in [10] a method of using particle swarm optimization algorithm to optimize equipment configuration

for monitoring points with voltage sag; Spertino, F. proposed a monitoring point configuration algorithm using an improved particle swarm optimization algorithm by reasonably setting the minimum number of monitoring points [11]; Karngala, A. K. proposed an equipment configuration optimization plan that takes into account the severity of voltage sag in substations, addressing the issue of existing power quality monitoring point layout schemes not taking into account the type, management requirements, and equipment configuration sequence of each monitoring point [12].

Electricity information collection data is usually stored as historical data in historical databases, and some scholars have begun to explore the value of this data; Zhou, M. analyzed electricity information collection data and proposed an electricity theft identification method based on electricity feature analysis, which can be used to screen suspected electricity theft users [13]; Ma, J. studied the fast clustering and anomaly detection techniques for power data flow in large-scale power information collection, and designed and implemented a flow clustering algorithm based on the clustering characteristics of power behavior in vertical and horizontal spaces, achieving fast clustering and anomaly detection [14,15].

The author reviews the quality related data of key links in electric energy meters and studies the method of extracting quality impact features; Compare various big data analysis technologies and establish a quality analysis model for smart energy meters; Use this model to predict and analyze potential quality hazards of smart energy meters, and conduct on-site verification. Continuously optimize the model based on the verification results.

3. Research Methods. The task of the electricity meter quality analysis model is to mine the patterns of faults in dismantled electricity meters based on relevant data of key links of electricity meters, predict the probability of faults in operating electricity meters with the same characteristics, and conduct on-site data verification.

The research and development design, material procurement, production and manufacturing, acceptance testing, installation and operation, and dismantling and scrapping processes that have a significant impact on the quality of electric energy meters are defined as key links. The data situation of each link is sorted out to facilitate subsequent data selection [16].

For key link data, use Pearson correlation coefficient and chi square test to conduct correlation analysis on data fields. Based on the threshold reference given by business experts, delete some fields with correlation coefficients greater than 0.5, and finally use the key link data of the electricity meter. After cleaning and transformation, generate data that can be used for modeling and analysis. The sample data of the model training set is based on historical data from Henan. In the data selection stage, a total of 130 fields were selected from the original data.

Analyze the 132 original features based on the key links of the electricity meter according to the following steps:

The first step is data visualization. In order to visually present the relationship between the characteristics and whether the electricity meter is faulty, these 132 original features were used to draw the distribution graphs of the faulty electricity meter and the normal electricity meter in each feature [17]. The distribution of fault table and normal table on several typical features is shown in Figure 3.1. As shown in the figure, there is no significant difference in the distribution of faults in each feature of the energy meter, and further feature extraction is needed through quantitative indicators.

The second step is to select features based on the Gini impurity method. The calculation formula is shown in equation 3.1.

$$I_G(f) = \sum_{i=1}^{m} f_i(1 - f_i) = \sum_{i=1}^{m} f_i - f_i^2 = 1 - \sum_{i=1}^{m} f_i^2 \tag{3.1}$$

In the formula, m represents the total number of categories; f_i is the probability that the sample points belong to class i.

Calculate the Gini importance of each feature by taking the reciprocal of Gini impurity, as shown in equation 3.2, and the results are shown in Table 3.1.

$$gini = \frac{1}{I_G(f)} \tag{3.2}$$

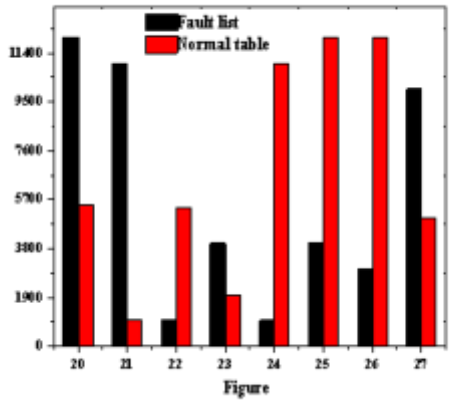


Fig. 3.1: Comparison of distribution of fault table and normal table on various features

Table 3.1: Feature Importance (Partial)

Name	Characteristic	Importance
The number of days the electric energy meter ran during the first abnormal collection	DAYS_BEFORE_FIRST_FA	122
The region code of the payment terminal	AREA_CODE	102
Running time of electric energy meter	OPS_MONTHS	88
Asset model	MODEL_CODE	76
Manufacturing unit	MANUFACTURER	64

Table 3.2: Preserved Features

Feature Name	describe
A_1	The number of days the electric energy meter ran during the first abnormal collection
A_2	The region code of the payment terminal
A_3	Running time of electric energy meter
...	...
A_n	The wiring method of the electricity meter

We calculate the proportion of the importance of each feature in the total importance of all features as shown in equation 3.3.

$$P_{importance} = \frac{gini_j}{\sum_{i=1}^{n} gini_i} \tag{3.3}$$

In statistics, events with a probability of less than 4% are generally considered as low probability events. Here, 4% is selected as the proportion threshold for feature selection, and features with an importance greater than 4% are retained [18]. Model by retaining 12 features through filtering. All features are shown in Table 3.2, represented by symbol A.

Step three, construct features. The construction feature is based on business and expert experience, constructing new features for warning records and abnormal code records of electric energy meters according to business logic, and dividing the features into one vote veto feature and important feature [19].

Table 3.3: Construction features (partial)

Feature Name	describe
B_1	Is there any abnormality in the electricity meter
B_2	Does the abnormal setting of electricity meter rates occur
B_3	Does voltage exceeding the limit occur
...	...
B_n	Does the abnormality of phase B of the high supply and high meter occur

Table 3.4: Construction features (partial)

Feature Name	describe
C_1	Does the meter fly away
...	...
C_n	Does the electricity meter stop running
ALARM_CODE_0201	Does voltage phase failure occur

If there is an abnormality corresponding to a veto feature in an electric energy meter, then the meter must have malfunctioned; If there have been anomalies corresponding to important features in an energy meter, it is possible that the meter has malfunctioned. The construction features include 13 veto features and 30 important features. Some features are shown in Table 3.3, represented by symbol B.

Step 4 summarize a total of 55 feature fields mentioned above. This part of the features is shown in Table 3.4, represented by the symbol C, where $C = A \cup B$.

Based on the key links of the electricity meter, establish a fault rate prediction model according to the above characteristic data, and complete the batch fault prediction of the electricity meter. There are two solutions to predicting batch failure rates: One is to directly predict the failure rate of batch energy meters; The second is to predict whether a single meter has failed, and then calculate the failure rate of the batch of electricity meters based on the number and total number of failures in the batch.

The direct prediction of failure rate can be achieved by: (1) Using a regression model to fit the linear relationship between the features obtained and the batch failure rate. This method can obtain the optimal weight relationship between the batch failure rate and each feature; (2) By using the data from the split table to obtain failure rate data at different times, and applying a time series model, the trend prediction of batch failure rate on the timeline can be obtained.

Another approach is to first predict single table failures, and then divide the number of failures by the total number of batches to obtain the failure rate of the batch [20]. Predicting whether a single table is faulty is a binary classification problem, and simple classifier models such as decision trees, SVM, Bayesian, etc. can be used. The results of these models are easy to interpret, but their accuracy is average and they are prone to overfitting; Ensemble learning models can also be used, including random forest algorithm, XGBoost algorithm, lightgbm algorithm, etc, such models are integrated on the basis of simple models. Compared with a single model, they are often more accurate and can effectively avoid over fitting. However, the calculation rules are complex, and the interpretability of the model is poor.

Recording the process of random event changes and developments in chronological order constitutes a time series. Observing and studying the time series, searching for its patterns of change and development, and predicting its future trends is called Time Series Analysis.

Time series prediction only requires a set of historical data of the variables to be predicted. Compared with regression prediction models, this method does not require the effort to determine the causal relationship between variables, but only needs to extend the historical trend determined by the time series model outward to predict future changes. Time series prediction is often suitable for situations where the independent variable data required for regression models is relatively scarce, and the historical data of the variables to be predicted is relatively complete, which is sufficient to reflect their changing trends.

Regression analysis is a statistical analysis method aimed at determining the quantitative relationship of interdependence between two or more variables. According to the type of relationship between the independent and dependent variables, it can be divided into linear regression analysis and nonlinear regression analysis. In big data analysis, regression analysis is a predictive modeling technique that studies the relationship between the dependent variable (target) and the independent variable (predictor) [21]. This technique is commonly used for predictive analysis, time series modeling, and discovering causal relationships between variables.

Ensemble learning is a framework of machine learning that combines multiple models to improve their overall generalization ability. There are three types of ensemble learning: Bagging, Boosting, and Stacking. The XG Boost algorithm is an improved gradient boosting learning algorithm, which is a method in Boosting. The algorithm principle is different from the traditional GBDT algorithm. Traditional GBDT only utilizes first-order derivative information during the training process, while the XG Boost algorithm performs second-order Taylor expansion on the loss function and adds a regularization term outside the loss function to obtain the optimal solution. This not only ensures model accuracy but also limits the complexity of the model, avoiding overfitting. The XGBoost algorithm is based on a tree model. The XGBoost algorithm is an additive model composed of k base models. Assuming that the tree model we want to train in the t-th iteration is $f_t(x)$, the prediction result $widehaty_i^{(t)}$ of sample i in the t-th iteration satisfies equation 3.4. The model loss function satisfies the equation $\sum_{i=1}^{n} l(\widehat{y_i}, y_i)$. In the formula, n represents the number of samples.

$$\widehat{y_i^{(t)}} = \sum_{i=1}^{n} f_k(x_i) = \widehat{y_i^{(t-1)}} + f_t(x_i) \tag{3.4}$$

The definition of the model objective function is shown in equation 3.5. In the formula, t represents the number of trees, and Ω represents the regularization term.

$$Obj = \sum_{i=1}^{n} l(\widehat{y_i}, y_i) + \sum_{i=1}^{t} \Omega(f_i) \tag{3.5}$$

Performing a second-order Taylor expansion on equation 3.3 and removing the constant term yields the objective function for the t-th iteration as shown in equation 3.6. In the formula, g_i is the first derivative of the loss function, and h_i is the second derivative of the loss function.

$$Obj^{(t)} \backsimeq \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{3.6}$$

According to equation 3.6, it can be seen that the XGBoost algorithm's loss function can be customized (there must be first and second derivatives), and the use of second derivatives makes gradient convergence faster and more accurate.

Establish linear regression, time series, and XGBoost algorithm models using existing Henan split table data. The hyperparameters of the three models are set to default values, and 98% is selected as the fault probability threshold. If the prediction result exceeds the threshold, it is judged as a fault Table, among them, linear regression and time series are used to determine batch fault tables, while XGBoost algorithm is only used for single Table fault determination. The model selected a total of 1190582 data from the first quarter of 2021 for training. Take archive information, R&D design data, material procurement data, production and manufacturing data, collected abnormal data, and measurement abnormal data as independent variables, and dismantle the sorting data to determine whether the electricity meter is faulty as the dependent variable to input into the XG Boost algorithm and linear regression model; Model batch failure rate as a time series. The model obtains the optimal joint probability distribution of the data through training, and applies this distribution to determine whether the smart energy meter is faulty in operation, achieving quality analysis of key links in the smart energy meter.

Under the same judgment criteria, the accuracy comparison of the three models is shown in Figure 3.2.

The XGBoost algorithm model has much higher prediction accuracy than other models, so the XGBoost algorithm model is chosen for subsequent analysis and prediction of electricity meter quality.

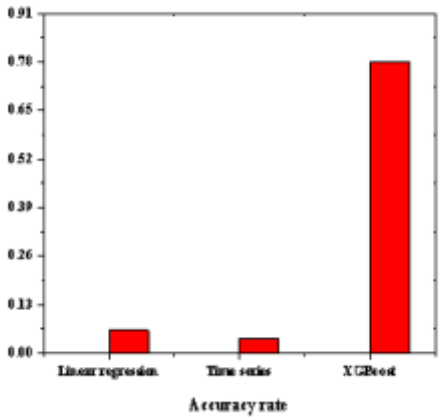


Fig. 3.2: Comparison of Model Accuracy

Table 4.1: Model Validation Results

Index	Value/piece	Proportion /%
Total	973621	
Actual number of faults	289178	8.00
Predict the correct quantity	695527	71.62
Number of prediction errors	276136	28.35
Predict the fault table as a fault Table(TP)	39342	4.06
Predict the fault table as a normal Table(FP)	48075	4.94
Predict a normal table as a faulty Table(FN)	61542	6.32
Predict the number of faults	100802	10.37

4. Result analysis. Using a total of 21157686 historical faulty electricity meter data from May 2019 to May 2021 in a certain area, a quality analysis model for intelligent electricity meters is established. The model predicts the fault data for three quarters from May to December 2021 in a certain area, and compares and verifies it with the actual dismantled data at the end of 2021.

The verification situation is as follows. The total number of electricity meters participating in the prediction is 3251317, involving 6112 arrival batches. Each meter is predicted for failure and compared with the actual results at the end of 2021 [22].

The validation data for the pre training stage of the model is that in the first quarter of 2021, there were 973621 electricity meters and 2313958 correctly predicted ones, accounting for 71.62%; The number of prediction errors is 917347, accounting for 28.35%. The detailed results are shown in Table 3.5.

After initial training, the accuracy of the model reached 0.73 and the recall rate reached 0.38. After verification, the model can recognize 44.65% of the fault tables, but there is also a 28.35% misjudgment situation. The accuracy of the model needs to be improved, and the evaluation indicators of the model performance are shown in Table 3.6 [23]. The definitions of model accuracy P and recall R are shown in equation 3.7:

$$P = TP/(TP + FP) \tag{4.1}$$

$$R = TP/(TP + FN) \tag{4.2}$$

After the first introduction of the model, the evaluation of the model results of training and test set, the accuracy of the training process is close to 1, while the accuracy of the test the index is stable around 0.70.

Table 4.2: Model Performance Evaluation Indicators

Evaluating indicator	Value
Accuracy	0.73
Accuracy	0.46
Recall	0.38

Finally, the validation process was evaluated and an accuracy of 0.46 was obtained. The results confirm that the general model is weak and overfitting may occur, requiring the optimization of the model parameters.

In order to improve the training model's performance, increase the generality, and reduce the risk of overwork, the model is gradually added to the electronic damage test data. set in the third and fourth quarters of 2021. pattern matching. Generalizability of the model to unknown data is verified using data augmented validation. Follow the pattern in two steps below:

1. The basic idea of good modeling based on K-fold cross-validation method is to group the original data in one way, one part is used as training set and the other factor is used as the test, and the classifier is first trained using the training method, and then as a performance test to evaluate the model to evaluate the training model using the test procedure. During the initial stage of model fitting, K-fold cross validation was used to improve the generalizability of the model [24].
   The main idea of K-fold cross-validation is to divide the original data equally into K sections, divide the data into K sections, select section i as the test set for section 3, and use the section K-1. based on the training set [25]. The average of the evaluation results of the K index was taken as the final evaluation of the model. The model is limited by this parameter in order to find the best combination for the model. Here, 5-fold cross-validation was used to select the correct one as the measurement parameter.
   With K-fold cross-validation, the precision of the training process was 0.66, the recall rate was 0.63, the measurement precision was 0.56, the recall rate was 0.55, and the acceptance precision was 0.46, and the reproducibility was increased. rate increased to 0.46. Up to 0.43, although the accuracy of training and testing has improved, the accuracy of the system is below 0.50, which is difficult to meet the needs, and it is necessary to take advantage of the model's hyperparameters;
2. Optimize the model a second time according to the network search method. Use the grid search method to optimize the hyperparameter values of the model.

The mesh search method involves partitioning the hyperparameters of the model into finite-valued elements. The program iterates over the composite values of all hyperparameters and selects the best model parameters based on parameter values as negative parameters.

Using the network search method, a precision of 0.82 for the training set, a recall rate of 0.84, a precision rate of 0.81, and a best return value of 0.78 for the parameters were obtained. 0.74 for the validation process and 0.82 for the recovery rate. A comparison of the effects before and after model development is shown in Figure 3.3.

The validation data of the model for each quarter from Q2 to Q4 2021 are shown in Figure 3.4.

Through optimization and adjustment, the model's parameters can quickly converge during the training phase, demonstrating excellent fitting ability in the training set. The accuracy rate in the validation set reaches 0.74, and the overall evaluation effect is relatively ideal, which can meet the actual needs of power grid business.

5. Conclusion. The author mainly focuses on the quality data of key links in electric energy meters, predicts the occurrence patterns of faults, and constructs an electric energy meter quality analysis model to study the quality analysis methods of key links in intelligent electric energy meters. The main research content includes the following two aspects:

1. Sort out the quality related data of key links in electric energy meters, study the key link data and quality impact feature extraction methods that affect the quality of electric energy meters, extract the regular features of faults in dismantled electric energy meters, use XG Boost algorithm model to learn the rules in dismantled electric energy meters, and construct a fault prediction model;

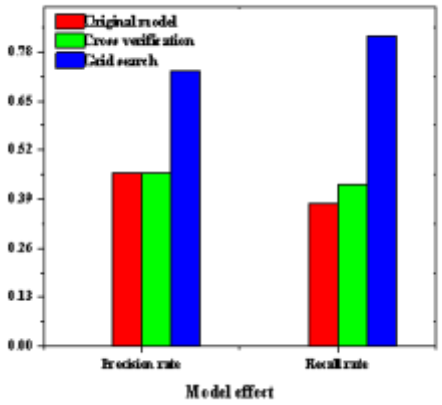


Fig. 4.1: Comparison of model effects before and after optimization

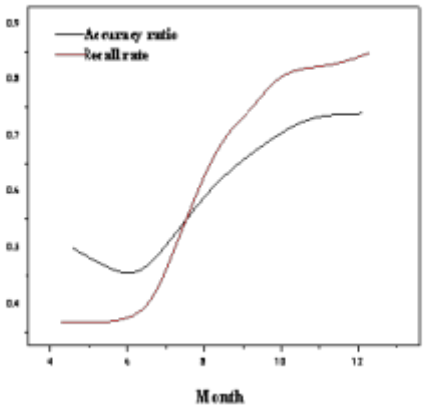


Fig. 4.2: Monthly Model Validation Results

2. Use the quality analysis model to identify the quality problem of the smart meter, model using the historical data, predict the explosion data at a specific location in April 2020, compare with the actual split data, and check. In May 2020, the model is optimized according to the confirmed results. The model was refined by using cross-validation and network research, and the final model achieved an accuracy of 0.74 and a recall rate of 0.82 of the validation process, which can be based on the real economy of the electricity project.

The author suggests a method to analyze and estimate the quality of smart meters during operation. This method is based on the important information from the main connection of the energy meter, and the selection of energy meter information for research and development, production, production, manufacturing, certification, installation, operation, demolition and disposal. sample data for design. The classification method of XGBoost algorithm is used to create intelligent models for the effective evaluation of power meters. After proving the accuracy, the results show that the accuracy of this method reaches 0.74, which can show the true value of the connection between smart meters.

Acknowledgement. The study was supported by Science and Technology Project funded by State Grid Co., LTD. (5700-202227226A-1-1-ZN).

REFERENCES

[1] Himeur, Y., Alsalemi, A., Al-Kababji, A., Bensaali, F., & Amira, A. (2020). Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations. Information Fusion, 64, 99-120.
[2] Ghazal, T. M. (2022). Energy demand forecasting using fused machine learning approaches. Intelligent Automation & Soft Computing, 31(1), 539-553.
[3] Kashinath, S. A., Mostafa, S. A., Mustapha, A., Mahdin, H., Lim, D., Mahmoud, M. A., ... & Yang, T. J. (2021). Review of data fusion methods for real-time and multi-sensor traffic flow analysis. IEEE Access, 9, 51258-51276.
[4] Muzammal, M., Talat, R., Sodhro, A. H., & Pirbhulal, S. (2020). A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. Information Fusion, 53, 155-164.
[5] Sahu, A., Mao, Z., Wlazlo, P., Huang, H., Davis, K., Goulart, A., & Zonouz, S. (2021). Multi-source multi-domain data fusion for cyberattack detection in power systems. IEEE Access, 9, 119118-119138.
[6] Zhang, Y., Jiang, C., Yue, B., Wan, J., & Guizani, M. (2022). Information fusion for edge intelligence: A survey. Information Fusion, 81, 171-186.
[7] Jiang, L., Wang, X., Li, W., Wang, L., Yin, X., & Jia, L. (2021). Hybrid multitask multi-information fusion deep learning for household short-term load forecasting. IEEE Transactions on Smart Grid, 12(6), 5362-5372.
[8] Himeur, Y., Rimal, B., Tiwary, A., & Amira, A. (2022). Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. Information Fusion, 86, 44-75.
[9] Chen, Y. (2021). College English teaching quality evaluation system based on information fusion and optimized RBF neural network decision algorithm. Journal of Sensors, 2021, 1-9.
[10] Nakutis, Ž., Rinaldi, S., Kuzas, P., & Lukočius, R. (2020). A method for noninvasive remote monitoring of energy meter error using power consumption profile. IEEE Transactions on Instrumentation and Measurement, 69(9), 6677-6685.
[11] Spertino, F., Chiodo, E., Ciocia, A., Malgaroli, G., & Ratclif, A. (2020). Maintenance activity, reliability, availability, and related energy losses in ten operating photovoltaic systems up to 1.8 MW. IEEE Transactions on Industry Applications, 57(1), 83-93.
[12] Karngala, A. K., & Singh, C. (2021). Reliability assessment framework for the distribution system including distributed energy resources. IEEE Transactions on Sustainable Energy, 12(3), 1539-1548.
[13] Zhou, M., Zhao, Y., Li, Q., Liu, X., Wang, Y., Sun, Y., ... & Yan, W. (2022). Smart Electricity Meter Reliable Life Prediction Based on Harmonic Factor. Electric Power Components and Systems, 50(13), 727-737.
[14] Ma, J., Teng, Z., Tang, Q., Qiu, W., Yang, Y., & Duan, J. (2021). Measurement error prediction of power metering equipment using improved local outlier factor and kernel support vector regression. IEEE Transactions on Industrial Electronics, 69(9), 9575-9585.
[15] Akhtar, I., Kirmani, S., & Jameel, M. (2021). Reliability assessment of power system considering the impact of renewable energy sources integration into grid with advanced intelligent strategies. IEEE Access, 9, 32485-32497.
[16] Zanghi, E., Brown Do Coutto Filho, M., & Stacchini de Souza, J. C. (2024). Collaborative smart energy metering system inspired by blockchain technology. International Journal of Innovation Science, 16(2), 227-243.
[17] Berka, A. L., MacArthur, J. L., & Gonnelli, C. (2020). Explaining inclusivity in energy transitions: Local and community energy in Aotearoa New Zealand. Environmental Innovation and Societal Transitions, 34, 165-182.
[18] Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360-3379.
[19] Chen, J., Zhao, F., Sun, Y., & Yin, Y. (2020). Improved XGBoost model based on genetic algorithm. International Journal of Computer Applications in Technology, 62(3), 240-245.
[20] Chakraborty, S., & Bhattacharya, S. (2021). Application of XGBoost algorithm as a predictive tool in a CNC turning process. Reports in Mechanical Engineering, 2(1), 190-201.
[21] Azmi, S. S., & Baliga, S. (2020). An overview of boosting decision tree algorithms utilizing AdaBoost and XGBoost boosting strategies. Int. Res. J. Eng. Technol, 7(5), 6867-6870.
[22] Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. Computational Economics, 61(2), 715-741.
[23] Song, P., & Liu, Y. (2020). An XGBoost algorithm for predicting purchasing behaviour on E-commerce platforms. Tehnički vjesnik, 27(5), 1467-1471.
[24] Li, S., & Zhang, X. (2020). Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. Neural Computing and Applications, 32(7), 1971-1979.
[25] Sankar, S., Potti, A., Chandrika, G. N., & Ramasubbareddy, S. (2022). Thyroid disease prediction using XGBoost algorithms. Journal of Mobile Multimedia, 18(3), 1-18.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 17, 2024
Accepted: Jun 10, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 673–681, DOI 10.12694/scpe.v26i2.3958

HYBRID DATA PUBLISHING BASED ON DIFFERENTIAL PRIVACY

TAO WANG*, KAINING SUN †, RUI YIN‡, TENG ZHANG§, AND LONGJUN ZHANG ¶

Abstract. The advent of the information and intelligence era has led to explosive growth of data. The author proposes a hybrid data model based on differential privacy. The main content of this model is based on the study of differential privacy, processing the data through a noise mechanism, using the calculation of tuple attribute differences and noise addition, and finally constructing a mixed data model based on differential privacy through experiments. The experimental results indicate that: as the value of k increases, the clustering results tend to be optimal, verifying that clustering the original data can reduce noise addition. However, ICMD-DP anonymizes the original dataset, resulting in much higher information loss than DCKPDP and prototype algorithms. A mixed data model based on differential privacy enables better clustering performance of the original dataset, thereby utilizing differential privacy to better protect the data.

Key words: Differential privacy, Mixed data, Information, Clustering

1. Introduction. In the era of big data, the release and utilization of data are key to promoting knowledge economy and social progress. Relevant research institutions will utilize these data resources for mining and analysis, in order to provide better services to the public. However, while providing significant benefits, publishing personal data to the public poses a significant threat to user privacy. In order to ensure user privacy and security, it is necessary to protect them. However, how to ensure that the published data is both usable and does not leak the privacy information contained in the data has become a major challenge in research on data publishing privacy protection.

The explosive growth of data, the release of which can provide scientific decision-making, predict market trends, and promote social development, truly promoting the flow of data value [1]. However, these data often contain a large amount of sensitive information, and direct publication will inevitably lead to user privacy leakage. Therefore, how to protect sensitive user information and maximize the availability of published data during the data publishing process has become an urgent problem to be solved. In recent years, some methods have been proposed to address privacy protection issues in data publishing, mainly based on data anonymity publishing methods and data distortion publishing methods. Although using such methods can to some extent protect sensitive information in the data, they require the assumption that the attacker does not have background knowledge, and therefore cannot resist background knowledge attacks and combination attacks. With the rapid development of the Internet, big data analysis technology and cloud computing, individuals, enterprises and institutions will generate a continuous stream of massive data every day. These massive amounts of data, when applied to research, can improve people's lives, promote development, and bring great convenience to their lives. However, while enjoying convenience, people are also facing the problem of personal privacy being violated. How to protect the privacy and security of user data while meeting the research needs of providing reasonable data is one of the hot topics of discussion and research in today's era (Figure 1.1).

*State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (Corresponding author, TaoWang65@163.com)

†Xinjiang Energy Internet Big Data Laboratory,Urumqi, Xinjiang,832000, China. State Grid Xinjiang Electric Power Co., Ltd, Urumqi, Xinjiang, 832000, China (KainingSun6@126.com)

‡State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (RuiYin17@163.com)

§State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (TengZhang3@126.com)

¶State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 832000, China. Xinjiang Energy Internet Big Data Laboratory, Urumqi, Xinjiang, 832000, China (LongjunZhang7@163.com)

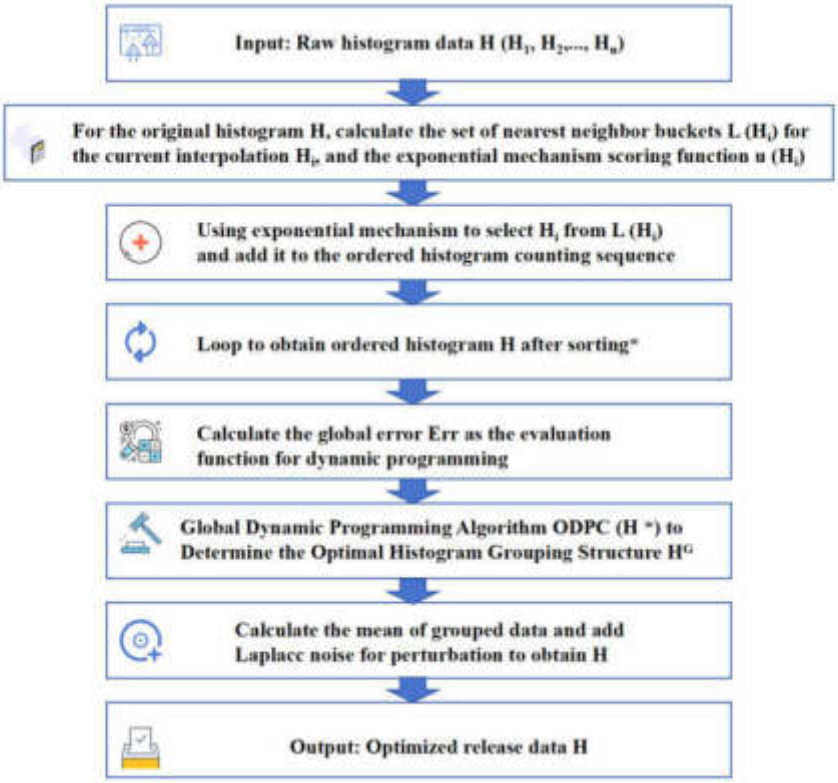


Fig. 1.1: Mixed data based on differential privacy

2. Literature Review. With the rapid development of mobile communication Internet and cloud computing technology, smart phones, wearable devices, sensors and other IP mobile devices with GPS chips can submit users' location information to location aware applications, providing consumers with convenient personalized services. But these mobile terminals can expose users' location tracks and personal information on the Internet in real time, which poses a great threat to people's privacy and security. Therefore, how to obtain valuable information from massive data while preventing privacy breaches is a current research hotspot in the field of data mining. With the rapid development of the Internet era, people find that deep learning models have great advantages in image processing, speech recognition and other fields. Deep neural networks have appeared in all aspects of human life. Among them, generative adversarial networks is one of the most promising deep learning models in the field of unsupervised learning. It consists of two parts: a generator and a discriminator. The generator generates "fake" data, which is then dynamically adjusted with the discriminator to ultimately generate the data that users need. But generating adversarial network models not only brings convenience to users, but also provides attackers with an opportunity to steal sensitive user data. For example, when a cancer diagnosis model is attacked by member inference, attackers can analyze the data information in the deep learning model to infer whether a patient has a certain type of cancer, and the user's privacy information is likely to be stolen by the attacker. Currently, there have been some research results on differential privacy data publishing methods, but these methods all have certain problems. Yang, J. proposed a new CMFD algorithm with the following workflow. Firstly, use the keypoint extraction method with the lowest contrast threshold to extract more keypoints from the input image. Secondly, a new technique, gradient hash matching, uses a hash table to quickly and effectively find similar pairs of key points, where the hash value is calculated using the gradient of the key points. Subsequently, a new method called simplified clustering filtering utilizes the density pattern of key points in the copy move region to remove mismatched key point pairs [2]. Huang, Z. et al.

believe that the accelerated mode matrix splitting method and the recently proposed generalized accelerated mode matrix splitting method are special cases.

Compared with existing methods, this method can use more information in each iteration, thereby improving computational efficiency. And the convergence of the method was studied, and the convergence of the method was proved under certain assumptions [3]. Zhang, P. proposed a data level fusion model that involves the integration of multiple information sources and unsupervised attribute selection of fused data [4]. Panfeng Zhang believes that excessive gradient perturbation noise in deep model differential privacy protection can lead to decreased usability, and proposes a differential privacy deep learning model based on particle swarm optimization algorithm. According to the particle swarm optimization strategy, the position of particles is mapped to network parameters to search for individual and global historical optimal positions. After perturbing the gradient obtained from the global optimal particle position, the model is re trained [5]. Liang, W proposed a differential privacy data publishing method based on DBSCAN clustering, but this method is also suitable for publishing numerical attribute data [6].

In response to the above issues, in order to ensure that the published data does not affect personal privacy and security, the author proposes a mixed data model research based on differential privacy. The differential privacy model, as a well-known privacy protection model, can provide privacy assurance by adding a certain amount of noise to the data query or analysis results without making any assumptions about the attacker's background knowledge. In a non interactive framework, data managers can publish datasets processed using differential privacy protection technology for researchers to mine and analyze.

The author reviews the quality related data of key links in electric energy meters and studies the method of extracting quality impact features; Compare various big data analysis technologies and establish a quality analysis model for smart energy meters; Use this model to predict and analyze potential quality hazards of smart energy meters, and conduct on-site verification. Continuously optimize the model based on the verification results.

3. Research Methods.

3.1. Differential privacy protection.

3.1.1. Definition of Differential Privacy. Differential privacy was initially used to limit the disclosure risk when returning query answers on a database, but its application in interactive scenarios strictly limits data analysis, because it only allows a limited number of queries to be answered, it promotes privacy protection research for data publishing in non cross five scenario scenarios.

Differential privacy is a model that provides strong privacy protection. In a non interactive framework, data managers can publish datasets processed using differential privacy protection techniques for researchers to conduct mining and analysis.

The main method for publishing differential privacy datasets in non interactive scenarios is based on histogram publishing. However, as the number of attributes increases, histogram based methods have serious limitations: For fixed attribute granularity, the number of histogram intervals increases exponentially with the number of attributes, which has a serious impact on computational cost and accuracy [7]. In addition, the histogram publishing method only provides approximate counts of partitioned data and cannot provide data details, thus limiting the utility of data analysis. Therefore, this limitation can be overcome by generating a universal dataset that satisfies differential privacy. The simplest method is to collect a set of query results that satisfy differential privacy, which requires querying each individual record in the original dataset. However, such queries require too much noise to meet the requirements of differential privacy, making it impossible for differential privacy datasets to maintain availability. The availability of differential privacy protection data can be improved by reducing query sensitivity and reducing the amount of noise added [8].

Differential privacy protection technology perturbs data by adding quantitative noise to the query results, ensuring that the insertion, modification, and deletion of records in any dataset will not affect the query results, thereby achieving privacy protection [9].

Differential privacy has a random algorithm K, as well as any adjacent datasets D1 and D2. If algorithm K satisfies differential privacy, it can be expressed as formula 3.1:

$$Pr[K(D_1 \in S)] \leqslant exp(\epsilon) Pr[K(D_2) \in S] \tag{3.1}$$

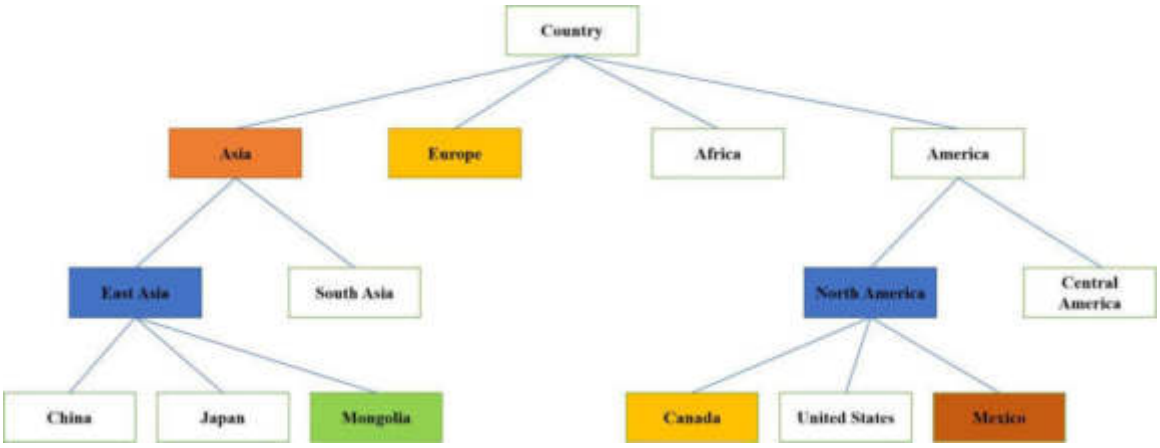


Fig. 3.1: Country Attribute Generalization Tree

3.1.2. Noise mechanism. Sensitivity refers to the maximum amount of change in the query result when the dataset changes and only one record changes. Differential privacy typically perturbs the return value of the query function with noise, and the magnitude of the added noise is closely related to the global sensitivity of the query function [10,11].

In practical applications, commonly used noise mechanisms include Laplace mechanism and exponential mechanism. The amount of noise can affect data security and availability, and is closely related to global sensitivity. Global sensitivity can be expressed as formula 3.2:

$$\Delta f = max_{D_1,D_2}||f(D_1) - f(D_2)|| \tag{3.2}$$

3.2. Differential Privacy Hybrid Data Publishing Algorithm.

3.2.1. Calculation of Tuple Attribute Differences. In the context of privacy data dissemination, data dissemination can be viewed as the collection of answers to continuous queries for each record in the dataset. The author proposes a hybrid data publishing algorithm based on differential privacy: using the k-prototype clustering algorithm, the initial class center is randomly selected first . Cluster the dataset based on an improved method for calculating the difference in tuple attributes, and then calculate the cluster center for each numerical attribute in each cluster based on the best clustering result. Generate a set of attribute values for categorical attributes. Next, traverse each data record and determine its clustering category. Replace numerical attributes with cluster center values and use Laplace mechanism to add noise. Use exponential mechanism to select categorical attributes. Finally, generate a differential privacy dataset. Due to the sensitivity of the query function being differentiated into k records in each set of data, it can reduce the amount of noise added and improve data availability.

Most existing data tables are mixed data Tables, which means that the data attributes in the tables are divided into numerical and categorical types. There are different methods for calculating attribute differences for data with different types of attributes [12].

Unlike numerical attributes, categorical attributes require the establishment of a generalized hierarchical tree to calculate attribute differences. Each subtyping attribute needs to establish a generalized hierarchical tree [13]. Figure 3.1 shows the generalized hierarchical tree of the Country attribute, with leaf nodes representing the values of each attribute on the Country attribute.

3.2.2. Noise addition method. For numerical attributes, the Laplace mechanism is used to add noise to the cluster center, which can be expressed as formula 3.3:

$$Centroid'(C_m(A_i^q)) = Centroid(C_m(A_i^q)) + Lap(\frac{\Delta f}{\epsilon}) \tag{3.3}$$

Table 3.1: adult Dataset

Attribute	Attribute type	Number of attribute values	Attribute	Attribute type	Number of attribute values
Age	Numerical type	75	Education level	Classification	15
Weekly working hours	Numericaltype	88	Gender	Classification	3
Education duration	Numerical type	18	Occupation	Classification	15
Marital status	Classification	8	Original nationality	Classification	42

Unlike numerical attributes, categorical attributes obtain values from a limited set of categories. Since adding Laplacian noise to cluster centers is meaningless, another method of obtaining differential privacy output is to select cluster centers in a probabilistic manner, which can be achieved through an exponential mechanism [14]. This mechanism selects the closest optimal center point based on input data, differential privacy parameters, and quality standards. In this case, the quality standard is the probability of each subtype attribute value appearing.

The differential privacy static data publishing algorithm based on k-prototype clustering is mainly used to publish mixed datasets containing numerical and subtype attributes. The algorithm is divided into a clustering and grouping stage and a data publishing stage. In the first stage, an improved k-prototype clustering algorithm is used to cluster and partition the data. In the second stage, differential privacy data publishing is achieved. If it is a numerical attribute, replace it with the cluster center value, and then use the Laplace mechanism to independently add noise to each attribute value; If it is a categorical attribute, the output attribute value is selected using an exponential mechanism based on the selection criteria of the central candidate of the attribute in the cluster attribute value set to which it belongs [15].

3.2.3. Algorithm Description. The DCKPDP algorithm is designed for publishing datasets containing mixed attributes, and the process mainly includes two parts: a) Cluster the original dataset using an improved k-prototype algorithm; b) Using differential privacy technology to add noise to the clustered dataset and output a dataset that satisfies differential privacy.

3.3. Experimental research.

3.3.1. Experimental Environment and Datasets. The experimental dataset used the adult dataset from the UCI machine learning database, which contains a total of 48943 data records. After deleting records with missing attributes, a total of 31257 records were obtained, due to the author's publication on a mixed attribute dataset, three numerical attributes and five categorical attributes were selected from the adult dataset as experimental attributes. The adult dataset is shown in Table 3.1.

3.3.2. Algorithm performance evaluation criteria. The author improved the k-prototype clustering algorithm to achieve better clustering performance on the original dataset, thereby incorporating less noise and improving data availability when using differential privacy protection. Therefore, the main purpose of this experiment is to demonstrate that the algorithm proposed in the article can improve data availability while ensuring a lower risk of data leakage [16].

Regarding the DCKPDP algorithm, adjust the privacy budgetϵ values are set to {0.02, 0.2, 2, 6}, and the number of attributes q is taken as 3 and 7 for comparative experiments, among them, when q is set to 3, two numerical attributes (age, age, weekly working hours) and two subtype attributes (original nationality, education level) are taken, and the information loss SSE caused by the DCKPDP algorithm is shown in Figure 3.2 (a) (b).

As shown in Figure 3.2, when q=3, the value of SSE is much smaller than when q=7. This is because as the number of attributes increases, more and more noise is added to the original data, resulting in more information loss in the data and a larger corresponding SSE value. Whenϵ taking 0.02, although SSE shows a downward trend, the change is not significant because a large amount of noise is added when the privacy budget is too small. Using the author's improved k-prototype clustering algorithm to process the original dataset resulted in very low data availability; Whenϵ taking 0.2, the change in SSE is most significant; Whenϵ taking 2 and 6,

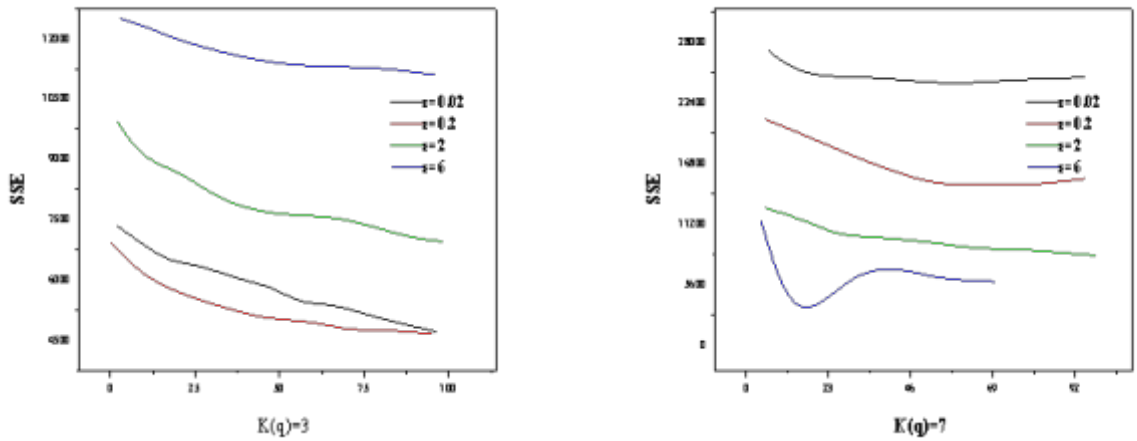


Fig. 3.2: Changes in SSE when q takes different values

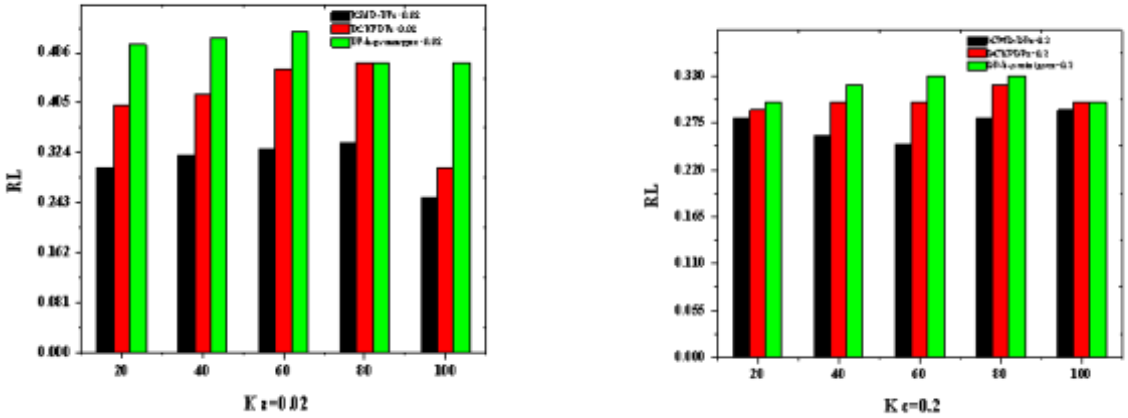


Fig. 3.3: When ϵ taking 0.002 and 0.2, the variation of RL with k value

the SSE values of the two are lower and the difference is small, because whenϵ taking a larger value, the added noise is small and has little impact on the SSE value of the data.

As the value of k increases, the number of clustering clusters increases. Data records with lower dissimilarity are divided into the same cluster, and the clustering effect is close to optimal. The less noise is added, so the overall trend of SSE is decreasing, which also proves the feasibility of the algorithm [17].

Compare the RL values of DCKPDP, ICMD-DP, and DP k-prototype algorithms, as shown in Figures 3.3 (a) (b) and 3.4 (a) (b).

From Figures 3.3 and 3.4, it can be seen that the privacy budget whenϵ taking different values, the RL value of ICMD-DP is lower compared to DCKPDP and DP k-prototype, indicating a lower risk of privacy leakage. This is because ICMD-DP anonymizes the dataset and applies differential privacy protection to the anonymized dataset, which will inevitably provide stronger protection for the data. The RL values of DCKPDP and DP-k-prototype are not significantly different, with a difference of about 3%. However, from the experimental results,

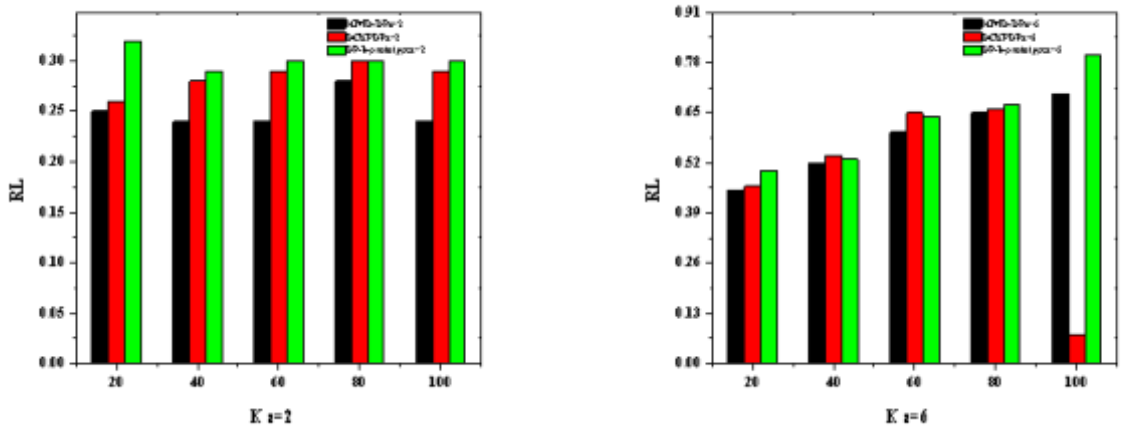


Fig. 3.4: When ϵ taking 2 and 6, the variation of RL with k value

even if ICMD-DP uses anonymization and differential privacy to process the original dataset, the difference in RL values between them and DCKPDP is still controlled within 7%. Therefore, from the perspective of data leakage risk alone, the dataset processed by DCKPDP can still meet the requirement of ensuring data privacy is not leaked. When privacy budgetϵ value of RL is 0.02, and the value of RL is the smallest, which means the risk of privacy leakage is minimized. This is because a large amount of noise is added to the data, and the data availability is also the lowest at this time; When privacy budget ϵ values are 0.2 and 2, it can be seen that the RL values of both DCKPDP and DP k-prototype are not significantly different, with a difference of about 4% and a difference of 7% compared to ICMD-DP; When privacy budgetϵ value of k is 6, as the value of k increases, the RL value increases significantly, and the risk of data leakage also increases. Therefore, considering the DCKPDP comprehensively whenϵ is taken 2, the algorithm performance is optimal [18].

4. Result analysis. The experimental settings for q are 3 and 7, when ϵ value is set to 2, the experimental results are shown in Figure 4.1(a) (b) to compare the changes in information loss of DCKPDP, ICMD-DP, DP k-prototype algorithm, and standard difference privacy algorithm on the adult dataset.

As shown in Figure 4.1, when ϵ taking 2, as the value of k increases, the clustering results tend to be optimal. The information loss of DCKPDP, ICMD-DP, and DP k-prototype algorithms gradually decreases, and the information loss is much lower than that of the standard difference privacy algorithm. This verifies that clustering the original data can reduce noise addition, however, ICMD-DP anonymizes the original dataset, resulting in much higher information loss than DCKPDP and DP k-prototype algorithms. The clustering algorithm proposed by the author adaptively selects the initial center point and improves the dissimilarity calculation formula compared to the DP-k-prototype algorithm. The clustering effect is improved, and the information loss caused by adding noise to it through differential privacy is also reduced, resulting in improved data availability [19,20]. Therefore, from the experimental results, it can be concluded that compared to ICMD-DP and DP kprotype algorithms, DCKPDP can reduce information loss and significantly improve data availability while ensuring a lower risk of information leakage, proving the superiority of the DCKPDP algorithm.

5. Conclusion. The author studied the privacy protection issue of mixed attribute data publishing and proposed a new data publishing protection method. In response to the research question, the author first improved the traditional k-prototype clustering algorithm's dissimilarity calculation method and proposed a method that can adaptively select the initial clustering center point, improving the accuracy and stability of clustering. Finally, differential privacy was applied to the classified dataset to ensure data privacy was not leaked. Through experimental verification, the DCKPDP algorithm can improve the availability of data while ensuring a lower risk of data leakage compared to similar algorithms. However, the author used a lower data

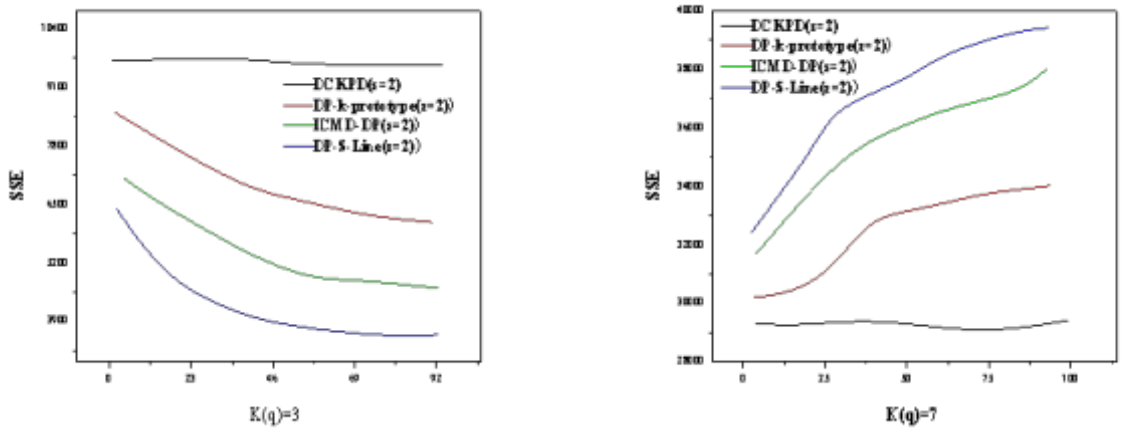


Fig. 4.1: When ϵ taking 2 and q with different values, the variation of SSE

dimension during the experiment, which may lead to efficiency issues when publishing high-dimensional data; And the author has adopted the principle of equal distribution for the allocation of privacy budget, which may result in waste of privacy budget and loss of data information, reducing the utility of data.

REFERENCES

[1] Li, L., Guo, L., Zhong, D., Huang, X., & Zhang, J. . (2023). Reliability analysis of deep-water explosion test vessel based on fuzzy interval. International Journal of Pattern Recognition and Artificial Intelligence, 37(04).

[2] Yang, J., Liang, Z., Li, J., Gan, Y., & Zhong, J. . (2023). A novel copy–move forgery detection algorithm via gradient-hash matching and simplified cluster-based filtering. International journal of pattern recognition and artificial intelligence(6), 37.

[3] Huang, Z., & Cui, J. . (2024). Accelerated relaxation two-sweep modulus-based matrix splitting iteration method for linear complementarity problems. International Journal of Computational Methods, 21(02).

[4] Zhang, P., Li, T., Yuan, Z., Luo, C., Wang, G., & Liu, J., et al. (2022). A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data. Information Fusion(80-), 80.

[5] Panfeng ZHANG, Danhua WU, & Minggang DONG. (2023). Differential privacy deep learning model based on particle swarm optimization. Computer Engineering, 49(9), 144-157.

[6] Liang, W., Chenyang, H., Jiangning, S., & Jianhua, Y. . (2024). Ctec: a cross-tabulation ensemble clustering approach for single-cell rna sequencing data analysis. Bioinformatics(4), 4.

[7] Ju, H., Ding, W., & Gu, P. Y. X. . (2023). Bi-directional adaptive neighborhood rough sets based attribute subset selection. International journal of approximate reasoning, 160(9), 1.1-1.18.

[8] (2022). Investigators from university of texas austin release new data on machine learning (3d microseismic monitoring using machine learning). Robotics & Machine Learning Daily News(4), 6-7.

[9] Bristow, N. R., Best, J., Wiggs, G. F. S., Nield, J. M., Baddock, M. C., & Delorme, P., et al. (2022). Topographic perturbation of turbulent boundary layers by low-angle, early-stage aeolian dunes. Earth Surface Processes and Landforms: The journal of the British Geomorphological Research Group(6), 47.

[10] Jeon-Young Kang, Michels, A., Crooks, A., Aldstadt, J., & Wang, S. . (2022). An integrated framework of global sensitivity analysis and calibration for spatially explicit agent-based models. Transactions in GIS: TG(1), 26.

[11] Akshay, V., Sunil, K., Raj, G. P., Tarique, R., & Arvind, K. . (2023). Enhanced cost and sub-epoch based stable energy-efficient clustering algorithm for heterogeneous wireless sensor networks. Wireless personal communications: An Internaional Journal(4), 131.

[12] Zheng, T., QiGe, Xiong, F., Li, G., Xue, Y., & Deng, X. . (2023). Study of the flexural performance and a novel calculation formula for the degree of composite action for precastc oncrete sandwich panels. The structural design of tall and special buildings(18), 32.

[13] Kim, H., Strang, A., & Sanz-Alonso, D. . (2023). Hierarchical ensemble kalman methods with sparsity-promoting generalized gamma hyperpriors. Foundations of Data Science, 5(3), 366-388.

[14] Viviana Elizabeth Zárate-Mirón, & Serrano, R. M. . (2023). The impact of smart specialization strategies on sub-cluster efficiency: simulation exercise for the case of mexico. Competitiveness Review: An International Business Journal , 33(2),

364-394.
[15] Sun, Z., & Zhao, J. . (2023). Comprehensive performance evaluation of landing gear retraction mechanism in a certain model of aircraft based on rpca method. Journal of Circuits, Systems and Computers, 32(11).
[16] Benrhouma, O., Alzahrani, A., Alkhodre, A., Namoun, A., & Bhat, W. A. . (2022). To sell, or not to sell: social media data-breach in second-hand android devices. Information & computer sceurity(1), 30.
[17] Chumnangoon, P., Chiralaksanakul, A., & Chintakananda, A. . (2023). How closeness matters: the role of geographical proximity in social capital development and knowledge sharing in smes. Competitiveness Review: An International Business Journal , 33(2), 280-301.
[18] Zhou, T., Hu, Z., Su, Q., & Xiong, W. . (2023). A clustering differential evolution algorithm with neighborhood-based dual mutation operator for multimodal multiobjective optimization. Expert Systems with Applications, 216, 119438-.
[19] Wang, W., Li, G., Wang, Y., Wu, F., Zhang, W., & Li, L. . (2022). Clearing-based multimodal multi-objective evolutionary optimization with layer-to-layer strategy. Swarm and Evolutionary Computation(68-), 68.
[20] Kaho, T., Watanabe, S., & Sakakibara, K. . (2022). Multi-objective branch and bound based on decomposition. IEEJ Transactions on Electronics, Information and Systems, 142(3), 373-381.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 17, 2024
Accepted: Jun 15, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 682–690, DOI 10.12694/scpe.v26i2.3959

GREEN DIGITAL OPERATION AND MAINTENANCE TECHNOLOGY OF POWER EQUIPMENT BASED ON DEEP LEARNING

QIA YANG*, GE QU †, QIZHEN SUN ‡, XIAOFEI DING§, AND JUE YANG¶

Abstract. In order to solve the problem of low accuracy in traditional digital operation and maintenance of power equipment, the author proposes a research on green digital operation and maintenance technology of power equipment based on deep learning. The author first analyzed the text characteristics and segmentation difficulties of work orders, summarized seven types of entities, and manually annotated 3452 work orders to form a training set, Secondly, pre train the BERT module using relevant equipment testing and fault analysis reports to obtain power word vectors, Then use the BiLSTM module to predict entity labels, Finally, the CRF module was introduced to optimize the prediction labels and conduct Chinese entity recognition experiments on 1000 work orders. The experimental results indicate that: Compared with LSTM, BiLSTM, and BiLSTM-CRF models, the BERT BiLSTM-CRF model has better entity recognition performance in power primary equipment operation and maintenance work order texts, with an F1 value of 85.7%, which is 11.6%, 7.3%, and 4.8% higher than the other three models, respectively. This system can effectively improve the work efficiency of operation and maintenance personnel, reduce their communication costs, achieve efficient management of on-site equipment, and assist enterprises in achieving production goals of improving quality and efficiency.

Key words: Power primary equipment operation and maintenance work order, BERT model, Bidirectional Long Short Term Memory Network, Conditional random airport

1. Introduction. In the construction process of the new power system, we will vigorously promote the research and development of ultra-high voltage power equipment and the application of power electronic devices, and bear the long-term impact of impact energy loads such as wind power and photovoltaic in the new power system. Conducting on-site operation and maintenance tests on them will incur a significant cost [1]. Applying digital twin technology to the field of state perception and operation and maintenance of ultra-high voltage power equipment in the "dual high characteristic" new power system should overcome numerous key links [2]. Multi domain simulation of high-voltage power equipment involves structural analysis models, flow field analysis models, multi-body dynamics models, fatigue analysis models, acoustic analysis models, electromagnetic simulation models, and electronic heat dissipation analysis models [3]. This physical model realizes the virtual construction and visual analysis of high-voltage power equipment, and also has the ability to simulate and calculate multiple physical fields. It can achieve real-time simulation calculation and dynamically display the distribution status of electrical/thermal/mechanical/fluid multiple physical fields through external loads and boundary conditions [4]. Physical space mainly includes performance, functionality, and structural strength; The digital space mainly includes models, algorithms, and visualization [5]. Flow field and electric field distribution can be effectively achieved in twin bodies, while visualizing and analyzing vibration, deformation, fluid, electric field potential current, fatigue, and life information.

Since the launch of the Power Production Management System (PMS) in the power grid, the operation and maintenance work order data of primary power equipment has shown explosive growth [6]. These data contain rich information on the health of power assets, but they are mainly stored in unstructured text format, which is difficult to effectively utilize. Therefore, it is particularly important to structure unstructured operation and maintenance work orders. With the development of named entity recognition technology in natural language processing, possible solutions have been provided for the above-mentioned problems.

*China Southern Power Grid Co., Guangzhou, 510000, China. (Corresponding author, QiaYang5@163.com)
†China Southern Power Grid Co., Guangzhou, 510000, China. (GeQu32@126.com)
‡China Southern Power Grid Co., Guangzhou, 510000, China. (QiZhenSun8@163.com)
§China Southern Power Grid Co., Guangzhou, 510000, China. (XiaoFeiDing7@126.com)
¶China Southern Power Grid Carbon Asset Management Co., Guangzhou, 510000, China. (JueYang82@163.com)

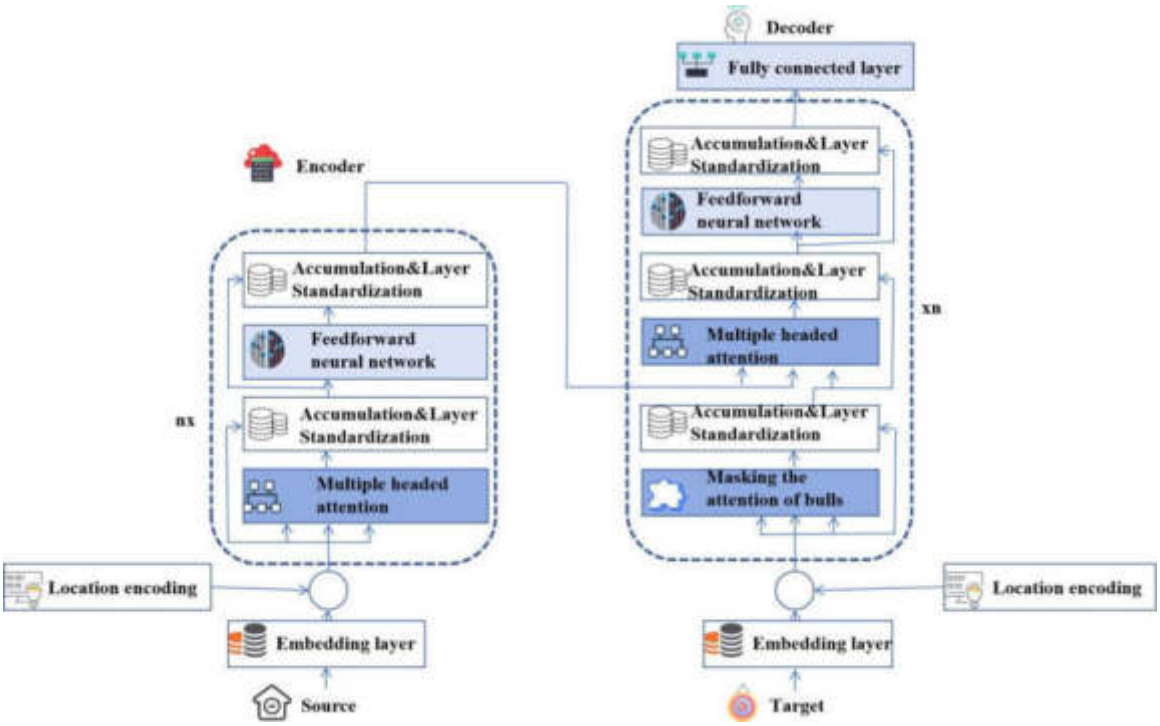


Fig. 1.1: Power Equipment for Deep Learning

Through the research and development of intelligent control technology for power equipment, a digital operation and maintenance management system for equipment is formed to enhance the information and intelligent management capabilities of key equipment in the installation process, testing and experiments of enterprises, assist in enterprise transformation and upgrading, complete delivery on time, and improve the core competitiveness of enterprise production activities [7] (Figure 1.1).

2. Literature Review. In the new situation of economic development, the construction of the power grid remains at a high level. As the core equipment of power grid construction, power equipment is a prerequisite for ensuring the safe and stable operation of the power grid [8]. Against the dual background of the "dual carbon" goal and digital transformation strategy, power equipment is a key component of the power industry and an important support for achieving carbon reduction and digital transformation goals [9]. As a leading enterprise in the power industry, strengthening the management of power equipment suppliers is of practical significance for power grid enterprises to operate themselves and collaborate with suppliers to achieve their goals [10]. On the basis of analyzing the value chain of power grid enterprises, the author aims to build an evaluation system for power equipment suppliers from the perspective of value chain synergy, based on the simulation results of the system dynamics model, optimization suggestions for power equipment management under this evaluation system are obtained, providing a basis for the evaluation and management of power equipment suppliers in power grid enterprises. Gilani, H. When selecting green water plants, one of the best methods is to use sustainable selection methods. In addition, the model solves the problem of improving all communications, choosing the right equipment for the treatment plant by choosing different equipment, and controlling the water loss [11]. Zahedi, R. believes that the demand for power equipment in power grid enterprises is complex and vast, and they face a large and uneven number of power equipment suppliers. The evaluation and selection of suppliers in the process of power equipment management become the key to selecting equipment. Building an efficient and reasonable power equipment evaluation system is conducive to comprehensively and effectively feedback information and behavior of power equipment suppliers in the big

data environment, thereby forming data resources and helping power grid enterprises to do a good job in power equipment supplier management [12]. With the operation and development of enterprises, the activities of realizing value appreciation in the value chain extend from product production activities to services and other activities. Mehmet Mithat ü ner focuses on organizing internal audits of power grid enterprise business processes from a value chain perspective, with value creation, value-added activities, and corresponding business processes as the main audit objects, in order to achieve value added for the enterprise [13]. Ali, Z. predicts the value chain operation of power equipment manufacturing enterprises from multiple aspects such as production, marketing, and service, effectively improving the accuracy of business risk prediction for power equipment manufacturing enterprises [14].

At present, named entity recognition technology has begun to penetrate into the field of electricity. The author constructed a model based on BERT BiLSTM CRF for unstructured operation and maintenance work order data of primary power equipment, and achieved named entity recognition for operation and maintenance work orders of primary power equipment [15]. Firstly, the text characteristics and segmentation difficulties of operation and maintenance work orders were analyzed, and seven types of entities were summarized: "equipment name", "equipment voltage level", "equipment line", "equipment substation", "equipment damaged parts", "equipment maintenance situation", and "equipment maintenance time". More than 3400 work orders were selected for entity annotation, forming a training set; Then, a BERT model suitable for the power field was pre trained, and BiLSTM was used as the entity label prediction layer and CRF was used as the processing layer for the global optimal solution of the label. Together, a power equipment operation and maintenance work order entity recognition model was constructed; Finally, a high accuracy recognition experiment for work order text entities was completed.

3. Method.

3.1. Overall demand analysis. In order to make the functions of the platform more in line with production reality, the designers started from the daily work of municipal station operation and maintenance management personnel, fully considered the business needs of three types of roles: decision-making, management, and execution. Based on 3D visualization and web related technologies, the key production elements of individual and various production equipment in the station were mapped to the 3D visualization scene, making it closely related to the management of daily production processes, and achieving dynamic monitoring and digital management of the entire life cycle of assets and equipment in the station. As shown in Figure 3.1, the system needs to achieve the following functional requirements.

Integrating multi-dimensional data with 3D models can dynamically display the overall and constituent units of the model, and users can query information such as process flow, individual size, and spatial position through interaction with the model.

3.1.1. Information on the use and maintenance of basic electrical equipment. Textual information on electronic equipment operation and control activities is obtained from 350,000 work order records in the company's PMS. The author only considers the operation and maintenance data of main transformers, isolating switches and circuit breakers for three-phase main electrical equipment of 120kV, 210kV and 510kV. After cleaning the data, more than 12000 work orders were selected as data.

3.2. Characteristics and segmentation difficulties of operation and maintenance work order text. After analyzing the contents of the above work notes, compared to the general calculations, we found that the work notes for the operation and maintenance of the main electronic equipment have the following characteristics.

1. The content of the operation usually includes information such as electronic equipment, equipment name, equipment line, electrical equipment, non-functional equipment, cleaning equipment, cleaning time, etc. However, because of the differences in writing supervision, there is a difference in the structure and execution of the content of the staff and the work order.
2. Due to the different types of equipment failure and the difference in cleaning process, the length of work order is very different. According to the available data, the shortest worksheet can be 11 words, and the longest is 354 words. Difficulty in the segmentation of work orders for the operation and maintenance of the main electrical equipment is:

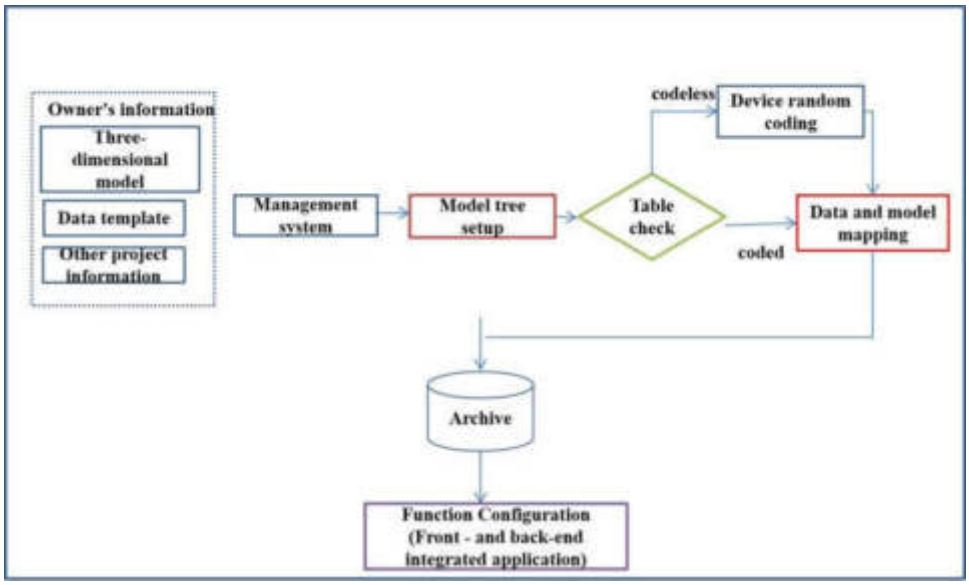


Fig. 3.1: Overall Business Requirements

(a) Due to the different writings of work and maintenance workers, there are some differences in their descriptions for the same product or equipment, for example, "main transformer" is abbreviated as "main transformer";
(b) for numerical data in the text of the work order, such as voltage level "120kV" and device model "capacitor model BAMH2", it is necessary to rely on the elements statement of the meaning for the conclusion of the enterprise proportion;
(c) For various work orders, the description of the maintenance of explosives and equipment is long and there are many devices, inconsistencies, and unclear segmentation boundaries, so the model should be able to be generalized.

3.3. BERT BiLSTM CRF based operation and monitoring operation certification standard. By analyzing the above mentioned O&M work orders and segmentation problems, the author developed an O&M work order recognition model based on BERT BiLSTM CRF.

First, the BERT module is prepared to use the interference measurement and error analysis of three main types of electrical equipment: the main transformer, circuit breaker and isolation; Second, to get the correct vectors, enter the text of the operation order into the pre-prepared BERT module in the sentence; Then use BiLSTM to extract the local features of the words and predict the object names; Finally, the correlation between the adjacent labels is done by CRF to get the best prediction for classification [16].

3.4. BERT module. In natural language processing, the term embedding is used to obtain a low-level representation of a language, i.e., a vector representation of the language.

Currently, mixed language models include Word2Vec and GPT. But these models have some difficulties in recognizing the work of the original equipment and management work force. The Word2Vec model generates static word embeddings and cannot represent polysemy. GPT, on the other hand, is a one-way language that cannot accept the content of that language.

The results of the BERT model reduce the above problems by identifying character-level and language-level features to improve the representation of language vectors and improve the accuracy of the model from previous training [17]. Therefore, the author chooses the BERT model to capture the message vectors of operations and control operations. To improve the accuracy of the model, the author first trained the BERT model.

Specifically, each word input by BERT is represented by the superposition of three vectors: Token Embed-

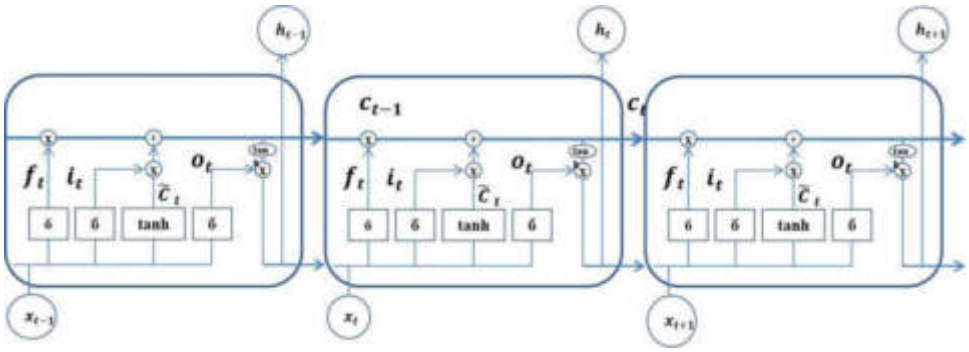


Fig. 3.2: Unit structure of LSTM

ding, Seg Element Embedding, and Position Embedding, which means word embedding, sentence embedding, and position embedding. In the figure, [CLS] is used to mark the beginning of a sentence, and [SEP] is used to mark the boundary between two sentences.

Its basic structure is a combination of various transformer encoders. The encoder has a multihead self-guided mechanism, a residual module, a normalization module, and a feedforward neural network.

In the encoder, the most crucial part is the self attention mechanism, which is expressed as:

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}}) \tag{3.1}$$

In the formula, Q, K, and V represent the query vector, key vector, and value vector, respectively, calculated by the word embedding vector through linear transformation matrices W_Q, W_K, and W_V. d_k is the dimension of the key vector, used to adjust the inner product of QKT to prevent gradient instability during training due to excessive inner product. The multi head self attention mechanism is a linear combination of self attention mechanisms, which can enable the model to learn more relevant information in different representation subspaces, increasing the diversity of model information sampling. Its formula is expressed as equation 3.2.

$$MultiHead(Q, K, V) = Concat(head_1, \cdots, head_n)W^0 \tag{3.2}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^K) \tag{3.3}$$

After each self attention module, residual and normalization processing (Add&Norm) is required on the results to improve the vanishing of model gradients and accelerate network convergence. Finally, link the feedforward neural network to enhance the model space and consolidate the original encoding information.

3.5. BiLSTM module. Due to the inability of fully connected neural networks to mine contextual semantic information of input sequences. Therefore, the author proposes Recurrent Neural Network (RNN), which has a certain memory function but suffers from gradient vanishing and exploding problems [18]. Subsequently, Long Short Term Memory (LSTM) networks introduced threshold mechanisms to address the issues of long-term dependencies and gradient vanishing [19]. The unit structure of LSTM is shown in Figure 3.2.

The core of LSTM is a forget gate, input gate, output gate, and a memory unit. Forgetting Gate f_t determines how much historical information affects C_t. The input gate i_t determines how much current input information affects C_t. The output gate o_t controls how much information is visible to the outside world, that is h_t. The calculation formula for updating the unit status of LSTM is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.4}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.5}$$

$$\widetilde{C_t} = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3.6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3.7}$$

$$C_t = f_t C_{t-1} + i_t \widetilde{C_t} \tag{3.8}$$

$$h_t = o_t tanh(C_t) \tag{3.9}$$

In the formula, σ represents the sigmoid excitation function, W represents the weight matrix, b represents the bias vector, h_{t-1} represents the output of the LSTM unit at the previous time, and x_t represents the input at the current time.

Due to the inability of unidirectional LSTM to obtain text features through posterior input information, the author uses Bidirectional Long Short Term Memory (BiLSTM) network to extract text features and improve the model capability of power entity recognition.

3.6. CRF module. BiLSTM extracts text features based on context, outputs the score for each character corresponding to each label, and outputs the label with the highest score for that character as the final label. However, there is a label mismatch issue. Therefore, a Conditional Random Field (CRF) is introduced to handle the dependency relationship between labels and obtain the optimal prediction sequence, aiming to improve the accuracy of prediction as much as possible.

For any input sequence $X = (x_1, x_2, \cdots, x_n)$, assuming that the corresponding output label sequence trained by BiLSTM is $Y = (y_1, y_2, \cdots, y_n)$, the score function is:

$$S(X, Y) \sum_{i=1}^{n} (A_{y_{i-1}, y_i} + P_{i,y}) \tag{3.10}$$

In the formula, n is the number of characters, A_{y_{i-1}, y_i} represents the transition score matrix between adjacent character labels in the text, and $P_{i,y}$ is the y_i label score of the i-th character in the text. The probability of predicting sequence Y is:

$$P(X, Y) = \frac{e^{s(X,Y)}}{\sum_{\widetilde{Y} \in Y_X} S(X, \widetilde{Y})} \tag{3.11}$$

In the formula, $\widetilde{Y}$ represents the true annotation sequence, and Y_X represents the set of all possible label sequences. Finally, the set of labels with the highest score will be used as the optimal label output.

3.7. Value Chain Theory Analysis. Analyzing the operational status of enterprises from the perspective of value chain, dividing enterprise activities according to their value appreciation status, and analyzing value activities can provide a more comprehensive understanding of the enterprise's business process, as well as the conditions and problems faced by each link. It can discover how to improve the quality and efficiency of each value link activity, thereby enhancing product value, maximizing enterprise value, and promoting sustainable development of the enterprise [20].

3.8. Experimental Preparation.

3.8.1. Annotation System and Datasets. The common named entity recognition annotation systems mainly include BIO, BIOES, BIOES+and other modes, and the author adopts the BIO annotation system. In this system, B (Began) represents the first character of an entity word, I (Inside) represents the middle and last characters of an entity word, and O (Outside) represents a non entity word. Filter out 3462 work orders from the operation and maintenance work order data source, and use annotation tools to manually annotate the text using seven types of labels: "Voltage Level", "Equipment Name", "Line Name", "Transfer for Sta", "Damage Part", "Repair Condition", and "Time".

Finally, the annotated results are transcoded into the BIO annotation system and divided into training and testing sets in a 7:4 ratio. An example of text annotation for operation and maintenance work orders is shown in Table 3.1.

Table 3.1: Example of Work Order Text Annotation

Character	Dimension	Label	Dimension
Li	B-EquipmentName	shut	I-EquipmentName
1	I-EquipmentName	hair	O
2	I-EquipmentName	heat	O
7	I-EquipmentName	Follow	B-RepairCondition
Septum	I-EquipmentName	Tracking	I-RepairCondition
leave	I-EquipmentName	measure	I-RepairCondition
open	I-EquipmentName	temperature	I-RepairCondition

Table 3.2: Training Environment Configuration

Operating system	Linux
CPU	Intel(R)Xeon(R)Silver4114
GPU	TeslaV100
Python	3.5

3.8.2. Evaluation indicators. The author uses recall rate R, accuracy rate P, and F_1 values to evaluate the performance of the model. The calculation method for each evaluation index is as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{3.12}$$

$$R = \frac{T_P}{T_P + F_N} \tag{3.13}$$

$$F_1 = \frac{2PR}{P + R} \tag{3.14}$$

In the model, T_P represents the number of actual positive events and predicted positive events, F_P represents the number of negative events and predicted positive events, and F_N represents the number of actual positive events and predicted negative events.

3.8.3. Experimental Environment. Build an experimental environment based on the PyTorch platform, and the specific training environment configuration is shown in Table 3.2.

3.8.4. Experimental parameter configuration. During the training process, the Transformer encoding part of the BERT model has 10 layers, 10 self attention mechanisms, and an output of 400 dimensions. Other specific hyperparameters are shown in Table 3.3.

4. Results and Discussion. In order to verify the recognition performance of the model used in the operation and maintenance work order text of primary power equipment, 1200 work order texts from the test set were used to test the model. The specific experimental results are shown in Table 4.1.

According to Table 4.1, compared with LSTM, BiLSTM, and BiLSTM-CRF models, the BERT BiLSTM-CRF model has better entity recognition performance in power primary equipment operation and maintenance work order texts, with an F1 value of 85.7%, which is 11.6%, 7.3%, and 4.8% higher than the other three models, respectively. Compared to the LSTM model, the BiLSTM model considers the posterior input in terms of input sequence, so the results are slightly better. Compared to the BiLSTM model, the BiLSTM-CRF model has added a CRF module, which can better handle the dependency relationships between labels, so the evaluation index is slightly higher. Compared to the BiLSTM-CRF model, the BERT BiLSTM-CRF model uses the BERT module, which can enhance the model's representation ability in the power field through pre training and output higher quality word vectors, thus improving the evaluation indicators.

Table 3.3: Experimental Parameter Settings

Parameter	value
Number of Transformer Layers	10
hidden_dim	210
optimizetr	Adam
learning_rate	0.0021
max_sequence_dim	400
batch_size	30
dropout	0.6
epoch	60

Table 4.1: Test set test results

Model	P	R	F_1
LSTM	0.814	0.706	0.742
BiLSTM	0.810	0.775	0.785
BiLSTM-CRF	0.835	0.804	0.808
BERT-BiLSTM-CRF	0.863	0.825	0.857

5. Conclusion. The author presents research on green digital marketing and electronic device technology based on in-depth research. Three-dimensional production equipment and maintenance, the operation and maintenance of workers can help improve productivity and maintenance, stabilize the station, improve business level and quality control. It helps achieve the goals of efficient management of equipment, maximum utilization of resources, automation of operations and management, and optimization of business operations. At the same time, it provides important knowledge for the development of other digital industries and future car maintenance.

REFERENCES

[1] Liu, J. Y., Liu, Y., Sun, T., Gao, H. Z., & Liu, Y. (2022). Ideas and key technologies of collaborative design for underwater weapon power system. journal of unmanned undersea systems, 30(4), 413-421.

[2] Xie, D., Xie, D., Wei, L., Wei, L., Wei, Z., & Wei, Z., et al. (2022). Water-induced dual ultrahigh mobilities over 400 cm 2 v 1 s 1 in 2d mos 2 transistors for ultralow-voltage operation and photoelectric synapse perception. Journal of Materials Chemistry C, 10(13), 5249-5256.

[3] Lei, X., Ji, H., Xu, Q., Ye, T., Zhang, S., & Huang, C. (2022). Research on data diagnosis method of acoustic array sensor device based on spectrogram. Global Energy Internet: English, 5(4), 16.

[4] Gheraba, A. R., Pati, D., Fedler, C. B., Schmidt, M., Molina, M. S., & Nejat, A., et al. (2023). A comparative analysis of visualization methods in architecture:employing virtual reality to support the decision-making process in the architecture,engineering,and construction industry. Civil Engineering and Architecture: English Version, 17(2), 73-89.

[5] Kviat, A. (2022). Post-digital prosumption and the sharing economy of space: the pay-per-minute cafe:. Journal of Consumer Culture, 22(3), 801-822.

[6] Dong, L., Chen, N., Liang, J., Li, T., Yan, Z., & Zhang, B. (2023). A review of indoor-orbital electrical inspection robots in substations. Industrial Robot: the international journal of robotics research and application, 50(2), 337-352.

[7] Ding, S., Zhang, T., Sheng, K., Chen, Y., & Yuan, Z. (2023). Key technologies and applications of intelligent dispatching command for high-speed railway in china. Railway Science (English), 2(3), 336-346.

[8] Ma, S., Yu, N., Wang, X., Mei, S., Zhao, M., & Han, X. (2023). Multi-criteria decision-making for power grid construction project investment ranking based on the prospect theory improved by rewarding good and punishing bad linear transformation. Energy Engineering (English), 120(10), 2369-2392.

[9] Yuan, Z., Jia, S., Liang, D., Wang, X., & Yang, Y. (2022). Research on slot-pole combination in high-power direct-drive pm vernier generator for fractional frequency transmission system. Journal of Electrical Machinery and Systems, Chinese Electrotechnical Society, 6(4), 445-453.

[10] Chukwu, O. A., & Adibe, M. (2022). Quality assessment of cold chain storage facilities for regulatory and quality management compliance in a developing country context. The International Journal of Health Planning and Management, 37(2), 930-943.

[11] Gilani, H., Shobeiry, S., Kami, M. B., & Sahebi, H. (2023). A sustainable redesign model for the water/wastewater supply

network: a water–energy nexus approach. Kybernetes, 52(5), 1842-1860.
[12] Zahedi, R., Ghodusinejad, M. H., & Gitifar, S. (2022). Threats evaluation of border power plants from the perspective of fuel type and providing solutions to deal with them: a case study of iran. Transactions of the Indian National Academy of Engineering, 8(1), 55-67.
[13] Mehmet Mithat üner, Cigdemoglu, C., Wang, Y., Yalcin, A., & Cavusgil, S. T. (2023). A review of the evolving conceptualization of internationalization from a global value chain perspective. Review of International Business and Strategy, 33(4), 549-580.
[14] Ali, Z. (2023). Predicting smes performance through green supply chain practices: a mediation model link of business process performance. Asia Pacific Journal of Marketing and Logistics, 35(2), 432-450.
[15] Daikun, W. U., Qi, F., & Shuaishuai, C. (2023). Research of bim technology in urban rail operation and maintenance and application. International Equipment Engineering and Management: English version, 28(3), 180-192.
[16] Lixia, Y. (2022). G3i equipment maintenance auxiliary tester design solution. Equipment for Geophysical Prospecting, 32(3), 199-203.
[17] Zhang, J., & Hua, Q. I. (2022). Data mining and spatial analysis of social media text based on the bert-cnn model to achieve situational awareness:a case study of covid-19. Journal of Surveying and Mapping: English Edition, 5(2), 11.
[18] Pan, S., Yan, K., Yang, H., Jiang, C., & Qin, Z. (2022). A sparse spike deconvolution method based on recurrent neural network like improved iterative shrinkage thresholding algorithm. Geophysical Prospecting for Petroleum, 58(4), 533-540.
[19] Duan, H., Meng, X., Tang, J., & Qiao, J. (2023). Prediction of no_(x)concentration using modular long short-term memory neural network for municipal solid waste incineration. Chinese Journal of Chemical Engineering: English Edition, 56(4), 46-57.
[20] Mehmet Mithat üner, Cigdemoglu, C., Wang, Y., Yalcin, A., & Cavusgil, S. T. (2023). A review of the evolving conceptualization of internationalization from a global value chain perspective. Review of International Business and Strategy, 33(4), 549-580.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 17, 2024
Accepted: Jun 18, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 691–698, DOI 10.12694/scpe.v26i2.3962

RESEARCH AND IMPLEMENTATION OF CAMPUS NETWORK INTRUSION DETECTION SYSTEM BASED ON DATA MINING AND IMAGE PROCESSING

ZHE ZHANG*

Abstract. In order to solve the problem of traditional intrusion detection system programs usually being manually written, with a large workload and certain limitations, the author proposes the research and implementation of a campus network intrusion detection system based on data mining and image processing. Its hardware components include a data warehouse, sensors, checkers, generators, etc; The software design includes a packet capture module, a data preprocessing module, and an event analyzer module. Experimental comparison with traditional methods. The experimental results show that the campus network intrusion detection system designed by the author is far superior to traditional system design in detecting intrusion behavior, approaching 100% infinitely, and has high effectiveness. The system has a certain degree of adaptive ability and can effectively detect external intrusions.

Key words: Image processing, Data mining, Campus network, Intrusion detection

1. Introduction. In recent years, with the rapid development of image acquisition and storage technology, we have been able to easily obtain a large amount of useful image data (such as remote sensing image data, medical image data, etc.). But how to fully utilize these image data for analysis and extract useful information from them has become the biggest problem we face. Image data mining has emerged as an emerging field in data mining [1,2]. Image data mining is a technology used to mine hidden knowledge, relationships within or between images, and other patterns hidden in image data on a large scale. It is still in the experimental research stage and is an emerging but highly promising research field. Introducing data mining techniques into image processing is not something that can be achieved overnight. Early image data mining was only focused on certain preprocessing of images, such as image segmentation based on data mining, image feature extraction based on data mining, and so on. With the rapid development of image processing and data mining technology, the early exploratory application of data mining technology in image processing can no longer meet the needs of practical applications. Therefore, the author proposes the idea of further integrating the two disciplines. At this stage, the image is first preprocessed, striving to use a unified representation model and method to process the image data; Then, data mining is carried out on the processed image data to more effectively extract relevant useful data. Among them, the latter stage is currently the focus of research and development in image data mining technology [3,5].

As the Internet continues to expand rapidly, it has become an integral part of daily life, bringing with it a myriad of challenges to network security. The persistent threat of hackers infiltrating computer networks and the onslaught of virus attacks have presented significant hurdles for networks and information systems. In response, diverse measures have been implemented to safeguard network systems. These include digital signature technology, biometric technology, content filtering technology, network isolation technology, and more. While these defense mechanisms offer a degree of protection against external interference and ensure swift and efficient information transmission, they also suffer from drawbacks such as high rates of false positives and false negatives[6]. Consequently, accurately and effectively detecting attacks remains a formidable challenge. Traditional intrusion detection systems are constrained by their ability to only detect data from established databases. As the frequency and complexity of attacks escalate, these databases expand, leading to decreased detection efficiency and system overload. In contrast, network intrusion detection utilizing data mining techniques delves into vast datasets to uncover hidden and previously unknown valuable insights. By accurately

*School of Computer Science and Technology, Nanyang Normal University, Nanyang, Henan, 473061, China (Corresponding author, ZheZhang56@126.com)

identifying correlations among intrusion data, it sidesteps the need for labor-intensive manual analysis and minimizes the associated workload and human intervention. This approach ultimately curtails operational and maintenance expenses [7].

2. Literature Review. Image data mining refers to extracting or mining useful information or knowledge from large-scale image sets. Therefore, we can understand image data mining as an application of data mining in the field of images [8]. The two fundamental concepts of image data mining are "large-scale image sets" and "extracting and mining useful information and knowledge.". From the perspective of "large-scale image sets", it involves fields such as image acquisition, image storage, image compression, and multimedia databases; From the perspective of "mining useful information and knowledge", this approach encompasses a range of disciplines including image processing and analysis, pattern recognition, computer vision, image retrieval, machine learning, artificial intelligence, and knowledge representation. Therefore, image data mining is a multidisciplinary emerging field, and most of the other fields it involves are also in the development stage, and it itself is also in the experimental stage. Intrusion detection, as a key component of computer network security, involves multiple technologies such as daily log parsing, system vulnerability checking, and network link detection. Intrusion detection systems can be categorized into two main types: anomaly-based detection technology and misuse-based detection technology. Anomaly-based detection technology assesses current system user actions against typical user behavior to identify any deviations indicative of intrusion activity [9-10]. Brahma, A. et al. devised a collaborative neural fuzzy inference system that integrates genetic algorithm-based database intrusion detection systems. This innovative approach proves adept at identifying malicious transactions within databases. Through rigorous experimental analyses and comparative assessments against established statistical database intrusion techniques, the efficacy of the proposed system was convincingly demonstrated [11]. Attou, H. et al. introduced a cloud-based intrusion detection model leveraging random forest and feature engineering. The model incorporates a random forest classifier, enhancing the accuracy of the detection system [12]. Wang et al. introduced an innovative intrusion detection approach based on deep learning principles. Their method utilizes two separate in-memory autoencoders trained on both normal network traffic and attack patterns. This allows for the capture of dynamic relationships between traffic features, particularly in the context of imbalanced training data. Subsequently, the original data is inputted into a triplet network for training, with triplets formed by the reconstructed data from the two encoders. Ultimately, the distance relationship within these triplets serves as the basis for determining whether a given traffic instance constitutes an attack [13-14].

With the gradual emergence of various security threats, it also drives continuous innovation in information network security. A large number of information security protection technologies have emerged, including host security protection technology, data encryption technology, network firewall, website security access technology, identity authentication technology, and vulnerability scanning technology. Due to the widespread potential information security issues in computer networks, the above technologies are unable to detect and respond to intrusion attacks in a timely manner, resulting in serious economic losses. Therefore, it is particularly important to construct a proactive and proactive security protection system that can actively prevent attacks and intrusion behaviors. Unlike passive defense firewall technology, as an active attack security protection technology, it can monitor network data in real-time, detect various attacks or abnormal behaviors inside and outside the network, and actively respond and block attacks before they occur or cause serious economic losses. Intrusion detection systems are a good choice for addressing popular information security threats such as network attacks and cross site scripting attacks.

The author mainly designs a high-level template library for a low code management system based on the results of system requirements analysis. Firstly, the overall architecture of the system is designed according to the overall requirements of the system. Then, based on the results of functional requirements analysis, the system functional modules are designed, and the user roles and permissions existing in the system are divided. Finally, the system directory structure is designed.

3. Method.

3.1. Intrusion detection technology and classification. In the 1980s, intrusion detection technology (IDS) emerged as a proactive defense mechanism. By gathering crucial data from computer systems and subjecting it to analysis, IDS can identify potential security policy breaches and network attacks. Intrusion

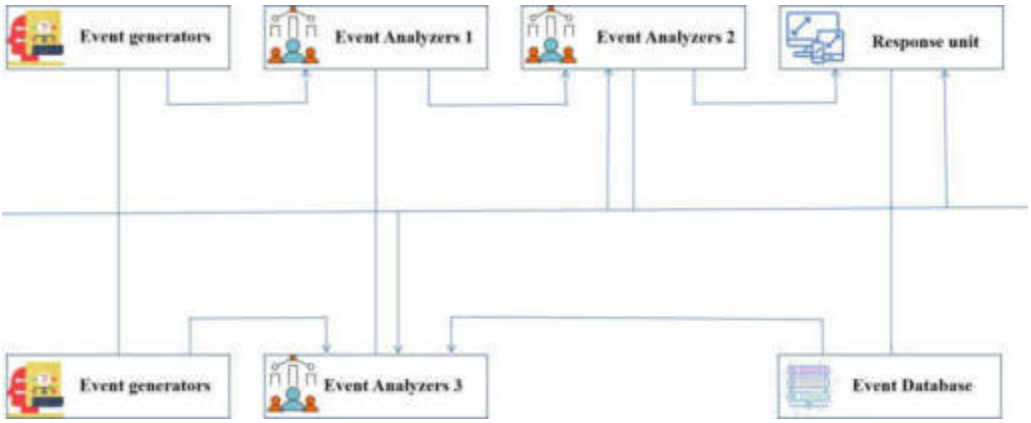


Fig. 3.1: Intrusion Detection System Model

behavior typically falls into two categories: internal and external. Internal intrusion pertains to unauthorized actions by legitimate users, whereas external intrusion involves the infiltration of hackers or unauthorized individuals. Such intrusive activities pose risks to the integrity, functionality, and confidentiality of network data, and various intrusion detection technologies prioritize different aspects of this threat landscape. Intrusion detection can be categorized based on the outcomes they aim to achieve, resulting in anomaly detection and misuse detection, which employ distinct detection methods[15]. Additionally, they can be classified as network-based or host-based, depending on the data sources they utilize. Furthermore, intrusion detection systems can be distinguished by their detection timing, either real-time or non-real-time. Models such as the unified model and the Snort model are prevalent in this domain, typically comprising event generators, analyzers, response units, and databases, as illustrated in Figure 3.1.

3.2. Process Analysis of Intrusion Detection. When dealing with intrusion behavior, the detection of network systems typically involves four key stages:

1. Data Collection: This initial phase involves gathering information about the system through various means such as external sensors or proxy hosts. It includes capturing user behaviors, system statuses, and fundamental network data.
2. Data Processing: In this stage, the collected data undergoes processing and transformation into a standardized format recognized by computers. This step aims to enhance the efficiency and speed of detection.
3. Data Analysis: Here, the processed information is analyzed. This may involve comparing patterns with known databases, conducting statistical analyses using probability theory, and identifying potential threats. Uncertain or suspicious data is forwarded to the control module for further assessment.
4. System Response: Based on the analysis and in accordance with predefined rules, the system responds accordingly. This could involve actions such as reconfiguring routers, isolating intruder IPs, or modifying file properties to mitigate the intrusion.

This process is illustrated in Figure 3.2.

3.3. Data mining techniques. Data mining involves uncovering potentially significant patterns or rules within vast amounts of complex, noisy, and inconsistent data. The implementation of data mining typically encompasses three primary processes: (1) Data Preparation: This initial stage involves various tasks such as selecting appropriate data targets, identifying operational entities, preprocessing data to remove noise, and reducing the dimensionality of the data. These steps aim to make the data suitable for analysis. (2) Data Mining: In this phase, utilizing different data mining models, suitable mining algorithms are chosen to extract valuable insights from large, incomplete, and irregular datasets. The goal is to uncover patterns or relationships that can help predict outcomes or discover hidden knowledge. (3) Data Representation and Evaluation:

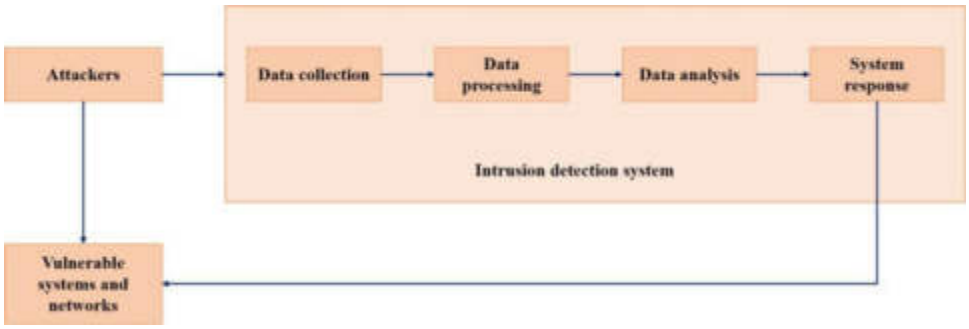


Fig. 3.2: Intrusion Detection Process

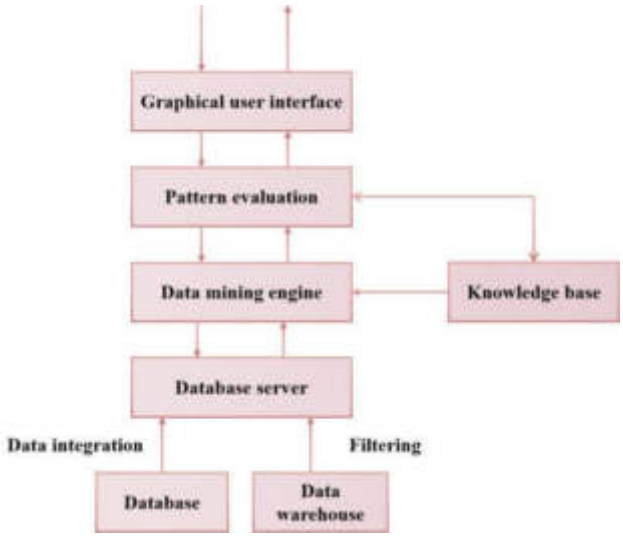


Fig. 3.3: Intrusion Detection Process

Following data mining, the acquired information undergoes analysis through techniques such as association rule mining, classification, and clustering. This process aims to derive meaningful insights and present them in a comprehensible manner, often through data visualization techniques. The structure of a data mining system is depicted in Figure 3.3.

3.4. Data Mining Algorithms for Intrusion Detection. Intrusion detection systems heavily rely on data mining algorithms for effective threat detection. Various algorithms come with distinct pros and cons, suitable for different models. Statistical analysis, feature analysis, change and deviation analysis, and clustering are frequently employed methods in data mining. Among these, association rules play a central role, revealing the connections within data. Mathematically, association rules can be defined as follows: Suppose there's a database D containing m pieces of information, denoted as Ti, where each piece of information comprises n units I. The relationship between these units is depicted as $D = \{T_1, T_2, \cdots, T_m\}, T = \{I_1, I_2, \cdots, I_n\}$, the subset of the database is represented by A. If there is | A |=K, then A is called the project set of K. The support of database D containing project set A is represented by Cx. If there are project sets A and B, then their association rules are:

$$Support(A \Rightarrow B) = P(A \cup B) \tag{3.1}$$

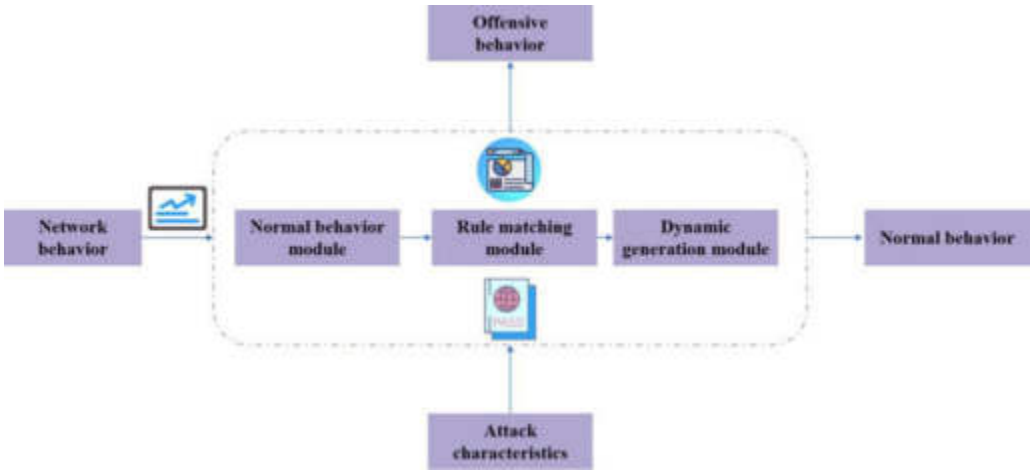


Fig. 3.4: Intrusion Detection Process

$$Confidence(A \Rightarrow B) = P(A|B) \tag{3.2}$$

In equations 3.1 and 3.2, Support(A) signifies the frequency of occurrence of itemset A within the database D, while Confidence(A) denotes the strength of the association rule for A. The calculation method:

$$Support(A)/\% = \frac{C_x}{|D|} \times 100 \tag{3.3}$$

Hence, by examining the support and confidence values between project sets A and B, we can ascertain the association rules between them. If the confidence of association rules A and B exceeds the minimum confidence threshold for the project set and the support of A and B is equal to or greater than the minimum support threshold, it indicates strong association rules between the project sets A and B. Conversely, if these conditions are not met, it suggests weak association rules between them. We can also gauge the correlation between project sets A and B through another metric:

$$r_{A,B} = \frac{\sum(A - \overline{A})(B - \overline{B})}{(n-1)\sigma_A \sigma_B} \tag{3.4}$$

In equation 3.4, the average values of A and B are $\overline{A}$, $\overline{B}$, and the standard deviation is $\sigma_A \sigma_B$.

3.5. Overall Design of Intrusion Detection System. To ensure timely and effective analysis of network data, the author emphasizes the essence of intrusion detection system design: mining association rules and sequence rules between data, and implementing classification recognition based on these rule definitions. Since different system models require different data mining algorithms, selecting a suitable intrusion detection system becomes crucial. For this purpose, the author suggests using the lightweight open-source intrusion detection system, Snort. Despite its effectiveness against most network attacks, Snort suffers from limitations such as low efficiency, false alarms, missed reports, and the inability to perform dynamic real-time detection. Consequently, there's a need to enhance the Snort detection model[16]. The enhanced Snort system model is illustrated in Figure 3.4.

3.6. Hardware composition of campus network intrusion detection system based on data mining. With the comprehensive analysis of hardware requirements for intrusion detection systems, the author designed the hardware part of a campus network intrusion detection system based on frequency hopping data mining. This part mainly consists of a data warehouse, sensors, inspectors, generators, etc. Its advantages are stable performance, good scalability, and the ability of multiple intrusion detection components in the

network to work together. Sensors can timely capture raw data appearing in computer networks, obtain basic characteristics of system data analysis, and upload formatted data to inspectors. The inspector uses a detection model to detect the data structure of the sensor uploaded data, detect whether there is intrusion behavior, and upload the results to the data warehouse. As the application storage center of the entire system, a data warehouse is mainly used to store historical data information. A generator refers to the ability to process detected abnormal data or intrusion behavior in a timely manner, generate a special type of data, and then connect this data with the historical information stored in the data warehouse to obtain new intrusion features, and put them into the intrusion feature storage library, which is conducive to early detection when encountering the same attack. When the system detects abnormal data, the alarm will alert [17].

3.7. Software composition of campus network intrusion detection system based on data mining.

3.7.1. Packet Capture Module. Wincap is mainly composed of three parts, namely the packet filtering driver Packet.dll, the lower level data dynamic connection library, and API that can directly access driver data. Wincap can capture a large number of data packets on computer networks and filter them according to pre-set patterns. The execution steps for capturing data packets using Wincap are as follows. Firstly, obtain the adapter sequence table and the model of the network adapter in the system. Secondly, select a default adapter number 1 from all adapters and set it to mixed mode. Once again, set the packet filtering driver in BPO to filter the original data packets, such as interfaces, IP addresses, etc. The buffer pool needs to have 564 K of storage space, allocate a packet object, and link to the allocated buffer pool; Assign packet matching objects and link them to a pre allocated buffer pool to capture multiple driver packets; And receive the target data packet from the network Lap Adaptation, and put the data packet into the packet structure targeted by the Lap Adaptation. Once received, it immediately returns True, otherwise it returns FALSE. Finally, with the help of a triggering device and a mathematical function model, the captured data packets that comply with the filter rules are uploaded to the network protocol analysis module for analysis. After completion, the processing object of the data packet is released, and the network card instrument is shut down to restore the network card to its normal state.

3.7.2. Data preprocessing module. Data processing refers to the large-scale cleaning, collection, transformation, discretization, and classification of raw data, providing clear and reasonable data preparation for data mining. Data processing refers to the pre-processing of captured data packets for subsequent data matching and intrusion detection processes. The preprocessing module mainly serves BPO filtering and decoding of network protocol code. Protocol decoding refers to the programmatic processing of data packets. The captured raw network packets cannot be directly applied and can only be used after being manipulated. This is because network packets are divided into several data slices with their own feature codes before uploading, but Internet networks are based on group conversion programs, so the order and sequence of uploading this data slice to the host is irregular, and the data uploaded to the host system is an unordered data stream. In order to effectively reply to the data stream, it is necessary to recombine the data stream. Therefore, the processing methods vary for different network protocols [18,19].

3.7.3. Event Analyzer Module. By processing the aforementioned packet capture module, the original network packets were obtained, but data mining cannot be directly used to apply them in intrusion detection systems. Before application, the initialization data packet must be processed according to data mining methods to extract main features. After data preprocessing, it must be checked according to the classifier of the data structure. The event analysis module mainly includes two processing stages, the training stage and the actual testing simulation stage. During the training phase, a pre selected combination of data is used, followed by a predetermined plan. After extracting the main features and preprocessing the data, the best classification results can be obtained through training; After capturing the basic data packets in the actual running network during the testing simulation phase, the same feature extraction and data preprocessing operations will be performed on them, and then the classifier obtained during the training phase will be used to monitor the system for intrusion behavior [20].

Table 3.1: Test Environment Configuration Information

Project	Configuration information
Database CPU	Intel(R)Core(TM)i7 @2.26 GHz 2.26 GHz
Business processing server memory	DDR38 G
Monitoring center bandwidth	1000 M/s
Switch model	Cisco 3350
storage	Flash 3G

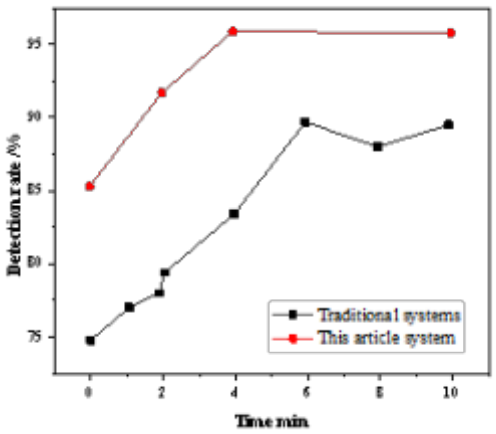


Fig. 4.1: Comparison of experimental results

3.8. Experimental Preparation. In order to more clearly and specifically demonstrate the practical application effect of the campus network intrusion detection system designed by the author based on frequency hopping data mining, a comparison is made with traditional campus network intrusion detection systems to compare their detection rate capabilities. In order to ensure the accuracy of the experiment, two campus network intrusion detection system designs were placed in the same testing environment for detection rate capability experiments. The test environment configuration is shown in Table 3.1.

4. Results and Discussion. During the experiment, two different campus network intrusion detection systems were designed to work simultaneously in the same environment, and the changes in detection rate and capability were analyzed. The comparison of experimental results is shown in Figure 4.1.

The experimental results show that the campus network intrusion detection system designed by the author based on frequency hopping data mining is far superior to traditional system design in detecting intrusion behavior, approaching 100% infinitely, and has high effectiveness.

5. Conclusion. The author advocates for developing and deploying a campus network intrusion detection system that leverages both data mining and image processing techniques. The rapid development of the network not only enriches people's lives, but also brings many network security issues. As an important part of network protection, network intrusion detection systems play a very important role. As the network scale grows and the variety of network attacks increases, conventional intrusion detection systems are struggling to effectively adapt to the evolving environment. The author first analyzed the classification and models of intrusion detection, as well as commonly used data mining algorithms. By integrating data mining techniques into intrusion detection, the author conducted comprehensive data analysis and successfully developed a detection system. This system's accuracy and detection speed were evaluated against other existing systems, determining its

practical significance. At the same time, considering the multitude of intrusion behaviors, the future focus of intrusion detection technologies lies in enhancing application layer intrusion detection and implementing adaptive intrusion detection mechanisms.

REFERENCES

[1] Ekpotu, W. F. , Akintola, J. , Obialor, M. C. , & Philemon, U. . (2023). Historical review of hydrogen energy storage technology. World Engineering and Technology (in English), 11(3), 454-475.
[2] Kim, E. , An, J. , Cho, H. C. , Cho, S. , & Lee, B. . (2023). A sensor data mining process for identifying root causes associated with low yield in semiconductor manufacturing. Data Technologies and Applications, 57(3), 397-417.
[3] Qu, Y. , Samarati, P. , Benslimane, A. , & Yu, S. . (2022). Call for papers special issue on privacy-preserving data mining for artificial intelligence of things. Big Data Mining and Analysis, 5(1), 1.
[4] Zhi-wenZHANG, Tian-geLIU, & Peng-juNIE. (2022). Real-time semantic segmentation for road scene based on data enhancement and dual-path fusion network. Acta Electronica Sinica, 50(07), 1609-1620.
[5] Saesi, N. , & Taleghani, M. . (2023). Linking competitors' knowledge and developing innovative products using data mining techniques. Computer and Communication (English), 11(7), 37-57.
[6] Xue-songWANG, Han-linZHANG, & Yu-huCHENG. (2022). Autoencoder and hypergraph-based semi-supervised broad learning system. Acta Electronica Sinica, 50(03), 533-539.
[7] Li, Q. Q. , Wang, G. K. , Liang, Z. X. , & Hu, Z. J. . (2022). Highly transparent and adhesive poly(vinylidene difluoride) films for self-powered piezoelectric touch sensors. Chinese Journal of Polymer Science, 40(7), 726-737.
[8] Kumar, A. , Abhishek, K. , Ghalib, M. R. , Shankar, A. , & Cheng, X. . (2022). Intrusion detection and prevention system for an iot environment. Digital Communication and Networking: English Version, 8(4), 540-551.
[9] Al-Kahtani, M. S. , Mehmood, Z. , Sadad, T. , Zada, I. , Ali, G. , & Elaffendi, M. . (2023). Intrusion detection in the internet of things using fusion of gru-lstm deep learning model. Intelligent Automation and Soft Computing, 37(8), 2279-2290.
[10] Cui, L. . (2022). Information security management measures for college archives under the network environment. Electronic Research and Applications, 6(6), 15-19.
[11] Attou, H. , Guezzaz, A. , Benkirane, S. , Azrour, M. , & Farhaoui, Y. . (2023). Cloud-based intrusion detection approach using machine learning techniques. Big Data Mining and Analytics, 6(3), 311-320.
[12] Brahma, A. , Panigrahi, S. , Samal, N. , & Gountia, D. . (2022). Rule-based database intrusion detection using coactive artificial neuro-fuzzy inference system and genetic algorithm. International journal of business intelligence and data mining, 6(1), 32-43.
[13] Attou, H. , Guezzaz, A. , Benkirane, S. , Azrour, M. , & Farhaoui, Y. . (2023). Cloud-based intrusion detection approach using machine learning techniques. Big Data Mining and Analytics, 6(3), 311-320.
[14] Wang, W. , Li, J. , Zhao, N. , & Liu, M. . (2023). Mem-tet: improved triplet network for intrusion detection system. Computer, material, and continuum (English) (7), 471-487.
[15] Akanni, A. , Akanni, W. , & Daso, O. H. . (2023). Captcha-based honey net model against malicious codes. Computer and Communication (English), 11(3), 159-166.
[16] Ding, S. , Kou, L. , & Wu, T. . (2022). A gan-based intrusion detection model for 5g enabled future metaverse. Mobile Networks and Applications, 27(6), 2596-2610.
[17] Alsemairi, S. S. . (2023). Factors influencing employees on compliance with cybersecurity policies and their implications for protection of information and technology assets in saudi arabia. Intelligent Information Management, 15(4), 259-283.
[18] Wang, C. , Chen, R. L. , & Gu, L. . (2023). Improving performance of virtual machine covert timing channel through optimized run-length encoding. Journal of Computer Science and Technology, 38(4), 793-806.
[19] Nagamunthala, M. , & Manjula, R. . (2023). Implementation of a hybrid triple-data encryption standard and blowfish algorithms for enhancing image security in cloud environment. Computer and Communication (English), 11(10), 135-149.
[20] Gihonia, S. A. , Mabela, R. M. , René Gilles Bokolo, Kimba, E. , Katshitshi, M. , & Kalombo, M. , et al. (2022). Intrusion detection system with remote signalling for vehicles using an arduino controller and radio-frequency technology. Software Engineering and Applications (English), 15(4), 14.
[21] Sezgin, A. , & Boyac, A. . (2023). Aid4i:an intrusion detection framework for industrial internet of things using automated machine learning. Computers, Materials, and Continuum (in English), 76(8), 2121-2143.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 17, 2024
Accepted: Jun 25, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 699–707, DOI 10.12694/scpe.v26i2.3976

INFORMATION MONITORING OF TRANSMISSION LINE OPERATING ENVIRONMENT BASED ON INTERNET OF THINGS TECHNOLOGY

SIJIA ZHENG,* WEI DU,† JIAYING WANG,‡ GUOZHU YANG,§ AND YAJIE ZHAO¶

Abstract. In order to solve the problem of real-time monitoring of transmission line status in smart grid construction, the author proposes an information-based monitoring method for the operating environment of transmission lines based on Internet of Things technology. This approach implements a monitoring system for transmission line status that leverages Internet of Things (IoT) technology, Passive Optical Network (PON) technology, and fuzzy expert diagnosis techniques. Extensive experimentation and field trials have yielded promising outcomes, facilitating the real-time monitoring of transmission line statuses and precise diagnostic assessments. The experimental results indicate that this system has shortened inspection time, reduced costs, and improved inspection accuracy. The accuracy can reach 99.9%, with high reliability. The accuracy and feasibility of the system have been verified, and it has the characteristics of comprehensive state information collection, stable network, high output accuracy, and simple deployment.

Key words: PON technology, Internet of Things, Fuzzy Analytic Hierarchy Process, Smart grid, Monitoring of transmission line status

1. Introduction. The Internet of things means the Internet connecting things. The Internet of Things technology enables objects to be intelligent and realize information communication and interaction between people and things, things and things by installing electronic tags (RFID), sensors, two-dimensional codes and other sensing devices on various objects, and then sharing the dynamic and static information of objects acquired by various sensing devices on the Internet [1-2]. The Internet of Things is not a new technology. On the basis of the existing Internet system, it is necessary to standardize and increase the means of perception and information processing, and extend Internet technology to objects. The Internet of Things technology has been widely applied in various fields related to national economy and people's livelihood (such as electricity, mobile communication, petrochemicals, etc.), and the power industry is one of the more widely used fields of Internet of Things technology, for example, online monitoring of transmission lines, intelligent substations, intelligent computer rooms, etc., the power Internet of Things covers various links such as power production, transmission, consumption, and management. Its industrial chain covers chips, sensors, terminal devices, information and communication networks, software applications, value-added services, and operation and maintenance. It is necessary to rely on the needs of power business to develop practical, serialized, and low-cost power Internet of Things products and systems, expand the industry application scope of Internet of Things products, improve product market share, and actively explore international markets [3].

Transmission lines are the power links with the largest and most widely distributed assets in the power grid, complex and harsh equipment operating environments, and frequent external factors. Due to the safety issues of transmission line equipment, such as high operating temperature of conductors, sag changes, wind bias discharge, gentle wind vibration, tower tilting, etc., most of them cannot be detected by the naked eye [4]. Therefore, it is urgent to install effective sensing equipment on transmission lines, achieve timely response to abnormal transmission line equipment through fault online monitoring systems, track and judge the operating status of transmission lines in real time, provide timely warning for line faults, and ensure the safe and stable

*State Grid Electric Power Space Technology Company Limited, Beijing, 102211, China (SijiaZheng7@163.com)
†State Grid Electric Power Space Technology Company Limited, Beijing, 102211, China (Corresponding author, WeiDu32@126.com)
‡State Grid Electric Power Space Technology Company Limited, Beijing, 102211, China (JiayingWang3@163.com)
§State Grid Electric Power Space Technology Company Limited, Beijing, 102211, China (GuozhuYang8@126.com)
¶State Grid Electric Power Space Technology Company Limited, Beijing, 102211, China (YajieZhao16@163.com)

operation of the power grid, applying IoT technology to the online monitoring system of transmission lines, obtaining equipment parameters of transmission lines, including wire tension, line temperature, grounding resistance, wire to ground distance, wire dancing frequency, pollution, tower pole stress, insulator wind deviation, etc, the parameters of various devices are transmitted to the data processing center through the Internet of Things gateway, and then advanced data analysis technology is used to process the data of various devices, determine whether there is a fault in the transmission line, and provide a basis for power grid safety decision-making [5-6]. With the increasing demand, scope, and types in the power supply market, the construction of supporting facilities for power transmission quality, power supply safety, and real-time monitoring of power lines is becoming increasingly strict. It is urgent to combine Internet of Things technology to build a multi element micro automatic meteorological station suitable for monitoring power lines and cables, to achieve microclimate observation, mobile meteorological observation posts, the seasonal ecological monitoring of power transmission line operational environments involves gathering meteorological data like temperature, humidity, air pressure, wind velocity, and precipitation (rain and snow). This data collection aims to enable the monitoring and early warning of meteorological conditions affecting power transmission line operations, timely formulate and implement preventive measures, and reduce the potential impact of meteorological factors on power transmission lines [7,8].

2. Literature Review. The Internet of Things (IoT) refers to a network that links various objects to the internet using information-sensing devices like RFID, GPS, infrared sensors, and laser scanners. These devices operate according to established protocols, facilitating intelligent functions such as identification, positioning, tracking, monitoring, and management of objects. The emergence of IoT technology as a next-generation information and communication technology has garnered considerable attention. Liu, L. et al. developed an online monitoring system for power fiber optic transmission networks utilizing IoT technology. The hardware components include sensors and microcontrollers, while the software aspects entail analyzing the system's functional requirements. They established a data acquisition module for power fiber optic transmission networks based on IoT, successfully implementing the system's software functions. Comparative experiments have verified the system's effectiveness in online monitoring, demonstrating significant potential for wider adoption[9]. Shen, Y. et al. employed wireless communication network technology for environmental monitoring in animal farms, devising a real-time reporting system aimed at continuously monitoring the health of animals[10]. Xu, X. et al. investigated remote monitoring of network public opinion within IoT applications. They focused on understanding campus public opinion dynamics and influencing its dissemination and direction. This exploration and research into online public opinion monitoring align with the objectives of smart campuses, and its theoretical advancement continues to evolve[11]. Innovatively integrating IoT technology, researchers have devised an image encryption transmission method tailored for power monitoring, capitalizing on the attributes of surveillance videos from stationary cameras. Experimental findings underscore the system's prowess across power monitoring image processing, encryption, and transmission, outperforming conventional algorithms by a significant margin [12].

The principal component analysis method is one of the widely used methods based on multi-dimensional statistical variable online monitoring technology. It is combined with Internet of Things technology to apply it to the online monitoring system of transmission lines. The basic idea is to first use Internet of Things technology to construct an information model of the normal operation status of the line, including monitoring statistical variables and thresholds, then, real-time collection of various parameters of the operating equipment of the transmission line is carried out, and they are projected onto the IoT information model. The statistical variable values are monitored at the online calculation point, and the presence of faults in the transmission line equipment is determined by comparing the online calculation values and threshold values. The author applied the combination of Internet of Things technology and LiDAR measurement technology to the online monitoring system of transmission lines, and achieved good results through practical application, proving the effectiveness of the method proposed by the author.

3. Method.

3.1. Internet of Things Technology. The system for monitoring the status of transmission lines, which combines IoT technology, mainly includes the application layer, transmission layer, and perception layer. In the

transmission circuit section, including towers and transmission lines, intelligent sensors are equipped. These intelligent sensors collect equipment information on the line and micro meteorological information of the surrounding environment, which includes: Data on line icing, insulator contamination, line dancing, temperature, humidity, and more are gathered and sent to a monitoring agent via short-range wireless communication. Subsequently, this data is relayed to the Optical Line Terminal (OLT) at the substation through optical fiber, then further transmitted to a remote control center via a dedicated power network. Utilizing an expert diagnostic system, the current line status is accurately assessed, and results are disseminated to relevant departments for prompt action to prevent accidents [13,14].

3.2. Perception layer. Wireless sensor networks consist of numerous small nodes equipped with sensing, computing, and wireless communication capabilities, primarily designed for environmental monitoring rather than communication. These sensor nodes are strategically positioned in the monitored area to gather specific environmental data, which is then transmitted to aggregation nodes for further analysis[15]. The benefits of wireless sensor networks are manifold: they enable the acquisition of long-term, high-resolution environmental data at close range, a feat unattainable by conventional monitoring equipment; the computing and storage capacities of sensor nodes empower them to execute tasks like data filtering, compression, and application-specific processing; inter-node communication facilitates collaborative completion of intricate tasks such as target tracking; and finally, task reassignment capability allows for flexibility in altering the network's purpose as needed. The problems that wireless sensor networks need to solve include self-organization and self configuration of the network; Communication protocols; Distributed data management (data collection, storage, querying, retrieval, etc.); Specific data fusion processing for various applications; Energy conservation should be integrated into all designs. When designing the system, the main considerations include interference, power supply, transmission distance, and other issues. Intelligent wireless sensors will be deployed along transmission lines and towers to monitor both the condition of the lines and the surrounding environment. These sensors will collect data and transmit it to a central monitoring device, which acts as an agent overseeing the system's status. Communication between the sensor devices and the monitoring agent occurs via the ZigBee protocol.

3.3. Network layer. The network layer serves as a conduit for transferring data collected by the perception layer to the application layer. Traditional data transmission systems rely on wireless public networks like GPRS and 3G, but their speed and bandwidth limitations hinder efficient transmission of large data volumes in smart grids. However, with the introduction of OPGW (overhead ground wire composite optical cable) in transmission lines, optical fiber communication has become feasible. PON technology, particularly EPON, emerges as a mature and practical solution for optical communication. Within this system, EPON technology is employed to relay real-time monitoring data. PON comprises three main components: OLT, ODN, and ONU. OLT equipment is situated in the substation, while ODN and ONU are positioned on the line towers. In this setup, the ONU acts as a state monitoring agent device, responsible for collecting and aggregating data from intelligent sensors. Communication between the status monitoring agent device and the intelligent sensors is wireless, with options including ZigBee or WiFi. At the same time, considering the on-site environment and the difficulty of obtaining power, it is required that the status monitoring agent device must be able to adapt to outdoor work with low power consumption, generally requiring a power consumption of less than 5W. The power supply method should be solar panel power supply, and there should also be battery energy storage to prevent prolonged cloudy and rainy weather. The data communication architecture is shown in Figure 3.1.

3.4. Application layer. At the application layer, a fuzzy expert diagnostic system is employed. Developing expert systems entails integrating extensive expert knowledge and feature vectors. Transmission line status information encompasses four key aspects: fundamental details, operational management data, disaster alerts, and environmental monitoring data. This data is fed into the fuzzy expert system for analysis and decision-making[16]. Steps for establishing a fuzzy expert system:

1. Build a multi-level structural model as needed.
2. Create a fuzzy consistent judgment matrix to assess the relative importance of elements and derive a set of weights. This matrix represents pairwise comparisons among elements to gauge their significance

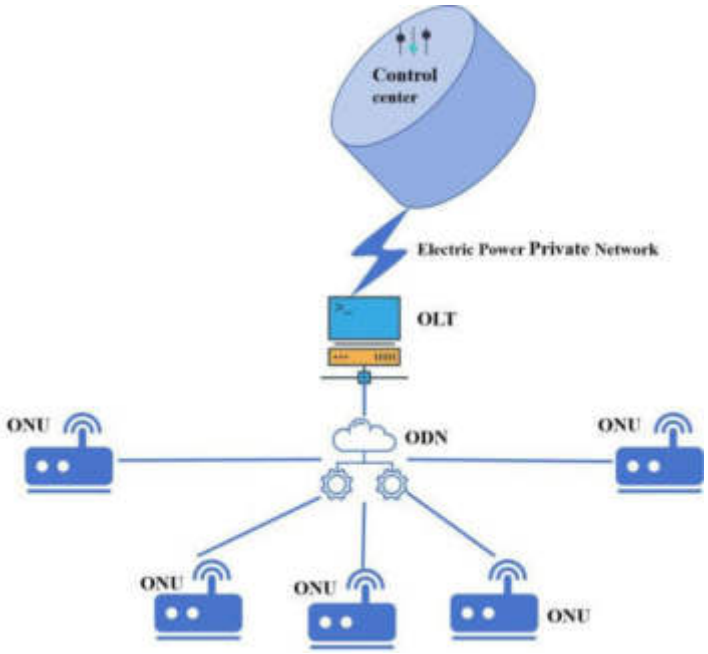


Fig. 3.1: Communication Network Model

$a_1, a_2, \cdots, a_n$ be:

$$F = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix} \tag{3.1}$$

In the formula, f_{ij} represents the membership degree of the i-th element in the layer that has a fuzzy relationship with the j-th element. The larger f_{ij}, the more important a_i is than a_j. When f_{ij}=0.5, it indicates that a_i is equally important as a_j.

3. Sum fuzzy matrices by row:

$$W_i = \sum_{h=1}^{n} f_{ij}, i = 1, 2, \cdots, n \tag{3.2}$$

4. Perform the following mathematical transformation on W:

$$W_{ij} = \frac{w_i - w_j}{2n} + 0.5 \tag{3.3}$$

Obtain the weight matrix:

$$F = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix} \tag{3.4}$$

5. The weight set $A = (A_1, A_2, \cdots, A_n)$ of the elements in the fuzzy consistency judgment matrix W represents the weight allocation of each factor. The sum of the elements in each row of the fuzzy

consistency matrix (excluding the diagonal element) indicates:

$$L_i = \sum_{j=1}^{n} W_{ij} - 0.5, i = 1, 2, \cdots, n \tag{3.5}$$

The total without diagonal elements is:

$$\sum_{j=1}^{n} L_i = \frac{n(n-1)}{2} \tag{3.6}$$

In the formula, n is the order of the matrix.
Since L_i represents the importance of indicator i relative to the upper level objectives, normalizing L_i can obtain the weights of each indicator:

$$A_i = \frac{L_i}{\sum_{j=1}^{n} L_i} = \frac{2L_i}{n(n-1)} \tag{3.7}$$

6. Determine the set of solutions $M = (M_1, M_2, \cdots, M_n)$, where n is the number of solutions. Mapping each element in the structure to M can determine a fuzzy relationship R.

$$F = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \tag{3.8}$$

In the formula, r_{ij} represents the importance of a single factor i in scheme j.

3.5. System functions and components. The transmission line operation environment monitoring system based on LiDAR measurement technology mainly consists of a spatial data management display module, inspection data processing module, and terrain and landform change analysis module (see Figure 2). Among them, the spatial data management display module realizes the management display of ledger data, inspection data (video images, point cloud data, etc.), and basic spatial data. The inspection data processing module integrates the independently developed IMOS cloud point cloud processing software, which is used to process laser point cloud data collected during inspections, obtain basic data for terrain and landform change analysis, and display terrain image maps. The analysis module of terrain and landform changes is the core part of the system function, which is used to analyze point cloud data and obtain the impact of environmental changes on transmission lines [17].

3.6. Working principle and process. The transmission line operation environment monitoring system based on LiDAR measurement technology is supported by high-precision point cloud data and high-definition image data collected. The raw data is filtered, classified, and extracted using the IMOS-cloud point cloud processing software to obtain result data. Combined with POS (Point of Sale, Flight Position, Sensor Attitude Information) data and image data, DEM (Digital Elevation Model), DOM (Document Object Model) and other result data are generated. The processing and analysis process is shown in Figure 3.3. The main steps are introduced below [18].

3.7. Calculation of distance between ground features and conductors. Process the generated laser point cloud data using the IMOS-cloud point cloud processing software and store it as a las file based on single span partitioning. The calculation process for the distance between objects and wires is as follows:

1. Determine the distance between buildings such as houses and trees within the passage and each point of the conductor (that is obtain a rectangular range of specified distances on both sides of the tower).
2. Classify and extract ground points and feature points (crossing objects, buildings, vegetation) from point cloud data for calculation.

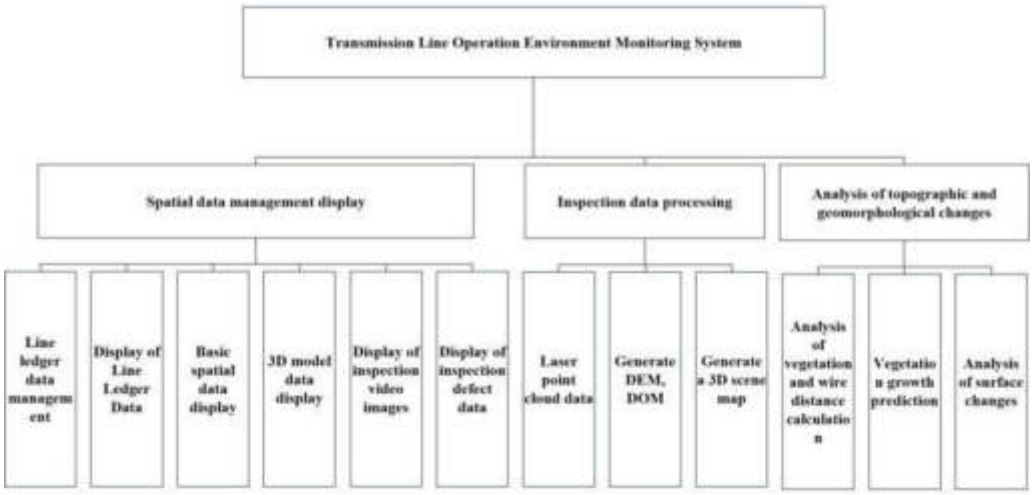


Fig. 3.2: Composition of Transmission Line Operation Environment Monitoring System

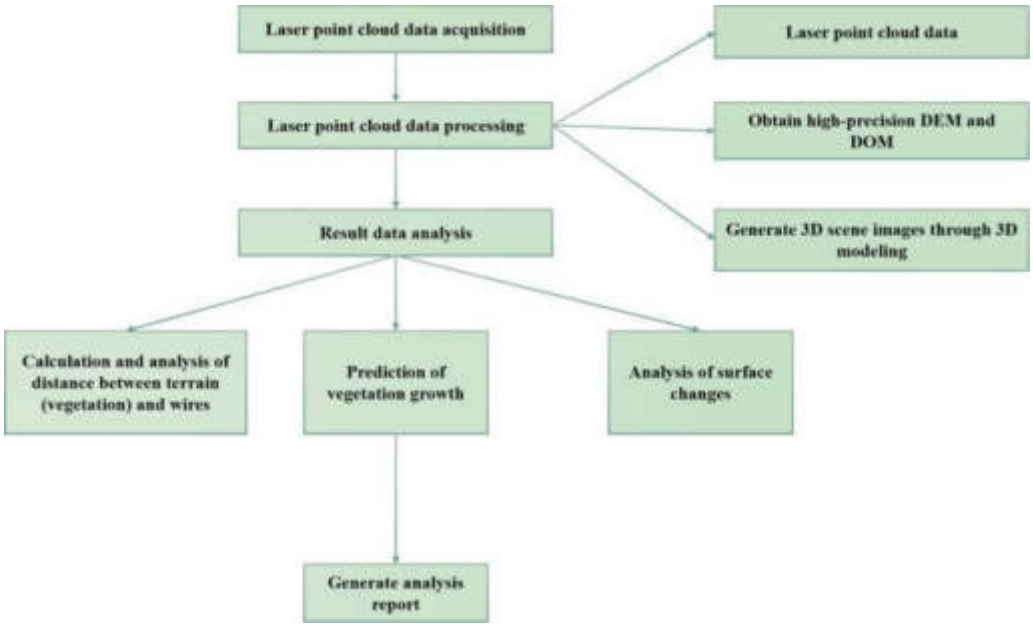


Fig. 3.3: Processing and Analysis Process of Transmission Line Operation Environment Monitoring System

3. Construct irregular triangular meshes on the classified point sets using the Crust algorithm.
4. Obtain the highest point in each triangular grid, and then establish a horizontal circular area with the highest point as the center and the set safe distance as the radius. Filter out the power line points that are horizontally projected within this circular area, and calculate the distance between each point and the highest point. If it is less than the safe distance, it indicates a safety hazard, and record the spatial information of the hazard points.

Table 4.1: Grid data corresponding to the first collection of DEM

Number	Longitude x/(°)	Latitude y/(°)	Altitude z/m
1	109.902078	40.506983	981.30
2	109.902078	40.507008	981.22
3	109.902006	40.507034	981.15
4	109.902016	40.507061	981.58
5	109.902015	40.507087	981.40
6	109.902025	40.507113	981.37

3.8. Prediction of vegetation growth curve. Due to the gradual growth of vegetation over time, tree flash faults can occur when the distance between vegetation and conductors within the transmission line corridor is less than the safe distance, affecting the safe and stable operation of the transmission line. The prediction of vegetation growth curve only considers the relationship between tree height and time; The grey algebraic curve model proposed on the basis of the grey model is suitable for predicting linear and nonlinear processes with low growth rates. Therefore, this system introduces this model to predict the growth of trees. Due to the fact that the vegetation in the transmission line corridor is mainly dominated by poplar trees, the tree growth model is obtained based on the height values of poplar trees at different age stages when establishing the model; Then, the tree height value obtained from each inspection record is fed into the prediction model to obtain the height prediction value for a certain period of time in the future, calculate its distance from the wire, generate a vegetation growth change curve, and provide warning information [19].

3.9. Analysis of topographical changes. Terrain change detection is the process of comparing information data from different periods in the same area to identify differences and understand the changes in terrain. DEM is a solid ground model that represents ground elevation in the form of an ordered numerical array. The DEM structure of a regular grid is simple and computationally concise, allowing for quick comparative calculations.

1. Firstly, the data space registration of different phase DEMs is used to unify them into the same coordinate system. Then, stacking and comparison are performed, and the elevation difference is calculated. After removing errors, the areas where the terrain has changed can be identified. The area of the changed areas is calculated, and the spatial information of the changed areas is recorded.
2. After image classification, the comparative method is used to classify image data from different periods. Pixels are compared one by one to determine the position of changes, and the trend of changes in the horizontal direction is found to obtain three-dimensional terrain change information [20].
3. Calculate the inclination and base height of nearby towers using point cloud data, and compare them with existing tower information to determine the impact of terrain changes on the towers. If the change area is located below the transmission line, it is necessary to calculate whether the net clearance distance of the transmission wires in the change area is within the safe range to determine whether it will affect the wires.

4. Results and Discussion. Convert the DEM data obtained from the original data collected during inspection and the corresponding data from the same area to rectangular grid data, as shown in Tables 4.1 and 4.2. By overlaying and subtracting DEM, the location of subsidence or protrusion on the ground was determined. After profile analysis (as shown in Figure 4.1), it was found that the terrain had protrusions, resulting in a distance less than the safe distance from the conductor.

During the patrol inspection, 13 trees endangering the superelevation safety distance of the line, 6 hidden dangers of the passage, 3 desert terrain changes, 1 river crossing width and 1 mudflat landform were found; Simultaneously calculating the settlement of nearby tower foundations, tower inclination, and predicting the growth cycle of trees, as well as other terrain change parameters, can provide early warning of their impact, shorten inspection time, reduce costs, and improve inspection accuracy. The accuracy can reach 99.9%, with high reliability.

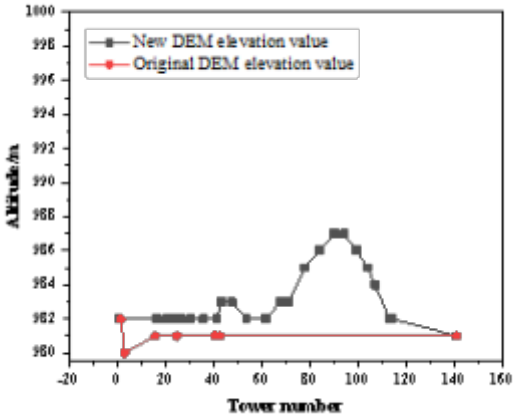


Fig. 4.1: Terrain Change Profile

Table 4.2: Grid data corresponding to the second collection of DEM

Number	Longitude x/(°)	Latitude y/(°)	Altitude z/m
1	109.902086	40.506994	981.32
2	109.902096	40.507016	981.24
3	109.902105	40.507044	981.15
4	109.902113	40.507077	981.56
5	109.902123	40.507096	981.40
6	109.902135	40.507128	981.37

5. Conclusion. The author proposes an information-based monitoring of the operating environment of transmission lines based on Internet of Things technology. Online monitoring is an important means to ensure the safe operation of the power grid and improve its transmission capacity. Due to the fact that most equipment safety issues on transmission lines cannot be detected with the naked eye, there is an urgent need to build an online fault monitoring system. The Internet of Things technology involves deploying a large number of sensing devices (such as smart sensors, electronic tags, RFID modules, etc.) to include all key equipment parameters of transmission lines in the monitoring scope. By establishing a unified information model at the sensing layer, it is compatible with data sensing formats from different manufacturers, after data preprocessing and storage by the aggregation controller, the data is transmitted to the power grid through a unified communication protocol. At the application layer, the principal component analysis method is used to analyze the real-time monitoring data. By setting appropriate alarm thresholds, it is determined whether there is a fault in the transmission line. The author's research did not involve the study of a unified information model and unified communication protocol for the Internet of Things. The next step of research will mainly focus on obtaining a unified information model suitable for parameter perception of transmission line equipment.

Acknowledgement. Management Technology Project of State Grid Power Space Technology Co., LTD. : Deepening Application Research of 3D System based on laser scanning Data (5100/2022-44005B)

REFERENCES

[1] Xu, Y., Wang, Z., Gao, H., Jiang, Z., Yin, Y., & Li, R. (2023). Towards machine-learning-driven effective mashup recommendations from big data in mobile networks and the internet-of-things. Digital Communication and Networking: English Version, 9(1), 138-145.

[2] Wang, X., Bao, X., Huang, Y., Zheng, Z., & Fei, Z. (2023). On optimization of cooperative mimo for underlaid secrecy industrial internet of things. Frontiers of Information Technology & Electronic Engineering, 24(2), 259-274.
[3] Nie, Z. X., Long, Y. Z., Zhang, S. L., & Lu, Y. M. (2022). A controllable privacy data transmission mechanism for internet of things system based on blockchain:. International Journal of Distributed Sensor Networks, 18(3), 303-315.
[4] Chen, X., Tang, X., & Xu, X. (2022). Digital technology-driven smart society governance mechanism and practice exploration. Frontiers of Engineering Management, 10(2), 319-338.
[5] Daouda KonaneWend Yam Serge Boris OuedraogoToussaint Tilado GuinganeAbdoulaye ZongoZacharie KoalagaFranois Zougmoré. (2022). An exact solution of telegraph equations for voltage monitoring of electrical transmission line. Energy and Power Engineering, 14(11), 669-679.
[6] Zhen-tao ZHONG, Hong-hui ZHANG, Geng LIU, & Wei-ling LUO. (2022). Research on model system of monitoring-evaluation-warning for implementation supervision of territory spatial planning. Journal of Natural Resources, 37(11), 2946-2960.
[7] Tachibana, Y., & Okano, K. (2022). Regulatory developments in the uk broadband market: focusing on competition/investment, organizational structure and co-regulation. Journal of Information and Communications Policy, 5(2), 95-116.
[8] Khan, T., Withers, C., Martin, E., & Bonilla, N. (2022). Approaches for effective negative pressure isolation space control to minimize airborne transmission of contaminants in residential homes:. Indoor and Built Environment, 31(5), 1405-1417.
[9] Liu, L., & Wang, L. (2022). On Line Monitoring System of Power Optical Fiber Transmission Network Under Internet of Things Technology. International Conference on Multimedia Technology and Enhanced Learning. Springer, Cham, 30(5), 72-92.
[10] Shen, Y. (2022). Information monitoring of animal husbandry industry based on the internet of things and wireless communication system. Computational and Mathematical Methods in Medicine, 2022, 18(2), 288-305.
[11] Xu, X., & Sun, B. (2022). The dissemination and evaluation of campus ideological and political public opinion based on internet of things monitoring. Computational intelligence and neuroscience, 3(3), 282-293.
[12] Nan, Y., Zhen, W., Jun, Z., Xueqiong, Z., & Hai, X. (2023). Unsupervised model-driven neural network based image denoising for transmission line monitoring. Optoelectronic Express: English version, 19(4), 248-251.
[13] Ferreira, C. C., & Lind, F. (2023). Supplier interfaces in digital transformation: an exploratory case study of a manufacturing firm and iot suppliers. Journal of Business & Industrial Marketing, 38(6), 1332-1344.
[14] Wei, J., Hua, Z. D., Shuangbao, M., Gaocheng, Y., & Wei, C. (2022). Dynamic walking characteristics and control of four-wheel mobile robot on ultra-high voltage multi-split transmission line:. Transactions of the Institute of Measurement and Control, 44(6), 1309-1322.
[15] Oshikawa, K., Matsushima, T., Kuwabara, N., Wakisaka, T., & Fukumoto, Y. (2022). Communication distance estimation for power line communication using various differently constituted power cables. IEEJ Transactions on Electrical and Electronic Engineering, 17(4), 498-505.
[16] Ren, Q., Sun, Y., Li, S., Wang, B., & Yu, Z. (2023). Energy-efficient data collection over underwater mi-assisted acoustic cooperative mimo wsns. China Communications: English version, 20(11), 96-110.
[17] Luo, H., Wang, X., Xu, Z., Liu, C., & Pan, J. S. (2022). A software-defined multi-modal wireless sensor network for ocean monitoring:. International Journal of Distributed Sensor Networks, 18(1), 28-4.
[18] Lu, Z., & Tan, X. (2022). 3d sensor network location spatial positioning technology based on machine learning. International Journal of Emerging Electric Power Systems, 24(1), 13-23.
[19] Zhang, X., Liu, Y., & Qu, T. T. P. (2022). Research on remote online firmware upgrade system for embedded devices. Journal of Internet Technology, 23(7), 1587-1596.
[20] Zai, W. (2023). New-energy vehicle transmission system optimization and design. Electronic Research and Applications, 7(3), 20-26.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 18, 2024
Accepted: Jun 25, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 708–718, DOI 10.12694/scpe.v26i2.3978

APPLICATION OF INTELLIGENT ROBOT IN THE INTELLIGENT RECOMMENDATION OF MOBILE APPLICATION CONTENT

PING ZHOU*, BOWEI DUAN†, FEILONG LIU‡, XIN JING§, YAN LEI¶, AND XIJING LI‖

Abstract. In order to bring a new perspective of human-computer interaction to TV program hosts, increase program highlights, and promote the integration and development of new technologies and new media, the authors propose the application of intelligent robots in mobile application content intelligent recommendation. Designed and implemented an artificial intelligence virtual assisted hosting robot system, including overall system design, knowledge base design, and robot interaction control system implementation. A "smart investment advisory model" has been established based on multidimensional calculations and analysis of project areas, investment amounts, and on-site attention. The results showed that the final system did not recommend the 6th and 7th investors with the highest matching degree in the field, nor did it recommend the 8th investor with the highest matching degree in investment amount, nor did it recommend the 4th investor with the highest matching degree in communication. Instead, it selected the 1st investor by combining three factors, the entrepreneurs of this period ultimately followed Little A's advice and chose investor 1 as their entrepreneurial investor. In 46 investment recommendation applications, there were 6 times when Mr. A's investor selection suggestions were not the highest bidding investors on the program site. In these 6 times, 5 entrepreneurs followed Mr. A's suggestions, and his analysis received recognition from different entrepreneurial mentors multiple times, demonstrating the rationality of Mr. A's recommendations, with a recommendation adoption rate of 97.72% (45/46). The system has achieved intelligent data analysis and recommended reasonable investors for entrepreneurs, with a high recommendation adoption rate. The comprehensive application of technologies such as speech recognition, facial recognition, and natural language processing has achieved interactive functions for program hosting, ensuring consistency and fluency in program hosting. The practical application effect is good, promoting the application of artificial intelligence robots in television media.

Key words: Artificial intelligence, Hosting robots, Natural language processing, Recommendation algorithm, Q&A system

1. Introduction. With the development of technologies such as artificial intelligence, big data, the Internet of Things, and 5G, "everything is a medium" is no longer just a beautiful vision. Intelligent technology has already integrated into every aspect of people's lives and profoundly changed and shaped the news and communication industry as a core driving force, giving birth to real-time responsive media services. A truly user centered and interconnected ecosystem is emerging, continuously providing dynamic and infinitely close to media scenarios [1]. The popularity and development momentum of smart media is strong. According to a report by International Data Corporation (IDC), the global smart voice assistant market is expected to reach a scale of 7.7 billion US dollars by 2025. The Global Data Journalism Survey shows that data news is showing a rapid growth trend globally, with over 70% of surveyed news organizations reporting that they have started using data news for reporting in the past five years [2]. This data indicates that data news, as one of the important directions for the development of intelligent media, has gradually become the mainstream form of news reporting.

In the perspective of intelligent media, traditional news production faces a series of problems and challenges: Firstly, information overload. In the Internet era, information sources are extremely rich, and traditional news may be difficult to meet the public's demand for high-quality news content when screening and integrating a large amount of information. Secondly, the issue of timeliness [3]. The production and dissemination speed of traditional news is relatively slow, making it difficult to meet the modern people's pursuit of real-time information. With the rise of emerging platforms such as social media, traditional news faces serious challenges

*Aostar Information Technologies Co.,Ltd.,Sichuan,610041,China. (Corresponding author, PingZhou7@163.com)
†Aostar Information Technologies Co.,Ltd.,Sichuan,610041,China. (BoweiDuan@126.com)
‡Aostar Information Technologies Co.,Ltd.,Sichuan,610041,China. (FeilongLiu9@163.com)
§Aostar Information Technologies Co.,Ltd.,Sichuan,610041,China. (XinJing2@126.com)
¶Aostar Information Technologies Co.,Ltd.,Sichuan,610041,China. (YanLei982@163.com)
‖Aostar Information Technologies Co.,Ltd.,Sichuan,610041,China. (XijingLi7@163.com)

in terms of timeliness. Thirdly, the objectivity and accuracy of news [4]. Due to the influence of human factors, political stance, commercial interests, and other factors, traditional news reporting may sometimes encounter issues such as bias and inaccuracy, which can affect the objectivity and accuracy of news. Fourthly, low audience engagement. The traditional news production process is mostly top-down, with relatively low public participation [5]. However, audiences increasingly hope to participate in the production and dissemination of news, so traditional news may not be able to meet this demand. Fifth, the presentation form is singular. The traditional form of news presentation is relatively single, mainly consisting of text, images, and videos, which is difficult to meet the needs of modern audiences for diverse and visual information [6]. Data news is an innovative form of news that has emerged from the perspective of intelligent media. Based on data elements, it provides an effective way to solve problems in traditional news production by fully utilizing interdisciplinary knowledge such as big data, artificial intelligence, and computer science. A way of conveying news information to readers through visualization and interaction.

2. Literature Review. Until now, many researchers have conducted various studies on robot content recommendation in open domain human-computer interaction systems [7]. Meng et al. proposed an intelligent recommendation method based on multi interest networks and adversarial deep learning, which applies multi-source behavioral information to multi view embedding extraction for better predictive performance. Specifically, multi view preference embedding, including self embedding, interaction aware embedding, and neighbor based embedding, is combined to build user interests at a finer granularity [8]. Zheng, F. et al. proposed a personalized education system based on hybrid intelligent recommendation. Specifically, a hybrid framework for artificial intelligence has been proposed, with a focus on providing targeted recommendations for implementing a comprehensive standard curriculum plan. This will be the main tool for creating flexible differentiated teaching plans that fully meet the individual needs and specificity of each student [9]. Zhang, X. et al. proposed a collaborative filtering recommendation algorithm that improves the user model. Firstly, the algorithm takes into account the rating differences caused by different user rating habits when expressing preferences, and adopts a decoupling normalization method to normalize the user rating data; Secondly, considering the forgetting transfer of user interests over time, a forgetting function is used to simulate the forgetting law of scores, and the weight of time forgetting is introduced into user scores to improve the accuracy of recommendations; Finally, improvements were made to the similarity calculation when calculating the nearest neighbor set [10]. Ghahramani, M. et al. proposed an intelligent method for route recommendation in an IoT waste management system under given spatial constraints. It conducted a thorough analysis based on artificial intelligence methods and compared the corresponding results [11].

The author introduces an artificial intelligence virtual assisted hosting robot, little A brother, designed and developed based on the needs of the CCTV Entrepreneurial Investment Program. Little A appeared on the TV screen as a cartoon character, able to perform corresponding robot actions while engaging in intelligent conversations. Its use in programs mainly includes two aspects: Firstly, partnering with real hosts to assist on-site hosts and increase intelligent interaction; The second is to recommend investment projects and investors reasonably through data analysis.

3. Research Methods.

3.1. Related Work.

3.1.1. Knowledge Q&A System. Knowledge question answering systems are one of the important research directions in the field of natural language processing, with the aim of enabling machines to communicate, think, and answer questions like humans. In 1950, the famous British mathematician Turing published a paper titled "Computers and Intelligence" in Mind, proposing the concept of machine intelligence and the experimental method for machine intelligence testing, the Turing Test. Afterwards, a number of question and answer systems for professional fields emerged in Western countries, such as ELIZA for mental illness and Baseball question and answer system for the American basketball league. The START Q&A system developed by the Massachusetts Institute of Technology in the 1990s was the world's first open domain Q&A system. In the 21st century, with the emergence of IBM's Watson and Apple's Siri, the application of question answering systems has experienced explosive growth, mainly including knowledge-based question answering systems, community-based question answering systems, and web-based question answering systems [12].

A question answering system based on a knowledge base, which uses the content in the knowledge base as the source of knowledge. Knowledge is mainly stored in a knowledge graph or structural document in the form of triplets "entity, relationship, entity" or "entity, attribute, attribute value". How to construct a knowledge graph and how to query a knowledge graph are two important research directions. A community-based question answering system, with question answering communities as the source of knowledge, mainly focuses on problem classification, question matching, and answer screening. The purpose of problem classification is to narrow down the search scope of questions and improve the accuracy of answer acquisition. Question matching is the process of comparing the actual question text with the question answer, which can be further divided into text similarity comparison and semantic similarity comparison, in order to find similar questions and obtain answers. Answer screening is the process of selecting the best answer from multiple sources. The web-based question and answer system uses the Internet as the source of knowledge, and uses search engines to search for answers to questions contained in web pages. The advantage is that there is no need to create and prepare knowledge data in advance, and the latest knowledge content can be obtained at any time. The disadvantage is that web search answers are not precise enough, making it difficult to ensure the effectiveness of question answering compared to knowledge bases or community question answering systems. The required research content includes problem keyword extraction, problem rewriting, reading comprehension, and named entity recognition.

The author comprehensively considers the advantages and disadvantages of question answering systems with different answer sources and constructs an auxiliary program hosting robot question answering system suitable for TV program recording scenes.

3.1.2. Recommendation System. Recommendation systems are an application research direction in the field of data mining. Traditional recommendation algorithms can be divided into five categories: Content-based recommendation algorithms, collaborative filtering based recommendation algorithms, knowledge-based recommendation algorithms, association rule-based recommendation algorithms, and hybrid recommendation algorithms [13,14].

Content based recommendation algorithm is based on the user's previous rating of the project, learning the user's preferences, and recommending content that is similar to the user's preferences to the user. The advantage is that there is no need for user historical data, and there is no cold start problem with new content. The disadvantage is that there is a cold start problem with new users, and the diversity of recommendation results is poor.

Recommendation algorithms based on collaborative filtering are divided into user collaborative filtering algorithms and project collaborative filtering algorithms. Based on the premise that similar users have similar interests in content, the former recommends the same content to similar users, while the latter recommends similar content to the same user. The advantage is that recommendations have diversity and can recommend unknown new content to users, which can be used in the multimedia field. The disadvantage is that there are problems with new users and cold start of new content, and there is a strong dependence on user rating data.

Knowledge based recommendation algorithms convert user requirements into rules and recommend corresponding content based on the rules. The advantage is that there are no new users or cold start issues with new content, and the recommendation results are interpretable. The disadvantage is that a knowledge base needs to be built, and the algorithm has poor scalability.

A recommendation algorithm based on association rules, if a user is interested in a certain content, they may be interested in other content related to that content and recommend the relevant content to the user. The advantage is that it is easy to discover new content points of interest and does not require domain knowledge. The disadvantage is that the degree of personalization is low and the extraction of association rules is complex. A hybrid recommendation algorithm is a combination of two or more of the above recommendation algorithms to generate better recommendation results [15].

3.2. Design of Virtual Hosting Robot System. The system design of the hosting robot for this program mainly includes: overall network architecture design, interaction control process design, and knowledge base system design. The system is implemented using JAVA web technology, which facilitates the updating and management of backend data as well as the calling of various functional modules.

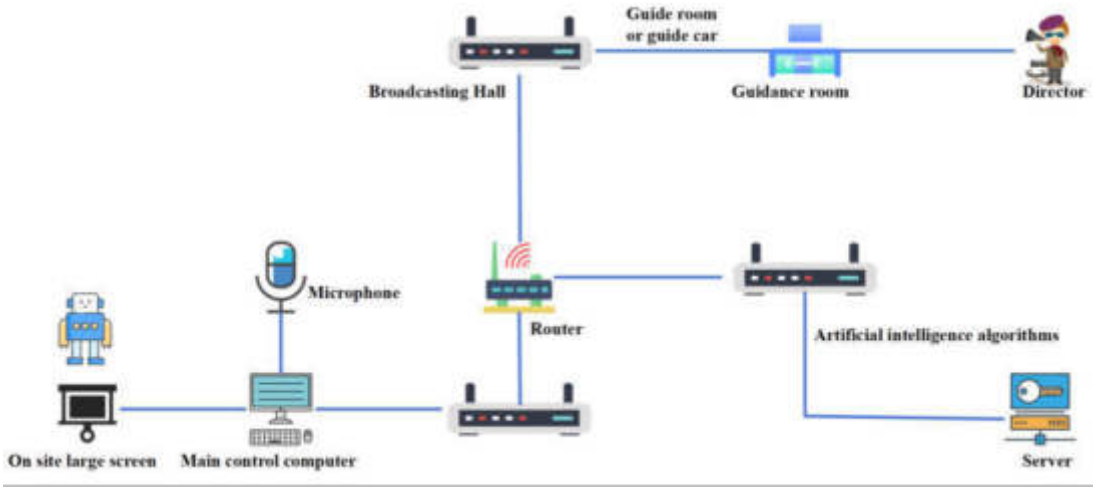


Fig. 3.1: Network Architecture of Program Hosting Robot

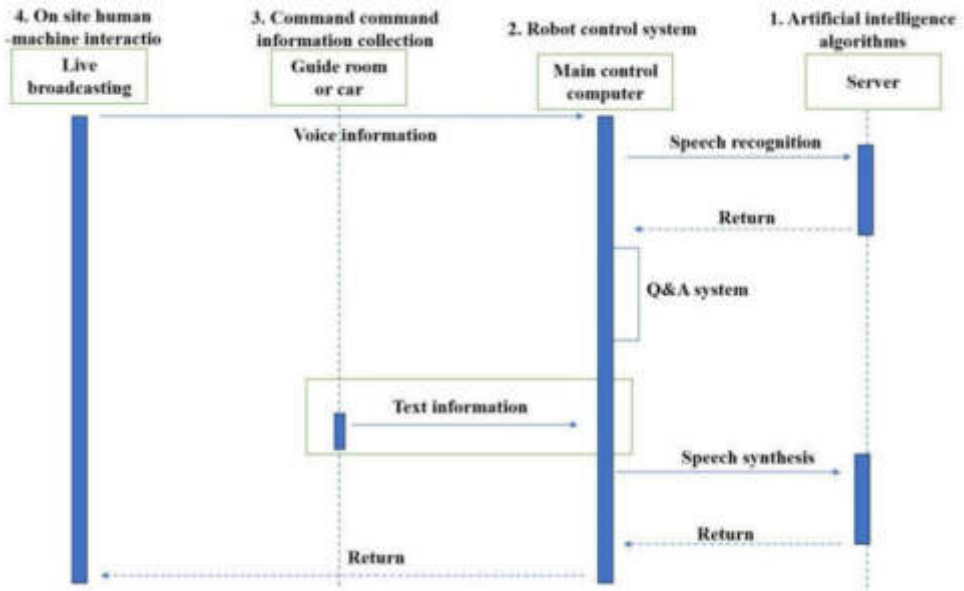


Fig. 3.2: Interaction sequence of program host robots

3.2.1. System Network Architecture Design. According to the network environment requirements of the program recording site, a network system for hosting robots has been added to the existing network system, as shown in Figure 3.1. The interaction sequence between each part is shown in Figure 3.2.

It mainly includes the following four parts:

1. Artificial intelligence algorithm server, which is installed with artificial intelligence speech recognition, facial recognition, naming recognition, knowledge base system, etc;
2. The robot control system is connected to the server and broadcasting command system respectively, and real-time data analysis is carried out based on the on-site situation of the program;
3. The command command information collection system in the guidance room issues action commands

Table 3.1: Program Hosting Robot Knowledge Base System

Serial Number	Knowledge base name	Include content examples
1	Common Corpus for Hosts	Hello everyone, hello, welcome everyone. I'm glad I'm here. I'm here. Okay, no problem, take a break. I'm sorry, your question is beyond my scope of knowledge. It's a bit difficult, won't it bother me? Thank you, no need to thank you. Goodbye, goodbye, etc
2	Entrepreneur Information Library	Entrepreneur's name, gender, age, graduation school, education, major, entrepreneurial experience, work experience, nationality, family background, overall situation of the entrepreneurial team, etc
3	Entrepreneurial Project Information Database	Project name, industry, expected investment amount, current progress of the project, intellectual property situation, existing investment situation of the project, market situation, project development plan, etc
4	Investor information	Investor's name, gender, age, education, major, previous investment experience, development status of invested projects, field of invested projects, etc
5	Investment learning case library	Establish a case library of past investment projects, analyze investment situations, and lead the robot system to conduct self-learning. If necessary, automatically adjust the corresponding weight coefficients to improve and enhance the reliability of recommendation algorithms.
6	Robot expression and action library	Smile, cross your waist, nod, applaud, somersault, forward, backward, spin, kick, punch, etc

to the robot based on the guidance command information;

4. The on-site human-computer interaction system obtains real-time information based on the dialogue and communication between the host, investors, and entrepreneurs on site, and provides corresponding responses and interactions after analysis [16].

3.2.2. Knowledge Base System Design. Analyze the "roles" of this investment program, which mainly include: Host, entrepreneur, and investor. Establish corresponding knowledge bases based on these roles, as shown in Table 3.1.

3.3. Implementation of Question and Answer Analysis System. The program hosting robot system utilizes multiple natural language processing technologies to deeply understand and analyze the text of speech recognition results. Firstly, text correction technology is used to automatically correct the words that are recognized incorrectly in speech recognition, and preliminary analysis of the text is carried out using segmentation, part of speech tagging, named entity recognition, and dependency syntax. Afterwards, conduct semantic understanding of the text and clarify the dialogue intention. When the conversation intention is identified as answering a question, the system retrieves the knowledge base, knowledge map or the Internet according to the question type to obtain the corresponding answer or knowledge, and then combs the answers, selects the best answer from the candidate answers, and returns to the program scene through speech synthesis. In addition, virtual robots can also pose questions to entrepreneurs like hosts, which are pre inputted into the system before the program is broadcasted. When the system recognizes the dialogue intention as requiring questioning, the robot will automatically appear on the screen and ask questions for entrepreneurs or investors to answer [17].

The architecture of a robot question answering system based on natural language processing is shown in Figure 3.3, which mainly uses techniques such as text error correction, lexical analysis, naming recognition, and question answering retrieval.

1. Text correction: After speech recognition, text correction is an important guarantee for correct recognition of question and answer intentions. Errors in speech recognition will lead to errors in subsequent steps. After extensive system testing, it was found that the main errors in speech recognition are homophone errors and approximate speech recognition errors caused by accent issues. In errors, entities are primarily named with the names of individuals and institutions. The system has carried out targeted

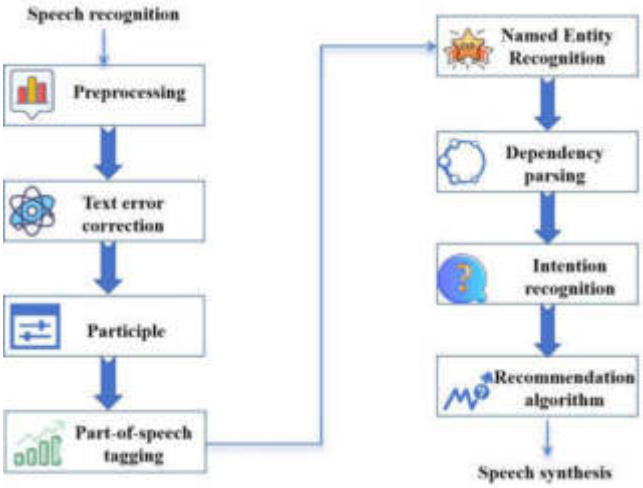


Fig. 3.3: Architecture diagram of robot question answering system

error detection and correction through rule templates and deep learning techniques.

2. Lexical analysis: Mainly includes Chinese word segmentation, part of speech tagging, named entity recognition, and dependency syntax analysis. Chinese does not have space separation between words in English, which affects the semantic analysis of Chinese due to word boundary errors. Word segmentation technology can make Chinese sentences have semantically reasonable word sequences. Part of speech tagging is to label unique parts of speech for each segmented word, which helps with subsequent dependency syntactic analysis. Named entity recognition specifically identifies entities with specific meanings in a statement, which are often crucial information in question answering. Dependency syntax is mainly used to recognize the interrelationships between words in a sentence, such as subject verb relationships, verb object relationships, and definite center relationships. Through syntactic analysis, the semantic backbone of a sentence can be extracted, helping intelligent robot systems achieve precise understanding of dialogue intent.
3. Naming recognition: Naming recognition is a part of lexical analysis in the field of natural language processing, mainly solving the problem of unregistered noun recognition in robot systems. It mainly includes person names, place names, institution names, or professional domain nouns. These words appear less frequently in general corpora, but are crucial for question answering and information retrieval tasks [18]. This technology was applied in the program hosting robot system, which obtained 49270 investment field news headlines through web crawler technology, and used domain dictionaries for initial annotation. Serious annotation errors were manually corrected, achieving automatic recognition of named entities in communication problems by the robot system, and finally generating corresponding knowledge graphs.
4. Q&A retrieval: When the conversation intention is recognized as an answer to a question, the system will search the knowledge base or knowledge graph based on different types of questions. The system sets a high threshold to ensure the accuracy of the answer, and the results in the knowledge base will be directly returned. The results in the knowledge graph will be packaged into templates according to different contents and then returned, if no results can be found in the knowledge base and knowledge map, the system will use the search engine to search the corresponding answers on the Internet. If there are structured answers, the system will give priority to the structured answers. If there are no structured answers, the system will crawl a large number of relevant search results, and use the in-depth learning method of reading comprehension to summarize and extract the correct answers. In the previous program recording, the knowledge base and knowledge graph we built successfully answered the vast majority of questions, with only a few questions applied to the network search step, and the

system application effect was good.

An example of actual interaction between the host and the robot in a certain episode of the program is as follows:

Host: "Little A Brother".

Robot: Hello everyone, I'm here.

Host: Did you see the underwater robot just now? Are you strong or was the underwater robot strong just now?

Robot: I'm too difficult! As both robots, it's not too urgent to argue with each other. I can speak but not swim, it can swim but not speak, so let me summarize that both of us have our own achievements in different fields.

Host: You're still quite impressive, okay, thank you, little brother A.

Robot: Goodbye.

3.4. Intelligent Investment Advisor Recommendation Algorithm Model. Through data mining and analysis of the program materials provided by the program team, there are three main factors that have a significant impact on the final successful signing, namely:

1. Investors themselves pay attention to whether the investment field matches the entrepreneurial field of entrepreneurs;
2. Does the amount of investment made by the investor match the required funds for the entrepreneurial project;
3. On site communication and interaction between investors and entrepreneurs.

An "intelligent investment advisory algorithm model" was established by analyzing these three factors. During the program recording process, the matching degree between investors and entrepreneurial projects was calculated in real-time on-site. When entrepreneurs consulted, recommendations were given for choosing investors.

3.4.1. Investment advisory algorithm calculation model. Based on the analysis of multiple entrepreneurial investment programs, an investment advisor recommendation model was established. The calculation formula is divided into four parts, and the final score is obtained by adding the scores of the four parts according to their weights. This set of five formulas together forms the investment recommendation algorithm model. The following equations 3.1, 3.2, 3.3, 3.4 and 3.5:

$$x_i = \epsilon + \sum_{n_i} \gamma \tag{3.1}$$

$$y_i = \frac{g_i}{e} \tag{3.2}$$

$$z_i = \frac{d_i}{max(d) + \eta} \tag{3.3}$$

$$m_i = \lambda \frac{(e - g_i)}{e} \tag{3.4}$$

$$s_i = \alpha x_i + \beta y_i + \delta z_i + m_i \tag{3.5}$$

The above algorithm is explained in steps 1-6:

Step 1: First, calculate the fit xi between the investor's focus area and the entrepreneur's entrepreneurial field, as shown in equation 3.6.

$$x_i = \epsilon + \sum_{n_i} \gamma \tag{3.6}$$

Among them: ϵ to establish the minimum level of fit in the field of venture capital, γ to add points to single field matching, n_i represents the number of areas that the i-th investor focuses on and are related to the entrepreneurial project of the entrepreneur.

Step 2: Calculate the matching degree y_i between the investor's willingness to contribute and the entrepreneur's financing needs, using the formula shown in equation 3.7.

$$y_i = \frac{g_i}{e} \tag{3.7}$$

Among them, g_i represents the amount that the i-th investor is willing to contribute, and e represents the financing needs of the entrepreneur.

Step 3: Calculate the matching score z_i for on-site communication between investors and entrepreneurs, as shown in equation 3.8.

$$z_i \frac{d_i}{max(d) + \eta} \tag{3.8}$$

Among them, d_i represents the number of times the i-th investor interacts with the entrepreneur, and max (d) represents the number of times the investor interacts with the entrepreneur the most among all on-site investors, η to establish a foundation for communication.

Step 4: Calculate the additional score mi when the investor's investment amount exceeds the financing needs of the entrepreneur. The calculation formula is shown in equation 3.9.

$$m_i = \lambda \frac{(e - g_i)}{e} \tag{3.9}$$

Among them: γ for single field matching, e represents the financing needs of the entrepreneur, and g_i represents the amount of investment that the i-th investor is willing to contribute.

Step 5: The calculation results of equations 3.6-3.9 are weighted and summed according to the importance of investment factors to obtain the final matching score Si. The investor with the highest matching score is the recommended investor by the system, and the calculation formula is shown in equation 3.10.

$$s_i = \alpha x_i + \beta y_i + \delta z_i + m_i \tag{3.10}$$

Among them, x_i, y_i, z_i, and m_i respectively have the aforementioned meanings, α β δ weight coefficients of each focus item are respectively, based on past statistical analysis results, it is generally assumed that α=0.2, β=0.5, δ=0.3, adjusted appropriately according to different project types.

Step 6: Perform a final rationality check on the matching score si obtained in step 5.

If it falls within the range of the project's entrepreneurs and investors, the final result is determined to be the recommended value. If there is a deviation beyond the expected range, adjust the corresponding coefficients in the previous steps and recalculate the s_i result until a relatively more reasonable investor is selected as the main recommendation.

In addition, there are also individual abnormal situations, such as when all investors give up their investment and turn off the lights, the project ultimately fails to invest. In this case, the robot recommendation system can also provide investment matching scores, but the score is extremely low. When the matching score is below 0.5, the system does not recommend choosing investors [19].

3.4.2. Analysis of Focus Areas. According to incomplete statistics, 642 entrepreneurs from different fields have stepped on the stage of a TV station's entrepreneurship program through layer by layer selection. According to statistics of the entrepreneurial fields of these 642 entrepreneurs, the 8 fields with the largest entrepreneurial projects are 372 Internet projects, 250 software applications, 181 life services, 131 intelligent manufacturing projects, 88 education projects, 73 medical and health projects, 72 cultural and creative projects, and 64 new materials, as shown in Table 3.2.

The system regularly updates the investor information knowledge graph to obtain information on the investment areas of interest for all investors on site. Through the introduction of the entrepreneur's project, a

Table 3.2: Program Hosting Robot Knowledge Base System

	Area	Number	Proportion/%
Statistics of	internet	372	57
Internet	outside the Internet	270	41
(including software)	software applications	250	38
other areas	outside of software applications	391	60
classification	life services	181	27
statistics	Intelligent manufacturing	131	20
	Educational technology	88	13
	medical hygiene	73	11
	cultural and creative	72	10
	new materials	64	10
	other areas	27	3
	amount	642	100

pre trained classification model is used to automatically identify the entrepreneur's entrepreneurial field. If the entrepreneurial field of the entrepreneur is the same or related to one or more of the investor's areas of concern, the system will give a higher score. If the field is not relevant, or if investors have little or no investment in the entrepreneurial field where the entrepreneur is located, the system will give a lower score, but not zero, because investors are likely to invest in this project in this episode.

3.4.3. Analysis of Investment Amount. Through sorting and analyzing the program materials, it was found that the financing needs of previous entrepreneurs ranged from 2 million to 30 million. Compared to the financing needs of entrepreneurs, the amount of investment provided by investors was relatively low, at least 500000 yuan, in the successfully signed projects, the final price of 500000 to 2 million (excluding) accounts for 3.07%, 2 to 6 million (excluding) accounts for 38.51%, 6 to 10 million (excluding) accounts for 52.48%, and 10 million and above accounts for 5.94%. The main reason for financing failure is that there is a significant difference between the amount of investment from investors and the financing needs of entrepreneurs, and the amount of investment has a significant impact on whether the contract can be successfully signed. At the live broadcast of the program, real-time information on investor bids is obtained, and the financing needs of entrepreneurs are compared [20]. When the amount is insufficient, the system calculates the score proportionally. When the bid exceeds the financing needs, the system will give investors an additional score when calculating the total score.

3.4.4. On site communication analysis. The system utilizes speech recognition and facial recognition technology to recognize and analyze the speech and facial expressions of investors, determine the number of interactions each investor has when facing entrepreneurs, and analyze their level of investment willingness. The number of times investors speak and interact with entrepreneurs reflects their level of attention to the project. The more interested investors are in the project, the more they will ask entrepreneurs about project information in order to tap into the development potential of the project.

4. Result analysis. A virtual hosting robot (Little A) has participated in the recording of 24 episodes of a venture capital program on a certain TV station. There have been no incidents where the recording of the program was affected by issues with the robot. While ensuring the accuracy of answering questions during hosting interactions, little A's response speed gradually improves with the continuous optimization of the system. The actual application case analysis is as follows.

4.1. Application examples of recommendation systems. At the recording site of the entrepreneurship program, the "Intelligent Investment Advisor Algorithm" will calculate the "investment matching degree" based on the project field, investment amount, and on-site communication situation, providing entrepreneurs with advice on selecting reasonable investors. Figure 4.1 shows the calculation results of the investor recommendation system for a certain period.

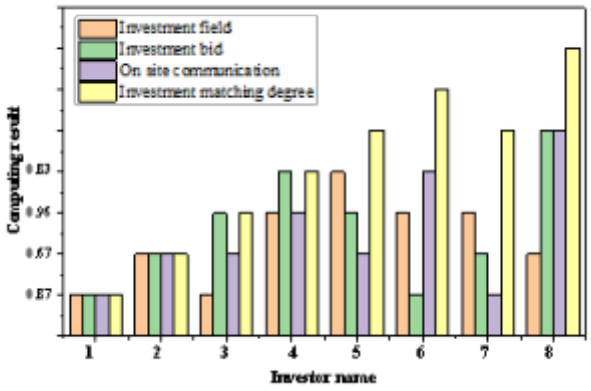


Fig. 4.1: Calculation results of investor recommendation system

From Figure 4.1, it can be seen that the final system did not recommend the 6th and 7th investors with the highest matching degree in the field, nor did it recommend the 8th investor with the highest matching degree in investment amount, nor did it recommend the 4th investor with the highest matching degree in communication. Instead, it selected the 1st investor by combining three factors, the entrepreneurs of this period ultimately followed little A's advice and chose investor 1 as their entrepreneurial investor. When a TV program is actually broadcasted, when a host or entrepreneur consults a robot, little A will broadcast the recommended results on the big screen: According to artificial intelligence calculations, the investor recommended by brother A has a matching degree of 0.86 in his investment field, 0.74 in on-site communication, 0.82 in investment amount, and 0.82 in comprehensive matching. The highest matching degree among investors is investor 1.

4.2. Reasonability analysis of recommendation algorithms. Regarding the investor recommendation system, out of a total of 73 entrepreneurial projects in 24 episodes of the program, 40 entrepreneurs actively sought advice from Mr. A, while another 6 entrepreneurs consulted Mr. A's opinions on the host's advice. In 46 investment recommendation applications, there were 6 times when Mr. A's investor selection suggestions were not the highest bidding investors on the program site. In these 6 times, 5 entrepreneurs followed Mr. A's suggestions, and his analysis received recognition from different entrepreneurial mentors multiple times, demonstrating the rationality of Mr. A's recommendations, with a recommendation adoption rate of 97.72% (45/46). For one of the entrepreneurs who did not choose the investor recommended by Brother A, the main reason for consulting is to consider that "the chosen investor's location can provide greater market resources and quickly help the project generate sales." In this sense, the "region and market" dimensions can be further considered in the recommendation model, which can be further improved and perfected in the future.

5. Conclusion. The author introduces the implementation method of artificial intelligence virtual assisted hosting robot, and proves the effectiveness of the method through successful application in a TV program. Artificial intelligence and robotics technology are being widely applied in various industries, and television programs are no exception. The author designed and implemented a virtual intelligent hosting robot system, established an intelligent investment advisory algorithm model, and implemented intelligent question answering and investment recommendation functions during the program hosting process. This is an innovative application of artificial intelligence robot technology in television programs, which is of great significance for promoting the intelligent development of the film and television media industry.

REFERENCES

[1] Pang, G., Yang, G., Heng, W., Ye, Z., Huang, X., Yang, H. Y., & Pang, Z. (2020). CoboSkin: Soft robot skin with variable stiffness for safer human–robot collaboration. IEEE Transactions on Industrial Electronics, 68(4), 3303-3314.
[2] Tella, A. (2020). Robots are coming to the libraries: are librarians ready to accommodate them?. Library Hi Tech News, 37(8), 13-17.
[3] Pang, G., Yang, G., & Pang, Z. (2021). Review of robot skin: A potential enabler for safe collaboration, immersive teleoperation, and affective interaction of future collaborative robots. IEEE Transactions on Medical Robotics and Bionics, 3(3), 681-700.
[4] Engwall, O., Lopes, J., & Åhlund, A. (2021). Robot interaction styles for conversation practice in second language learning. International Journal of Social Robotics, 13(2), 251-276.
[5] Tanwani, A. K., Anand, R., Gonzalez, J. E., & Goldberg, K. (2020). RILaaS: Robot inference and learning as a service. IEEE Robotics and Automation Letters, 5(3), 4423-4430.
[6] Seyitoğlu, F., & Ivanov, S. (2021). Service robots as a tool for physical distancing in tourism. Current issues in tourism, 24(12), 1631-1634.
[7] James, S., Ma, Z., Arrojo, D. R., & Davison, A. J. (2020). Rlbench: The robot learning benchmark & learning environment. IEEE Robotics and Automation Letters, 5(2), 3019-3026.
[8] Meng, S. , Li, Q. , Qi, L. , Xu, X. , Yuan, R. , & Zhang, X. . (2023). An intelligent recommendation method based on multi-interest network and adversarial deep learning. Computers & Security, 47(5), 705-712.
[9] Zheng, F. . (2022). Personalized education based on hybrid intelligent recommendation system. Journal of Mathematics, 5(2), 3019-3026.
[10] Zhang, X. . (2022). Intelligent recommendation algorithm of multimedia english distance education resources based on user model. Journal of Mathematics, 39(2), 412-434.
[11] Ghahramani, M. , Zhou, M. , Molter, A. , & Pilla, F. . (2022). Iot-based route recommendation for an intelligent waste management system, 64(1), 38-68.
[12] Sun, M., Hu, W., & Wu, Y. (2024). Public perceptions and attitudes towards the application of artificial intelligence in journalism: From a China-based survey. Journalism Practice, 18(3), 548-570.
[13] Dwivedi, R., Mehrotra, D., & Chandra, S. (2022). Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: A systematic review. Journal of oral biology and craniofacial research, 12(2), 302-318.
[14] Yufeia, L., Salehb, S., Jiahuic, H., & Syed, S. M. (2020). Review of the application of artificial intelligence in education. Integration (Amsterdam), 12(8), 1-15.
[15] Cao, L. (2021). Artificial intelligence in retail: applications and value creation logics. International Journal of Retail & Distribution Management, 49(7), 958-976.
[16] Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49, 30-50.
[17] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. Ieee Access, 8, 75264-75278.
[18] Mohamed, E. S., Belal, A. A., Abd-Elmabod, S. K., El-Shirbeny, M. A., Gad, A., & Zahran, M. B. (2021). Smart farming for improving agricultural management. The Egyptian Journal of Remote Sensing and Space Science, 24(3), 971-981.
[19] Haleem, A., Javaid, M., Qadri, M. A., Singh, R. P., & Suman, R. (2022). Artificial intelligence (AI) applications for marketing: A literature-based study. International Journal of Intelligent Networks, 3, 119-132.
[20] Huang, X., Zou, D., Cheng, G., Chen, X., & **e, H. (2023). Trends, research issues and applications of artificial intelligence in language education. Educational Technology & Society, 26(1), 112-131.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 18, 2024
Accepted: Jun 19, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 719–728, DOI 10.12694/scpe.v26i2.3984

RESEARCH ON DATA FUSION METHOD OF MATHEMATICAL CREATIVITY EDUCATION BASED ON AHP HIERARCHICAL ANALYSIS

PINGPING WANG *

Abstract. Dempster-Shafer (DS) evidence theory is widely employed in the real world as a primary instrument for modeling uncertainty. However, when applying Dempster's combination rule, seemingly conflicting pieces of data could be merged to yield surprising results. There are a number of proposed scales for measuring the level of discrepancy between pieces of evidence. However, the presented methods only use a single metric to analyze the conflicting results. However, it is usually unwise to rely on a single criterion to measure the extent to which the facts conflict with one another. For the reason that flaws, differences, differences, and ambiguity all contribute to a higher level of conflict between the evidence. The objective of this study is to propose a novel data fusion method based on AHP for enhancing mathematical creativity education. This work stands out by effectively integrating multiple criteria to improve decision-making processes in educational settings. Given the efficacy of the approach proposed in this study in bringing together seemingly disparate data, we want to eventually extend its application to other forms of uncertainty theory, such as fuzzy set theory and imprecise probabilities. The concept of direct inquiry into the total ambiguity of evidence within the context of discernment will also be explored. Several criteria factors are used to determine the level of disagreement between the information presented here. A proposed analytic hierarchy technique uses multiple criteria to properly weight each piece of data. The initial stage is to assign a numerical value to each piece of evidence based on how well it satisfies each of the criteria. The criterion layer's covariance matrix can be derived by examining the relationship between the criteria's numerical values. After collecting data on each criterion's quantitative change, a fuzzy preference relation matrix can be built. The scheme layer makes use of a fuzzy preference relation matrix in place of a traditional pairwise comparison matrix. Each piece of evidence is given an overall weight based on its combined scheme weight and criteria weight. Two numerical experiments are offered to illustrate the effectiveness of the proposed method after the final weights have been applied to the primary evidence and the evidence has been combined using Dempster's rule. Results show that the proposed methodology outperforms alternative approaches to dealing with contradictory evidence discussed in the literature.

Key words: Analytic hierarchy process; Data Fusion; Mathematical education; Conflict Measurement; Dempster-Shafer Evidence Theory

1. Introduction. Since data from a single sensor may not be adequate to obtain the desired information of target identification, in real applications it is often necessary to supply all of the information of target estimation with several sensors [1]. As a result of this possibility, it is typically necessary to supply all of the information of target estimation with many sensors. Multi-sensor systems frequently provide data that is inconsistent with one another as a result of variations in the precision and resistance to interference of the individual sensors. There have been numerous more hypotheses put forward as potential approaches to modeling uncertainty [2,3]. There is the theory of probability [4], the theory of DS evidence [5], the theory of possibility [6], the theory of fuzzy sets [7], the theory of rough sets [8,9] and so on. The concept of DS evidence was initially proposed by Dempster, and it was then popularized and expanded upon by Shafer. Since its inception in the 1960s, the DS evidence theory [10] has been widely accepted as a powerful instrument for dealing with uncertainty information that cannot be effectively handled using conventional probability theory. The fact that traditional probability theory does not adequately take into account uncertainty and imprecision is yet another issue that can be remedied by DS evidence theory. It has been useful in a variety of fields, including information fusion, decision analysis, the identification of defects, and others [11]. Because it makes it easier to combine data that comes from a number of sources in different locations, It is difficult to meet this condition in many uses of Dempster's rule since it requires ensuring that all of the bodies of evidence share the same dependability, which can be a time-consuming process. Dempster's combination rule has the potential to yield superior fusion outcomes in situations when the level of disagreement between the evidence is relatively low. When we apply Dempster's combination rule to these extremely contradictory pieces of data, however, we obtain two very different sets of findings. The Zadeh

*Soochow College, Soochow University, Suzhou, Jiangsu, 215031, China (Corresponding author, ply_yoyolight@163.com)

paradox [12] is an illustration of the type of results that are paradoxical. Another illustration of a situation in which the total is fair but does not help in making a decision [13]. Numerous amendments to Dempster's combination rule have been suggested by industry professionals after they invested a significant amount of time and effort investigating this issue. Some researchers believe that Dempster's combination rule is to fault for the unexpected findings, and as a result, they have offered alternative rules of combination [13]. Smets put up a credible evidence-based belief paradigm that may be shared [14]. In this model, the normalising phase is skipped and the contrasting mass is instead assigned to a null set. In circumstances that call for evidence that can be relied upon, however, it may be challenging to achieve total satisfaction. Lefevre et al. [15] came up with the idea of applying a unified belief function approach to the rules of combination, which proportionally distributes potentially opposing facts across all parts that are concerned. This technique, on the other hand, requires a greater willingness to take risks where there is significant controversy over the facts. Instead of mandating that inconsistent data be normalized, the new method that was developed by Yager assigns the status "unknown" to the data [16]. The method is extremely sensitive to changes in the data, and the degree of uncertainty in its results increases significantly, when contradictions are resolved by allocating them all to the unknown state. Inagaki first introduced the conflict coefficient k by fusing Yager's technique with Dempster's combination rule [17]. This led to the creation of the notion. The conclusion of a combination is not independent of the order in which the evidence was presented; this is true even if the physical importance of this coefficient is unclear. Sun et al. propose a method that addresses both of these issues and measures the credibility of conflicting pieces of evidence [18]. Even when presented with evidence that strongly contradict one other, their approaches are sound; however, they do not satisfy the commutativity and associativity standards. This study aims to address the gap in the literature regarding the integration of AHP with mathematical creativity education, particularly in the context of data fusion methods. The paper addresses the challenge of effectively fusing data in mathematical creativity education using the Dempster-Shafer (DS) evidence theory, particularly in scenarios where evidence presents a high level of conflict.

1.1. Literature Review. Some academics have claimed that the flawed data used to support Dempster's counter-intuitive results is the main problem, not the combination rule itself. Dempster's combination rule has been used successfully by certain researchers, while others have discovered its faults [19]. Before Dempster's combination rule was employed, Murphy originally established the procedure that changes the original evidence by calculating the fundamental arithmetic average [20]. The evidence-based similarity metric was suggested by Deng et al, who adapted Murphy's method [21]. This was accomplished by first disclosing the evidence distance, and then modifying the evidence presented in accordance with the strength of the evidence's support. Each piece of evidence gets assigned its own discount rate based on how different it is from the others. However, since they only employed one criterion, we can't know how much evidence disagrees with each other. According to Burger, distance and conflict are two entirely distinct ideas that cannot be interchanged. To account for this problem, An et al. modified the similarity measure to include fuzzy logic [22]. After that, they considered how much room for contradiction and uncertainty the evidence gave them. Despite taking into account two criteria, both methods have limitations and cannot accurately measure various forms of conflict. Previous research has explored various data fusion techniques such as DS evidence theory and fuzzy set theory. However, there is a paucity of studies that integrate AHP with mathematical creativity education, which this study aims to fill.

Numerous researchers have looked into the advantages of the multi-criteria decision-making approach over the more traditional single-criterion evaluation [23]. Using the multi-criteria aggregation method PROMETHEE II, we may quantify the evidence conflict by assigning weights to the various criteria [24]. These two methods were used in tandem. Each piece of evidence is assigned a discounting factor based on a set of criteria developed after considering the features of the evidence conflict. This is done to determine if the evidence may be ignored without consequence. The superior outcomes are a direct result of the multi-criteria idea's design, which takes into account multiple fusion levels. In an effort to lessen the subjectivity and contingency of decision making, a number of performance indicators are used to assign relative importance to the various pieces of information. Silva and de Almeida proposed using a multi-criterion technique to quantify the level of disagreement between the pieces of evidence [25]. The ECTRE TRI multi-criteria decision-making approach was evaluated for this purpose. Through the process of modeling, three separate conflict criteria are combined to provide a quantitative assessment of the evidence discrepancies. Mathematical creativity is crucial for developing innovative problem-

solving skills. Existing literature emphasizes the need for advanced methods to accurately assess and enhance these skills, thereby highlighting the relevance of the proposed AHP-based data fusion method.

In an effort to address conflicts and problems that come from a lack of data, we apply multiple criteria to conduct a comprehensive examination of the degree of difference between the evidence. The final outcome is a state-of-the-art, comprehensive framework for assessing legal disputes. To account for the wide variety of conflicts that can occur between the pieces of evidence, it is necessary to take into consideration their uncertainty, imprecision, imperfection, dissimilarity, and disparity. To solve these criterion problems thoroughly, we present a revised analytical hierarchy process. To begin, we determine how well each piece of evidence meets each of the evaluation criteria and assign a score. The quantitative values of each criterion are used to calculate the covariance between them, and this information is stored in a covariance matrix that constitutes the criteria layer. In order to get the criterion layer weights, we invert the covariance matrix into the pairwise comparison matrix. Absolute value dissimilarity between criteria is used to build a fuzzy preference relation matrix. The scheme layer uses fuzzy preference relation matrices in place of pairwise comparison matrices to calculate weights. The sum of everyone's opinion on how important each piece of evidence is. Several numerical examples are provided to show how the suggested method works and how effective it is.

2. Methodology.

2.1. Evidence Theory Dempster and Shafer. The DS evidence theory and its associated theories are briefly introduced. 2 = F1, F2, FN, a set of exhaustive and mutually exclusive hypotheses, is the framework for judgments. The two groups' merger is 22.

$$2^{\emptyset} = \{\emptyset, F_1, F_2 \dots F_N, \{F_1, F_2\}, \{F_1, F_2, F_3\} \dots \emptyset\} \tag{1}$$

Let's pretend 2 is a framework for analysis, and that the set 22 of propositions is the power set in 2, where A 2 is any subset in 2. The mass function is a good fit for these parameters because it maps 22 to [0, 1].

$$\begin{cases} m\{\emptyset\} = 0 \\ \sum_{A \in \emptyset} m(A) = 1 \end{cases} \tag{2}$$

Basic probability assigns m.

The fundamental probability assignment function on 2 is m.

$$Bel\ (A) = \sum_{B \in C} m(A), A \in \emptyset \tag{3}$$

Bel(A) values indicate entire belief. A single set proposition A has a belief function that meets the following conditions:

$$\begin{cases} Bel(A) = 0 \\ Bel(A) = 1 \end{cases} \tag{4}$$

The belief function constraint rule is equation (4).

2.2. Conflict Metrics. Sensor data is unreliable, therefore information sources conflict. Dempster's data fusion rule prioritises consistency above conflict. Dempster's combination rule won't apply in circumstances of strong evidence disagreement, therefore the fusion will be startling. Solving this problem requires quantifying evidence disagreement. Conflict measurement methodologies vary. The research design and data collection method involves the following: First it is required to provide each piece of evidence with a numerical value that will represent how effectively each specific item satisfies the proposed criteria. The covariance matrix for the next layer may be acquired by evaluating the relationship between the numerical values of each criterion. Once the acquired data is carefully analyzed, a fuzzy preference relation matrix can be constructed. The fuzzy preference relation matrix is constructed by comparing each pair of elements based on their relative importance, allowing for degrees of preference rather than binary judgments. It helps in capturing the uncertainty and imprecision in decision-making.

2.3. Analysis. Shafer initially used the conflict coefficient k to quantify the evidence's disagreement [26]. Here are two examples:

Example 1. Two independent bodies of evidence m1 and m2 in the whole frame of discernment 2 = 'A, B, C, D' have the following basic probability assignment:

$$\text{m1: } 0.25(\text{A}) \; 0.25(\text{B})$$
$$\text{m1(C)} = 0.25 \; \text{m1(D)} = 0.25$$
$$\text{m2: A}=0.25 \; \text{B}=0.25$$
$$\text{m2(C)} = 0.25 \; \text{m2(D)} = 0.2$$

Calculating the conflict coefficient yields 0.75. The conflict analysis shows that the evidence disagrees proportionally to k. Thus, m1 and m2 evidence disagree greatly. These two evidence sets are identical, hence there is no conflict. Intuition cannot support this conclusion.

Example 2. Consider two fundamental probability assignments m1 and m2 in the frame of perceiving 2 = -1, -2, -3, which is complete:

$$\text{m1: } \text{m1}(\theta 1, \theta 2) = 0.6 \; \text{m1}(\theta 1, \theta 3) = 0.4$$
$$\text{m2: } \text{m2}(\theta 2, \theta 3) = 0.5 \; \text{m2}(\theta 1, \theta 2, \theta 3) = 0.$$

Calculation yields the conflict coefficient k = 0. Conflict analysis shows that m1 and m2 evidence are consistent, hence there is no conflict. The following example shows why the conflict coefficient k cannot appropriately assess how much the two sets of evidence disagree. Jousselme measured evidence distance in vector space [27].

If there are two bodies of evidence m1 and m2 in the second frame of discernment, their evidence distance is defined as

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(\bar{m}1 - m2)^T \underline{D}(m1 - m2)} \tag{5}$$

where the two primary probability assignment functions, mE 1 and mE 2, are represented by vectors. A and B are the m1 and m2 proofs for power set 22. The focus element cardinality is represented by the || symbol. When two sets of evidence are diametrically opposed to one another, their distance is defined as one, and when they are similar to one another, it is close to zero. Conflict coefficient's oversimplification can be corrected by increasing evidence distance. It is possible that unreal situations will be exposed by the evidence distance test.

2.4. Conflict Assessment Criteria. We can show that the level of disagreement between pieces of evidence cannot be reliably established using a single criteria evaluation methodology to quantify conflict by studying the aforementioned conflict measurement techniques. One measure of evaluation cannot possibly capture the full picture. Some measuring procedures reveal just slight disagreement at the same level of evidence, whereas others show substantial discrepancy. The disagreement in the evidence needs to be measured using multiple criteria. We include flaws like similarity, dissimilarity, disparity, imprecision, and ambiguity to form a more grounded assessment of the evidence's reliability. The reasons for these occurrences are then discussed.

Number one, Overview. The evidence isn't very convincing because it lacks detail. Doubt is viewed as a complex set in the DS evidence theory. As the size of the subset grows, so does the quality of the evidence and the trustworthiness of the beliefs.

Second, Inequality. The discrepancy in the evidence provides insight into the different base probabilities attached to the various claims. Mutually adjusting their perspectives may be facilitated by concentrating on shared interests. They also calculated the concordance between the primary points of evidence. This is the strongest evidence of disagreement that can be found between the two camps. Here, we may utilize the evidence distance to quantify the discrepancies between pieces of evidence.

Third, Distinction. Burger made a geometric analogy between the importance of proof in the law. Incorporating the Pignistic probability transform into vector space is a key part of Zhao et al. The function of the Pignistic vector angle evidence dispute measured the degree of similarity between two pieces of evidence by using the sine of the angle between them. Medium-range distance and angle measurement technique.

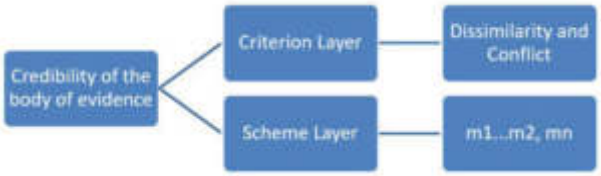


Fig. 2.1: Evidence Credibility Evaluation's Hierarchical Structure Model

Fourth. It is usual practice to utilize the Shannon entropy derived from probability theory when assessing the reliability of evidence. If one wants to make accurate calculations, this method should be used instead of the DS evidence theory method. Probability theory is another approach. Please see Fig. The uncertainty interval is a statistical tool used to quantify degrees of knowledge. Information entropy measures of uncertainty only consider m(A), Bel(A), and Pl(A) information. Uncertainty data from measurements have been greatly reduced, making precise descriptions impossible. The degree of doubt is quantified by averaging the gap between the most definite case and each singleton's confidence interval [58]. This procedure takes into consideration the inherent uncertainty in each piece of evidence. According to the theory of belief intervals, the interval between zero and one is the most suspect.

Criteria-based decision making. The previous section covered the first five of the evidence conflict factors. A combination of these factors provides a reliable evaluation of the evidence's degree of disagreement. In practice, there is a disagreement between the various evidence standards because they all offer something different and have different preferences. This means that the same data set can give contrasting images of conflict according to different criteria. By calculating the weight given to each criterion, we may analyze the degree of discrepancy in the data. The multi-criteria decision-making process determines how heavily various criteria should be weighted. Using AHP and the notion of fuzzy preference relations, we rank each criterion in this analysis. Using AHP and the theory of fuzzy preference relations facilitates the making of multi-criteria judgments. Below you can find the instructions.

The AHP method of hierarchical weighted decision making was created to making decisions based on a number of different factors. By taking this tack, officials can simplify difficult situations. A scheme's weights can be calculated by comparing and weighing its constituent pieces. It provides an organizational structure for identifying the optimal strategy [28]. This method incorporates quantitative computation with qualitative analysis, assigns appropriate weights to each criterion for each choice scheme based on the experience of the decision-makers, and establishes whether or not the goal is attainable.

Use of the AHP Technique Steps. We make a hierarchical order. Modeling decision-making problems effectively requires layering them as part of an analytic hierarchy. We sort problems by importance and take a closer look. On the basis of their interplay, the decision objective, the criterion, and the object are placed on the top, middle, and bottom layers, respectively. The lowest level is the scheme, the middle level is the criteria, and the highest level is the objective. In Figure 2.1 we see the study's hierarchical structure model. The 'Hierarchical Structure Model' consists of three levels: the goal at the top, criteria in the middle, and alternatives at the bottom. Each component represents a step in the decision-making process, from setting objectives to evaluating options based on weighted criteria. A comparison matrix with each pair consisting of two cells. Over and over, you can refer to aij to indicate the relative impact of xi and xj on C. A = (aij) $(n \times n)$ represents the comparison matrix. Pair comparison Matrix A. The ratio of the effects of xi on component C's aji to those of xj is 1/aij. Pairwise comparison matrix A has the scores on Santy's 1-9 scale as its entries (aij and aj).

We verify that the weights are consistent and perform a simple computation. In order to assign relative importance to each criterion, the pairwise comparison matrix A was used. Matrix If aijajk = aik and i, j, and k are 1, 2, and n, then the elements of A must be accurate. If and only if the value of the biggest eigenvalue of the consistency matrix A is equal to n, then A is a positive reciprocal matrix of order N. Here is how we

characterize the CI:

$$CI = \frac{\gamma_{max} - n}{n - 1} \tag{6}$$

Find the data's mean random consistency index RI. CR formula:

$$CI = \frac{CI}{RI} \tag{7}$$

When the CR is less than 0.1, pairwise comparison matrix discrepancies are acceptable. If not, modify the pair comparison matrix.

To make sure everything is in order, we need to add up the weights. Using the findings from the previous stage, we can now determine how much weight to give to each criterion and the contribution of each scheme layer. The results are verified at last.

Although subject matter experts typically produce pairwise comparison matrices for use in AHP, these matrices are not immune to the influence of bias.

2.5. The Connection Between Unclear Preferences. Multi-attribute decisions benefit from the use of the fuzzy preference relation. The decision-making process necessitates the comparison of the schemes in pairs and the expression of preferences. The preference relation is a useful tool for conveying and understanding the preferences of decision-makers.

2.6. Approach. Weights can only be allocated with some subjectivity due to the random nature with which comparison matrices for pairs are formed. Based on the notion of fuzzy preference relations, the authors of this study propose a refined version of the AHP approach. This allows us to avoid the issues that have plagued more conventional approaches. The fuzzyness of the preference relation matrix reflects the changes in the quantitative criteria. Next, we give each piece of evidence a weight that takes into account how persuasive it is at both the criterion and scheme levels. Combining these methods reduces the impact of subjectivity and variability.

The generation of the pairwise comparison matrix is difficult because of the complexity of the objective items, and it does not always pass the consistency test. Because of the intricate nature of the objects. The covariance is used in this paper to determine the importance of each criterion. This was quantified by statisticians. More covariance between random variables indicates that they are not truly independent. The covariance matrix was constructed from the numerical values of the criteria characteristics. The parwise comparison matrix is built in the following way:

1. Determine the numerical index value for all n qualifying characteristics.
2. Use numerical data to compute covariance. Let X = x1, x2, xn and Y = y1, y2, yn be two random variables, and let x and y stand for their respective expectations.

The fuzzy preference relation matrix measures attribute index value dissimilarity across evidence sources and evaluation criteria. Judgements are less biassed. Computing Methods:

1. Calculate variance Quantifying the attribute index of each body of evidence for each criteria attribute. The ith and n-1 body of evidence have Vi attribute index value differences. Vi's weight reduces system-i gap.
2. Determine the dissimilarity between the evidence body attribute and the pijth element of the fuzzy preference relation matrix.

$$P_{ij} = \frac{V_i}{V_i + V_j} \tag{7}$$

Due to differences in information collection and accuracy amongst sensors, the evidence in multi-sensor data fusion is inconsistent. To determine the weight and credibility of each piece of evidence, we must first determine the extent to which they diverge. This article suggests adopting a modified AHP approach for in-depth conflict analysis to evaluate the reliability of each piece of evidence. The weighted average can be calculated after giving weights to the various pieces of evidence. Combine the evidence using Dempster's weighted average method.

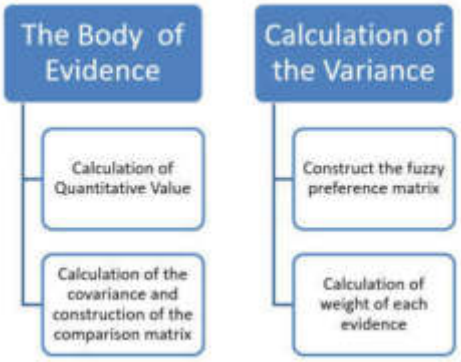


Fig. 2.2: Weight Determination

Table 2.1: Quantitative numerical

	N(m)	DifS(m1)	Diss(m1)	Conf(m1)	TU(m1)
M1	0.065	0.664	0.965	0.563	0.045
M2	0.40	0.465	0.413	0.655	0.055
M3	0.030	0.600	0.310	0.420	0.66

Table 2.2: The Covariance Matrix

	C1	C2	C3	C4	C5
C1	0.0242	0.532	0.463	0.0643	0.04343
C2	0.0423	0.643	0.0432	0.03245	0.06145
C3	0.0543	0.302	0.0463	0.0435	0.0735
C4	0.0513	0.342	0.0543	0.0654	0.0554
C5	0.0234	0.534	0.0364	0.0643	0.04413

A. Establishing Ranking Systems. In this analysis stage, you will build a three-tiered hierarchy, the top level of which will be evidence weight (A measure of the credibility and importance assigned to each piece of evidence in the analysis, used to calculate the weighted average in decision-making). Finally, we'll use weight to indicate how credible each piece of evidence is. Using this dual-pronged method, targets can be located. The organizational chart is depicted in Figure 2.2.

B: Weighing in. The pairwise comparison matrix evaluates each criterion. Pairwise comparison matrices are created by correlating numerical values for each criterion for each piece of evidence. Five things:
C1: Non-specific measurement N(mi);
C2: difS(mi) discrepancy;
C3: Dissimilarity (diss(mi));
C4: Conf (mi).
C5: Uncertainty measurement, TU(mi);

We divide each column component by the column's major diagonal to update the covariance matrix C (in cij/cii). Matrix B, the relative covariance, is then available. Table 2.3 shows the findings.

A criteria covariance pairwise comparison matrix eliminates expert knowledge-related judgement differences.

3. Result and Discussions. Let's imagine many sensors are unable to identify targets, despite the fact that each sensor is providing valuable context for the type of target being identified. This table depicts the

Table 2.3: The relative Covariance Matrix

	C1	C2	C3	C4	C5
C1	1	0.4522	0.463	0.5443	0.4743
C2	0.4123	0.56423	0.4173	0.6143	0.6845
C3	0.5443	0.2352	0.2443	0.2143	0.7385
C4	0.5137	0.3742	0.8237	0.727	0.5554
C5	0.2134	0.1831	0.5184	0.5584	0.4413

Table 3.1: Basic Probability Distribution

	O1	O2	(O1, O2)	(O2, O3)	O
m1	0.8	0.4	0.4	0.5	0.4
m2	0.4	0.5	0.4	0.6	0.6
m3	0	0.2	0.2	0.2	0.7
m4	0.5	0.3	0.8	0.7	0.5
m5	0.2	0.1	0.5	0.5	0.3

basic probability distribution for each sensor reading within the context of the system's acquired data from five different types of sensors, where 2 = 1, 2, 3. According to Table 3.1, the majority of the sensors (m1, m2, m4, and m5) believe in H1, while sensor m3 believes in H2. It appears that sensor m3's data does not correspond to the others. The standard deviation of your answers will be extremely high if you utilise Dempster's combination rule as written. The evaluation index value associated with each criterion is used to calculate the covariance between the criteria.

The difference in evaluation indices across all accessible bodies of evidence for each criterion must be computed before building the fuzzy preference relation matrix or the consistency fuzzy preference relation matrix. Consistency matrix for the evidence's P1 fuzzy preference relations is as follows:

$$P_1 = \begin{pmatrix} 0.4533 & 0.254 & 0.654 \\ 0.3645 & 0.342 & 0.543 \\ 0.534 & 0.522 & 0.555 \end{pmatrix}$$

$$\overline{P}_1 = \begin{pmatrix} 0.4533 & 0.254 & 0.654 \\ 0.3645 & 0.342 & 0.543 \\ 0.534 & 0.522 & 0.555 \end{pmatrix}$$

Table 3.2 shows the results of applying criterion C2 to the consistency fuzzy preference relation matrix, which is subsequently used to weight the evidence presented in the scheme layer (A hierarchical level in the AHP method where different schemes or plans are evaluated and weighted based on their consistency and preference relations.).

$$P_1 = \begin{pmatrix} 0.4731 & 0.5534 & 0.3524 \\ 0.3725 & 0.4462 & 0.1143 \\ 0.5132 & 0.3212 & 0.1255 \end{pmatrix}$$

$$\overline{P}_1 \begin{pmatrix} 0.5132 & 0.3051 & 0.3752 \\ 0.6511 & 0.5012 & 0.3513 \\ 0.3656 & 0.4372 & 0.3154 \end{pmatrix}$$

Applying the following formula to the weights assigned to each piece of evidence yields the weighted average: The values for the first dimension are: m(1) = 0.5949, m(2) = 0.1617, m(3) = 0.0515, m(1, 2) = 0.0772, m(2, 3) = 0.0153, and m(2) = 0.0994.

Table 3.2: The Weight of each body

	C1	C2	C3	C4	C5
m1	0.38	0.54	0.2	0.8	0.2
m2	0.44	0.55	0.2	0.5	0.4
m3	0.6	0.25	0.8	0.1	0.3
m4	0.55	0.76	0.4	0.2	0.5
m5	0.20	0.14	0.2	0.5	0.2

The evidence will be combined four times with a weighted average in accordance with the Dempster rule before a decision is reached. Dempster's combination rule states that although most evidence points to target 1, target 2 is more likely and target 1 is impossible. Since the majority of evidence is in favour of H1 and only one piece of evidence is in favour of H2, the result is unreasonable. This Evidence m1 provides the most convincing backing for the m1 hypothesis. Conflicting evidence receives less weight in the final combination outcomes when more criteria are utilised to set weights for the body of evidence. Our method works best when there are conflicting pieces of evidence. A probability estimate of the underlying distribution of sensor data for a task involving fault diagnostics. To establish the validity of the findings, we thoroughly analysed five distinct measures of conflict. The suggested procedure has the potential to generate more accurate and reasonable fusion findings than alternative methods, even in highly contested circumstances.

This demonstrates that Dempster's method, which can even yield counterintuitive results, cannot be used to the strongly contradictory data. It is challenging to make firm inferences from the available evidence due to the intricacy of the scenario. These results validate the usefulness of the suggested methodology for determining the evidence weighting factor. A more believable fusion outcome can be achieved by precisely characterising the degree of disagreement between the pieces of evidence, as discussed in the analysis offered in the third part. The proposed approach not only successfully modifies the conflict evidence, but also has improved convergence and precision. Its treatment of the multiple conflicting facts is both remarkable and successful when compared to other methods. The proposed method in this paper demonstrates remarkable efficacy in reconciling data that initially appear incongruous, opening avenues for its application in other realms of uncertainty theory, including fuzzy set theory and imprecise probabilities.

4. Conclusion. The research introduces an advanced method, bolstered by a refined Analytic Hierarchy Process (AHP), adept at harmonizing discordant datasets. This approach surpasses traditional measures by integrating a nuanced fuzzy preference relation matrix, significantly reducing subjective bias and aligning more closely with intuitive assessments of evidence conflict. Our method has been validated through a numerical example that underscores its robustness, even amidst scenarios rife with conflict. However, the method's efficacy is not without its vulnerabilities; it may falter in the face of starkly contradictory data, and the initial subjectivity in pairwise comparison matrices could influence the accuracy of evidence weighting. Looking ahead, the scope of this method will be expanded to encompass other domains within uncertainty theory, including fuzzy set theory and imprecise probabilities. Additionally, we aim to delve deeper into the overall uncertainty of evidence, refining our technique to better accommodate the complexities of discernment processes. Despite these challenges, the experimental validation through a fault diagnosis application confirms the method's substantial potential, offering a reliable and informed decision-support framework.

Acknowledgments. Project Source: higher education reform research subject in Jiangsu Province, Project Title: Practical Research on the Construction of Advanced Mathematics Gold Course Based on Self-learning Network Platform (2019JSJG335)

REFERENCES

[1] Zhou K. Water Richness Zoning and Evaluation of the Coal Seam Roof Aquifer Based on AHP and Multisource Geological Information Fusion. Geofluids. 2021 Oct 29;2021:e1097600.

[2] Shabani A, Ziaii M, Monfared MS, Shirazy A, Shirazi A. Multi-Dimensional Data Fusion for Mineral Prospectivity Mapping (MPM) Using Fuzzy-AHP Decision-Making Method, Kodegan-Basiran Region, East Iran. Minerals. 2022 Dec;12(12):1629.
[3] Yu W, Xing J. Sports Event Level Measurement Indicator System Using Multisensor Information Fusion. J Sens. 2021 Nov 1;2021:e9330438.
[4] Li Y. Hierarchical Analysis of Intelligent Fusion Data of College Training Theory based on Virtual Cloud Classroom. In: 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). 2023. p. 245–8.
[5] Qin Z, Liao H, Chen L, Zhang L. Enterprise Performance Management following Big Data Analysis Technology under Multisource Information Fusion. Secur Commun Netw. 2021 Dec 16;2021:e7915670.
[6] Boix-Cots D, Pardo-Bosch F, Pujadas P. A systematic review on multi-criteria group decision-making methods based on weights: Analysis and classification scheme. Inf Fusion. 2023 Aug 1;96:16–36.
[7] Li Y, Kou G, Li G, Hefni MA. Fuzzy multi-attribute information fusion approach for finance investment selection with the expert reliability. Appl Soft Comput. 2022 Sep 1;126:109270.
[8] Torkayesh AE, Yazdani M, Ribeiro-Soriano D. Analysis of industry 4.0 implementation in mobility sector: An integrated approach based on QFD, BWM, and stratified combined compromise solution under fuzzy environment. J Ind Inf Integr. 2022 Nov 1;30:100406.
[9] Zhu L. Urban Landscaping Landscape Design and Maintenance Management Method Based on Multisource Big Data Fusion. Comput Intell Neurosci. 2022 Aug 30;2022:e1353668.
[10] Yue H, Liao H, Li D, Chen L. Enterprise Financial Risk Management Using Information Fusion Technology and Big Data Mining. Wirel Commun Mob Comput. 2021 Dec 15;2021:e3835652.
[11] Al-Zubaidi SQD, Fantoni G, Failli F. Analysis of drivers for solving facility layout problems: A Literature review. J Ind Inf Integr. 2021 Mar 1;21:100187.
[12] Wei R, Chen S, Zhang S, Zhang J, Ding R, Mi J. An AHP-ME-Based Vehicle Crash Prediction Model considering Driver Intention and Real-Time Traffic/Road Condition. Math Probl Eng. 2022 Jul 21;2022:e4371305.
[13] Ouyang L, Zhu Y, Zheng W, Yan L. An information fusion FMEA method to assess the risk of healthcare waste. J Manag Sci Eng. 2021 Mar 1;6(1):111–24.
[14] Tang Y, Wu S, Zhou Y, Huang Y, Zhou D. A New Reliability Coefficient Using Betting Commitment Evidence Distance in Dempster–Shafer Evidence Theory for Uncertain Information Fusion. Entropy. 2023 Mar;25(3):462.
[15] Yang G. Low-Carbon Awareness Information Technology of Enterprise Executives Based on Big Data and Multimodal Information Fusion. Mob Inf Syst. 2022 Jun 28;2022:e1534440.
[16] Wang J. SOA-based Information Integration Platform for Educational Management Decision Support System. Math Probl Eng. 2022 May 2;2022:e7553333.
[17] Li L, Lei B, Mao C. Digital twin in smart manufacturing. J Ind Inf Integr. 2022 Mar 1;26:100289.
[18] Yu L, Ye Y, Guo W, Chen S, Zhang H. The Application of Multisensor Information Fusion Technology in Environmental Restoration. J Sens. 2022 Sep 28;2022:e9514361.
[19] Sáenz-Royo C, Chiclana F, Herrera-Viedma E. Intentional bounded rationality methodology to assess the quality of decision-making approaches with latent alternative performances. Inf Fusion. 2023 Jan 1;89:254–66.
[20] Wu B, Qiu W, Huang W, Meng G, Huang J, Xu S. A multi-source information fusion approach in tunnel collapse risk analysis based on improved Dempster–Shafer evidence theory. Sci Rep. 2022 Mar 7;12(1):3626.
[21] Shen L, Wang G, Gao H. Regional Information Management of Higher Education Based on Network Security and Grey Relational Analysis. Secur Commun Netw. 2022 Mar 31;2022:e6638408.
[22] Luo Z, Jingying C, Guangshuai W, Mengyi L. A three-dimensional model of student interest during learning using multimodal fusion with natural sensing technology. Interact Learn Environ. 2022 Jul 1;30(6):1117–30.
[23] Miao H. Intervention Methods of College Counselors on Students' Psychological Crisis under the Background of Deep Learning. Math Probl Eng. 2022 Sep 30;2022:e9966484.
[24] Zhou L. Research on Evaluation of Art Education Effect in Colleges and Universities Based on Big Data Technology. Math Probl Eng. 2022 Sep 28;2022:e5671785.
[25] Ali U, Shamsi MH, Bohacek M, Purcell K, Hoare C, Mangina E, et al. A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making. Appl Energy. 2020 Dec 1;279:115834.
[26] Wang S ming, Vu LH. The integration of digital twin and serious game framework for new normal virtual urban exploration and social interaction. J Urban Manag. 2023 Jun 1;12(2):168–81.
[27] Zhang J, Feng H, Liu B, Zhao D. Survey of Technology in Network Security Situation Awareness. Sensors. 2023 Jan;23(5):2608.
[28] Liang W, Li T. Research on human performance evaluation model based on neural network and data mining algorithm. EURASIP J Wirel Commun Netw. 2020 Sep 10;2020(1):174.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 19, 2024
Accepted: Jul 15, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 729–738, DOI 10.12694/scpe.v26i2.4012

RESEARCH ON BIG DATA VISUALIZATION TECHNOLOGY BASED ON MULTI-SOURCE VIBRATIONAL DATA ACQUISITION

HAOJIE LING*, XIANG WAN†, YI GOU‡, AND YUAN HUANG§

Abstract. In order to scientifically and reasonably monitor the soil environment of green spaces in urban residential areas, the layout and sampling methods of soil monitoring points for green spaces in residential areas were studied, including the selection of representative residential areas, determination of monitoring point sampling positions, and determination of the number of points. The author proposed a research on the layout and sampling of soil monitoring points for green spaces in urban residential areas based on multi-source data collection and big data visualization. By using multi-source big data visualization methods, representative residential areas of a certain city were selected to monitor heavy metals (cadmium, mercury, arsenic, lead, copper, chromium, zinc, and nickel) in the green soil of their residential areas. The study reveals variations in heavy metal concentrations in the soil across residential areas of differing building ages. To ensure thorough monitoring of soil environmental conditions in residential areas, it's recommended to include neighborhoods of varying building ages as monitoring sites. Our findings indicate that the choice of sampling locations within these areas does not substantially affect the heavy metal content in soil samples. Therefore, it's preferable to prioritize sampling from residential areas rather than focusing solely on large green spaces within them, There are differences in samples from different monitoring points within the same residential area, and at least 3-4 monitoring points should be set up in each residential area to represent the soil environmental conditions of that residential area. The application of multi-source big data has a positive effect and advantage on the distribution of urban soil monitoring points.

Key words: Multi-source data collection, Urban residential areas, Soil monitoring, Big data, Layout points

1. Introduction. With the rapid development of IT technology worldwide, various industries and industries have generated their own information data, and corresponding data centers have also been established. From the current situation, the direction of information technology development in various industries is shifting from discrete to centralized. The amount of information that needs to be managed in data centers has sharply increased, and the information in small data centers cannot reflect comprehensiveness in a timely manner and is difficult to control. Therefore, it is necessary to centrally manage distributed data centers to further promote the development of data centers.

As an important IT infrastructure, enterprises and institutions are highly concerned about its importance. Furthermore, data centers are currently evolving towards a wider range of shared resources. Contemporary data centers are no longer simply using management commands or tools, but also adopting security technology, standardization technology, virtualization technology, and automation management technology. The current data center is constantly advancing, step by step towards a more complete data center. In the view of more mature data centers, standardization and automation are their basic characteristics, visualization and virtualization are their essential attributes, and security is an important consideration direction [1]. Among them, standardization and automation refer to the ability of data centers to have unified standards and fully automated network maintenance and operation management, standardized scripts and software configurations, visualization and virtualization refer to the use of computer 3D technology and virtual storage technology to change the appearance of data centers, express the large amount of digital information generated during their monitoring process in a simpler and more reliable way, and reduce labor costs.

In the data center industry, finance and telecommunications were the earliest to invest in construction, with large investments and a high market share. The importance of data centers in this industry is at the core. After entering the mobile Internet, the financial and telecommunications industries have developed rapidly, which

*Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610041, China. (Corresponding author, HaojieLing5@126.com)
†Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610041, China. (XiangWan283@163.com)
‡Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610041, China. (YiGou7@126.com)
§Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610041, China. (YuanHuang3731@163.com)

has led to the management and construction of the data center. After more than 20 years of development, the construction of data centers has shown exponential growth [2].

Driven by the mobile Internet and the rapid development of wearable technology, a large amount of information data has been generated. With the help of these, the demand for bandwidth has been growing, which has refreshed enterprises' understanding of data centers. In recent years, the development of domestic data centers has shown a trend of large-scale. Data centers have developed from small scale to large scale. Under the influence of "Internet Plus", two operators have strengthened the construction of big data centers and the support of cloud services.

In the past, the focus of data centers was on information stability, mainly to be able to recover data in a timely manner when problems occurred. Therefore, data in one place would be backed up in multiple locations according to different geographical locations, and data standardization processing would also be carried out at certain intervals. This way, in case of accidents, data could be recovered within a certain period of time, ensuring the normal operation of services.

Today, the focus of data centers is no longer on data recovery. With the rise of big data, data analysis, real-time processing, and real-time transmission have become key research directions in contemporary data centers [3].

In particular, the network interconnection of mobile Internet and wearable devices has put forward unprecedented requirements for the network delay of contemporary data centers. Netizens hope to access cloud data in real-time as quickly as accessing local data, which requires a wider and more uniform distribution of data centers. According to this development trend, data centers will become more efficient, achieve higher density, and at the same time, obtain higher computing processing capabilities. The entire network will have new computing concepts, content distribution will become more important, and data information will be redundant and scattered throughout the network to reduce the difficulty for users to obtain data and reduce latency [4].

2. Literature Review. Eldahshan, K. A. et al. tackled the issue of visualizing large datasets by employing feature selection techniques to reduce data volume and minimize model training time (Tt) while preserving data integrity. They utilized the Random Forest Importance Algorithm (RFI) and the Embedding Method based Selection from Model (SFM) method to identify the most crucial features, comparing them with the Chi2 tool based on Selection Percentile (SP) method. Subsequently, logistic regression (LR) and k-nearest neighbor (KNN) algorithms were applied for classification. The study concludes that by eliminating redundant and irrelevant data, feature selection methods significantly enhance data analysis and visualization capabilities [5].

Barik, R. K., and others introduced GeoTCloud, a cloud-based geospatial big data infrastructure model tailored for visualizing geospatial data within the tourism sector. This model facilitates the storage, analysis, and presentation of large-scale geospatial data, offering valuable support for the advancement of smart city initiatives and the enhancement of tourism-related services [6].

Yang, H. P. et al. introduced a novel approach for efficiently transmitting and visualizing meteorological big data using WebGIS technology. Leveraging open-source tools like HTML5 and Mapbox GL, their solution optimizes data compression and transmission on the server side, as well as distributed requests and rendering on the browser side. They developed a high-low 8-bit compression method capable of compressing 100MB files to megabyte-level files with 90% compression rate, maintaining data accuracy to two decimal places. Additionally, their approach integrates pyramid tile cutting, concurrent domain name request processing, and texture rendering. Experimental results demonstrate rapid transmission and display of up to 100MB grid files within milliseconds, supporting multi-terminal service applications. This grid data visualization model offers insights for big data and technology centers, with potential applicability across various industries [7].

The author's research is based on multi-source big data visualization technology to select representative residential areas from tens of thousands of residential areas in the city, and investigate the soil of different types of representative residential green spaces in the central urban area of a northern city; Quantify the concentrations of heavy metals, including cadmium (Cd), mercury (Hg), arsenic (As), lead (Pb), chromium (Cr), copper (Cu), zinc (Zn), and nickel (Ni); Using methods such as non parametric testing and principal component analysis, this paper analyzes the selection of representative residential areas, establishing the spatial distribution and number of monitoring sites for soil analysis in urban residential areas is crucial. This framework serves as a foundation for monitoring heavy metal levels in urban soil and offers technical assistance in accurately assessing

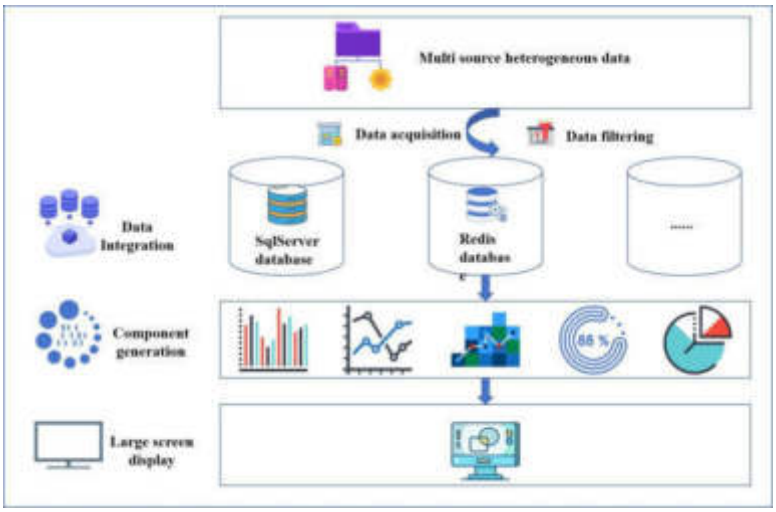


Fig. 3.1: Basic ideas for data visualization

and evaluating the environmental quality of urban residential areas.

3. Research Methods.

3.1. Overview of Data Visualization Content. With the continuous progress of information technology, traditional manual statistics, calculations, lists, and graphs can no longer be effective interpretation methods for massive heterogeneous data from multiple sources. Scientists have also exclaimed, "What we do is only collect data." In order to solve such problems, a new technological field - data visualization has emerged [8]. The basic idea of data visualization is shown in Figure 3.1.

Data visualization refers to organizing information points based on a proposed framework, dividing them into business modules, selecting and matching appropriate charts according to different data types, and depicting information from multiple perspectives and levels. It can also provide a fast human-computer interaction experience. Due to the fact that people perceive graphics more quickly and deeply than text or numbers, data visualization technology utilizes computer graphics to replace large, dense, and complex data with vivid graphic forms, mining valuable information from the vast ocean of data and presenting it intuitively.

Data visualization not only saves people time receiving information, but also provides rich visual impact, which is more conducive to gaining insight into the essence of data, so as to better serve the next step of work decision-making and play a certain role in the survival and development of the enterprise.

3.2. Data Visualization Design Steps. Data visualization is the last step in the data analysis process and an important step in presenting the analysis results. An excellent data visualization work needs to comprehensively consider three factors: indicator content, aesthetic design, and user interaction experience. It closely combines and flexibly applies statistics, computer technology, and art design. The specific design steps are shown in Figure 3.2.

1. Clearly define design objectives. When facing visualization requirements, it is necessary to conduct research on product requirements and indicator content, and determine the two main directions of "what to display" and "how to display". At the beginning of data visualization design, only by fully understanding the business content can we grasp the design goals. Taking the visualization design of shipbuilding accuracy data as an example, it is necessary to have a deep understanding of the entire shipbuilding process and the focus of precision control work in the early stage. Only in this way can data and graphics be better combined, playing a crucial role in how to explore the value of data in the future.
2. Reasonably divide the sectors. After the goal is determined, it is necessary to preliminarily divide the

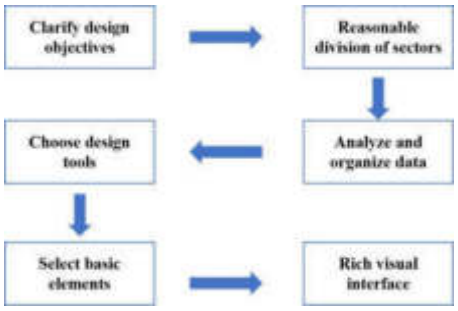


Fig. 3.2: Data Visualization Design Steps

business modules and develop a page layout based on the work content. Place important information in the middle and above, and lower priority information on both sides and below, highlighting the hierarchical structure. Then organize the information points according to the divided sections, and start organizing and mining the data accordingly [9].

3. Analyze and organize data. After preliminary preprocessing of the obtained multi-source heterogeneous data, data indicators are analyzed and mined from multiple dimensions to determine the perspectives and contents that need to be expressed for each indicator. A multi-level system and correlation are established for the data, and the calculation methods for each indicator are summarized. These data are then integrated into multiple data tables according to different classifications, which serve as the data source for visual components and lay the foundation for component generation.
4. Choose a design tool. In the context of big data, data visualization technology is becoming increasingly mature and can now integrate various emerging technologies such as user interaction technology and data interaction technology. The available tools also have richer presentation forms, simpler operation methods, and support more data sources. Choosing the appropriate data visualization based on functional requirements and the programming ability of designers can help reduce the difficulty of design work and improve the ideal level of final presentation. Data visualization tools can generally be divided into chart libraries, business intelligence analysis, visualization screens, and professional categories, which can meet various functional requirements in various industries. Give examples to illustrate the main differences in the characteristics of three widely used data visualization tools.
5. Select basic elements. The common basic elements of data visualization include column charts, tables, maps, etc. Select specific elements under each element based on data relationships, render the organized data into visual graphics, and complete data mapping and component generation. At the same time, make good use of extended functions such as chart linkage and highlight dynamic analysis to enrich the visual hierarchy and enhance the user experience. There are various types of data visualization charts, including bar charts, line charts, pie charts, scatter charts, radar charts, and their variations. Each chart has its own applicable situation, and only by comprehensively considering factors such as data relationships, project requirements, and audience orientation can we make the most of everything and truly play the role of the chart [10]. When the types of information are different, the requirements for data relationships will vary, and the chart forms presented will also be different. List six main data relationships and their corresponding chart types.
6. Rich visual interface. The final step is to handle the overall details, and the rational design of the visual aspect can also provide assistance in the optimal presentation of data results. Attention should be paid to the reasonable arrangement of layout, uniformity and coordination of fonts, color matching, and other visual expressions of the visual interface, in order to complement the content and complete the overall construction of the visual interface. A reasonable spatial layout can make the content of the data visualization interface more hierarchical, highlight key data information, and thus improve the efficiency of user interpretation of information. The layout design of data visualization interfaces should generally follow three principles: Ensure that critical information is centrally located and readily

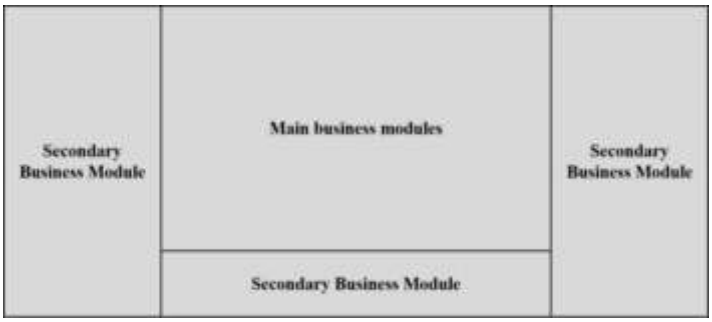


Fig. 3.3: Data Visualization Layout Scheme

discernible, facilitating the capture of key points. Maintain a balanced spatial arrangement of various page elements to enhance design aesthetics. Emphasize data presentation in the layout while avoiding unnecessary complexities or elements that hinder the effectiveness of conveying information. Based on the above principles, multiple layout design schemes can be developed, and the scheme shown in Figure 3.3 can cover most application scenarios. The central position of the page is the main module used to display the main indicators, while the left, right, and bottom of the page are secondary modules that can display a large number of secondary indicators. It has the characteristics of clear content hierarchy and simple and beautiful design.

3.3. Materials and Methods.

3.3.1. Overview of the research area. The research area is a residential area in the central urban area of a northern city. As an important urban area, this area has a long history of construction and development, high population density, and mature socio-economic development.

3.3.2. Multi source big data visualization. In order to monitor the soil environmental quality of green spaces in urban residential areas and scientifically reflect their impact on the health of residents, priority is given to selecting residential areas with high population density and frequent population activities as monitoring objects. Visualize multi-source big data and select representative residential areas. Visualize the city's POI points, weibo posts, and dianping website evaluation information through spatial processing using geostatistical software. Using web scraping technology to obtain urban residential area information on Baidu Maps and Lianjia.com, vectorizing and mapping it to obtain a spatial distribution map of residential areas. According to the results of the sixth national population census in 2010, the population density spatial distribution map was obtained by vectorizing the population of each street [11].

3.3.3. Soil Sample Collection. Using the method of judging and distributing points uniformly in space, set up sampling points in green spaces within residential areas, with 5 sampling points set up in each residential area. The sampling points are set up in large concentrated green spaces to ensure that samples are collected from the lawns, shrubs, and trees in each residential area. Each sample is composed of 5 surface samples with a depth of 0-20 cm and evenly mixed in equal quantities using the diagonal method, with a sampling amount of 2 kg. Sampling is carried out using the Global Positioning System (GPS) to locate and record geographic coordinates. The sample is stored in a cloth bag and labeled with a sample label. Based on the monitoring results of this study, the pH range of soil samples is 7.85-8.58, with an average of 8.19. The soil samples in residential areas of the urban center are alkaline.

3.3.4. Sample pretreatment and analysis. Soil cadmium was determined by solid direct injection and graphite furnace atomic absorption spectroscopy. The mercury content is determined by solid direct injection and cold atomic absorption method. The total amount of arsenic, lead, chromium, copper, zinc, and nickel was determined by powder pressing method and X-ray fluorescence spectroscopy. During the analysis process, no less than 10% of soil standard samples are inserted into each batch and analyzed in parallel in the laboratory.

The analysis results of all standard samples are within the uncertainty range, and the relative deviation of parallel sample determination is 0% to 7%, which meets the technical specification requirements [12].

3.3.5. Data statistical analysis methods. Discriminate outliers based on data. Use independent sample K-W test to test for differences in soil samples from different types of residential areas. Use Principal Component Analysis (PCA) to analyze soil samples within residential areas. The Kruskal Wallis test is a non parametric method aimed at testing whether the median of multiple samples is equal. Kruskal Wallis ranks high, so it remains invariant to any monotonic transformation of the measurement range. Known names: Kruskal Wallis H-test, one-way analysis of variance.

Calculate the average rank of each group of samples as shown in equation 3.1:

$$\overline{R_j} = \frac{R_j}{n_j} = \frac{\sum_{i=1}^{n_j} R_{ij}}{n_j} \tag{3.1}$$

Among them, R_{ij} is the sorting number, and n_j is the number of samples in the jth group.

Calculate the Kruskal Wallis statistic, as shown in equation 3.2:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i(\overline{R_i} - \overline{R})^2 \tag{3.2}$$

Among them, N is the sum of the number of samples in each group, ni is the number of samples in the i-th group, and k is the number of sample groups.

3.3.6. Data statistics. ArcGis 10.1 software was used for spatial data processing, and SPSS 18.0 was used for statistical and analytical analysis of soil heavy metal content, correlation, and differences; Canoco 5.0 conducts principal component analysis and plotting of heavy metal content in soil samples.

4. Result analysis.

4.1. Selection of Residential Areas. Perform geographic information processing on urban residential area information to obtain a distribution map of residential areas. Perform geographic information processing on the 2010 census results to obtain a population density distribution map. The above results are spatially overlaid with urban POI point data spatial distribution, Sina Weibo sending volume data distribution, and Dianping evaluation information distribution data [13,14]. From the distribution map of residential areas and population density, it can be seen that the residential areas and population of the city are concentrated in the central area. Among them, the density of residential areas in the urban center is higher than that in the suburban areas, and its density gradually decreases from the urban core to the outer periphery; The highest population density in the urban core area exceeds 2×104 people/km^2. It can be seen that the distribution of residential areas and population density is most concentrated in the urban center. For example, selecting residential areas for soil sample collection in concentrated areas can have a larger radiation population and better reflect the impact of soil on the population, making it more representative.

The distribution map of urban POI, Weibo sending volume, and Dianping network evaluation reflects the laws and spatial distribution characteristics of human social activities. The POI, Weibo posts, and Dianping comments of the city are all concentrated in the central area of the city, which is basically consistent with the distribution of residential areas and population density. It can be seen that crowd activities are most concentrated in the central area of the city. After data space integration and visualization, the distribution of residential areas with high population density, dense POI points, and frequent crowd activities was obtained. These residential areas were selected as alternative residential areas, and priority was given to selecting residential areas in the central urban area. In addition to considering the population density and frequency of activities in the area, residential area screening is conducted based on factors such as the distance of construction, distribution of administrative jurisdiction, and green space area within the residential area [15]. In order to consider that the selected residential areas can represent different situations of urban residential areas, the selection principle takes into account residential areas of different building ages and administrative jurisdictions. At the same time, the concentrated green space area in residential areas should not be less than 10% of the

Table 4.1: Residential Area and Monitoring Point Information

Residential area name	Architectural Era	Monitoring points
A community	1986	A1 A5
B community	2008	B1 B5
C community	1980	C1 C5
D community	2003	D1 D5
E community	2000	E1 E5
F community	1982	F1 F5
G community	2001	G1 G5

Table 4.2: Significance of K-W test for heavy metals in soil samples from different sampling locations

Heavy metal	Cd	Hg	As	Pb	Cr	Cu	Zn	Ni
Progressive significance/P	0 365	0 201	0 210	0 301	0 763	0 962	0 614	0 880

total area of residential areas, facilitating the collection of soil samples. Based on the above considerations, 7 residential areas were selected. The construction period of residential areas varies from 1980, 1990, 2000, and 2010. Research on these residential areas can represent the soil environmental characteristics of the main residential areas in the city. The selected residential areas are shown in Table 4.1.

The selected residential areas have different building ages and are distributed in various administrative districts in the central area of the city. The area of concentrated green space in residential areas shall not be less than 10% of the land area occupied by residential areas.

4.2. Abnormal value discrimination. The Grubbs method was used to identify outliers in soil heavy metal content, and the G values of all samples did not exceed the critical values given in the Grubbs Table, with no outliers. All monitoring data in this study were used for subsequent analysis and calculation [16].

4.3. Differences in Soil Heavy Metals in Residential Areas of Different Building Ages. Designate neighborhoods built before 2000 as "old neighborhoods" (A, C, F neighborhoods), and neighborhoods built after 2000 as "new neighborhoods" (B, D, E, G neighborhoods). Perform principal component analysis on the heavy metal content of soil samples from two groups of residential areas to determine the differences in heavy metal content among soil samples from different types of residential areas. Principal component analysis is obtained through Euclidean distance projection. The distance between the sample points in the figure represents the similarity in the composition of heavy metal content between them. The closer the sample points are, the more similar they are. From the results, it can be seen that there is a difference in the heavy metal content of soil samples between the "old community" and the "new community". The soil samples from "old communities" such as A and C are relatively concentrated in the first and fourth quadrants of the map [17]. The soil samples of new communities such as B, D, E, and G are mainly concentrated in the second and third quadrants. The dominant elements in the soil samples of community A are Pb and Cu; The dominant elements in the soil samples of community C are Cu, Cr, Ni, and Zn. F community belongs to the "old community", but the soil samples in F community are more similar to the "new community", which may be due to the location of F community on the edge of the urban center and the slow development of facilities around the residential area. Considering the comprehensiveness and representativeness of soil environment monitoring in residential areas, it is advisable to consider different building ages when selecting residential areas.

4.4. Differences in Soil Heavy Metal Content at Different Sampling Positions. Differential analysis was conducted on the heavy metal content of samples at three different sampling locations: lawn, under trees, and shrubs. The analysis results are shown in Table 4.2. The progressive significance of the K-W test for heavy metal content in three soil samples was greater than the given test level by 0.05, indicating that there was no significant difference in heavy metal content among the samples obtained from the three sampling positions [18]. It can be seen that the sampling location has no significant impact on the monitoring results

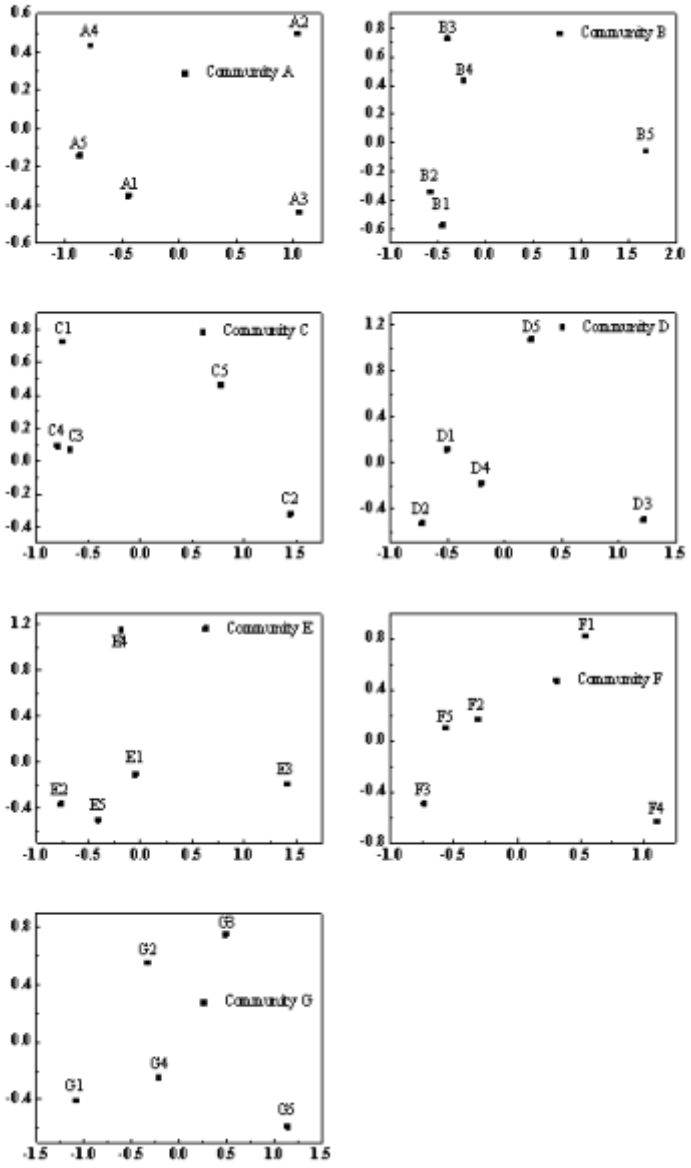


Fig. 4.1: Principal component analysis of soil samples from various residential areas

of soil samples in residential areas. Considering the convenience and representativeness of on-site sampling for soil monitoring in residential areas, priority should be given to sampling from green spaces with a relatively large area in residential areas.

4.5. Sampling quantity. In order to determine the differences in different locations within each residential area, principal component analysis was performed on the heavy metal content of soil samples from each residential area, as shown in Figure 4.1. There are certain differences in the heavy metal content among dif-

ferent soil sampling points in different residential areas. Except for B and E communities, the sampling points in other residential areas are scattered in four quadrants, indicating that at least four sampling points need to be set up in these residential areas to represent the heavy metal content characteristics of the soil environment in that residential area. B1 and B2 points in community B are relatively similar, while B3 and B4 points are relatively similar, indicating that retaining only one point per pair of these two points in community B can represent the soil heavy metal content characteristics of each pair of sample points. E1, E2, and E5 in E community have high similarity, meaning that one of these three sampling points can represent the heavy metal content of these three sampling points. For B and E communities, three soil monitoring points can represent the heavy metal content in the soil within that community. Based on the analysis results of all residential areas, at least 3-4 soil sampling points should be set up in residential areas. In special circumstances such as complex green space types or pollution accidents, the number of points can be appropriately increased.

5. Conclusions.

1. Multi source big data visualization technology can be applied to the monitoring of green soil environment in urban residential areas. It can quickly and accurately determine regional boundaries and urban structure related to crowd activity and population density. It has the advantages of fast, comprehensive, and representative selection for residential areas with frequent crowd activity and high radiation population.
2. The heavy metal content in soil samples from residential areas of different building ages varies to a certain extent. If in the actual monitoring process, considering the comprehensiveness and representativeness of soil environment monitoring in urban residential areas, accurately and objectively reflecting the soil environment status of the entire city's residential areas, residential areas of different building ages should be selected as monitoring objects.
3. Different sampling positions have no significant impact on the heavy metal content of soil samples. In actual monitoring work, while ensuring the accuracy and representativeness of monitoring, the convenience and timeliness of on-site sampling should be considered. Priority should be given to collecting soil samples from green areas with a dominant area in residential areas.
4. There are differences in the content of heavy metals in the soil within the same community. At least 3-4 soil sampling points should be set up in a residential area. In special circumstances such as complex types of green spaces or pollution accidents, the number of points can be appropriately increased.

REFERENCES

[1] Gao, F., Yue, P., Cao, Z., Zhao, S., Shangguan, B., Jiang, L., ... & Liang, Z. (2022). A multi-source spatio-temporal data cube for large-scale geospatial analysis. International Journal of Geographical Information Science, 36(9), 1853-1884.

[2] Li, Y., & Gan, H. (2021). Tourism information data processing method based on multi-source data fusion. Journal of Sensors, 2021, 1-12.

[3] Manjunatha, S., & Annappa, B. (2020). Real-time big data analytics framework with data blending approach for multiple data sources in smart city applications. Scalable Computing: Practice and Experience, 21(4), 611-623.

[4] Yang, B., & Liao, Y. M. (2022). Research on enterprise risk knowledge graph based on multi-source data fusion. Neural Computing and Applications, 34(4), 2569-2582.

[5] Eldahshan, K. A. , Alhabshy, A. A. A. , & Mohammed, L. T. . (2023). Filter and embedded feature selection methods to meet big data visualization challenges. Computer, material, and continuum (in English)(1), 23.

[6] Barik, R. K. , Tripathy, S. , Nayak, A. , & Roy, D. S. . (2022). Cloud gis model for geospatial bigdata visualization towards smart city: a case study of bhubaneswar, odisha. 2022 OITS International Conference on Information Technology (OCIT), 1-5.

[7] Yang, H. P. , Sun, Y. R. , Chen, N. , Jiang, X. W. , Chen, J. H. , & Yang, M. , et al. (2022). Quick compression and transmission of meteorological big data in complicated visualization systems. Complexity, 27(7), 692-711.

[8] Dong, Y., Wen, C., & Wang, Z. (2023). A motor bearing fault diagnosis method based on multi-source data and one-dimensional lightweight convolution neural network. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 237(2), 272-283.

[9] Yang, X., Stewart, K., Fang, M., & Tang, L. (2022). Attributing pedestrian networks with semantic information based on multi-source spatial data. International Journal of Geographical Information Science, 36(1), 31-54.

[10] Hu, J., Cai, S., Huang, T., Qin, X., Gao, Z., Chen, L., & Du, Y. (2021). Vehicle travel destination prediction method based on multi-source data. Automotive Innovation, 4, 315-327.

[11] Mills, M. C., & Rahal, C. (2020). The GWAS Diversity Monitor tracks diversity by disease in real time. Nature genetics, 52(3), 242-243.

[12] Sethi, S. S., Bick, A., Ewers, R. M., Klinck, H., Ramesh, V., Tuanmu, M. N., & Coomes, D. A. (2023). Limits to the accurate and generalizable use of soundscapes to monitor biodiversity. Nature Ecology & Evolution, 7(9), 1373-1378.
[13] Perera, S., Wilson, R. W., Butterley, T., Osborn, J., Farley, O. J., & Laidlaw, D. J. (2023). SHIMM: a versatile seeing monitor for astronomy. Monthly Notices of the Royal Astronomical Society, 520(4), 5475-5486.
[14] Jiang, W., Fan, J., Li, L., Xu, K., Liu, J., & Zhao, X. (2023). Study on Human's Working Memory for Short-term Stationing in Plateau Airport. Journal of Human Movement Science, 4(1), 51-59.
[15] Gil, J. (2020). Army to activate two Iron Dome batteries at Ft. Bliss. Inside Missile Defense, 26(24), 2-2.
[16] Agrawal, R., Potter, R., Saikia, S. J., Longuski, J. M., Davis, R. M., & Collom, B. (2021). Conceptual design and assessment of a mars orbital logistics node for sustainable human exploration. Acta Astronautica, 189, 199-215.
[17] Luo, Q. C., Sun, K. W., Chen, T., Zhang, Y. F., & Zheng, Z. W. (2024). Trajectory planning of stratospheric airship for station-kee** mission based on improved rapidly exploring random tree. Advances in Space Research, 73(1), 992-1005.
[18] Buyong, M. R., Larki, F., Hakimi Zainal, M. I., Almahi, A. Y. A., Ismail, A. G., Hamzah, A. A., ... & Yeop Majlis, B. (2020). Implementation of capacitance as simultaneous sensing and actuating tool in tapered microelectrode arrays for dielectrophoresis-on-a-chip application. Microelectronics International, 37(4), 215-224.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 22, 2024
Accepted: Jul 10, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 739–748, DOI 10.12694/scpe.v26i2.4070

APPLICATION OF GENETIC ALGORITHM IN MESHLESS OPTIMIZATION OF ELASTIC FOUNDATION WITH RIBBED PLATES AND BEAMS

XIAOMEI LIU*

Abstract. In order to overcome the mesh dependency of finite element method, the author proposes the application of genetic algorithm in meshless optimization of elastic foundation with ribbed plates and beams. The ribbed plate is regarded as a combination of plates and beams. Based on the meshless method and combined with genetic algorithm, the rib arrangement position of the rectangular ribbed plate is optimized to minimize the deflection of the center point of the ribbed plate under lateral load. Compared to traditional finite element methods, using the author's meshless method for rib position optimization analysis of ribbed plates does not require mesh reconstruction, and the nodes discretized on the plate and ribs always do not need to be changed. The results indicate that the deflection values of the center points corresponding to the second generation individuals are more concentrated, and there are also many individuals with smaller deflection values compared to the first generation. The hybrid genetic algorithm is indeed effective. The author added the constrained random direction method to form a hybrid genetic algorithm based on genetic algorithm, which accelerates convergence speed, reduces computational repetition rate, and significantly reduces the computational algebra of genetic algorithm to two or three generations, resulting in better results.

Key words: Meshless method, Ribbed plate optimization, Minimum moving multiplication, Hybrid genetic algorithm

1. Introduction. As a commonly used structural form in engineering, the mechanical performance analysis of plate and shell structures has always been a hot topic in computational mechanics. Since the initial classical mechanical analysis theory of plates and shells, many experts and scholars have shown great interest in plate and shell structures. With the improvement of industrial production level and the gradual deepening of Internet technology, research on plate and shell structures has emerged in endlessly [1]. Plate shell structure on elastic foundation is also a common structural style in engineering, such as raft foundation of buildings, rigid concrete pavement, runway for aviation flight, etc. In practical engineering, plate and shell structures based on this foundation are usually subjected to single or combined loads, such as uniformly distributed loads, moving loads, and so on. Maintaining the stability of plate and shell structures on elastic foundations under complex load conditions is crucial for engineering. Compared to previous engineering structures, flat plates often encounter problems of insufficient stiffness and bearing capacity in practical applications [2]. If the stiffness is improved by increasing the thickness of the plate, it will also result in excessive weight of the plate and low economic benefits. Ribbed plate is a method of increasing the local stiffness and stability of the foundation plate by adding ribs [3]. Therefore, combining the ground substrate with ribs is an effective form of shell structure to resist excessive local deformation. At present, many scholars at home and abroad have conducted a lot of research and analysis on the mechanical calculation of plate and shell structures on elastic foundations. However, there is relatively little research on the addition of ribbed plates to elastic foundations. Due to the presence of ribs, problem analysis will be more complex than plate problem analysis, and different numbers and positions of ribs will affect the overall mechanical performance of ribbed plates [4]. For different load conditions, the reasonable arrangement and optimization of the rib positions of ribbed plates on elastic foundations. Currently, many scholars at home and abroad have conducted a lot of research and analysis on the mechanical calculation of plate and shell structures on elastic foundations, but there is relatively little research on the aspect of ribbed plates on elastic foundations, due to the presence of ribs, problem analysis will be more complex than flat plate problem analysis, and different numbers and positions of ribs will affect the overall mechanical performance of the ribbed plate [5]. The rational arrangement and optimization of the rib positions for adding ribbed plates on elastic foundations under different load conditions can enhance the local stiffness

*Xuchang Vocational Technical College, Xuchang Henan, 461000, China (Corresponding author, XiaomeiLiu5@126.com)

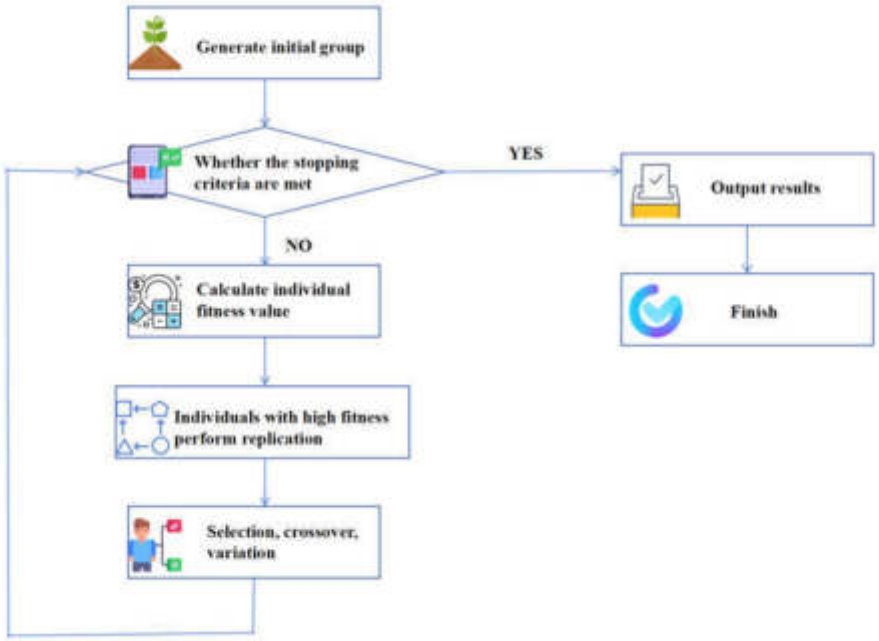


Fig. 1.1: Genetic Algorithm

and safety performance of the structure, which has certain theoretical significance and engineering reference value for the research of adding ribbed plates on elastic foundations [6]. As shown in Figure 1.1:

2. Literature Review. Many scholars at home and abroad have conducted extensive research on ribbed plates. In the early stages of the research, the research model on ribbed plates went through a period of exploration. In the initial research, many scholars regarded the ribbed plate as a uniformly dense plate, and the ribs became attached and adhered to the initial flat plate. This model is called the orthogonal anisotropic plate model. Subsequently, it was discovered that the seed delivery model was only suitable for ribbed plates with dense rib arrangements, so improvements were made to this model, resulting in the Grillage model. Although the principle is clear and the formula is simple, satisfactory solutions cannot be obtained for most ribbed plates. Later scholars tended to consider ribbed plates as a hybrid structure, with the original plate being calculated as a plate and the ribs as a beam model.

In the bending analysis of ribbed plates, Tan, C. first analyzed the stability of ribbed plates under uniformly distributed loads using the energy standard [7]; In order to prevent local deformation of the ribbed plate, Ahmad, H., Hashim derived a relationship about the minimum size range of longitudinal ribs. With the development of technology, methods such as finite strip method and finite element method have also been applied to the bending of ribbed plates [8]. The Rayleigh Ritz method is also used in the bending analysis of ribbed plates, and many scholars have made significant contributions in this regard. Patel, V. G. solved the linear bending problem of rectangular ribbed plates using first-order shear deformation theory and moving least squares approximation [9]; Rasoulizadeh, M. N. analyzed the transverse bending of rectangular ribbed plates using Huber theory and considering the effect of mid plane strain, and calculated the bending problem of unidirectional ribbed plates under local loads using the mixed trigonometric series method [10]; Lin, J. conducted a structural analysis of ribbed plates under the combined action of compression and bending loads, based on the application of the principle of minimum potential energy and considering the comprehensive effect of shear lag and shear deformation in the calculation process, laying a good foundation for calculating the concrete main cable-stayed bridge [11].

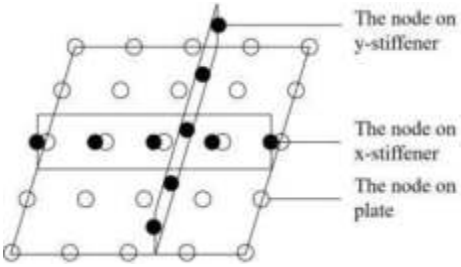


Fig. 3.1: Mesh free model with ribbed plates

The author combines genetic algorithm with constrained random direction method to form a hybrid genetic algorithm. The hybrid genetic algorithm combines the advantages of constrained random direction method and genetic algorithm, and optimizes the calculation results of each generation of genetic algorithm using constrained random direction method. The author uses a hybrid genetic algorithm and a meshless model based on a ribbed plate structure to optimize the rib arrangement of the ribbed plate. The effectiveness of the proposed method is verified through numerical examples.

3. Research Methods.

3.1. Mesh free model with ribbed plates. The meshless model with ribbed plates is shown in Figure 3.1.

From Figure 3.1, it can be seen that the arrangement of nodes on the x-axis and y-axis ribs does not follow the position of the nodes on the board, but is distributed between two rows or two columns of nodes on the board. According to the moving least squares approximation, the expression of the shape function can be obtained 3.1:

$$N(x,\overline{x}) = P^T(\overline{x})A^{-1}(x)B(x) \tag{3.1}$$

Among them: $P(x)$ is the basis function matrix; $A(x), B(x)$ is a matrix composed of a weight function and a basis function, respectively. The displacement field function of the plate can be obtained from the moving least squares approximation and first-order shear deformation theory as follows 3.2:

$$\begin{cases} u_p(x,y,z) = \sum_{l=1}^{n} N_l(x,y)u_{0pl} - z\sum_{l=1}^{n} N_l(x,y)\varphi_{pxl} \\ v_p(x,y,z) = \sum_{l=1}^{n} N_l(x,y)u_{0pl} - z\sum_{l=1}^{n} N_l(x,y)\varphi_{pyl} \\ w_p(x,y,z) = \sum_{l=1}^{n} N_l(x,y)w_{pl} \end{cases} \tag{3.2}$$

Among them: $u_{0pl}, v_{0pl}, \varphi_{pxl}, \varphi_{pyl}$ and w_{pl} are the node parameters of the board; n represents the number of board nodes. The displacement field function of the x-axis rib is as follows 3.3:

$$\begin{cases} u_{sx}(x,z) = u_{0s}(x) - z\varphi_{sx}(x) \\ = \sum_{l=1}^{N} \phi_{xl}(x)u_{0sl} - z\sum_{l=1}^{N} \phi_{xl}(x)\varphi_{sxl} \\ w_{sx}(x) = \sum_{l=1}^{N} \phi_{xl}(x)w_{sxl} \end{cases} \tag{3.3}$$

Among them, u_{0sl}, φ_{sxl}, and w_{sxl} are the displacement field parameters of the x-axis ribs; N is the total number of nodes on the ribs; ϕ_{xl} is a shape function. Similarly, similar expressions also apply to ribs in the y-direction.

3.1.1. Displacement coordination conditions. The displacement between the ribs and the plate on the ribbed plate has the following relationship. The following equations 3.4, 3.5, and 3.6:

$$[w_p]_p = [w_{sx}]_s \tag{3.4}$$

$$[\varphi_p]_p = [\varphi_{sx}]_s \tag{3.5}$$

$$[u_p]_p = [u_{sx}]_s \tag{3.6}$$

Among them, point p is the point on the mid plane of the plate corresponding to node s on the rib; The point c is the intersection of the line connecting node s on the rib and point p on the plate with the contact surface. Substituting equations 3.2 and 3.3 into equations 3.4 to 3.6 yields the following equations 3.7, 3.8, and 3.9:

$$\sum_{I=1}^{n} N_I(x_i, y_i) w_{pI} = \sum_{J=1}^{N} \phi_{xJ}(x_i) w_{sxJ} (i = 1, \cdots, N) \tag{3.7}$$

$$\sum_{I=1}^{n} N_I(x_i, y_i) \varphi_{pI} = \sum_{J=1}^{N} \phi_{xJ}(x_i) \varphi_{sxJ} (i = 1, \cdots, N) \tag{3.8}$$

$$\sum_{I=1}^{n} N_I(x_i, y_i) u_{0pI} + e \sum_{I=1}^{n} N_I(x_i, y_i) \varphi_{pxI} = \sum_{J=1}^{N} \phi_{xJ}(x_i) u_{0SJ} (i = 1, \cdots, N) \tag{3.9}$$

Represented by a matrix as follows 3.10, 3.11, and 3.12:

$$T_p \delta_{p\varphi} = T_{sx} \delta_{ex\varphi} \tag{3.10}$$

$$T_p \delta_{pw} = T_{sx} \delta_{exw} \tag{3.11}$$

$$T_p \delta_{pu} + e T_p \delta_{px\varphi} = T_{sx} \delta_{su} \tag{3.12}$$

where e represents the distance between the neutral plane of the plate and the neutral axis of the ribs. Multiplying both ends of the equation by T_{sx}^{-1} can simplify the three equations above into the following equation 3.13:

$$\delta_{sx} = T_{spx} \delta_p \tag{3.13}$$

Among them, equation 3.14 is as follows:

$$\begin{gathered} \delta_{sx} = [u_{0s1}, w_{sx1}, \varphi_{sx1}, \cdots, u_{0sN}, w_{sxN}, \varphi_{sxN}]^T \\ \delta_p = [u_{0p1}, v_{0p1}, w_{p1}, \varphi_{px1}, \varphi_{py1}, \cdots, u_{0pn}, v_{0pn}, w_{pn}, \varphi_{pxn}, \varphi_{pyn}] \end{gathered} \tag{3.14}$$

T_{spx} is a matrix of $3N \times 5n$, and when the position of the ribs changes, only the matrix needs to be recalculated, without the need to redistribute the plate nodes [12]. Optimizing the arrangement of ribs inevitably involves frequent changes in the position of the ribs. Based on the meshless model mentioned above, there is no need to redraw the plate mesh at each step of the optimization process like the finite element method, which can greatly reduce computational complexity. There is a similar relationship between y-shaped ribs and x-shaped ribs, which will not be repeated here.

3.1.2. Control equation. According to the first-order shear deformation theory, the potential energy of a ribbed plate can be expressed as follows 3.15:

$$U_p = \frac{1}{2} \iiint_{-h_p/2}^{h_p/2} \epsilon_p^T D \epsilon_p dz dx dy + \frac{1}{2} \iint \frac{G h_p}{k} (\gamma_{pxz}^2 + \gamma_{pyz}^2) dx dy \tag{3.15}$$

Among them, h_p represents the thickness of the ribbed plate; ϵ_p, γ_{pxz} and γ_{pyz} represent the strain matrix of the ribbed plate; D, G represents the elastic matrix and shear matrix of the ribbed plate, respectively; K=56 represents the shear correction coefficient.

The potential energy of x-shaped ribs can be expressed as follows 3.16:

$$U_{sx} = frac12 \iint_{-h_{sx}/2}^{h_{sx}/2} E(u_{0s,x} - z\varphi_{sx,x})^2 W_{sx} dzdx + \frac{1}{2}\int \frac{GA_{sx}}{k}(w_{sx,x} - \varphi_{sx})^2 dx \tag{3.16}$$

Among them: E represents the elasticity matrix; W_{sx} represents the width of the ribs; A_{sx} represents the cross-sectional area of the rib. The potential energy of the y-shaped ribs can be expressed similarly, so the total potential energy of the ribbed plate can be expressed as follows 3.17:

$$U = U_p + U_{sx} + U_{sy} \tag{3.17}$$

According to the principle of virtual work, the control equation for ribbed plates can be obtained as follows 3.18:

$$K\delta_p = f \tag{3.18}$$

where: K is the stiffness matrix; f is the load vector.

3.1.3. Complete conversion method. The meshless method using moving least squares fitting does not satisfy the Kronecker condition, so the displacement boundary cannot be directly applied. The author proposed this concept using the complete transformation method, which is a very effective method for handling boundary conditions in Galerkin's method. After being processed by the complete transformation method, boundary conditions can be directly applied like the finite element method [13].

3.2. Hybrid Genetic Algorithm. Genetic algorithm has the disadvantages of high computational complexity and high repetition rate, but it has the global optimization characteristics that numerical optimization methods do not possess, while constrained random direction method has the disadvantages of strong local optimization and weak overall optimization ability. By combining these two algorithms and taking advantage of each other's strengths, a new hybrid optimization method, namely the hybrid genetic algorithm, is obtained.

3.2.1. Genetic Algorithm. For optimization problems with constraints, they can generally be described as follows 3.19:

$$\begin{cases} maxf(x) \\ s.t.X \in R \\ R \subseteq U \end{cases} \tag{3.19}$$

Among them: X is the design variable; f(X) is the objective function; U is the basic design space; R is a set of feasible solutions composed of all solutions that satisfy the constraint conditions.

The optimization of ribs with ribbed plates is particularly effective in this case, as its sensitivity to the objective function is not easy to obtain and the computational scale is relatively small. The development of genetic algorithms relies on the improvement of computational level and the development of interdisciplinary fields, simulating the evolutionary process of biological populations - the crossover and mutation between chromosomes - to encode data in practical problems and perform a series of operations on the encoded data.

Genetic algorithm is a method of binary encoding design variables, followed by operations such as crossover, inheritance, and mutation to modify individuals, and then using the values of fitness functions to filter [14]. After multiple iterations, individuals who make the objective function better are obtained.

A pattern is a small combination that has undergone crossover and mutation without changing the encoding order. According to the pattern theorem of genetic algorithm, it can be known that good patterns will grow exponentially with the increase of iteration times, while according to the assumption of building blocks, better modules can be pieced together during the iteration process to form better individuals. The genetic algorithm, which adopts the strategy of preserving the best individual, accumulates excellent patterns and the probability of obtaining the best result is 1. From this, it can be inferred that as long as the algebraic selection of the genetic algorithm is large enough, the genetic algorithm that adopts the strategy of preserving the best individual will converge to the optimal solution.

3.2.2. Constrained Random Direction Method. The basic idea of the constrained random direction method is to generate random directions for searching within a feasible range that satisfies the constraint conditions. The basic operating steps are:

1. Generate a random point within the feasible domain;
2. Generate a series of random directions around a random point as the center;
3. Find points along these directions that can make the objective function better;
4. Repeat steps 1 to 3. The initial point is randomly generated by the computer.

The termination condition of optimization design iteration is an important component of selecting optimization design. The author chooses the function descent criterion as the termination condition for the iteration of the constrained random direction method [15,16]. Use the value of the objective function in two adjacent iteration processes to determine whether the program can be terminated. When the difference between two adjacent objective function values is less than the given acceptable iteration accuracy, the iteration terminates.

3.2.3. Using Hybrid Genetic Algorithm to Optimize the Program Flow of Ribbed Plate. The working principle of hybrid genetic algorithm is to first use genetic algorithm for calculation, obtain the contemporary optimal solution, and then use constrained random direction method to find the local solution near this point. Due to the strong dependence of the constrained random direction method on the selection of initial points, the probability of obtaining an approximate optimal solution from the constrained random direction method after obtaining a contemporary better solution through genetic algorithm will increase. The process of using hybrid genetic algorithm to calculate ribbed plates is as follows:

1. Determine design variables, set population size and iteration times, and select numerical values for crossover probability and mutation probability.
2. Scatter nodes and Gaussian points in the design area.
3. Loop the background grid points and calculate the shape function values at each calculation point.
4. Integrate the stiffness matrix of the plate and beam in the ribbed plate.
5. The position of the ribs during the optimization process is a design variable that constrains the movable range of the design variable. Set the binary encoding method to keep the design variables within the constraint range, and design the encoding length and sequence according to the required accuracy [17,18].
6. Utilize genetic algorithm for optimization, generate corresponding populations, and perform decoding operations on the populations to obtain actual design variable values.
7. Calculate the unique transformation matrix based on the values of the design variables, and combine it with the stiffness matrix of the ribs and plates calculated in step 4) to form the total stiffness of the ribbed plate.
8. Perform meshless analysis to obtain the deflection of the center point of the ribbed plate.
9. Compare the fitness function values of each individual to obtain the best individual in the current population and all populations so far. Adopt the optimal preservation strategy and replace the worst individual in the current population with the best individual so far.
10. Obtain the optimal individual and perform constrained random direction method optimization operation.
11. Perform selection operations on genetic algorithms. If the proportional selection method is used, the ratio of the fitness of each individual to the fitness of the entire population is the probability of an individual entering the next generation; If a deterministic sampling selection method is used, the fitness value is used to calculate the expected value of each individual's ability to survive in the next generation population, and this expected value is used to select individuals who can enter the next generation.
12. Perform crossover operations on genetic algorithms. Using a single point crossover method to exchange binary codes between two adjacent genes.
13. Perform genetic algorithm mutation operations. Using the method of basic positional variation, randomly assign gene values on one or several loci for mutation operations [19,20].
14. Generate the next generation of new individuals.
15. Determine whether the iteration termination condition is met. If it is met, the iteration stops. If it is not met, return to step 6) to continue the calculation.

Table 4.1: Analysis of Optimization Results of Vertical Rib and Rib Plate Hybrid Genetic Algorithm

Serial Number	x-direction rib position/m	y-direction rib position/m	Plate midpoint deflection/$10^{-5}m$	Relative error/(%)
1	0.498757	0.494321	7.837	0.06
2	0.507341	0.505817	7.843	0.08
3	0.495716	0.505312	7.841	0.08
4	0.517681	0.476218	7.885	0.64
5	0.509646	0.498761	7.844	0.08
6	0.497374	0.505842	7.838	0.07
7	0.497894	0.504423	7.838	0.06
8	0.502065	0.494346	7.837	0.07
9	0.510039	0.500435	7.843	0.08
10	0.494728	0.488582	7.846	0.14

Hybrid genetic algorithm introduces constrained random direction into the framework of genetic algorithm. Since the constrained random direction method can find better results each time than the previous one, it can accelerate the efficiency of genetic algorithm in finding better solutions without changing its convergence.

4. Result analysis.

4.1. Example verification. Take a ribbed plate with two perpendicular ribs for calculation. The ribbed plate has a thickness of 0.2m and a square plate with a length and width of 2m. The two ribs have the same size, with a height of 0.2m and a width of 0.02m. The meshless method arranges nodes as 8. In order to ensure population diversity, the population size is set to 21, the crossover probability is set to 0.54, and the mutation probability is set to 0.06. Ribbed plates are subjected to uniformly distributed lateral loads, with a magnitude of 2Pa and fixed boundary conditions on all four sides. Take the computational algebra of genetic algorithm as the second generation and optimize the placement of ribs in the ribbed plate. It is known that under the action of uniformly distributed loads, the optimal placement position for the ribs of a vertical ribbed plate is near the centerline of the plate. The optimization goal is to increase the deflection value of the center point of the ribbed plate, and the constraint is the length and width of the ribbed plate. As a control, the cross shaped ribs distributed at the centerline were calculated, with a deflection value of 7.836×10^{-5} at the center point. The results of hybrid genetic algorithm optimization are shown in Table 4.1 [21,22].

Through calculation, it can be seen that the results of using the hybrid genetic algorithm for two generations are both near the centerline position of the plate, and the error of the deflection at the center point of the cross intersecting rib and the rib position at the centerline of the plate is within an acceptable range, which proves the effectiveness of the hybrid genetic algorithm.

4.2. Optimization of Double Vertical Rib and Ribbed Plate Rib Layout under Local Load. The height of the two vertical ribs of a square with ribs is h_s=0.2m, the width of the ribs is w_S=0.02m, the edge length of the slab is L=2m, and the thickness of the slab is h=0.02m. The elastic modulus of both the plate and ribs is taken as E=18GPa, with a Poisson's ratio of μ=0.4, under the action of a local load (2Pa) of a quarter to the left and bottom of z. The meshless model adopts 8×8 uniformly distributed nodes for discretization, with fixed supports on all four sides. The optimization objective is to obtain the deflection value at the center point of the ribbed plate, with constraints on the length and width of the ribbed plate [23,24]. In the hybrid genetic algorithm, the crossover probability is set to 0.5, the mutation probability is set to 0.06, and the iteration number is two generations [25,26]. When the y-axis rib position is 0.73456 and the x-axis rib position is 0.433293, the deflection at the midpoint of the plate is the smallest (0.90896). The variation of its value is shown in Figures 4.1 to 4.2.

As shown in Figure 4.1, the optimal distribution is in the fourth quadrant. In the second generation, the distribution of numerical points in the fourth quadrant appears to be significantly denser than in the first generation, with more individuals approaching the optimal solution. The deflection values of the center points of each individual in each generation of the population are shown in Figures 4.3 to 4.4. As shown in Figure 4.2, the

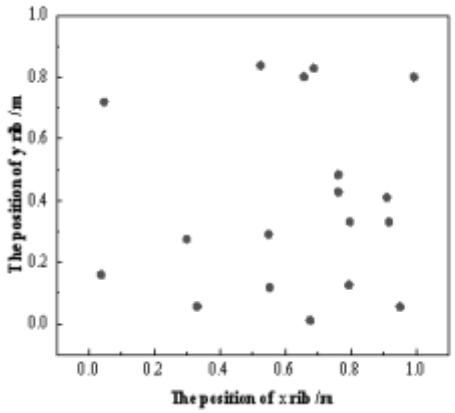


Fig. 4.1: Distribution of rib positions in the first generation of individuals

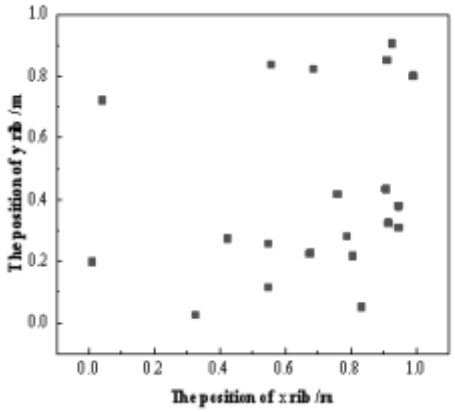


Fig. 4.2: Second generation individual rib position distribution map

deflection values of the center points corresponding to the second generation individuals are more concentrated, and there are also many individuals with smaller deflection values compared to the first generation. The hybrid genetic algorithm is indeed effective and can be gradually improved by adding computational examples in the future.

5. Conclusion. The author calculated the ribbed plate based on the meshless method, taking advantage of the advantage that the meshless method does not require element division to optimize the placement of ribs on the ribbed plate, and proposed a new optimization method. The main advantages and results of this method are as follows:

1. The use of meshless methods does not require reconstruction of the nodes distributed on the ribbed plate surface and ribs, greatly reducing the computational complexity of the optimization process.
2. The genetic algorithm was mixed with the constrained random direction method, and the global search was performed using the genetic algorithm. The local search characteristics of the constrained random direction method were used to optimize the placement of ribs when the ribbed plate was subjected

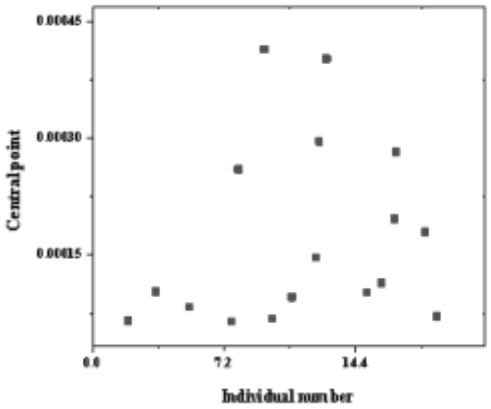


Fig. 4.3: Center point deflection values corresponding to the first generation of individuals

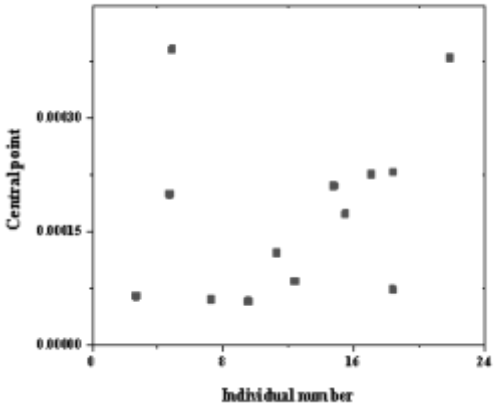


Fig. 4.4: Center point deflection values corresponding to the second generation individuals

to lateral loads, after obtaining the calculation results of the genetic algorithm, further optimization operations can be carried out using the constrained random direction method, which not only greatly reduces the computational complexity of the genetic algorithm, but also improves the computational effectiveness of the constrained random direction method. A simple genetic algorithm requires more computational algebra for optimization operations, while the author's example can achieve better results in fewer computational algebras, proving the effectiveness of the hybrid genetic algorithm.

REFERENCES

[1] Eşki, H., Sayın, B., & Güneş, B. (2020). The effect on structural behavior of different slab types for RC buildings. Journal of Structural Engineering & Applied Mechanics, 3(1), 41-48.

[2] Gong, Y., Shan, Y., Xiang, P., Huang, W., & Ding, F. (2020). Static experimental behavior of peripheral composite slab-beam joints. ACI Structural Journal, 117(4), 255-265.

[3] Mata Falcón, J., Bischof, P., Huber, T., Anton, A., Burger, J. J., Ranaudo, F., ... & Kaufmann, W. (2022). Digitally fabricated ribbed concrete floor slabs: a sustainable solution for construction. RILEM Technical Letters, 7, 68-78.

[4] Hasan, K. F., Fayadh, O. K., & Hasan, Q. F. (2021). Design and analysis of flat and grid slab system with conventional slab comparative approach. Design Engineering, 1(7), 1628-169.
[5] Hasenko, A., Semko, V., & Skliarenko, S. (2023). Calculation of the Constructive Nonlinearity of Ribbed Reinforced Concrete Slabs of Self-Stressed Covering Panels. Slovak Journal of Civil Engineering, 31(3), 12-23.
[6] Albarram, A., Qureshi, J., & Abbas, A. (2020). Effect of rib geometry in steel–concrete composite beams with deep profiled sheeting. International Journal of Steel Structures, 20, 931-953.
[7] Tan, C., Zhang, Y., Zhao, H., Zhang, B., & Du, T. (2022). Study on Shear-Lag Effect of Steel–UHPC Ribbed Slab Composite Structures Using Bar Simulation Method. Buildings, 12(11), 1884.
[8] Ahmad, H., Hashim, M. H. M., & Fozi, N. H. M. (2022). A BRIEF REVIEW OF STEEL FIBER SELF-COMPACTING CONCRETE IN RIBBED SLAB. ASEAN Engineering Journal, 12(3), 39-47.
[9] Patel, V. G., & Rachchh, N. V. (2020). Meshless method–review on recent developments. Materials today: proceedings, 26, 1598-1603.
[10] Rasoulizadeh, M. N., Ebadi, M. J., Avazzadeh, Z., & Nikan, O. (2021). An efficient local meshless method for the equal width equation in fluid mechanics. Engineering Analysis with Boundary Elements, 131, 258-268.
[11] Lin, J., Bai, J., Reutskiy, S., & Lu, J. (2023). A novel RBF-based meshless method for solving time-fractional transport equations in 2D and 3D arbitrary domains. Engineering with Computers, 39(3), 1905-1922.
[12] Ahmad, I., Khan, M. N., Inc, M., Ahmad, H., & Nisar, K. S. (2020). Numerical simulation of simulate an anomalous solute transport model via local meshless method. Alexandria Engineering Journal, 59(4), 2827-2838.
[13] Ahmad, H., Khan, M. N., Ahmad, I., Omri, M., & Alotaibi, M. F. (2023). A meshless method for numerical solutions of linear and nonlinear time-fractional Black-Scholes models. AIMS Math, 8(8), 19677-19698.
[14] Sobhani, E. (2022). Vibrational performance modeling for coupling of a full-ellipsoid shell with a cylindrical shell with a focus on flexibility at coupling and boundary conditions via the GDQ-meshless method. Engineering Analysis with Boundary Elements, 144, 329-351.
[15] Gu, Y., & Sun, H. (2020). A meshless method for solving three-dimensional time fractional diffusion equation with variable-order derivatives. Applied Mathematical Modelling, 78, 539-549.
[16] Rao, X., Zhao, H., & Liu, Y. (2024). A novel meshless method based on the virtual construction of node control domains for porous flow problems. Engineering with Computers, 40(1), 171-211.
[17] Ehsani, A., & Dalir, H. (2021). Multi-objective design optimization of variable ribs composite grid plates. Structural and Multidisciplinary Optimization, 63(1), 407-418.
[18] Ko, K. Y., Solyaev, Y., Lurie, S., Babaytsev, A., Rabinskiy, L., & Kondakov, I. (2023). Theoretical and experimental validation of the variable-thickness topology optimization approach for the rib-stiffened panels. Continuum Mechanics and Thermodynamics, 35(4), 1787-1806.
[19] Liu, Y., Kohrn, M., Wick, M., & Pischinger, S. (2022). Optimization of the bipolar plate rib structure in proton exchange membrane fuel cells with an analytical method. International Journal of Hydrogen Energy, 47(40), 17683-17698.
[20] Kumar, A., Priyanka, Kumar, S., & Kalia, S. (2023). Optimization and correlations development for heat transfer and fluid flow characteristics of ZnO/H2O-ethylene glycol-based nanofluid flow through an inclined ribbed square duct. Numerical Heat Transfer, Part A: Applications, 84(11), 1352-1367.
[21] Barcenas, L., Ledesma-Orozco, E., Van-der-Veen, S., Reveles-Arredondo, F., & Rodríguez-Sánchez, E. A. (2020). An optimization of part distortion for a structural aircraft wing rib: an industrial workflow approach. CIRP Journal of Manufacturing Science and Technology, 28, 15-23.
[22] Savine, F., Irisarri, F. X., Julien, C., Vincenti, A., & Guerin, Y. (2021). A component-based method for the optimization of stiffener layout on large cylindrical rib-stiffened shell structures. Structural and Multidisciplinary Optimization, 64(4), 1843-1861.
[23] Aziza, H., & Krichen, S. (2020). A hybrid genetic algorithm for scientific workflow scheduling in cloud environment. Neural Computing and Applications, 32(18), 15263-15278.
[24] Iranmanesh, A., & Naji, H. R. (2021). DCHG-TS: a deadline-constrained and cost-effective hybrid genetic algorithm for scientific workflow scheduling in cloud computing. Cluster Computing, 24, 667-681.
[25] Tsoulos, I. G., Anastasopoulos, N., Ntritsos, G., & Tzallas, A. (2023). Train RBF networks with a hybrid genetic algorithm. Evolutionary Intelligence, 16(1), 375-381.
[26] Grosch, B., Kohne, T., & Weigold, M. (2021). Multi-objective hybrid genetic algorithm for energy adaptive production scheduling in job shops. Procedia CIRP, 98, 294-299.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 30, 2024
Accepted: Jul 1, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 749–756, DOI 10.12694/scpe.v26i2.4072

MONITORING AND THREE-DIMENSIONAL NUMERICAL ANALYSIS OF ROCK AND SOIL DEFORMATION DURING CONSTRUCTION

JUN WU*

Abstract. In order to solve the problem of large monitoring errors in settlement deformation in construction engineering, the author proposes rock and soil deformation monitoring and three-dimensional numerical analysis under construction. The author proposes a settlement and deformation monitoring method for pile foundation construction in green building engineering. By setting settlement deformation monitoring points and setting the frequency of pile foundation contour monitoring, a contour line tracking edge monitoring model is constructed to obtain monitoring results, and longitudinal contour correction is used to correct monitoring errors, thus achieving settlement deformation monitoring. The experimental results show that compared with the traditional symmetrical settlement deformation monitoring group and the traditional compression observation settlement deformation monitoring group, the monitoring error obtained by the contour line tracking settlement deformation monitoring group designed by the author is controlled below 1.1. The author's monitoring method has better results, higher accuracy, and smaller errors, and has practical application value.

Key words: Contour line tracking, Green buildings, Engineering design, Construction settlement, Deformation monitoring, Monitoring methods

1. Introduction. With the development of modern society, the scale of engineering construction is gradually increasing. At this time, large-scale buildings will generate huge stress due to their own heavy weight during construction, and the stress will have an impact on the foundation rock and soil of large-scale buildings. If the foundation rock and soil undergo large-scale deformation, it will inevitably affect the stability of large-scale buildings. Therefore, in order to ensure the quality of large-scale buildings, many construction units now attach great importance to the application of rock and soil deformation monitoring technology [1]. The most critical indicator for evaluating the stability of soil and rock and whether buildings can be used reasonably is the deformation of soil and rock. The monitoring results can also verify the accuracy and suitability of the calculated parameter values or calculation methods. The monitoring data of deformation can also serve as a basis for the types of reinforcement methods used in actual engineering practice for rock and soil [2,3]. The author analyzes the current status and development trends of this technology in order to understand its specific applications. Due to the complex and diverse engineering characteristics of soil and rock, if there are omissions or incomplete grasp in survey and design, errors can occur and affect the quality of the project. Therefore, for some major projects with poor geological conditions, soil and rock deformation monitoring can timely and effectively detect problems, take appropriate measures, and ensure the normal and orderly progress of the project. This accumulates valuable practical experience, which has important and far-reaching implications for exploring the relevant fields of soil and rock engineering and improving the level of on-site survey tasks [4].

A large number of on-site engineering monitoring studies have shown that deformation monitoring of rock and soil has the characteristics of complex monitoring objects, long cycles, single means, and serious constraints on the on-site construction environment. The current deformation monitoring of rock and soil mainly uses inductive, resistive, and vibrating wire strain gauges, as well as displacement gauges and image scanners and other related measurement equipment. These sensor sensing components are mostly metal materials and mostly point collection, which are constrained by factors such as sensor material, applicable environment, massive data, signal transmission, etc. They have the disadvantages of easy corrosion failure, low survival rate, poor durability, low accuracy, real-time performance, and low level of intelligent monitoring, and can no longer meet the needs of complex rock and soil deformation monitoring [5,6]. Therefore, global researchers continue to

*Jiangxi Environmental Engineering Vocational College, Ganzhou, Jiangxi, 341000, China (Corresponding author, JunWu783@126.com)

explore and develop innovative technologies for monitoring rock and soil deformation. Based on the advantages of monitoring distance and accuracy, distributed fiber optic sensing technology has become the focus of attention for scholars worldwide. In terms of intelligent processing of monitoring data, as some rock and soil deformation monitoring technologies have achieved full real-time monitoring, how to intelligently process the collected massive data and quickly and accurately obtain key information on rock and soil deformation instability has become a bottleneck that urgently needs to be overcome in the development of monitoring technology. Artificial intelligence algorithms can be used to clean and extract massive data, using clustering, fuzzy algorithms, association rules, etc. for data mining, and combining machine learning to establish a deformation trend prediction model for rock and soil. This will be an important research direction for future monitoring and early warning of rock and soil (geological) engineering [7,8].

2. Literature Review. Under the guidance of sustainable development social goals, the number of green buildings is increasing year by year, which will also have a certain impact on traditional engineering buildings and processes. Usually, in the actual construction process, most construction workers pay more attention to controlling the bearing capacity of pile foundations to avoid a series of problems such as construction settlement and deformation to the greatest extent possible. However, due to the novel construction structure and model of green buildings, traditional construction methods are not applicable, which leads to some projects having building defects and being unable to continue advancing, forming obstacles [9]. Therefore, a more flexible monitoring method for settlement and deformation of construction pile foundation should be constructed by combining contour line tracking technology. Ge, C. and others studied a certain tunnel and used the discrete element method to simulate the construction steps of the three-step method and single-sided excavation method with and without systematic anchor support. Analyzed the results of tunnel surrounding rock stress, vertical displacement, and surface deformation under different working conditions, and analyzed the mechanism of system anchor support from a microscopic perspective. The results show that the single-sided excavation method can gradually release the load and deformation, better exerting the support capacity of the lining; The presence of system anchor rods improves the shear resistance of the surrounding rock, enhances the arching effect, and significantly reduces the vertical displacement and surface deformation of the surrounding rock [10]. Li, Z. and others studied the deformation and failure characteristics and acoustic emission response of concrete under dynamic loads under different soaking times. Revealed the failure mechanism and precursor characteristics of acoustic emission response of concrete under drop hammer impact [11]. Yu, X. et al. investigated the relationship between strain energy, volume fraction (VF), confining pressure, and failure mode of backfill surrounding rock through triaxial compression tests, acoustic emission (AE), and microscopic tomography (CT). The research results indicate that total strain energy and elastic energy are positively correlated with two factors (VF and constraint) [12].

The author combines contour tracking technology to construct a construction settlement deformation monitoring method, which measures the actual construction deformation or settlement status in complex environments. Reduce the probability of construction accidents, strengthen the control of the project after completing pile foundation construction, take multiple measures to address deformation or settlement issues, and minimize the increase of additional stress, laying the foundation for subsequent construction.

3. Method.

3.1. Application of three-dimensional numerical analysis. With the continuous improvement of modern technology, more and more monitoring technologies will be developed to provide practical needs. In the future, the engineering safety monitoring information management and monitoring data network system can also be applied to the management and processing analysis of engineering monitoring data, achieving remote real-time sharing of monitoring data and networked management and analysis, this is sufficient to reduce the labor intensity of manual data analysis, improve labor production efficiency, increase mechanized applications, reasonably avoid errors caused by human factors, and enable monitoring results to be timely and accurately returned to relevant survey and design personnel. It plays an important role in avoiding design and construction risks, ensuring construction progress and operational safety, and accurately grasping dynamic changes [13,14]. The combination of rock and soil deformation monitoring technology and computer application operation is relatively close. The signal information obtained through electric pulse signal feedback

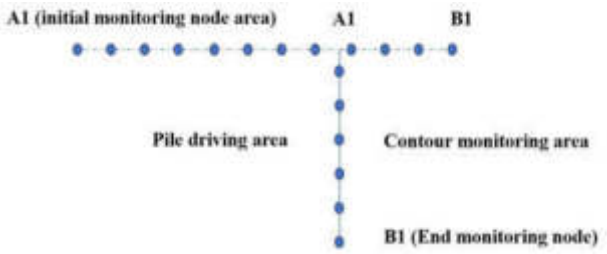


Fig. 3.1: Diagram of pile foundation contour monitoring and pile driving area

is not directly information data that can be translated by humans [15]. Therefore, it is necessary to use the conversion display function of computer technology for data processing, so that the electric pulse signal can be manually interpreted, in order to accurately and accurately explore the relevant changes in rock and soil deformation characteristics.

3.2. Layout of settlement deformation monitoring points. Usually, in order to improve the overall quality and effectiveness of green building construction, a certain number of additional measurement nodes are selected in certain areas, with the aim of obtaining numerical information of buildings in corresponding positions and providing theoretical basis for subsequent construction [16]. In order to meet the construction requirements of green building projects, it is necessary to first delineate the corresponding monitoring area and calculate the unit monitoring distance, as shown in formula 3.1.

$$H = (0.5k - \sum_{i=1} b^2 i \times \beta_1) + \beta_2 \tag{3.1}$$

In formula 3.1: H represents the monitoring distance of the unit, k represents the observation range, i represents the number of observation markers, b represents the depth of error, β_1 represents the preset number of nodes, and β_2 represents the measured number of nodes. Through the above calculation, the actual unit monitoring distance can be obtained. According to the specific monitoring scope defined, the fuzzy settlement position of the engineering construction is clarified, and the specific settlement points of the pile foundation are marked using precision leveling observation method and directional settlement monitoring method, arranged symmetrically in pairs. At this point, it is necessary to drill holes at the edge of the pile foundation, bury corresponding quantities of Φ25 type steel bars at the marked positions, and use high-strength concrete for reinforcement. At the same time, the monitoring device should be placed according to the number of deployed nodes, linked to the monitoring platform, to achieve the deployment of settlement monitoring nodes.

3.3. Frequency setting for pile foundation contour monitoring. After completing the layout of settlement monitoring points, the corresponding frequency of pile foundation contour monitoring should be set according to the actual engineering construction needs [17]. The monitoring frequency is a monitoring program set for the current construction situation, which is generally not fixed and changes accordingly with the progress of construction. The management level is level four. Firstly, determine the local pile driving area, as shown in Figure 3.1.

According to Figure 3.1, the setting of the pile foundation contour monitoring and driving area can be completed. At the same time, within the calibrated range, corresponding indicator parameters are set according to different pile driving soil layers, as shown in Table 3.1.

According to Table 3.1, it is possible to set the monitoring index parameters for the pile foundation profile. Next, based on the acquisition and summary of monitoring values and information, a specific frequency of change is set. When the equivalent value of the compression modulus is 0.972, the monitoring frequency is most balanced. Based on the empirical coefficient of pile foundation settlement, the standard monitoring mean is calculated, as shown in equation 3.2.

$$u = \alpha^3 + \int (m+n) \times \frac{m_1 m_2 - 1}{2} \tag{3.2}$$

Table 3.1: Setting of Monitoring Index Parameters for Pile Foundation Profile

Building monitoring of soil layers	depth/m	Reverse monitoring frequency ratio
Silty clay	10.14	1.02
Silty sand	11.23	1.25
Clay	11.56	1.16

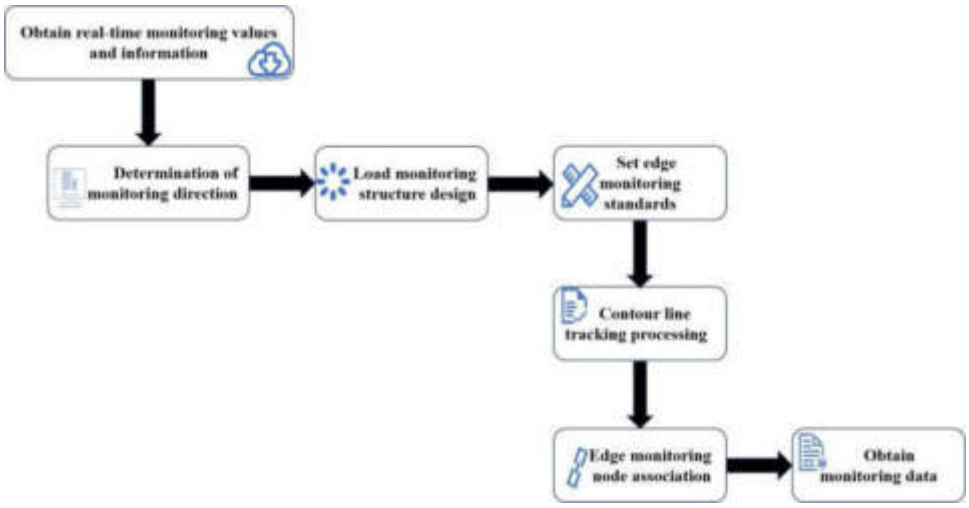


Fig. 3.2: Structural diagram of contour line tracking edge monitoring model

In Formula 3.2: u represents the monitoring mean, α represents the settlement coefficient, m_1 represents the monitoring distance of unit markers, m_2 represents the compression distance, and n represents the error distance. Through the above calculation, the actual monitoring mean can be obtained, and the specific balance monitoring frequency can be achieved.

3.4. Construction of contour line tracking edge monitoring model. After setting the monitoring frequency for the pile foundation contour, it is necessary to combine contour line tracking technology to construct a more diverse and flexible edge settlement deformation monitoring model. Edge monitoring is actually a means of obtaining numerical values, information, and images. Based on the changes in the grayscale values of the scenery and background during construction, associate the corresponding monitoring nodes set up. At the same time, set more intuitive and quantitative edge monitoring points based on the real-time data transmitted back from the monitoring nodes, combined with contour tracking technology. At this point, the variation vector ratio for measuring building settlement deformation is generally controlled between 1.03 and 5.2. Using C50 prefabricated concrete slabs, they are placed at the location of settlement deformation, and a total of 5 pressure bearing segments need to be set longitudinally to achieve multi-dimensional assembly monitoring. Subsequently, in order to expand the detection range of green buildings, the monitoring specifications can be adjusted and a specific monitoring model can be constructed, as shown in Figure 3.2.

According to Figure 3.2, the design of the contour line tracking edge monitoring model structure can be completed. Use the model to plan the corresponding monitoring structure, and at the same time, use monitoring nodes to achieve short distance monitoring control and obtain monitoring results. However, it should be noted that for local construction monitoring areas, centralized monitoring should be adopted to maximize the overall monitoring accuracy and better integrate with contour line tracking technology. With the assistance of edge monitoring points, a multi-dimensional contour line tracking edge monitoring model should be further constructed. Based on this, as the monitoring area changes, it is necessary to adjust the corresponding monitoring nodes at a distance of 300cm. Can complete the design and adjustment of the

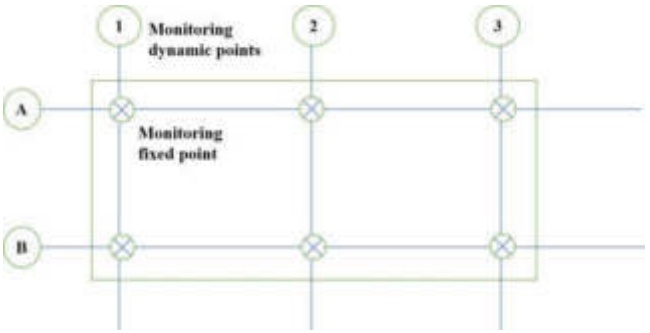


Fig. 3.3: Layout diagram of vertical contour correction piles

monitoring node position adjustment structure. Subsequently, within the determined green building, a related auxiliary framework is constructed, with specific monitoring cycles set. Based on the settlement and deformation status of the building, the monitoring position and effective area are adjusted, thereby optimizing and improving the overall ability of the monitoring model and expanding the actual monitoring range.

3.5. Longitudinal contour correction for settlement deformation monitoring. After completing the design of the contour line tracking edge monitoring model, combined with the longitudinal contour correction method, better settlement deformation monitoring can be achieved [18]. Using contour line tracking method for multi-dimensional monitoring correction processing. Using the WHPB platform, set 6 steel pipe piles at the designated positions, and each steel pipe pile needs to be equipped with a corresponding monitoring device and correlated with each other. The size of the pile type is D1.24m x 123.82m. The monitoring piles need to be inclined towards the interior of the building platform, with a depth of 101m. The actual depth is 90.51m, and the exposed length of a single pile is about 20.31m. The monitoring pile position is set, as shown in Figure 3.3. According to Figure 3.3, the layout and construction of longitudinal contour correction piles can be completed [19]. At the same time, combining contour line tracking technology, a longitudinal correction monitoring program is set up, a monitoring model is adopted, and a cyclic monitoring mechanism is formed. As the monitoring nodes change, the monitoring area is also constantly adjusted, achieving multi-dimensional dynamic correction monitoring.

3.6. Experimental preparation. The main purpose of this study is to analyze and study the practical application effect of the settlement and deformation monitoring method for green building pile foundation construction under contour line tracking technology [20]. Considering the accuracy of the final test results, it is necessary to conduct analysis in the same testing environment, setting up a traditional symmetrical settlement deformation monitoring group, a traditional compression observation settlement deformation monitoring group, and the contour line tracking settlement deformation monitoring group designed by the author. The final test results need to be discussed in a comparative form to ensure their reliability. Before analyzing and studying the practical application effect of the settlement deformation monitoring method for green building pile foundation construction under contour line tracking technology, a testing environment needs to be established. Firstly, set the edge contour line and use the Sobel operator to calculate the dynamic ratio of edge monitoring at the calibrated position, as shown in equation 3.3.

$$C = \sqrt{k^3} + \sum_{q=1}(\theta \times \frac{q^3}{f_1 f_2}) - 1 \tag{3.3}$$

In formula 3.3: C represents the dynamic ratio of edge monitoring, k represents the number of operators, θ represents the measured monitoring range, q represents the number of monitoring cycles, f_1 represents the preset monitoring range, and f_2 represents the actual monitoring range. Through the above calculation, the actual dynamic ratio of edge monitoring can ultimately be obtained. Based on the relevant monitoring platforms and nodes, set specific dynamic monitoring programs and use models to determine their applicability and

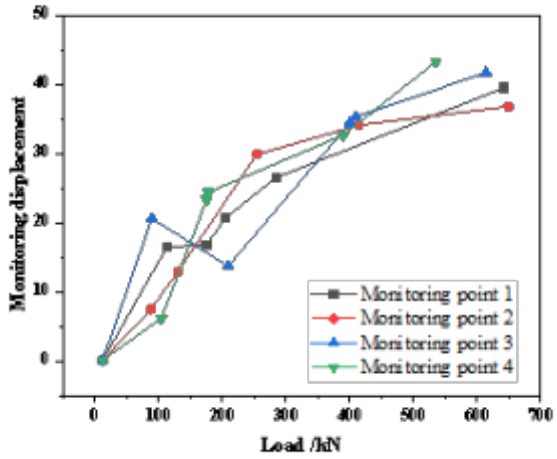


Fig. 3.4: Analysis of settlement and deformation at measurement points during construction

error probability. If within a reasonable range, it can ensure the final monitoring results. Subsequently, it is necessary to select D building as the main target object for testing, and obtain specific data and information for measurement to form a complete monitoring reference. By combining contour line tracking technology to obtain the original monitoring images, installing Gaussian space filters to help the equipment obtain monitoring signals more smoothly, setting zero crossing monitoring points, and combining with actual settlement deformation situations, calculating the specific unit monitoring distance, which generally needs to be controlled between 5.5m and 10.5m. After verifying whether the tested device is in a stable state, ensuring that there are no external factors interfering with it, and ensuring that it is in good condition, specific testing can begin.

Based on the testing environment built above, combined with contour tracking technology, conduct specific testing analysis. When the monitoring area is in the optimal smooth state, a second-order edge operator is used to construct a multi-dimensional construction settlement deformation monitoring matrix. By using cyclic detection, the monitored values and information are measured, and the binarization threshold is calculated, as shown in equation 3.4.

$$r = \frac{\int \zeta - \sqrt{t}}{0.2\eta^2 + \epsilon} - \chi^3 \tag{3.4}$$

In formula 3.4, r represents the binarization threshold, η represents the settlement depth, ζ represents the elevation distance, represents the offset length of the contour line, t represents the node difference, and χ represents the interference range.

3.7. Results and Discussion. Through the above calculation, the actual binarization threshold can be obtained. According to the changes in settlement deformation, the binarization threshold obtained will also change. Four main monitoring nodes are set, combined with the Roberts operator monitoring mode and changes in load force, to determine the actual monitoring situation, as shown in Figure 3.4.

According to Figure 3.4, the analysis and research on the settlement and deformation of the measurement points during construction can be completed. Based on professional equipment and monitoring nodes, obtain corresponding test values and information. At the same time, conduct the same test in different areas to determine the detection errors of four monitoring nodes. The test results are shown in Table 3.2.

According to Table 3.2, the analysis of the test results can be completed: compared with the traditional symmetrical settlement deformation monitoring group and the traditional compression observation settlement

Table 3.2: Comparison and Analysis of Test Results

Monitoring nodes	Monitoring error of traditional symmetrical settlement deformation monitoring group	Monitoring error of traditional compression observation settlement deformation monitoring group	Monitoring error of the settlement deformation monitoring group for tracking the contour line
Monitoring point 1	2.54	3.046	1.01
Monitoring point 2	3.03	3.16	1.04
Monitoring point 3	2.07	2.01	1.00
Monitoring point 4	3.41	2.45	1.08

deformation monitoring group, the monitoring error obtained by the contour line tracking settlement deformation monitoring group designed by the author is controlled below 1.1. Indicating that this monitoring method has better effect, high accuracy, small error, and practical application value.

4. Conclusion. The author proposes the monitoring and three-dimensional numerical analysis of rock and soil deformation under construction. Combining contour line tracking technology, the author constructs a settlement deformation monitoring method for green building engineering pile foundation construction, and analyzes and studies the practical application effect. For the construction of green buildings, the mode is relatively novel, so it is necessary to construct a monitoring mode with stronger characteristics, in complex environments, monitoring errors can be minimized as much as possible, and cyclic and multi-objective settlement and deformation monitoring methods can be constructed to achieve comprehensive monitoring and analysis. In addition, within a reasonable range, simplify the construction monitoring process as much as possible, save monitoring costs, and create a more stable monitoring environment to improve the construction speed and quality effect of the building.

REFERENCES

[1] Chen, B. G. , & Jia, Z. P. . (2023). Optimal strut position of deep foundation pit with convex corner under surcharge of adjacent building. Rock and Soil Mechanics, 44(8), 2400-2408.
[2] Carbonell, J. M. , Lluís Monforte, Ciantia, M. O. , Arroyo, M. , & Gens, A. . (2022). Geotechnical particle finite element method for modeling of soilstructure interaction under large deformation conditions. Journal of Rock Mechanics and Geotechnical Engineering: English Edition, 14(3), 17.
[3] Zhao, T. , Zhang, P. , Xiao, Y. , Guo, W. , Zhang, Y. , & Zhang, X. . (2023). Master crack types and typical acoustic emission characteristics during rock failure. International Journal of Coal Science and Technology: English Edition, 10(1), 73-86.
[4] Pan, J. J. , Sun, X. J. , Zuo, Y. Z. , Wang, J. P. , Yi-Wei, L. U. , & Han, B. . (2023). Effects of skeleton void ratio on the strength and deformation characteristics of coarse-grained soil. Rock and Soil Mechanics, 44(8), 2186-2194.
[5] Pan, J. J. , Sun, X. J. , Zuo, Y. Z. , Wang, J. P. , Yi-Wei, L. U. , & Han, B. . (2023). Effects of skeleton void ratio on the strength and deformation characteristics of coarse-grained soil. Rock and Soil Mechanics, 44(8), 2186-2194.
[6] Liu, X. J. , Yang, K. , Guo, F. , Tang, S. Q. , Liu, Y. H. , & Zhang, L. , et al. (2022). Effects and mechanism of igneous rock on selenium in the tropical soil-rice system in hainan province, south china. Chinese Geology, 5(1), 11.
[7] Ammirati, L. , Martire, D. D. , Bordicchia, F. , Calcaterra, D. , Russo, G. , & Mondillo, N. . (2022). Semi-real time systems for subsidence monitoring in areas affected by underground mining: the example of the nuraxi-figus coal district (sardinia, italy). International Journal of Coal Science & Technology, 9(1), 1-15.
[8] Cheng, T. , Guo, B. H. , Sun, J. H. , Tian, S. X. , Sun, C. X. , & Chen, Y. . (2022). Establishment of constitutive relation of shear deformation for irregular joints in sandstone. Rock and Soil Mechanics, 43(1), 51-64.
[9] Gupta, S. K. . (2023). Needs of green buildings innovation and implementation flexibility for prompt effectiveness in developing countries. Civil Engineering and Architecture: English Version, 17(7), 358-371.
[10] Ge, C. , Su, L. , Wang, L. , Xu, S. , Yu, P. , & Tao, Z. . (2022). Discrete element simulation and monitoring analysis of different construction methods of the shallow buried bias tunnel. Advances in civil engineering(Pt.13), 28(4), 335-346.
[11] Li, Z. , Li, X. , Yin, S. , Lei, Y. , Tian, H. , & Niu, Y. . (2024). Deformation failure and acoustic emission response characteristics of water-containing concrete under impact load. Construction & Building Materials(Jan.19), 412.
[12] Yu, X. , Song, W. , Tan, Y. , Kemeny, J. , & Wang, J. . (2022). Energy dissipation and 3d fracturing of backfill-encased-rock under triaxial compression. Construction & Building Materials(Jul.25), 3(1), 40-48.
[13] Qiu, M. , Yang, G. , Zhang, P. , & Duan, J. . (2022). Field test on the construction deformation characteristics for a loess highway tunnel at the shallow portal section. Hydrogeology & Engineering Geology, 48(3), 135-143.

[14] Golpaygani, A. T. , & Parand, F. A. . (2022). Design and development a mobile monitoring system for improving the occupational health and safety on road construction sites. Journal of Clinical Engineering, 47(1), 20-26.
[15] Yao, L. I. , & Jia-Ping, L. I. . (2023). Multi-directional cyclic simple shear behaviour of loose sand under complex initial stress states. Rock and Soil Mechanics, 44(9), 2555-2565.
[16] Rui, S. , Juanjuan, T. , Xin, Y. , & Xingyi, L. I. . (2023). Application research and case analysis of green technology in building engineering under the direction of energy conservation demands. Landscape research: English version, 15(2), 1-5.
[17] Jiang, S. , Zhu, Y. , Qing, L. I. , Zhou, H. , Hong-Liang, T. U. , & Yang, F. J. . (2022). Dynamic prediction and influence factors analysis of ground surface settlement during tunnel excavation. Rock and Soil Mechanics, 43(1), 195-204.
[18] Nan, Y. , Zhen, W. , Jun, Z. , Xueqiong, Z. , & Hai, X. . (2023). Unsupervised model-driven neural network based image denoising for transmission line monitoring. Optoelectronic Express: English version, 19(4), 248-251.
[19] Zhao, W. , Xiao, J. , Liu, S. , Dou, S. , & Liu, H. . (2023). Robotic direct grinding for unknown workpiece contour based on adaptive constant force control and human–robot collaboration. Industrial Robot: the international journal of robotics research and application, 50(3), 376-384.
[20] Tang, X. W. , Lin, W. K. , Zou, Y. , Liang, J. X. , & Zhao, W. F. . (2022). Experimental study of the bearing capacity of a drainage pipe pile under vacuum consolidation. Journal of Zhejiang University (English Edition) Series A: Applied Physics and Engineering, 23(8), 13.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 30, 2024
Accepted: Jul 15, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 757–765, DOI 10.12694/scpe.v26i2.4074

THE INCOME ALLOCATION MECHANISM OF TRUSTED "DUAL CONTRIBUTION" DATA ASSETS BASED ON BLOCKCHAIN TECHNOLOGY

QIANHUI CHEN*, WEIBIN DING†, AND HUAQIANG SHEN‡

Abstract. In order to solve the problem of researching the distribution of returns on trustworthy "dual contribution" data assets, the authors propose a research on the distribution mechanism of returns on trustworthy "dual contribution" data assets based on blockchain technology. Integrate and apply various levels in the blockchain system. In this model, a node authorization control mechanism is added in the network layer, which can be customized in the consensus layer, and the data query efficiency can be improved through the optimization of the structure and the establishment of the index, so that the data can be managed intelligently in the intelligent contract layer, and encrypts information with customizable encryption algorithms at the transaction layer. In simpler terms, the new data asset management model utilizing blockchain technology has made on-chain data queries 2.25 times faster compared to conventional models. The benefit distribution mechanism model of blockchain is more conducive to internal cooperation than the traditional model.

Key words: Data resources, Blockchain, Benefit distribution

1. Introduction. As technology keeps evolving, we're witnessing the dawn of the financial technology era. Innovations like big data, cloud computing, blockchain, and artificial intelligence are increasingly finding their way into different corners of the economy, reshaping traditional models including supply chain finance [1]. Despite the advancements in supply chain finance, small and medium-sized enterprises (SMEs) continue to grapple with the challenge of accessing affordable financing. This is primarily due to their limited resources, modest scale, and lack of established credit history. However, blockchain technology offers a promising solution. With its decentralized nature and features like encrypted, transparent, and tamper-proof data, blockchain presents a viable avenue for addressing SMEs' financing woes [2].

Reasonable distribution of benefits in research and development alliances is of great significance for maintaining their sustainability and stability [3-4]. When distributing benefits, it is necessary to first clarify the distribution principles and influencing factors to make the distribution plan more reasonable. From the perspective of profit distribution, the prerequisite for maintaining supply chain partnerships is that the total profits obtained by each company through establishing supply chain partnerships are greater than the total profits obtained by each company when they do not cooperate. Secondly, individual enterprises will gain more profits in long-term cooperation. If the distribution of benefits in the alliance does not reach a level of fairness and justice, members are likely to be dissatisfied for this reason, leading to the leakage of intelligence and resources within the alliance, which will have a huge impact on the alliance [5]. The contribution of a company to the supply chain also affects its competitiveness in the market. If members of the company can leverage their advantages, improve their business models, regulate their behavior, and enhance their competitiveness, they can enhance their advantages and improve their overall efficiency. Therefore, investing more members can earn more profits [6]. The chaos of multiple entities in the supply chain is quite common. It is integrated into the blockchain technology support supply chain supervision function, plays a leading role, conducts in-depth monitoring on the centralized allocation, procurement, operation, financing, and provides a strong guarantee for the promotion of blockchain technology in the supply chain [7]. Although blockchain investment requires an increase in alliance operating costs, it can enhance the competitiveness and sustainable development ability

*State Grid Zhejiang Electric Power Investment & Operation Co., Ltd., HangZhou, ZheJiang, 310007, China (Corresponding author, QianhuiChen6@163.com)
† State Grid ZheJiang Electric Power Co.,Ltd., HangZhou, ZheJiang, 310007, China (WeibinDing@126.com)
‡ State Grid ZheJiang Electric Power Co.,Ltd. JiaXing Electric Power Supply Company, JiaXing, ZheJiang, 314033, China (HuaqiangShen@126.com)

of the alliance. If alliance members actively introduce blockchain technology to bear the corresponding costs, they should receive more benefit distribution [8].

2. Literature Review. In the System of National Accounts (SNA), the application scope of the present value of future income method is relatively narrow, and its priority is ranked after the market price method and cost method [9]. This is because the future income flow of assets is highly uncertain, which may lead to unstable asset valuation results, and the valuation process is more complex than the cost method. When valuing data assets, asset returns can be stripped from the operating earnings of institutional units, eliminating the uncertainty of future asset revenue streams; When estimating asset returns, methods such as layered measurement and moving average can be used to reduce the instability of future revenue streams for data assets. In addition, in practice, the fluctuation of asset returns is a normal phenomenon, and asset market prices will also fluctuate with the fluctuation of asset returns. Therefore, it is reasonable for the valuation results of assets to have fluctuations. Therefore, although there are difficulties in valuing data assets using the present value of future income method, these difficulties can be overcome. Lyimo, A. J., and others introduced a better blockchain based IoT design that can display distributed control of IoT data collection, recording, auditing, and storage[10]. Josphineleela, R. et al. have introduced a novel optimized fuzzy architecture tailored for facilitating the exchange and communication of data among Internet of Things (IoT) devices through blockchain technology [11]. Taliep, N. et al. have proposed a method for managing private data stored on the blockchain. By using this method, even if the data is only available to authorized users, the transparency and tamper resistance of blockchain logs are maintained [12].

The present value of future income method reflects the profitability of data assets and demonstrates the advantages of data element income, which cannot be replaced by other methods. Using the present value of future income method to value data assets is a beneficial attempt to promote reasonable pricing of data assets and improve the distribution system of element income. The author will explore how to use the present value of future income method to value data assets based on the National Accounts system and relevant statistical accounting standards. Firstly, define the data assets discussed by the author and discuss their identification and classification issues; Secondly, explore the valuation approach of future income present value method for different types of data assets; Finally, analyze the forms of expected return functions and expected lifespan for different data. Therefore, the author provides practical ideas for the valuation of data assets, and expands the accounting and valuation of data assets.

3. Method.

3.1. Data Management, Data Resource Management, and Data Asset Management. As we move into the Big Data Age, data management is becoming more and more important. There are three major levels of data management at the moment: Basic Data Management, Resource Management, and Asset Management. Data encompasses symbols that capture and represent real-world events, including physical symbols or combinations thereof that describe the characteristics, states, and connections among objects. Data management refers to the use of computer technology to store, process, and utilize these data in an organized and efficient manner. Traditionally, data management has focused primarily on the physical aspects of data, emphasizing the structure of stored data and the relationships between different datasets. At this level of management, data is often seen merely as a vessel for information, with little attention given to the intricate relationships between data points. The International Association for Data Management characterizes data management as the governance of data resources, a concept closely aligned with data resource management. However, while data management tends to emphasize data processing, data resource management places greater emphasis on leveraging data for decision support. In simpler terms, data resource management builds upon the foundation of data management, focusing more on utilizing data for informed decision-making rather than just processing it [13].

Data asset management encompasses both data management and data resource management, but with a distinct focus on the strategic value of data to enterprises. Essentially, not all data is considered a valuable asset; only data resources that can contribute to future economic benefits qualify as such under data asset management principles. The relationship between data management, data resource management, and data asset management is shown in Figure 1. Enterprises must own and control these data assets. Data management primarily concerns

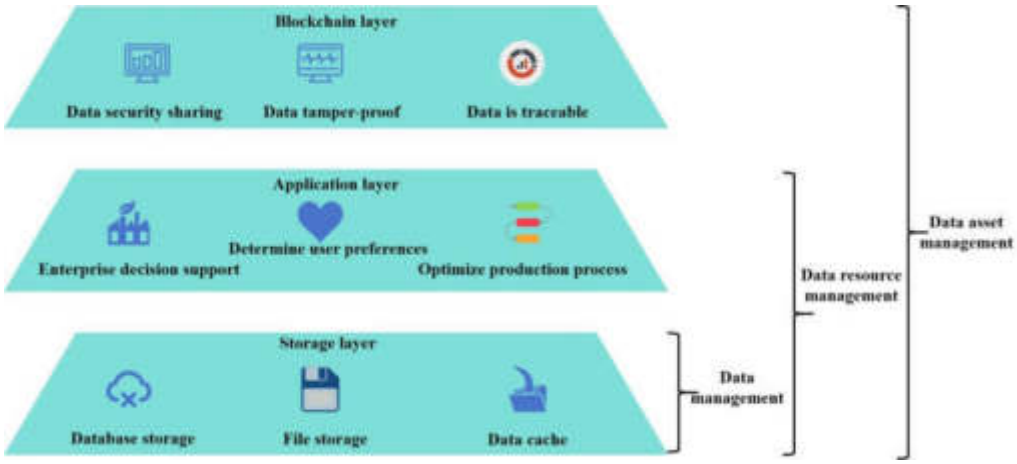


Fig. 3.1: Data Management Hierarchy

itself with the storage layer, encompassing tasks such as data storage and caching. On the other hand, data resource management goes beyond storage, recognizing data as a valuable resource in its own right. This perspective extends to the application layer, where the emphasis lies on leveraging data resources to drive decision-making, optimize production processes, and more effectively support enterprise objectives. In simpler terms, while data management deals with storing data, data resource management sees data as a resource to be utilized, and data asset management specifically focuses on maximizing the economic benefits derived from these valuable data resources. Data asset management represents a paradigm shift, elevating data from mere information to a strategic asset. It entails treating data as a unique asset class, capable of not only enhancing internal operations but also serving as a valuable commodity that can be traded for revenue generation. In this framework, data asset management goes beyond traditional data management and resource management by incorporating considerations of data security, privacy, and sharing mechanisms. An emerging approach in this regard involves leveraging blockchain technology. By harnessing blockchain's inherent features such as secure sharing, tamper resistance, and traceability, data assets can be safeguarded and effectively managed [14].

3.2. Data assetization process. In essence, not all data resources possess the potential to become data assets. The process of converting these resources into assets, thereby extracting value from them, is termed data resource assetization. Figure 3.2 illustrates the specific steps involved in this process. During the production phase, enterprises inevitably generate data resources linked to their products. Subsequently, these resources can be transformed into data assets through two primary methods. The first scenario entails data possessing inherent value. For instance, monetizing such data resources directly under appropriate and lawful conditions represents the most straightforward approach. Conversely, in the second scenario, the data itself might not hold standalone value. However, it can empower existing business operations. For example, various applications utilize data mining techniques to analyze user behavior, yielding insights into user needs. By scrutinizing these data resources, production and business methodologies can be refined, indirectly enhancing the revenue generated by current products. This constitutes an indirect avenue for monetizing data resources[15].

3.3. A New Data Asset Allocation Model Based on Blockchain Technology. The current DQM framework within the blockchain is only focused on some aspects, and there is considerable scope to improve the full integration of blockchain and data asset management. Hence, the author suggests a new model of data asset management utilizing blockchain technology. The new model not only makes use of specific layers in the blockchain, but also integrates and optimizes all the layers of the network, the agreement, the data, the intelligent contract, and the application. Figure 3.3 gives an overview of the proposed concrete framework. The security and sharing of data are enhanced at the network and the consensus level by the node authorization classification and the custom consensus mechanism. In the data layer, we can improve the efficiency of data

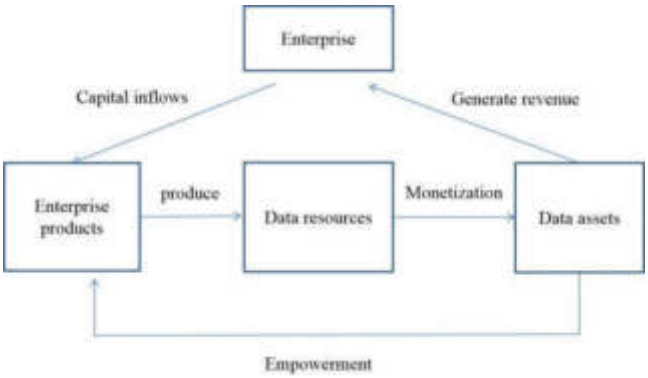


Fig. 3.2: Data assetization process

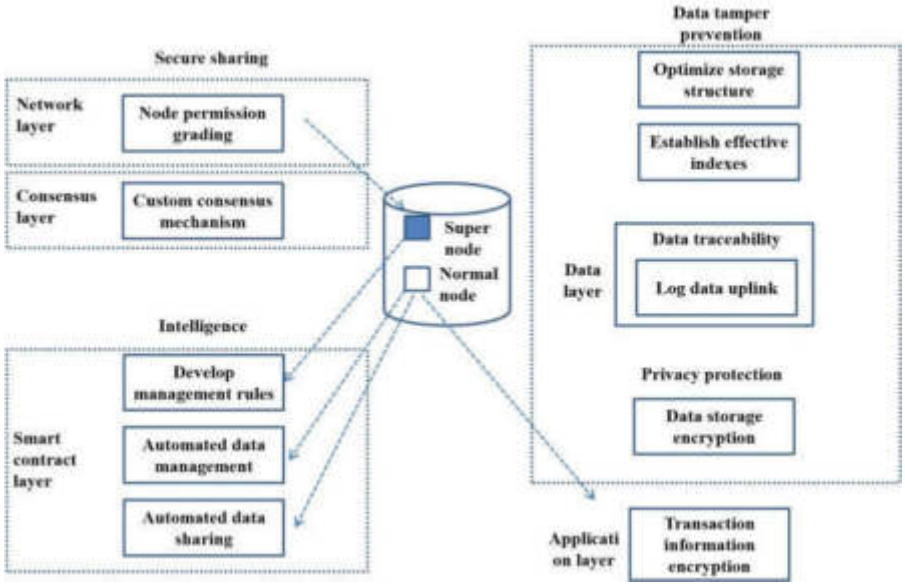


Fig. 3.3: A new model framework for data asset management based on blockchain

query through optimization of memory structure and query methods, including establishing efficient indexes. Additionally, data traceability is ensured through log data chaining, while privacy is maintained via encryption algorithms applied to data. In the Intelligent Contract Layer, the data management is partially automated by deploying Intelligent Contract Program Segments. Finally, security measures are put in place to protect sensitive user information, including transaction data encryption.

In the traditional blockchain data layer, data is organized using a linked list structure for block connection and a Merkle tree structure for data storage within blocks. This design grants blockchain its hallmark feature of resistance to data tampering. Moreover, the inclusion of a timestamp field ensures temporal coherence, allowing for the restoration and tracing of all historical operations based on immutable data. However, traditional blockchains feature a single data structure and straightforward query methods, resulting in suboptimal efficiency for on-chain data queries. In data asset management scenarios, where frequent data interactions occur, high query efficiency is paramount. Thus, in the data layer of the new data asset management model, blockchain technology is employed to uphold data tamper resistance while enhancing efficiency[16,17]. Figure 3.4 illustrates

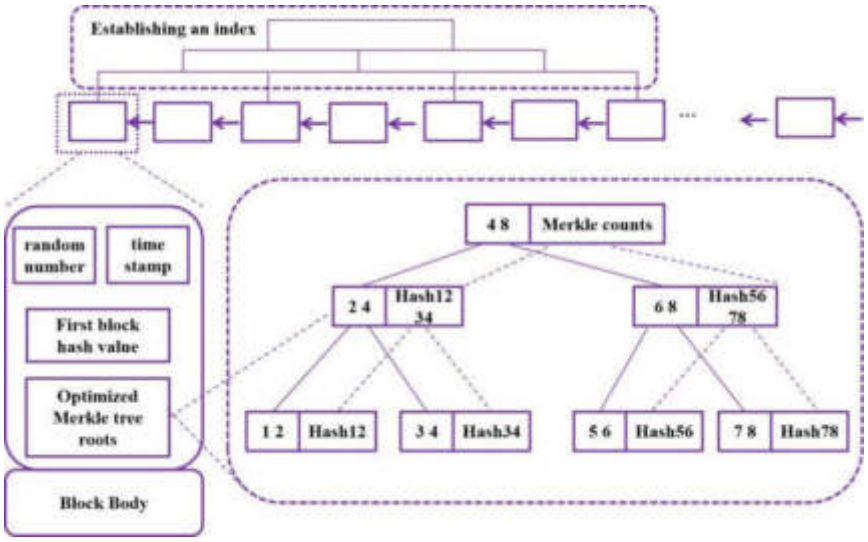


Fig. 3.4: Schematic diagram of data layer optimization

strategies for optimizing the blockchain data structure to enhance query efficiency without compromising data integrity. One approach involves modifying the Merkle tree structure within blocks and integrating it with other balanced tree structures. This hybrid structure ensures that data remains resistant to tampering while improving query efficiency. Furthermore, without altering the fundamental chain structure of the blockchain, query indexes like skip tables can be implemented. These indexes streamline data retrieval processes, making on-chain queries more efficient. Moreover, uploading log-type data onto the blockchain, which is inherently resistant to tampering, facilitates the traceability of all historical operations. This feature proves invaluable in defining responsibility for unexpected occurrences during data asset management and circulation processes.

In traditional blockchain systems, the smart contract layer ensures automated and trustworthy operations, free from human intervention. These operations are traceable and irreversible, thus maintaining the automated nature of the blockchain technology. In the new Data Asset Management Model, officials create a common transaction rule by making use of an automatic script code. These agreements also allow for authentication and verification of data by super nodes in the network, which guarantees the legality of data asset sharing transactions. Users, both official and ordinary, utilize smart contracts to automate data asset management processes. They can also facilitate automated data asset sharing among themselves. Similar to traditional blockchain systems, deploying smart contracts in the new model is treated as a special transaction on the chain. In this model, however, the super-nodes package their published smart contracts into blocks and upload them to the blockchain for deployment. On the other hand, contracts from normal nodes must be verified by super nodes before deployment on the blockchain. This ensures the integrity and security of the deployed contracts within the data asset management framework [18].

In the application layer of traditional blockchain, transactions can be encrypted using asymmetric encryption algorithms, making them highly private. In the scenario of data asset management, nodes participating in data exchange need to protect sensitive information during their transaction process, while data owners need to protect their own data assets. Therefore, in the new model of blockchain based data asset management, it is not only necessary to support the encryption of sensitive information in transactions, but also to support data encryption in the process of data asset sharing. For sensitive information in transactions, encryption algorithms can be used directly for encryption, and only encryption nodes can access the transaction information they participate in. For the shared data assets, the data owner needs to encrypt them using the public key related to the data asset sharee, and share them with the data asset sharee. In the new blockchain-based data asset management model, the recipient of a data asset utilizes their private key for decryption before

Table 3.1: Data Income Changes Table

Project	Revenue generated in the first year of data collection	Revenue generated in the second year of data collection	Revenue generated in the third year of data collection	.....
Data collected in phase T		-	-	-
Data collected in T-1 period			-	-
Data collected in T-2 phase				-
.....				

accessing the content of the asset. Furthermore, the model allows ordinary nodes to employ their encryption algorithms for managing data assets. This additional feature enhances the privacy protection measures compared to traditional blockchain methods [19].

In the new model of blockchain-based data asset management, data security sharing requirements are effectively met by implementing node permission control and customizing the consensus mechanism. This ensures that only authorized nodes can access and manipulate sensitive data, enhancing overall data security within the management framework. Moreover, by optimizing the block structure and implementing indexes, the model guarantees efficient querying of data stored on the blockchain. This optimization ensures that users can swiftly retrieve relevant information, enhancing operational efficiency in data asset management tasks. To address the need for data traceability, the model uploads log data onto the blockchain. This practice ensures that a comprehensive record of all transactions and operations is maintained, facilitating auditing and accountability within the data asset management system. The sensitive data protection needs in data asset management are met through mechanisms that can encrypt and customize data and transaction information.

3.4. Benefit distribution method. The MCRS method is a simple calculation method that only considers the maximum profit and minimum profit in a major alliance, without considering the initial stage of alliance formation. Therefore, it is more suitable for more stable alliances, that is, stronger alliances.

The simplified MCRS method, namely the maximum minimum cost method. The calculation formula for this method is:

$$\omega_i = X_i min \frac{X_i max - X_i min}{\sum_{i \in N}(X_i max - X_i min)}[V(N) - \sum_{i \in N} X_i min] \forall_i \in N \tag{3.1}$$

Among them, the expected benefits of each partner are allocated as their highest benefit, and the actual benefits that each partner should receive are allocated as their lowest benefit, that is

$$\begin{gathered} X_i max = V(N) - V(N-i) \forall_i \in N \\ X_i min = V(N)_i = X_i \end{gathered} \tag{3.2}$$

Generally speaking, the expected lifespan of a data asset is the expected time for the asset to generate significant economic benefits. According to Table 3.1, it can be determined that the data will no longer generate significant economic benefits after several years of collection. It should be noted that sometimes the emergence of new data can lead to the retirement of old data, but this does not mean that old data cannot create significant economic benefits anymore [20].

3.5. Experimental preparation. The experiment was performed on a PC with a 2.6 GHz Intel Core i7 processor and 16 GB memory. In the experiment, we build a blockchain platform with Python programming language, which aims to realize a new model of data asset management. The platform consists of a super node and 2, 4, 6 and 8 common nodes.

The concrete experiments are as follows: 2, 4, 6 and 8 common nodes upload data into the blockchain network within a given time frame, with the same number of data being uploaded per common node. There are two phases to upload data: the beginning and the end of the upload. Upload start is a command sent by a normal node to send data, and send end means that it has already been added to the block copy of every regular

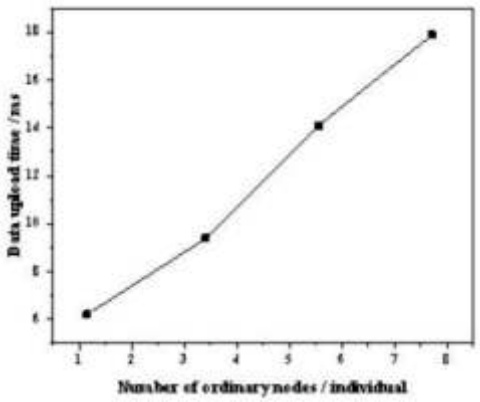


Fig. 4.1: Time for uploading data to the blockchain platform

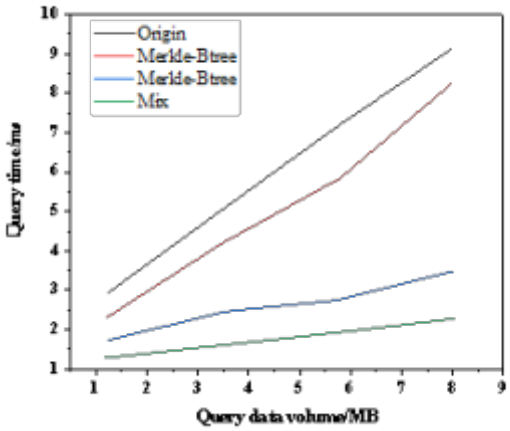


Fig. 4.2: Comparison of data query times using different methods

node. Thus, the time to upload the data to the blockchain platform is the time period from the beginning to the end of the upload.

4. Results and Discussion. In Figure 4.1, it's evident that as the number of nodes increases consistently, the data upload time demonstrates a stable and expected growth pattern, displaying a linear distribution overall. This observation validates the stability of the author's proposed model for data asset management leveraging blockchain technology. By examining the data upload times across multiple nodes, it becomes apparent that this new model significantly reduces upload times, showcasing its efficiency.

The experiment encompasses four distinct methods for querying data asset information: The first method employs conventional blockchain techniques for querying, referred to as the "origin method." The second method optimizes the existing data structure for querying by utilizing the Merkle B tree method. The third method employs the skip list structure for querying, denoted as the "skip list method." The fourth method, termed the "mix method," introduces the querying technique proposed in this article, integrating elements from blockchain methods. These methods are applied to query data asset information of varying capacities, with the query time recorded for each scenario. The results of the query time are illustrated in Figure 4.2.

Comparing the query times among the four methods reveals that the optimized structures (Merkle-B tree method, skip list method, mix method) outperform the origin method in terms of query speed. Notably, the mix method demonstrates the highest query efficiency, followed by the skip list method, with the Merkle B tree method trailing behind once more. This ranking can be attributed to several factors. The Merkle-B tree method, while optimizing the block data structure, only yields a modest increase in speed (17%) due to the relatively small amount of data within each block. In contrast, the skip list method's introduction of indexes between blocks significantly enhances query speed, resulting in a 1.24 times improvement. However, the mix method combines the strengths of both the Merkle-B tree method and the skip list method. By leveraging the advantages of both approaches, it achieves the fastest data query speed, performing 2.25 times faster than the origin method. This demonstrates the effectiveness of integrating optimized structures to enhance query efficiency.

5. Conclusion. The author suggests exploring a novel approach to distributing revenue from data assets, known as the "dual contribution" model, leveraging blockchain technology for enhanced trustworthiness. With the advent of big data, effective management of data assets becomes paramount, and blockchain offers promising solutions for ensuring robust security, privacy, and auditability in data asset management. While existing blockchain-based mechanisms for data asset management often focus on specific layers within the blockchain system framework, this proposed model aims to integrate multiple layers for a more comprehensive approach.

REFERENCES

[1] Le, T. D. . (2023). A shift towards household lending during the fintech era: the role of financial literacy and credit information sharing. Asia-Pacific Journal of Business Administration, 15(3), 466-485.
[2] Gaea, R. . (2020). Supply chain finance. Bankarstvo, 49(4), 100-111.
[3] Shemetov, V. V. . (2022). Some financial problems in the light of emm results:asset pricing and efficient portfolio allocation. Management Research: English Version, 10(5), 294-324.
[4] (2023). Swap variance hedging and efficiency: the role of high moments. Journal of Financial Research, 46(3), 681-709.
[5] Medri, I. , Aldroubi, A. , Martin, R. D. , Rohde, G. K. , & Thareja, S. . (2022). The signed cumulative distribution transform for 1-d signal analysis and classification. Foundations of Data Science, 4(1), 137-163.
[6] Melo, P. L. D. R. , Borini, F. M. , Isaac, V. R. , & Correa, V. S. . (2023). Regional development and the institutional environment for franchise chains: frontiers of small and medium-sized cities. Competitiveness Review: An International Business Journal , 33(2), 419-440.
[7] Yan, C. , Wang, X. , Zhang, X. , & Xu, R. . (2023). On incentive and coordination mechanism of service outsourcing based on principal-agent theory and blockchain technology. Journal of Artificial Intelligence Technology, 3(1), 1-9.
[8] Shi, Y. , Jiangyuanhong, L. I. , Wang, G. , Bai, S. , Bian, G. , & Boqiang, L. I. . (2022). Establish basic reseach funding navigation system, promote technical innovation to underpin development of national modern industrial system. Bulletin of Chinese Academy of Sciences, 36(5), 573-579.
[9] Shi, Y. , Jiangyuanhong, L. I. , Wang, G. , Bai, S. , Bian, G. , & Boqiang, L. I. . (2022). Establish basic reseach funding navigation system, promote technical innovation to underpin development of national modern industrial system. Bulletin of Chinese Academy of Sciences, 36(5), 573-579.
[10] Lyimo, A. J. , & Kumar, K. A. . (2022). Efficient way in sharing of iot data and uses of blockchain in auditing the stored data. International journal of internet technology and secured transactions: IJITST(1), 41(4), 351-379.
[11] Josphineleela, R. , Pellakuri, V. , Thanuja, R. , & Moses, D. . (2023). Secure internet of thing based data communication in blockchain model using novel teaching-learning optimized fuzzy approach. Transactions on Emerging Telecommunications Technologies, 3(3), 282-293.
[12] Taliep, N. , Lazarus, S. , Cochrane, J. , Olivier, J. , Bulbulia, S. , & Seedat, M. , et al. (2023). Community asset mapping as an action research strategy for developing an interpersonal violence prevention programme in south africa:. Action Research, 21(2), 175-197.
[13] Hughes, K. . (2022). Human trafficking, sm v croatia and the conceptual evolution of article 4 echr. The Modern Law Review, 85(4), 1044-1061.
[14] Neves, M. E. D. , Luís Baptista, António Gomes Dias, & Inês Lisboa. (2023). What factors can explain the performance of energy companies in portugal? panel data evidence. International Journal of Productivity and Performance Management, 72(3), 730-752.
[15] Shin, H. W. , Yoon, S. , Jung, S. , & Fan, A. . (2023). Risk or benefit? economic and sociocultural impact of p2p accommodation on community resilience, consumer perception and behavioral intention. International Journal of Contemporary Hospitality Management, 35(4), 1448-1469.
[16] Wang, Z. . (2022). Statistical thinking and new forms of contemporary art under the background of big data. Arts Studies and Criticism, 3(6), 346-350.
[17] Lai, Y. , Fang, X. , Han, R. , Xuan, Y. , & Huang, J. . (2023). Design and key technology of the energy consumption management system for the liquid cooling data center. Energy Science And Engineering, 11(3), 1284-1293.

[18] Stergiou, C. L. , & Psannis, K. E. . (2022). Digital twin intelligent system for industrial iot-based big data management and analysis in cloud. Virtual Reality & Intelligent Hardware, 4(4), 279-291.
[19] Chen, C. , Wang, L. , Liu, H. , Liu, J. , Xu, W. , & Huang, M. , et al. (2024). Construction of apricot variety search engine based on deep learning. Horticultural Plant Journal, 10(2), 387-397.
[20] Prabha, P. , & Chatterjee, K. . (2022). Design and implementation of hybrid consensus mechanism for iot based healthcare system security. International journal of information technology : an official journal of Bharati Vidyapeeth's Institute of Computer Applications and Management, 14(3), 1381-1396.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 31, 2024
Accepted: Aug 2, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 766–774, DOI 10.12694/scpe.v26i2.4075

DEFORMATION MONITORING AND ANALYSIS OF DEEP FOUNDATION PIT CONSTRUCTION PERIOD BASED ON INTERNET OF THINGS TECHNOLOGY

YUNBO XU[*] XIAONING DAI [†] HAOJIE ZHANG [‡] AND ZHANMING MA [§]

Abstract. In order to solve the problem of low detection accuracy in building structure safety monitoring, the author proposes a deformation monitoring and analysis during deep foundation pit construction based on Internet of Things technology. This method takes a large-scale construction project in a certain city as an example, applies neural network prediction models to deformation monitoring of foundation pits and main structures, and compares and analyzes the predicted results with actual measurement data. The experimental results show that the average MAE value of the predicted values is 0.15mm, and the average RMSE value is 0.17mm. The prediction accuracy of the neural network prediction model is high, which meets the accuracy requirements of deformation monitoring prediction. The use of Internet of Things technology can effectively ensure the safety of large buildings during construction, and has wider application value and prospects in deformation monitoring of future construction projects.

Key words: Neural network, Deformation prediction, Precision analysis, IoT technology

1. Introduction. In recent years, with the continuous development of urbanization and the influx of population, the utilization of urban land resources is no longer limited to flat areas, and the development of underground space has become an important part of urban construction [1,2]. During the long-term operation of building structures, they may be affected by environmental, geological changes, loads, and other factors, which may lead to structural safety hazards such as settlement, deformation, stress reduction, or stress concentration. In severe cases, it can lead to structural damage. With the support of Internet of Things technology, engineering managers can use sensors to collect specific data and use this data to determine the condition of building structures, thereby timely discovering and addressing hidden dangers. With the continuous development of underground space, the construction scale of foundation pit engineering is also constantly increasing. Due to complex geological conditions and possible improper construction techniques, there is a risk of serious accidents such as foundation pit collapse and surrounding building collapse during the construction process of foundation pits. For example, the collapse of the foundation pit of a certain subway line has caused huge social impact and economic losses [3]. Through the investigation of more than 160 excavation accidents, it was found that inadequate monitoring and untimely alarms are important reasons for excavation engineering accidents, so excavation monitoring is becoming increasingly important.

During the construction process of foundation pits (foundation pit enclosure, earthwork excavation, and basement construction), changes in the stress state of the soil inside and outside the pit can cause deformation of the foundation pit enclosure and soil. When the internal force or deformation of the enclosure structure exceeds the allowable limit, it can lead to instability and damage of the foundation pit, resulting in serious construction accidents. In order to timely obtain information on the stress and deformation of the retaining structure and surrounding soil, provide early warning and guidance for foundation pit construction, and monitor the entire construction process [4].

2. Literature Review. During the construction process, the main body of the building experiences significant settlement due to the continuous increase in load [5]. How to ensure the safety of foundation pits and main construction is increasingly becoming a focus of attention. In order to ensure construction safety, monitoring points must be set up in important areas for monitoring. In order to accurately predict

[*]Henan University of Engineering, Henan, Zhengzhou, 451191, China (Corresponding author, YunboXu6@163.com)
[†]Jiangsu Institute of Geological Exploration and technology, Nanjing, Jiangsu, 210049, China (XiaoningDai@126.com)
[‡]Henan University of Engineering, Henan, Zhengzhou, 451191, China (HaojieZhang9@163.com)
[§]Henan University of Engineering, Henan, Zhengzhou, 451191, China (ZhanmingMa@126.com)

the trend of changes in foundation pits and building structures, it is crucial to adopt effective prediction methods. Currently, the main prediction methods used include establishing regression models, singular spectrum analysis, empirical mode analysis, and establishing Kalman filtering models. Due to the influence of various factors such as foundation, load, precipitation, construction environment, and personnel on the deformation monitoring data of the foundation pit and main body, the established regression model is not significant and the accuracy of the prediction model is poor [6]. Singular spectrum analysis utilizes monitoring data from multiple periods for time series analysis, in order to obtain trend terms and useful high-frequency signals, achieve time series reconstruction, and then predict and analyze monitoring data. However, due to the many deformation inducing factors of the foundation pit and the main body, it is difficult to decompose and reconstruct useful monitoring information through singular spectrum analysis, resulting in low prediction accuracy. Empirical mode analysis is generally suitable for short-term time series analysis and prediction, while deformation monitoring of foundation pits and building structures is generally a long-term process, so its reliability is low. The Kalman filter prediction model has strong applicability for establishing linear models, but its performance is poor for nonlinear applications. In order to improve the prediction accuracy of deformation monitoring data for foundation pits and main structures, and ensure the safety of engineering operation, it is necessary to study a high-precision and efficient prediction model [7].

The application of neural networks is becoming increasingly widespread, not only in medical, biological, mechanical automation, electronic computers and other fields, but also in the engineering field. Wang, R. S. et al. investigated how precipitation and excavation impact the deformation characteristics of foundation pits. Their findings suggest that as dewatering and excavation proceed, the effect of dewatering on the support structure's deformation diminishes. Instead, the primary factors influencing support structure deformation gradually transition from dewatering to excavation activities within the foundation pit[8]. Jiang, M. et al. conducted an analysis of deep foundation pit excavation techniques, using a specific project as a case study. This analysis encompassed the project's characteristics, key technical measures employed in deep foundation pit construction, and the assessment of safety risk prevention and control measures. The objective was to offer valuable insights for ensuring the construction quality and safety of deep foundation pit engineering projects[9]. ZHANG Shubin et al. believe that deep foundation pit construction technology is one of the core technologies in construction engineering. The application of deep foundation pit construction technology can fundamentally improve the quality of engineering, compact the strength of the foundation, and ensure the good quality of construction projects [10]. XUFei et al. believe that when conducting geotechnical investigations of deep foundation pits, workers should also fully consider factors such as the geological and hydrological environment around the construction area, and try to avoid the adverse effects of these factors on the excavation process during the design process, in order to promote the smooth progress of deep foundation pit excavation and further improve the quality and reliability of construction projects [11].

Therefore, neural networks have unique advantages in processing non-stationary and nonlinear deformation monitoring data, and can better predict the trend of deformation data changes. The author intends to use IoT based technology for deformation monitoring and analysis during the construction period of deep foundation pits.

3. Method.

3.1. Overall design of monitoring system. The Internet of Things technology measures the operating data of bridges, houses and other building structures through sensors, and collects, stores, calculates, integrates and displays structural data with the help of Internet technology. In bridge structure monitoring, sensors that need to be used according to their monitoring content include vibration sensors, cable tension sensors, strain sensors, load sensors, etc. The overall structure of the monitoring system is shown in Figure 1, which includes Narrow Band Internet of Things (NB IoT), Analog to Digital Converter (ADC), Universal Asynchronous Receiver Receiver (UART), and various sensors. NB IoT is a cellular network technology used to achieve the Internet of Things, which has the advantages of low power consumption and low cost. ADC and UART are advanced peripheral interfaces of microcontrollers used to transmit data collected by sensors throughout the entire system. The software system is deployed on the server to process and display the bridge structure safety data collected by the system [12].

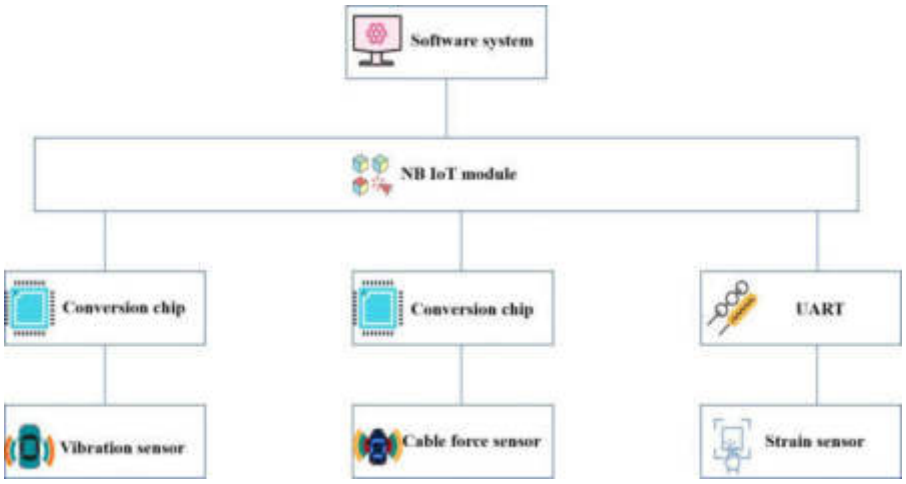


Fig. 3.1: Overall structure diagram of the monitoring system

3.2. Hardware Design of IoT Monitoring System.

(1) Sensor selection. Taking the YBJ-510 strain sensor as an example, in practical application, the sensor can be set at the pile foundation and other parts of the construction site to measure the initial frequency f_0, and then measure the real-time frequency f_i after a certain period of time. According to equation 3.1, the strain $\Delta\epsilon$ of the structure can be calculated.

$$\Delta\epsilon = K(f_i^2 - f_0^2) \tag{3.1}$$

In the formula, K is the calibration coefficient of the strain gauge. For vibration sensors, if the vibration speed is recorded as V, there are:

$$V = U/S_V \tag{3.2}$$

In the formula: U is the measured voltage value; S_V is the speed sensitivity of the sensor. For cable force sensors, the cable force is denoted as T, and the calculation method for T is shown in equation 3.3.

$$T = \frac{4ml^2 f_n^2}{n^2} - \frac{n^2 Eln^2}{I^2} \tag{3.3}$$

In the formula: I is the length of the steel cable; M is the weight per unit length of the steel cable; f_n is the nth frequency of the steel cable; EI is the bending stiffness of the steel cable; π is the pi. $n^2 EIn^2/I^2$ is the correction value of the bending stiffness of the steel cable to the cable force.

(2) The overall structure of the hardware system. After determining the main sensors of the monitoring system, other supporting hardware devices can be selected based on them, including Micro Controller Units (MCUs), power and voltage stabilization modules, and control circuits. The MCU of the system adopts STM32F103ZET6 chip, which integrates a digital to analog converter and can automatically convert voltage signals and digital signals. The chip transmits the data collected by the sensor to the NB IoT module through UART communication, and then transmits the data to the server of the software system through TCP/IP protocol [13,14].

3.3. Software Design of IoT Monitoring System.

(1) The overall structure of the software system. The software system consists of a database, front-end management interface, and server configuration. The database uses SQL Server, and the data Tables include strain threshold Table, vibration threshold Table, deformation threshold Table, cable force threshold table, load data Table, and others. Taking the cable force threshold Table as an example, its fields include cable number

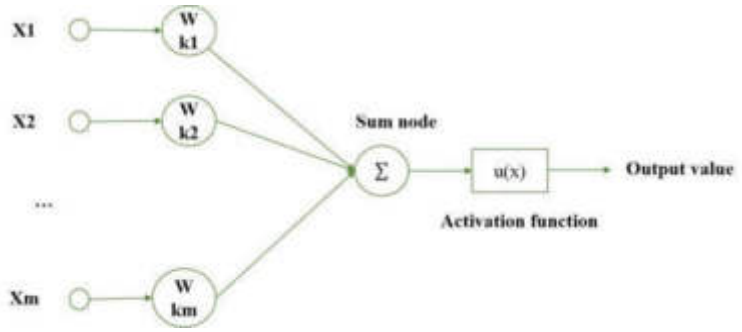


Fig. 3.2: Neuron Model

(Int type), cable length (Float type), unit length cable weight (Float type), cable force unit (Varchar type), cable force threshold (Float type), and input time (Timestamp type). The front-end management interface adopts MVC development mode, the back-end programming language is Java, and the system framework is SpringMVC [15].

(2) Calculation method for monitoring item threshold. The IoT monitoring system has an alarm function. Once the bridge tension, vibration effects, structural deformation, and structural strain monitored by the system exceed the threshold, the system will issue an alarm message. In order to design a reasonable monitoring alarm threshold, a finite element model of the construction site was constructed using Midas/Civil, with load types covering earthquake, vehicle, wind, and personnel activities.

(3) Design results of monitoring item threshold. The monitoring threshold includes cable tension, structural deformation, and structural strain.

3.4. Neural Network Models. Neurons are the fundamental information processing units of neural network models. The model diagram of neurons is shown in Figure 3.2.

The expression for neuron k is:

$$y_k = \varphi(u_K + b_K) \tag{3.4}$$

A neural network composed of multiple basic units of neurons has strong computing power. According to the classification of information flow within the network, it can be divided into feedforward networks and feedback networks [16].

(1) Feedforward network. In feedforward networks, information is first introduced from the outside by the input layer, flowing forward to the hidden layer and then to the output layer, layer by layer forward, so it is very simple to establish a multi-layer feedforward network (see Figure 3.3).

(2) Feedback network. In a feedback network, all nodes can process information and accept input from external information, as well as output to the outside world. Figure 3.4 shows a feedback network diagram of a single-layer fully connected structure [17].

3.5. Learning process of neural network prediction model. For neural networks, it is crucial to learn and improve their performance from the external environment. Learning is a process that requires external input, adjusting internal parameter structures (such as the number of hidden layer nodes), and adapting to the corresponding environment. The set used to solve learning problems is called a learning algorithm. The following text provides a brief introduction to several learning algorithms [18].

(1) Error correction learning. Assuming that neuron k is the unique node in the output layer of a certain network. The signal vector X(n) is the driver of neuron k, and the parameter n is the discrete time (the time step to adjust the weight of neuron k). The output signal $y_k(n)$ represents the actual output of neuron k, and $d_k(n)$ represents the expected response. This generates an error signal $e_k(n)$.

$$e_k(n) = d_k(n) - y_k(n) \tag{3.5}$$

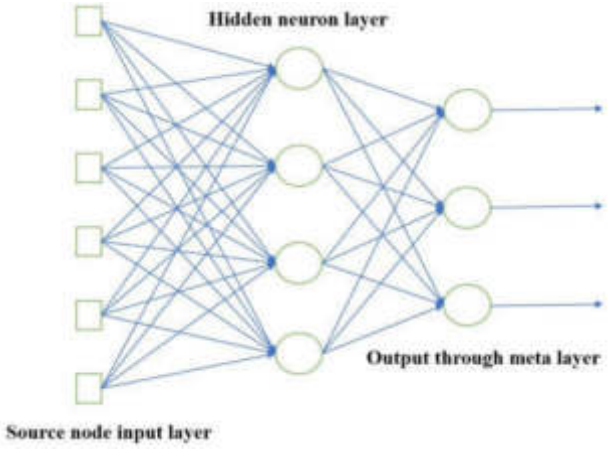


Fig. 3.3: feedforward network model

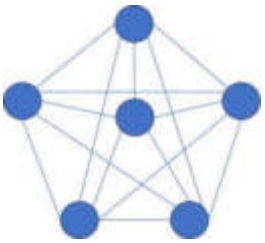


Fig. 3.4: Feedback network model of single-layer fully connected structure

By applying the error signal $e_k(n)$ to the weight of neuron k, the weight of neuron k is adjusted and adjusted until the network reaches a stable state (with almost no change in weight).

(2) Hebb Learning. If two neurons connected at both ends of a synapse are synchronously activated, the strength of that synapse is selectively enhanced; Otherwise, synapses will be selectively weakened or eliminated. The simplest mathematical expression Hebb learns is:

$$\Delta w_{k,j}(n) = \eta y_k(n) - x_k(n) \tag{3.6}$$

In the formula, x_k represents the presynaptic signal, y_k represents the postsynaptic signal, and $\Delta w_{k,j}$ represents the synaptic weight of neuron k, η represents a positive constant that determines the learning rate.

(3) Competitive learning. In competitive learning, output neurons in artificial neural networks compete with each other, with only one output neuron activated. Each input node of the winning neuron releases its weight in a certain proportion. The standard competitive learning rules are defined as:

$$\Delta w_{k,j} = f(x) = \begin{cases} \eta(x_j - w_{k,j}), \text{If neuron k fails to compete} \\ 0, \text{If neuron k fails to compete} \end{cases} \tag{3.7}$$

3.6. Design of neural network prediction model. According to the principles of neural networks, the design of neural networks mainly involves determining the number of neurons and initial parameters of the three-layer network [19].

(1) Number of network layers. The RBF neural networks constructed in this project are all three-layer networks, consisting of input layer, output layer, and hidden layer.

(2) Number of input layer neurons. Based on experience, the number of neurons in the input layer is equal to the dimensionality of the feature vectors of the training samples. Due to the various factors affecting the deformation of foundation pits and main buildings, including air pressure, temperature, precipitation, foundation, and building loads, which are difficult to obtain in practical work, deformation monitoring values are used as input variables and the number of variables is determined.

(3) Determination of the number of neurons in the output layer. Given the task at hand, which involves predicting deformation data for both the foundation pit and the main building structure, the output layer of the neural network comprises a single neuron. To address the inherent uncertainty in the output values, the output layer employs the Pureline function as its activation function.

(4) Number of hidden layer neurons. While there isn't a definitive algorithm for determining the number of neurons, it's widely recognized that the number of hidden neurons significantly impacts the performance of neural networks. Through iterative experimentation, researchers have derived a reference empirical formula:

$$m = \sqrt{n + k} + a \tag{3.8}$$

The formula incorporates several parameters: m represents the number of hidden layer neurons, n signifies the number of input layer neurons, k denotes the number of output layer neurons, and a is a constant ranging from 1 to 10. By varying the value of a, researchers can analyze the network learning error and ascertain an appropriate number of hidden layer neurons for optimal performance.

(5) Selection of error accuracy. The error accuracy needs to be determined through training. The initial value for this project is 10^{-4}, and adjustments will be made based on the convergence of the neural network.

(6) Selection of learning rate. Selecting an appropriate learning rate is paramount in training neural networks. The learning rate directly influences the magnitude of weight adjustments: higher learning rates result in larger weight changes. If the learning rate is small, the convergence rate of the network will also be affected. Typically, a learning rate of 1 is chosen to strike a balance between convergence speed and computational performance.

(7) Selection of learning frequency. The number of learning sessions needs to be determined based on specific circumstances. Generally, a larger number of learning iterations result in higher prediction accuracy.

3.7. Test preparation. This project is located in the central area of a certain city, with convenient transportation. The project is equipped with two levels of basement and a pile raft foundation. The excavation depth of the foundation pit is 10.0m, which belongs to a large-scale deep foundation pit project. The overall design scheme of the project adopts a sequential approach of drilling and grouting pile row, triaxial cement soil mixing pile waterproof curtain, and a diagonal brace combined with angle brace support system. The safety level of foundation pit support is Level 2. Monitor the displacement and settlement of the top of the foundation pit retaining piles and the settlement of surrounding roads during the construction process of the foundation pit. Among them, there are 48 monitoring points on the top of the retaining piles and 20 monitoring points on the settlement of surrounding roads. The main buildings of the project are two high-rise buildings, A1 # and A2 #, with 59 floors and a height of 303m, in zone A. Based on the structure, foundation, and load characteristics of the building, and in combination with relevant regulatory requirements, 24 observation points are set up for Building A1 # and Building A2 # [20].

4. Results and Discussion.

4.1. Analysis of experimental prediction data. According to the design of the network, the neural network structure for this experiment was determined. The cumulative changes of the first 40 monitoring points of foundation pit settlement CW6, CW15, CW22, CW36, and CW41 were selected as training samples for the network and trained separately to predict the settlement deformation for the next 10 times. In addition, the first 40 monitoring points of main settlement A107, A118, A201, A212, and A222 were selected to predict the settlement deformation for the next 10 times.

(1) Training results of point CW6. Compare and analyze the actual measured values of point CW6 with the predicted values of the neural network, as shown in Figure 4.1.

From the CW6 settlement change curve in Figure 4.1, it can be seen that the trend between the two is consistent, indicating a good prediction effect. In order to more intuitively reflect the comparison between

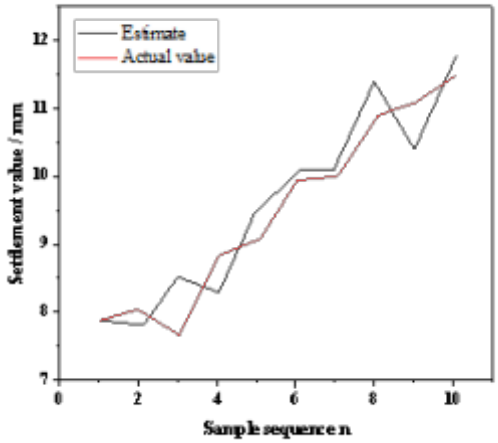


Fig. 4.1: Time series of predicted and actual measured settlement values at monitoring point CW6

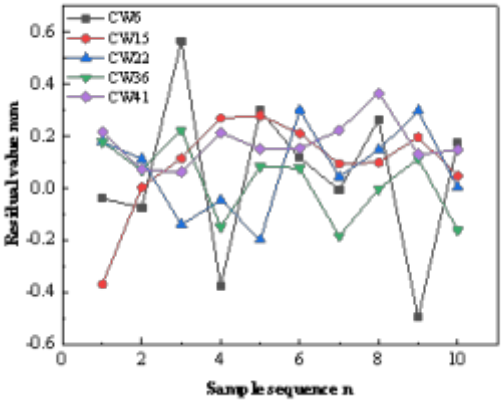


Fig. 4.2: Residual sequence diagram of foundation pit slope top settlement predicted by neural network

the actual and predicted values of the settlement at the top of the foundation pit slope, residual time series diagrams of monitoring points CW6, CW15, CW22, CW36, and CW41 were drawn. As shown in Figure 4.2.

In order to more intuitively reflect the prediction effect of building main body settlement, draw residual time series diagrams of neural network prediction values and observation values of monitoring points A107, A118, A201, A212, and A222 (as shown in Figure 4.3).

By analyzing the residual time series of the 5 monitoring points in the foundation pit and the 5 monitoring points for the main settlement, it can be seen that the residual values are all within 0.45mm, and the residual values are all around the 0 axis. After calculation, the maximum residual value is 0.43mm of the monitoring point CW6 Phase 3.

In order to further analyze the reliability and accuracy of neural network deformation prediction technology, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were introduced. The test analysis results are shown in Table 4.1.

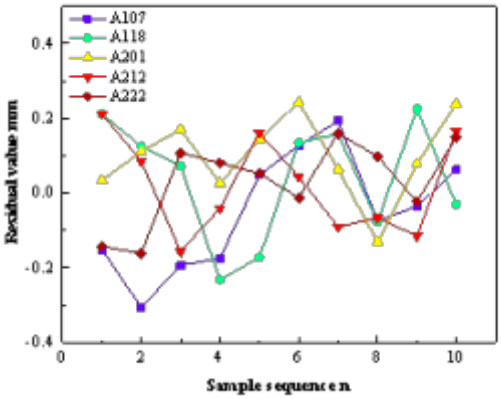


Fig. 4.3: Residual sequence diagram of main body settlement predicted by neural network

Table 4.1: Statistics of neural network prediction accuracy

Monitoring points	MAE/mm	RMSE/mm
CW6	0.21	0.29
CW15	0.16	0.19
CW22	0.14	0.16
CW36	0.10	0.12
CW41	0.16	0.18
A107	0.13	0.15
A118	0.13	0.15
A201	0.11	0.13
A212	0.11	0.12
A222	0.08	0.10

According to Table 4.1, the MAE and RMSE values predicted by the IoT neural network prediction model are all within 0.20mm and 0.30mm, with an average MAE of 0.15mm and an average RMSE of 0.17mm. The error value meets the accuracy requirements for measurement prediction. Which further demonstrates that the prediction model based on the Internet of Things neural network has good prediction accuracy, can effectively ensure the safety of high-rise buildings during construction and operation, and has more extensive application value and prospects in future prediction of foundation pit deformation and main body settlement deformation.

5. Conclusion. The author proposes a deformation monitoring and analysis method for deep foundation pit construction based on Internet of Things technology. Using an Internet of Things neural network prediction model, the deformation monitoring data of the main body and foundation pit of a super high-rise building in a certain city are predicted and analyzed. The predicted data from 5 monitoring points of the selected foundation pit and 5 monitoring points of the main settlement are analyzed. The results show that the average MAE value is 0.15 mm and the average RMSE value is 0.17 mm. The Internet of Things neural network has high prediction accuracy and can be well applied to predict deformation data of buildings and foundation pits, effectively ensuring the safety of high-rise buildings during construction and operation. It has a wider application value and prospects in future foundation pit deformation prediction.

6. Fund project. Enterprise entrusted Project approval unit: Jiangsu Provincial Institute of Geological Exploration and Technology Project name: Research on the key technology of early warning of super-large and super-deep subway foundation pit".

REFERENCES

[1] Zhang, L. , & Zhu, J. . (2022). Numerical simulation and field monitoring analysis for deep foundation pit construction of subway station. Structural Durability and Health Monitoring, 16(4), 20.
[2] Chen, B. G. , & Jia, Z. P. . (2023). Optimal strut position of deep foundation pit with convex corner under surcharge of adjacent building. Rock and Soil Mechanics, 44(8), 2400-2408.
[3] Jiang, M. . (2023). Construction technology and safety risk control measures of deep foundation pit excavation. World Architecture (Hundred Views), 7(2), 24-29.
[4] Jianhua, Y. , Jiyong, C. , Chi, Y. , Xiaobo, Z. , & Liansheng, L. . (2022). Discussion on blasting vibration monitoring for rock damage control in rock slope excavation. Earthquake Engineering and Engineering Vibration: English Version, 21(1), 53-65.
[5] Amiri, S. , & Dehghan, A. N. . (2022). Comparison of shallow tunneling method with pile and rib method for construction of subway station in soft ground. Frontiers of Structural and Civil Engineering: English Version, 16(6), 14.
[6] Zhang, Y. K. , Gong, G. F. , Yang, H. Y. , Chen, Y. X. , & Chen, G. L. . (2022). Towards autonomous and optimal excavation of shield machine:a deep reinforcement learning-based approach. Journal of Zhejiang University (English Edition) Series A: Applied Physics and Engineering, 23(6), 21.
[7] Oyama, M. , Kanaoka, S. , Komatsu, A. , & Ichimura, K. . (2023). Construction of a structural model of local settlement awareness the implementation of an educational program at the university of toyama and its usefulness of the course. Journal of The Japanese Institute of Landscape Architecture, 86(5), 567-572.
[8] Wang, R. S. , Guo, C. C. , Lin, P. Y. , & Wang, F. M. . (2023). Excavation response analysis of prefabricated recyclable support structure for water-rich silt foundation pit. Rock and Soil Mechanics, 44(3), 843-853.
[9] Jiang, M. . (2023). Construction technology and safety risk control measures of deep foundation pit excavation. World Architecture (Photo), 7(2), 24-29.
[10] ZHANGShubin. (2022). Brief discussion on construction technology and safety management countermeasures of deep foundation pit engineering. Engineering technology of foreign science and technology periodical database (abstract edition)(4), 74-78.
[11] XUFei. (2022). Key points of geotechnical engineering investigation in deep foundation pit engineering. Engineering technology of foreign science and technology periodical database (abstract edition), 51(4), 896-911.
[12] Xinya, L. , & Bo, J. . (2022). New directions of 5g for the development of blended learning models. Psychology Research, 12(10), 802-807.
[13] Cai, Z. B. , Chun-Lin, L. I. , You, L. , Chen, X. D. , Li-Ping, H. E. , & Cao, Z. Q. , et al. (2024). Prediction of contact resistance of electrical contact wear using different machine learning algorithms. Friction, 12(6), 1250-1271.
[14] Sun, W. F. , Tan, C. X. , Wang, Z. M. , Zhang, C. S. , & Shu-Ren, W. U. . (2022). Prediction of crustal stress of deep-buried tunnels based on bp artificial neural network. Journal of Geomechanics, 13(3), 227-232.
[15] Gan, L. I. , Wei-Bin, M. A. , Chang-Yi, Y. U. , Zhi-Gang, T. , & Feng-Nian, W. . (2023). Optimization of anchorage support parameters for soft rock tunnel based on displacement control theory. Journal of Mountain Science, 20(7), 2076-2092.
[16] Zhang, Z. , Qiu, X. , Shi, X. , & Yu, Z. . (2023). Chamber roof deformation prediction and analysis of underground mining using experimental design methodologies. Natural Hazards, 115(1), 757-777.
[17] Deng, Z. , Wei, P. , & Yuntao, H. . (2023). Resistance prediction of yacht based on cfd and profile deformation of bulbous bow. Journal of Chongqing Jiaotong University(Natural Sciences), 42(2), 144-148.
[18] Xie, C. , Nguyen, H. , Choi, Y. , & Armaghani, D. J. . (2022). Optimized functional linked neural network for predicting diaphragm wall deflection induced by braced excavations in clays. Frontiers of Geoscience: English Version, 13(2), 34-51.
[19] Zhou, H. , Che, A. , Shuai, X. , & Zhang, Y. . (2023). A spatial evaluation method for earthquake disaster using optimized bp neural network model. Geomatics, Natural Hazards and Risk, 14(1), 1-26.
[20] Lin, J. , Rijun, H. U. , Wang, P. , Song, Z. , Zhang, W. , & Niu, J. , et al. (2022). Surface sediment resuspension and suspended sediment transportation mechanism in the waters around miaodao strait. Marine Geology & Quaternary Geology, 42(3), 9-24.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 31, 2024
Accepted: Jul 17, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 775–787, DOI 10.12694/scpe.v26i2.4076

RESEARCH ON HETEROGENEOUS CROSS-DOMAIN IDENTITY AUTHENTICATION AND CONTROL IN CLOUD ENVIRONMENT

KAI XU*, FEIFEI YU†, ZHI YANG‡, JIANJUN ZHANG§, ZHIGUANG SONG¶, AND SHITAN LIANG‖

Abstract. To fulfill the need for cross-domain authentication in a hybrid cloud setting, the study focuses on identity authentication schemes that bridge various password systems, the author proposes a study on heterogeneous cross domain identity authentication and control in cloud environments. Introduce a multi center authentication management mechanism based on PKI to control and track the anonymous identities of users in different password system security domains. In the process of bidirectional authentication between users and cloud service providers, the scheme successfully negotiates session keys and converts anonymous identities across different password systems. Results indicate that the cloud-based cross-domain identity authentication scheme, without certificate signatures, involves three exponential operations during user registration, four exponential operations and three bilinear operations during the initial cross-domain authentication, and three bilinear operations during subsequent cross-domain stages. Meanwhile, the identity authentication scheme based on PTPM and certificateless public key requires three exponential operations during user registration, five exponential operations and three bilinear operations during the initial cross-domain authentication, and three bilinear operations during repeated cross-domain phases. This scheme achieves cross domain authentication in heterogeneous systems and uses lower computation time for dot multiplication and hash operations. Compared to other schemes, it achieves better computational efficiency while completing cross domain authentication in heterogeneous systems, while compared to the EIMAKP scheme, it has better computational efficiency. This approach effectively safeguards against replay, substitution, and man-in-the-middle attacks, ensuring secure cross-domain identity authentication across diverse password systems. It balances robust security measures with computational efficiency, thereby enhancing overall system reliability and integrity.

Key words: Hybrid cloud, Heterogeneous systems, Cross domain authentication, Anonymity, Bidirectional authentication

1. Introduction. With the rapid development of information technology in the cloud environment, cloud computing can scale, virtualize, automate and centralize services based on its own deployment complexity, network resource sharing, flexibility and portability, and security. Reliable network cloud services can be built through distributed features, and a communication structure between various servers can be established to complete massive cloud computing resource portability applications and access. Users can low-cost intercommunication network services [1]. For network security issues in cloud environments, cloud computing can provide reliable transmission and communication for server users to interact with information based on its distributed characteristics, and try to prevent external malicious attacks and illegal access to resources, playing a certain protective role. Due to the flexible sharing of cloud computing, it has led to the massive use of multi cloud environments and platforms. Through multi cloud environments, cross domain cloud services can be provided, as well as the emergence of private clouds [2]. The server side can control and lock the data centers of its cloud providers, making online cloud resource services have the characteristic of connectivity. This has brought great convenience to people's lives and the development of the online economy [3]. However, while generating benefits, it also makes many criminals eager to attack and break down network services to obtain greater illegal benefits, making network interactions more complex and increasing the risk of security breaches. The annual increase in network security breaches is 12% year-on-year, making our personal, community, and corporate information more transparent. Nowadays, criminals use illegal means to organize, purposefully, and systematically modify and obtain information data, bringing new huge security challenges to cloud services [4].

*Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610000, China (Corresponding author, KaiXu89@126.com)
†Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610000, China (FeifeiYu96@163.com)
‡Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610000, China (ZhiYang55@126.com)
§Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610000, China (JianjunZhang7@163.com)
¶Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610000, China (ZhiguangSong5@126.com)
‖Aostar Information Technologies Co., Ltd., Chengdu, Sichuan, 610000, China (ShitanLiang2@163.com)

Such incidents continue to grow every year. In January 2020, a massive amount of data from a giant cosmetics company was leaked online. The reason for this was that the company publicly disclosed an unsecured database online, which was discovered by security researchers and contained a total of 440336852 records, this includes port numbers, network IP addresses, references, etc. used within the company. Once leaked database information is obtained by malicious individuals for illegal operations, it will cause huge losses to the entire company [5]. In August 2020, shortly after the opening of the New Zealand Securities Exchange, there were several crashes, which not only disrupted the exchange's stock prices and index quotes, but also disrupted its debt market [6]. The reason is that the exchange was attacked by distributed denial of service (DDoS) attacks on its website.

2. Literature Review. In today's cloud computing environment, the mainstream authentication systems mainly include those based on PKI, IBC, and CL-PKC. Among them, the PKI mechanism is widely used due to its mature authentication mechanism, complete structure, and high security. Pradhan, R. et al. proposed an energy aware cloud task scheduling algorithm. It extracts concepts from traditional minimum, maximum, and minimum heuristics and integrates them with energy models. These heuristic algorithms are implemented in heterogeneous cloud environments. The EACTS energy model is designed to assess energy usage in cloud data centers. This algorithm predicts construction time, cloud utilization, and energy consumption based on benchmark data. Through experiments, the EACTS algorithm offers valuable insights into balancing energy efficiency and completion time. It provides a comparative analysis of different scheduling parameters to inform decision-making regarding optimization strategies [7]. Krishnadoss, P. et al. introduced an enhanced seagull optimization algorithm, amalgamating features from both cuckoo search (CS) and seagull optimization algorithm (SOA). This hybrid approach aims to optimize task scheduling in heterogeneous cloud environments by minimizing cost and time parameters. Through comparison with multi-objective ant colony optimization (MO-ACO), ACO, and Min Min algorithms using the Cloudsim 3.0 toolkit, the proposed algorithm's performance was evaluated. Simulation results indicate that the novel seagull optimization algorithm outperforms its counterparts, demonstrating its effectiveness in cloud computing task scheduling [8]. Pradhan, R. et al. introduced a novel approach to optimize task scheduling in cloud data centers, aiming to reduce both duration and energy consumption. This method employs genetic algorithms, where each chromosome represents a scheduling arrangement of independent tasks across available clouds or machines. Fitness functions are utilized to minimize overall execution time, with energy consumption assessed based on the achieved minimum completion time. The effectiveness of this approach was validated through testing on synthetic and benchmark datasets, demonstrating superior performance compared to conventional cloud task scheduling algorithms like Min Min, Max Min, and election heuristic algorithms in heterogeneous multi-cloud systems [9].

The author introduces a novel cross-domain identity authentication scheme tailored for mixed cloud environments, addressing the limitation of existing cloud-based authentication systems in supporting cross-domain authentication between disparate cryptographic systems. This proposed scheme, leveraging Public Key Infrastructure (PKI) and Certificateless Cryptography (CLC), enables secure identity authentication and access between users utilizing CLC and PKI public key cryptographic systems. Introduce a multi center authentication management mechanism based on PKI, with cloud authentication centers as the interaction center for the authentication process, in order to achieve bidirectional cross domain identity authentication between users and cloud service providers, and complete the control and tracking of user identity information throughout the entire process. Utilize temporary identities to achieve anonymity of user identities, and maintain traceability and controllability of user temporary identities and anonymous malicious behavior.

3. Research Methods.

3.1. Preparatory knowledge.

3.1.1. Hierarchical ID Tree Structure. The ID tree structure comprises a hierarchy where the root node represents the identity authentication center's identifier within the security domain, and the leaf nodes denote the identifiers of users and cloud service providers within the same security domain[10]. If the identity of the trusted third-party key generation center (KGC) in the security domain is IDKGC, the user identity in the security domain is $ID_{U_{ser}}$, and the identity form of the user in the security domain is defined as $ID_{KGC}||ID_{U_{ser}}$.

Fig. 3.1: Basic structure of PKI system

3.1.2. Related Difficulties and Assumptions.

CDH problem: G is an additive cyclic group of order q, and P is one of its generators. For any unknown $a, b \in Z_q^*$, given $aP, bP \in G$ calculate abP.

CDH assumption: For any algorithm A, there is no probability polynomial that the CDH problem can be successfully solved in time.

DLP problem: G is an additive cyclic group of order q, and P is one of its generators. For any unknown $a \in Z_q^*$, given $aP \in G$, calculate a.

DLP assumption: For any algorithm A, there is no probability polynomial that the DLP problem can be successfully solved in time [11].

3.1.3. Basic composition and structure of PKI system. In the PKI system, the authentication center is mainly responsible for verifying the authenticity of user identity information in the region, managing user digital certificates, and accepting services such as certificate revocation and updates. Users can access the certificate repository through the Lightweight Directory Access Protocol (LDAP) to query or download certificates. The basic structure of the PKI system is shown in Figure 3.1.

3.2. Cross domain identity authentication scheme based on heterogeneous systems.

3.2.1. Cross domain identity authentication model based on heterogeneous systems. The cross-domain identity authentication model based on heterogeneous systems involves four key entities:

1. Authentication Center: This entity, represented by CA (1), handles tasks such as application, issuance, revocation, and querying of certificates within its managed security domain. Additionally, the cloud authentication center CA manages these tasks within various security domains employing different password systems, as well as bidirectional authentication of user identities in different domains and temporary user identity conversion in different password systems;
2. Users U_p, authenticate their identity via the management center within their security domain and then use the cloud authentication center CA to validate their identity's legitimacy. The confirmation result is then relayed to the cloud service provider, enabling authentication of the visiting user through a trusted third party;

3. A cloud service provider provides users with various cloud service resources, which are authenticated by the management center within its own security domain[12]. The legitimacy of the cloud service provider within its security domain is confirmed through the cloud authentication center CA, and the verification results are returned to the visiting user, achieving authentication of the cloud service provider by a trusted third party.
4. The cloud key distribution center is mainly responsible for user authentication and the generation and distribution of partial keys in this security domain, and is responsible for tracing the true identity of malicious users.

For ease of description, let any two clouds be divided into Cloud 1 and Cloud 2. Cloud 1 is based on the PKI system, while Cloud 2 is based on the CLC system. In the initial stage, the cloud authentication center CA authenticates and issues certificates for its managed domains (Cloud 1 and Cloud 2). Users from various security domains initiate access requests to the remote cloud service providers they need to access. Upon receiving the request, the cloud service provider forwards the user's identity details to the cloud authentication center CA for authentication. If the authentication is successful, the cloud authentication center CA creates a compatible identity format for the user within the password system of the cloud service provider and returns it. In case of authentication failure, the result is directly communicated. Concurrently, the user submits the identity information of the cloud service provider to the cloud authentication center CA for authentication. If the authentication results are all passed, a trust connection is established. Although it is necessary to establish trust relationships between cloud service providers and users in various security domains through a cloud authentication center CA, as long as the trust relationship is established, users and cloud service providers no longer need a cloud authentication center CA to provide authentication for this trust relationship. Faced with the increasing number of security domains in cloud environments, adopting multi authentication centers between clouds to solve the security and performance bottlenecks of a single authentication center [13,14].

3.3. Specific Implementation Process of Certification Scheme.

3.3.1. System initialization. The Cloud Identity Management Center CA is responsible for managing the security and other related matters of various system authentication servers, while providing public parameters for the PKI and CLC password systems. For ease of description, $CA^{(1)}$ belongs to the PKI system and $KGC^{(2)}$ belongs to the CLC system; Input security parameters λ. the system selects the qth order additive cyclic group G1 and multiplicative cyclic group G2, defines a bilinear mapping $e : G_1 \times G_2 \to G_2$, selects the generator of group G1 as $P \in G_1$, and selects three secure hash functions H_1, H_2, and H_3; CA randomly selects a system master key $s \in \mathbb{Z}_q^*$ and calculates the system public key $P_{pub} = sP$; Publicly disclose system parameters $= \{q, G_1, G_2, e, P, H_1, H_2, H_3, P_{pub}\}$; $CA^{(1)}$ Randomly select a system master key $s^{(1)} \in \mathbb{Z}_q^*$ and calculate the system public key $P_{pub-1} = s^{(1)}P$; And publicly disclose the system parameters$s_1 = \{q, G_1, G_2, e, P, H_1, H_2, H_3, P_{pub-1}\}$; $KGC^{(2)}$ randomly selects a system master key $s^{(2)} \in \mathbb{Z}_q^*$ and calculates the system public key $P_{pub-2} = s^{(2)}P$; Public parameter $s_2 = \{q, G_1, G_2, e, P, H_1, H_2, H_3, P_{pub-2}\}$ [15].

3.3.2. User Registration. User registration specifically includes $CA^{(1)}$ - PKI User Registration and $PKG^{(2)}$ -CLC user registration.

1) $CA^{(1)}$ - PKI User Registration. If user $u_p^{(1)}$ randomly selects parameter $x_p^{(1)}, r_p^{(1)} \in \mathbb{Z}_q^*$, then the user's private key is $sk_p^{(1)} = x_p^{(1)}$ and the public key is $PK_P^{(1)} = x_p^{(1)}P$, and the user calculates the temporary identity $TID_p^{(1)} = H_1(ID_i||r_p^{(1)}P)$, users download their own root certificate through the certificate repository, extract the public key P_{pub-1}, and read the local timestamp $T_p^{(1)}$.

Send a certificate application $Encrypt(ID_P^{(1)}, ID_{CA}^{(1)}, TID_P^{(1)}, T_P^{(1)}, P_{pub-}, r_p^{(1)})_{P_{pub-}}$ encrypted by P_{pub-1}; $CA^{(1)}$ receives a message and decrypts the application message using its own private key to verify the legitimacy of $ID_P^{(1)}$ identity. It also checks whether $ID_P^{(1)}$ already exists in the registered user list, verifies whether the temporary identity is correct, and verifies the validity of $TID_P^{(1)} \stackrel{?}{=} H_1(ID_P^{(1)}||r_p^{(1)}P)$ and timestamp $T_P^{(1)}$, if the verification fails, the application failure information will be returned. If the verification passes, $CA^{(1)}$ randomly selects $z_p^{(1)} \in \mathbb{Z}_q^*$ and calculates $Z_P^{(1)} = z_p^{(1)}P$.

Issue certificate $Cert_P^{(1)} = \{m_p^{(1)}, T_{begin}^{(1)}, T_{end}^{(1)}, \delta_P^{(1)}, PK_P^{(1)}, ID^{(1)}, Z_P^{(1)}\}$ for user's temporary identity $TID_P^{(1)}$,

among them, $m_P^{(1)}$ is the user's certificate information, $T_{begin}^{(1)}$ and $T_{end}^{(1)}$ are the valid start and end dates of the certificate, $\delta_P^{(1)}$ is the signature information of $CA^{(1)}$ on the user's identity, and $\delta_P^{(1)} = s^{(1)}H_1(m_P^{(1)}||ID^{(1)}) + z_P^{(1)}$. When the authentication center $CA^{(1)}$ first verifies the user's identity as legitimate, it saves the registration list $\{ID_P^{(1)}, TID_P^{(1)}, T_{P1}^{(1)}, r_p^{(1)}P, PK_P^{(1)}\}$ for the user. $\{Cert_p^{(1)}, ID_P^{(1)}, TID_P^{(1)}, T_{P1}^{(1)}\}_{PK_P^{(1)}}$ is the timestamp of the user's certificate issuance time, places the certificate in the certificate repository for storage, and sends the certificate $u_P^{(1)}$ issued to the user $u_P^{(1)}$. The user EE downloads the certificate and verifies its validity [16]. Read the timestamp $T_P^{(1)}$, verify the validity of the timestamp, and determine whether equation 3.1 is valid. If it is not, the certificate will be rejected.

$$\delta_P^{(1)} P \stackrel{?}{=} P_{pub-1} H_1(m_p^{(1)}) + Z_P^{(1)} \tag{3.1}$$

2) $PKG^{(2)}$ -CLC user registration. Based on the ID tree structure, the identity of cloud service provider $CS_C^{(2)}$ within the $PKG^{(2)}$ system is $ID_{CS} = ID^{(2)}||ID_{CS}^{(2)}$, among them, $ID^{(2)}$ is the identity information of $KGC^{(2)}$, $ID_{CS}^{(2)}$ is the true identity information of $CS_C^{(2)}$, the user randomly selects the secret value $x_{CS}^{(2)}, r_{CS}^{(2)} \in \mathbb{Z}_q^*$, calculates the user's public key $OK_{CS}^{(2)} = x_{CS}^{(2)}$ and temporary identity $TID_{CS}^{(2)} = H_1(ID_{CS}||r_{CS}^{(2)}P)$; User $CS_C^{(2)}$ sends a registration application $Encrypt(ID_{CS}, TID_{CS}^{(2)}, OK_{CS}^{(2)}, r_{CS}^{(2)}P)_{P_{pub-2}}$ encrypted by P_{pub-2} to $KGC^{(2)}$; After receiving user messages, $KGC^{(2)}$ decrypts them using its own system master key $s^{(2)}$ to verify the legality of ID_{CS} identity and the correctness of temporary identity $TID_{CS}^{(2)} \stackrel{?}{=} H_1(ID_{CS}||r_{CS}^{(2)}P)$, if the verification fails, the application failure message will be returned; If the verification is successful, calculate $Q_{CS} = H_1(TID_{CS}^{(2)})$ and the user's partial private key $d_{CS}^{(2)} = s^{(2)}Q_{CS}$; Read the local timestamp $T_{CS}^{(2)}$ and save the user registration list $\{ID_{CS}, TID_{CS}^{(2)}, PK_{CS}^{(2)}, r_{CS}^{(2)}P, T_{CS}^{(2)}\}$; Return $Q_{CS}, d_{CS}^{(2)}$ to $u_{CS}^{(2)}$ through a secure channel; $ID_C^{(2)}$ calculates the private key $sk_{CS}^{(2)} = x_{CS}^{(2)} d_{CS}^{(2)}$.

3.3.3. PKI $\longleftrightarrow$ CLC cross domain authentication . This scheme uses a cloud based user identity management center (CA) in the cross domain authentication part to establish access identities for verified user identities. While verifying user identities, it also completes session key negotiation and determines the method of establishing access identities through password system identification sent through secure domains. Among them, CL is the identifier of the CLC cryptographic system, and PI is the identifier of the PKI cryptographic system. If the identity information of users or cloud service providers in the PKI domain is verified, it will be completed by the cloud authentication center CA to issue temporary access identities based on the CLC system for users. Similarly, if the identity information of users or cloud service providers in the CLC system is verified, it will be completed by the cloud authentication center CA to issue temporary access identities based on the PKI system for users [17]. Once a temporary access identity is established, users and cloud service providers can no longer rely on trust from cloud authentication centers and establish trust links for cross domain access between different password systems. The specific implementation process of cross domain identity authentication scheme is shown in Figure 3.2.

User $u_P^{(1)}$ randomly selects $n_P^{(1)} \in \mathbb{Z}_q^*$ and calculates the session key negotiation parameter $N_P^{(1)} = n_P^{(1)} sk_P^{(1)}$ $PK_{CS}^{(1)} N_P^{(1)} = n_P^{(1)} P^2$; User $u_P^{(1)}$ enters the temporary identity $TID_{CS}^{(1)}$ and password pw, and randomly selects $Cu_P \in \mathbb{Z}_q^*$; Calculate $w = H_3(TID_P^{(1)}||pw), h_p = H_2(mes_P||TID_P^{(1)}||w||T_P^{(1)}||PK_P^{(1)}||N_P^{(1)}||Cu_P)$, where Cu_P is a random parameter of the session message to maintain the freshness of the message, and T_p is the local timestamp; Send certificate information $\{cert_P, mes_P TID_P^{(1)}, w, N_P^{(1)}, T_P^{(1)}, PK_P^{(1)}, Cu_P, P_{pub-1}\}_{P_{pub2}}$ to $CS_C^{(2)}$.

After receiving the message, $cs_s^{(2)}$ obtains mes_p to determine whether it is an access request. If it is not an access request, it is rejected; Otherwise, according to $TID_P^{(1)}$, read user information from the access user list and verify the validity of the information; If the user information does not exist, upload $\{cert_P, ID_{CS}, L\}_{P_{pub}}$ to CA. CA obtains user $ID_P^{(1)}$ from $cert_P$ and verifies the legitimacy and $\delta_P^{(1)} P \stackrel{?}{=} P_{pub-1} H_1(m_P^{(1)}||ID^{(1)}) + Z_P^{(1)}$ of $ID_P^{(1)}$, if the verification is successful, a temporary access identity $ID_P = ID^{(2)}||TID_P^{(1)}$ will be established for the user, and $\alpha \in \mathbb{Z}_q^*$ will be randomly selected (Equation 3.2):

$$\delta_P = sH_1(ID_P) + \alpha \tag{3.2}$$

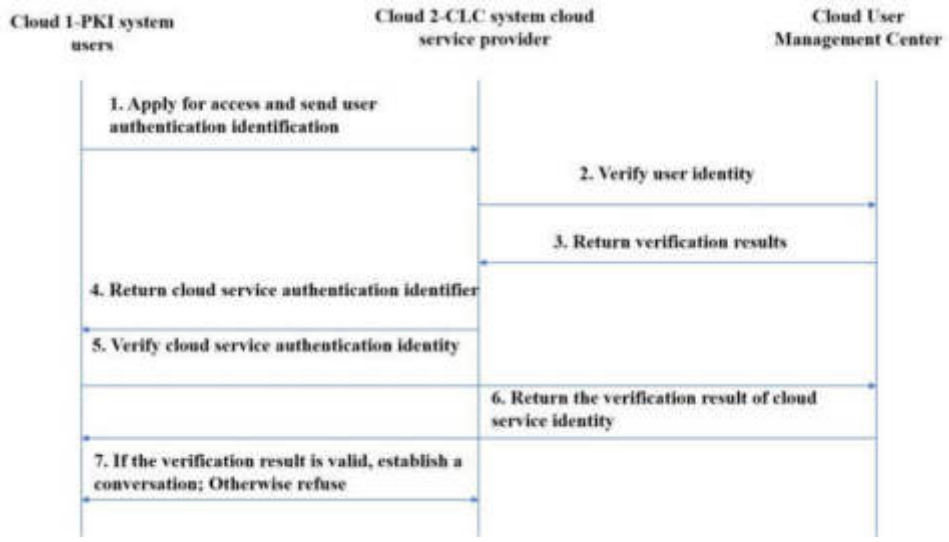


Fig. 3.2: Implementation process of cross domain identity authentication scheme based on heterogeneous systems

Send $Enc\{ID_P, result, \delta_P, \alpha P\}_{P_{pub-2}}$ to $KGC^{(2)}$; If the verification result does not pass, return "$\perp$".

If the $KGC^{(2)}$ judgment result is passed, verify the correctness of the signature $\delta_P \stackrel{?}{=} P_{pub}H_1(ID_P)+\alpha P$, determine the source of the message, and read the local timestamp T_{local}. In order to ensure security and establish a shorter effective duration of temporary identity access T_{Temp}, a temporary access list $\{ID_P.TID_P, T_{local}, T_{Temp}, T_P^{(1),PK_P^{(1)}}\}$ is established for users

$$K = \frac{M_P^{(1)} p^2}{x_{CS}^{(2)} PK_P^{(1)}} = N_P^{(1)} \tag{3.3}$$

$$h_P = H_2(mes_P||TID_P^{(1)}||w||T_P^{(1)}||PK_P^{(1)}||K||Cu_P) \tag{3.4}$$

Randomly select $f_C \in \mathbb{Z}_q^*$, calculate $F_C = x_{CS}^{(2)} f_C P$ and session key $CK = (h_P + f_C)PK_P^{(1)} x_{CS}^{(2)}$; After receiving the authentication result of the user's identity, $CS_C^{(2)}$ sends a $Encrypt(TID_{CS}^{(2)}, \{ID_{CS}, r_{CS}^{(2)}P\}_{P_{pub}}, d_{CS}^{(2)}, PK_{CS}^{(2)}, F_C, Cu_P)_{PK_P^{(1)}}$ to the user.

When user $u_p^{(1)}$ receives a message, verify whether $(d_{CS}^{(2)} + x_{CS}^{(2)})P \stackrel{?}{=} P_{pub-2}H_1(TI\ D_{CS}^{(2)}) + PK_{CS}^{(2)}$ is established, if the verification is successful, send $\{\{ID_{CS}, r_{CS}^{(2)}P\}_{P_{pub}}, TID_{CS}^{(2)}, d_{CS}^{(2)} + x_{CS}^{(2)}, PK_{CS}^{(2)}\}_{P_{pub}}$ to CA. After obtaining the message, CA verifies the legitimacy of the $ID^{(2)}$ security domain and checks its certificate validity and $TID_{CS}^{(2)} \stackrel{?}{=} H_1(ID_{CS}||r_{CS}^{(2)}\ P, d_{CS}^{(2)} + x_{CS}^{(2)})P \stackrel{?}{=} P_{pub2}H_1(TID_{CS}^{(2)}) + PK_{CS}^{(2)}$. After verification is completed, $\beta \in \mathbb{Z}_q^*$ is randomly selected and signature $\delta_C = sH_1(ID_P) + \beta$ is calculated.

Send $\{ID_P, result, \delta_C \beta P\}_{P_{pub-2}}$ to $CA^{(2)}$; If the verification result does not pass, return "$\perp$"; After receiving the result, if the result is valid, $CS_C^{(2)}$ calculates the user session key according to equation 3.5 and establishes the service [18].

$$CK = (h_P PK_{CS}^{(2)} + F_C)sk_P^{(1)} \tag{3.5}$$

3.4. CLC → PKI cross domain authentication. User $u_C^{(2)}$ under the certificateless public key cryptography system CLC wishes to access cloud resource $CS_P^{(1)}$ under the PKI system. Some steps are consistent with

PKI ↔ CLC cross domain authentication, so they will not be elaborated. Only elaborate on the core content of the differences. In the CA authentication section of the cloud authentication center, since at this time, user $u_C^{(2)}$ accessing cloud resource $CS_P^{(1)}$ under the user's uncertified public key cryptography CLC, therefore, at this point, the cloud authentication center CA will verify $u_C^{(2)}$'s identity information and issue a temporary certificate signed by CA for $u_C^{(2)}$, thereby establishing a connection between heterogeneous system user $u_C^{(2)}$ and cloud service providers. The other authentication modes are consistent with PKI ↔ CLC cross domain authentication.

3.5. Repeated cross domain authentication. Due to the establishment of a user identity list during the initial session, there is no need to interact with the cloud identity management center, which reduces the load on the cloud identity management center. The repeated cross domain authentication process is as follows.

(1). User $u_P^{(1)}$ randomly selects $n_{P_i}^{(1)} \in \mathbb{Z}_q^*$ and calculates the session key negotiation parameters $N_{P_i}^{(1)} = n_{P_i}^{(1)} sk_P^{(1)} PK_{CS}^{(1)}$ and $N_{P_i}^{(1)} = n_{P_i}^{(1)} P^2$; User $u_P^{(1)}$ enters temporary identity $TID_P^{(1)}$ and password pw, and randomly selects $C_{u_{P_i}} \in \mathbb{Z}_q^*$.

(2). Calculate $w = H_3(TID_P^{(1)}||pw)$ and $h_P = H_2(mes_P||TID_P^{(1)}||W||T_{P_i}^{(1)}||PK_{P_i}^{(1)}||$ $N_{P_i}^{(1)}||Cu_{P_i})$, where Cu_{P_i} is a random parameter of the session message, maintaining the freshness of the message, and T_{P_i} is the local timestamp.

(3). Send encrypted authentication information $Enc\{cert_P, mes_{P_i}, TID_{P_i}^{(1)}, w, N_P^{(1)},$ $T_{P_i}^{(1)}, Cu_{P_i}\}_{PK_{CS}^{(2)}}$ to $CS_C^{(2)}$. After receiving the authentication information, $CS_C^{(2)}$ obtains mes_{P_i} to determine whether it is an access request. Based on the user's temporary identity, $TID_P^{(1)}$ obtains the corresponding user information from the access user list and verifies whether w is the same as the list w saved in the list; Verify if $T_{P_i}^{(1)}$ has exceeded the specified validity period. If the verification is not successful, terminate the verification; Otherwise, calculate $K = \frac{M_P^{(1)} P^2}{x_{CS}^{(2)} PK_P^{(1)}} = N_{P_i}^{(1)}$ and $h_{P_i} = H_2(mes_{P_i}||TID_P^{(1)}||K||Cu_{P_i})$; Randomly select $f_{C_i} \in \mathbb{Z}_q^*$, calculate $F_{C_i} = x_{CS}^{(2)} f_{C_i} P$ and session key $CK = (h_{P_i} + f_{C_i}) PK_P^{(1)} x_{CS}^{(2)}$; After receiving the authentication result of the user's identity, $CS_C^{(2)}$ reads the local timestamp T_{C_i} and sends $Enc(TID_{CS}^{(2)}, \{ID_{CS}, r_{CS}^{(2)} P\}_{P_{pub}}, d_{CS}^{(2)} + x_{CS}^{(2)}, PK_{CS}^{(2)}, F_{C_i}, T_{C_i}, Cu_{P_i})_{PK_{CS}^{(1)}}$ to user $u_P^{(1)}$.

(4). After receiving a duplicate authentication return message, user $u_P^{(1)}$ checks the freshness of T_{C_i} and compares it with the message freshness parameter Cu_{P_i} to see if it is consistent. If the above judgment only fails once, the service will be terminated and the session will be stopped; Otherwise, calculate the session key $CK = (h_P PK_{CS}^{(2)} + F_C) sk_P^{(1)}$ and establish a service.

3.6. Safety and Performance Analysis.

3.6.1. Security analysis.

1) Bidirectional entity authentication. In each security domain, users within the domain achieve mutual authentication between users and resources within the security domain through the original authentication method. In the cross-domain authentication model for heterogeneous systems, users engage in bidirectional authentication with diverse cloud service providers through a request to the cloud identity management center CA. Initially, the cloud service provider authenticates the user's identity, ensuring the security and legitimacy of the user's domain via the cloud identity management center CA. Subsequently, the user authenticates the identity of the cloud service provider, verifying the security and legitimacy of the provider's domain through the same cloud identity management center CA. The authentication outcomes for each domain member are communicated solely by the trusted cloud identity management center CA, thereby completing bidirectional authentication between users and cloud service providers. Additionally, session negotiation keys are established during this bidirectional authentication process. Once the bidirectional authentication is completed, the cloud authentication center no longer provides trust support, reducing the burden on the cloud authentication center [19].

2) Anti replay attack. This scheme adds timestamps and random parameters to maintain session freshness in the cross domain authentication part for the message transmission process of authentication information exchange. Only when the timestamp and freshness parameters of the read message are valid and the random parameters that maintain session freshness during interaction are correct, will the authentication process participants consider the message to be valid. If a malicious attacker wants to replay the intercepted message to a new authentication interaction process to deceive the authentication system, but because the random parameters for maintaining session freshness obtained by the replay attacker from the intercepted authentication interaction message and the new authentication interaction message are different, the replay attacker cannot complete the authentication process through the intercepted authentication interaction message. Therefore, this scheme can effectively resist replay attacks.

3) Anti replacement attack. In this scheme, both w and h_P are bound to the temporary identity of user $u_P^{(1)}$, where the temporary identity $TID_P^{(1)} = H_1(ID_i||r_P^{(1)}P)$ is established based on the user's real identity IDi. In the initial conversation, $\delta_P^{(1)}$ is the signature information of $CA^{(1)}$ for the user's identity, $\delta_P^{(1)} = s^{(1)}H_1(m_P^{(1)}||ID^{(1)} + z_P^{(1)})$ is bound to the identity of $CA^{(1)}$, and $d_{CS}^{(2)} + x_{CS}^{(2)} = s^{(2)}H_1(TID_{CS}^{(2)}) + x_{CS}^{(2)}$ is bound to the identity information of $CS_C^{(2)}$; In repeated cross domain authentication, $w = H_3(TID_P^{(1)}||pw), h_P = H_2(mes_P||TID_P^{(1)}||W||T_{P_i}^{(1)}||PK_{P_i}^{(1)}||Cu_{P_i})$ has bound $u_P^{(1)}$ identity information, if the attacker replaces the user identity in the mutual authentication message, it cannot be verified by the receiving party. Therefore, this scheme can resist substitution attacks.

4) Anti Man in the Middle Attack and Key Security Analysis. Due to the fact that all parameters during interactive authentication are encrypted and sent through the destination public key, not only does it avoid the use of secure channels, but it also prevents intermediaries from obtaining parameters and avoiding intermediary attacks. Moreover, the calculations involved in authentication are based on difficulties such as discrete logarithms to ensure the security of parameter information, enabling effective protection of key negotiation parameters.

5) Traceability of bidirectional anonymity. In the authentication section, both $u_P^{(1)}$ and $CS_C^{(2)}$ use temporary identities $TID_P^{(1)}$ and $TID_C^{(2)}$ to replace the original identity form, achieving bidirectional anonymity between users and cloud service providers. If $u_P^{(1)}$ sends an illegal message, $CS_C^{(2)}$ will submit $TID_P^{(1)}$ to $CA^{(1)}$ in Cloud 1; $CA^{(1)}$ will query the registration list that saves the user's real information when issuing the certificate for the user, and confirm whether $ID_P^{(1)}$'s temporary identity is $TID_P^{(1)}$ again through $TID_P^{(1)} \stackrel{?}{=} H_1(ID_P^{(1)}||r_P^{(1)}\ P)$, if the verification is successful, it indicates that $ID_P^{(1)}$ is the owner of the temporary identity $TID_C^{(2)}$. When $CS_C^{(2)}$ provides malicious services, send its temporary identity $TID_C^{(2)}$ to the key distribution center $KGC^{(2)}$ of the CLC system; $KGC^{(2)}$ will query the registration list of $CS_C^{(2)}$ using the same calculation method. If verified, it will be determined that IDCS provides malicious services. Therefore, this scheme has bidirectional anonymity and traceability [20].

4. Result analysis.

4.1. Security comparison of cross domain identity authentication schemes. Members of a heterogeneous cryptographic system based on PKI and CLC can achieve bidirectional authentication in a cloud environment and have the feature of anonymous tracking. Throughout the authentication process, this scheme demonstrates robust resilience against replay attacks, man-in-the-middle attacks, and substitution attacks. Table 1 showcases a security comparison between this scheme and several other authentication models, including the identity-based multi-trust domain grid authentication model, wireless body area network anonymous authentication with provable security, and highly secure identity-based authentication key negotiation protocol. The results highlight that this scheme excels in security, as indicated by the checkmark (√) denoting compliance with security conditions, while the "x" symbol represents non-compliance.

4.2. Performance Analysis. For performance analysis, this scheme will be evaluated against EIMAKP, a cloud-based cross-domain identity authentication scheme utilizing certificateless signature, as well as identity authentication schemes based on PTPM and certificateless public key. Additionally, comparisons will be made

Table 4.1: Security comparison results of various cross domain identity authentication schemes

Programme	Anti replay attack	Anti Man in the Middle attack	Anti replacement attack	Mutual authentication	Anonymous tracking
Identity based multi trust domain grid authentication model	√	√	×	√	×
Anonymous authentication in wireless body area networks with provable security	√	√	×	√	×
A highly secure identity based authentication key agreement protocol	√	√	√	√	×
Author's proposal	√	√	√	√	√

with identity authentication schemes designed for multi-server environments. Due to the relatively low computational cost of point multiplication, only the computational costs of bilinear operations and exponential calculations with high computational costs are considered. Table 2 provides insights into computational efficiency, with Te representing bilinear operation time and TE representing exponential calculation time. Checkmarks (√) indicate conditions met, while "x" symbols denote unmet conditions. Regarding computational efficiency, the cloud-based cross-domain identity authentication scheme without certificate signatures involves three exponential operations during user registration, four exponential operations and three bilinear operations in initial cross-domain authentication, and three bilinear operations in repeated cross-domain phases. In comparison, the identity authentication scheme based on PTPM and certificateless public key entails three exponential operations during user registration, five exponential operations and three bilinear operations in initial cross-domain authentication, and three bilinear operations in repeated cross-domain phases. However, the cloud-based cross-domain identity authentication scheme based on certificateless signature, as well as the identity authentication schemes based on PTPM and certificateless public key, and those designed for multi-server environments, have not achieved cross-domain authentication for heterogeneous systems. The EIMAKP scheme achieves cross domain authentication for heterogeneous systems, but requires three bilinear operations when establishing cross domain authentication for the first time. This scheme achieves cross domain authentication in heterogeneous systems and uses lower computation time for dot multiplication and hash operations. When considering computational efficiency and achieving cross-domain authentication in heterogeneous systems, identity authentication schemes in multi-server environments outperform cloud-based cross-domain authentication schemes based on certificateless signatures and those relying on PTPM and certificateless public keys. Additionally, these schemes exhibit superior computational efficiency compared to the EIMAKP scheme.

4.3. Cross domain authentication execution efficiency. In order to verify the execution efficiency of the scheme, the author conducted simulation experiments in a Windows 10 − 64 bit, 16GB memory, 3.2GHz Intel Xeon i7 CPU, and vmware software environment. Multiple 64 bit Ubuntu 18.04 operating system virtual machines were installed in vmware for experimentation. Install the dependent GMP function library (GNU Multiple Precision Arithmetic Library) and glib function library in the Ubuntu virtual machine. Repeat the same experiment 20 times, and take the average value as the final experimental record. This experiment simulates 9 types of users as shown in Table 4.3, representing the authentication efficiency between different security domains.

To comprehensively assess the performance of the cross-domain authentication scheme presented in this chapter, we will analyze its efficiency in both initial and repeated cross-domain authentication scenarios. Furthermore, we will compare the efficiency of repeated authentication with the time consumption of messages of varying lengths. The experimental results will be compared with those of cutting-edge authentication protocol designs, particularly the identity-based cross-domain direct anonymous authentication mechanism and the proxy re-signature-based cross-domain authentication scheme. These protocols have demonstrated superior

Table 4.2: Comparison of computational efficiency of various cross domain identity authentication schemes

Programme	User registration stage	First cross domain authenti- phase	Repeated cross domain authentication -cation stage	cross- -domain authentication
EIMAKP	0	3Te	0	√
Cloud based cross domain identity authentication scheme based on certificate free signature	3TE	3Te+4TE	3TE	×
Identity authentication scheme based On PTPM and certificateless public key	3TE	3Te+5TE	3TE	×
Identity authentication scheme in a multi server environment	0	0	×	×
Author's proposal	0	0	0	√

Table 4.3: Scheme Execution Efficiency (s)

Customer type	User registration stage	Cross domain authentication stage	Repeated cross domain authentication stage
CLC->PKI	0. 018105	0. 058526	0. 00560
CLC->IBC	0. 018266	0. 027001	0. 00451
CLC->PKI,IBC	0. 017885	0. 074546	0. 00820
PKI->IBC	0. 022813	0. 033368	0. 00321
PKI->CLC	0. 023551	0. 037707	0. 00236
PKI->IBC,CLC	0. 023060	0. 423415	0. 00541
IBC->CLC,PKI	0. 017526	0. 083130	0. 00723
IBC->CLC	0. 018003	0. 035640	0. 00432
IBC->PKI	0. 017775	0. 065014	0. 00351

performance in terms of current user computational cost and message interaction rounds, making them highly relevant for cross-domain authentication. The results of these comparisons are depicted in Figures 4.1, 4.2, and 4.3.

Figure 4.1 illustrates that during the initial cross-domain authentication process, the identity-based cross-domain direct anonymous authentication mechanism and the proxy re-signature-based cross-domain identity authentication scheme operate seamlessly within the same authentication system. This eliminates the need for complex identity conversion calculations, streamlining the authentication process. Only the signature algorithm, verification algorithm, and bilinear mapping with high computational cost are required. On the contrary, the proxy re-signature-based cross-domain identity authentication scheme lacks a dedicated public key encryption algorithm, resulting in higher computational costs compared to the identity-based cross-domain direct anonymous authentication mechanism, this scheme does not adopt dual line down mapping, and a temporary identity issuance mode is designed in the identity conversion section, making it close to the proxy signature cost of certificate conversion. In the repeated cross domain authentication stage of this scheme, after the user establishes a session for the first time, the security domain of both parties in the established user access list does not require trust support and computation from the cloud authentication center, effectively reducing computational costs. As depicted in Figure 4.2, the repeated cross-domain authentication costs of the proxy re-signature-based cross-domain identity authentication scheme are consistently lower compared to the identity-based cross-domain direct anonymous authentication mechanism. Additionally, the cost remains

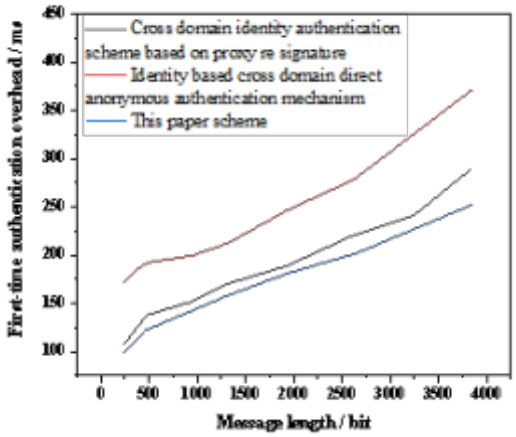


Fig. 4.1: Relationship between first cross domain authentication and message length

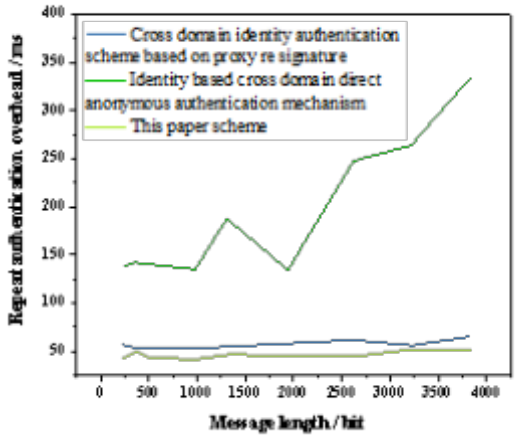


Fig. 4.2: Relationship between repeated cross domain authentication and message length

stable even when dealing with messages of varying orders of magnitude. Furthermore, Figure 4.3 demonstrates that this scheme exhibits higher efficiency in repeated authentication compared to the proxy re-signature-based cross-domain identity authentication scheme, despite the latter's similar efficiency in repeated cross-domain authentication [21].

5. Conclusion. The author proposes a cross domain identity authentication scheme based on heterogeneous systems in a hybrid cloud environment. Taking into account the distribution complexity of current cloud deployment patterns, a heterogeneous password system composed of PKI and CLC is used to manage various security domains and verify and convert temporary user identities across domains through an inter cloud identity authentication center. This scheme surpasses existing cross-domain identity authentication schemes by enabling cross-domain authentication across two types of password systems while maintaining security and achieving higher efficiency. Moving forward, the next phase of research will focus on developing cross-domain au-

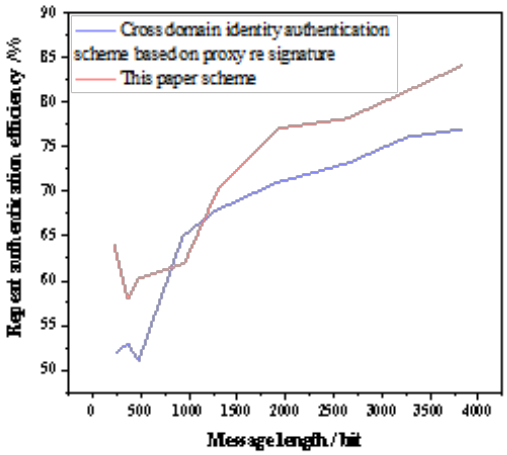


Fig. 4.3: Relationship between the efficiency of repeated cross domain authentication and message length

thentication schemes tailored for heterogeneous systems within cloud environments, without relying on trusted centers.

REFERENCES

[1] Lv, Y., Liu, W., & Wang, Z. (2020). Heterogeneous cross-domain identity authentication scheme based on proxy resignature in cloud environment. Mathematical Problems in Engineering, 2020, 1-12.
[2] Xuan, S., **ao, H., Man, D., Wang, W., & Yang, W. (2021). A cross-domain authentication optimization scheme between heterogeneous IoT applications. Wireless Communications and Mobile Computing, 2021, 1-14.
[3] Chen, J., Zhan, Z., He, K., Du, R., Wang, D., & Liu, F. (2021). XAuth: Efficient privacy-preserving cross-domain authentication. IEEE Transactions on Dependable and Secure Computing, 19(5), 3301-3311.
[4] Chen, R., Shu, F., Huang, S., Huang, L., Liu, H., Liu, J., & Lei, K. (2021). BIdM: A blockchain-enabled cross-domain identity management system. Journal of Communications and Information Networks, 6(1), 44-58.
[5] Wang, M., Rui, L., Yang, Y., Gao, Z., & Chen, X. (2022). A blockchain-based multi-CA cross-domain authentication scheme in decentralized autonomous network. IEEE Transactions on Network and Service Management, 19(3), 2664-2676.
[6] Feng, C., Liu, B., Guo, Z., Yu, K., Qin, Z., & Choo, K. K. R. (2021). Blockchain-based cross-domain authentication for intelligent 5G-enabled internet of drones. IEEE Internet of Things Journal, 9(8), 6224-6238.
[7] Pradhan, R. , & Satapathy, S. C. . (2022). Energy-aware cloud task scheduling algorithm in heterogeneous multi-cloud environment. Intelligent decision technologies: An international journal, 17, 3789-3800.
[8] Krishnadoss, P. , Poornachary, V. K. , Krishnamoorthy, P. , & Shanmugam, L. . (2023). Improvised seagull optimization algorithm for scheduling tasks in heterogeneous cloud environment. Computer, material, and continuum (in English)(2), 18.
[9] Pradhan, R. , & Satapathy, S. . (2022). Energy aware genetic algorithm for independent task scheduling in heterogeneous multi-cloud environment. Journal of Scientific & Industrial Research, 15(3), 3992-4002.
[10] Feng, C., Liu, B., Guo, Z., Yu, K., Qin, Z., & Choo, K. K. R. (2021). Blockchain-based cross-domain authentication for intelligent 5G-enabled internet of drones. IEEE Internet of Things Journal, 9(8), 6224-6238.
[11] Chen, J., Zhan, Z., He, K., Du, R., Wang, D., & Liu, F. (2021). XAuth: Efficient privacy-preserving cross-domain authentication. IEEE Transactions on Dependable and Secure Computing, 19(5), 3301-3311.
[12] Shen, M., Liu, H., Zhu, L., Xu, K., Yu, H., Du, X., & Guizani, M. (2020). Blockchain-assisted secure device authentication for cross-domain industrial IoT. IEEE Journal on Selected Areas in Communications, 38(5), 942-954.
[13] Tong, F., Chen, X., Wang, K., & Zhang, Y. (2022). CCAP: A complete cross-domain authentication based on blockchain for Internet of things. IEEE Transactions on Information Forensics and Security, 17, 3789-3800.
[14] Xuan, S., **ao, H., Man, D., Wang, W., & Yang, W. (2021). A cross-domain authentication optimization scheme between heterogeneous IoT applications. Wireless Communications and Mobile Computing, 2021, 1-14.
[15] Jia, X., Hu, N., Su, S., Yin, S., Zhao, Y., Cheng, X., & Zhang, C. (2020). IRBA: An identity-based cross-domain authentication scheme for the internet of things. Electronics, 9(4), 634.
[16] Wang, M., Rui, L., Yang, Y., Gao, Z., & Chen, X. (2022). A blockchain-based multi-CA cross-domain authentication scheme

in decentralized autonomous network. IEEE Transactions on Network and Service Management, 19(3), 2664-2676.
[17] Huang, C., Xue, L., Liu, D., Shen, X., Zhuang, W., Sun, R., & Ying, B. (2022). Blockchain-assisted transparent cross-domain authorization and authentication for smart city. IEEE Internet of Things Journal, 9(18), 17194-17209.
[18] Liu, Q., Gong, B., & Ning, Z. (2020). Research on CLPKC-IDPKC cross-domain identity authentication for IoT environment. Computer Communications, 157, 410-416.
[19] Zhang, Y., Luo, Y., Chen, X., Tong, F., Xu, Y., Tao, J., & Cheng, G. (2022). A lightweight authentication scheme based on consortium blockchain for cross-domain IoT. Security and Communication Networks, 2022, 1-15.
[20] Long, W., Wu, C. H., Tsang, Y. P., & Chen, Q. (2021). An end-to-end bidirectional authentication system for pallet pooling management through blockchain internet of things (BIoT). Journal of Organizational and End User Computing (JOEUC), 33(6), 1-25.
[21] Yan, X., Li, L., Chen, J., & Sun, L. (2023). Public key based bidirectional shadow image authentication without pixel expansion in image secret sharing. Frontiers of Information Technology & Electronic Engineering, 24(1), 88-103.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 31, 2024
Accepted: Jul 17, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 788–795, DOI 10.12694/scpe.v26i2.3898

THE PREDICTION AND EVALUATION OF MANUFACTURING TECHNOLOGY INNOVATION BASED ON MACHINE LEARNING AND BIG DATA ANALYSIS

FANG YANG*

Abstract. A data anomaly detection method was designed based on chemical manufacturing and oil refining units. Massive data storage and calculation are used in the cloud computing framework for petrochemical enterprises, refineries, and other large enterprises. The massive data is segmented based on the modified time series method, and the anomaly analysis is carried out. Thus, the abnormal data of the chemical manufacturing and oil refining units can be monitored. The practice proves that the algorithm proposed in this paper is a feasible, simple and effective data correction scheme.

Key words: Chemical manufacturing; Data correction; Machine learning; Fault error detection; Modified timing method

1. Introduction. Process data in the petroleum refining industry are generally flow-related operating parameters, such as flow, concentration, temperature and other field operating parameters. Under normal circumstances, accurate process test data is an essential basis for operation analysis, improvement of production process control and improvement of factory production and management [1]. However, the test results obtained in practical applications often contain some randomness and errors, which are inconsistent with the characteristics of the manufacturing process itself. The error of process test data can be divided into two kinds: one is random error, and the other is fault error. The random deviation in the system is caused by the noise of the test signal and the random variation during operation. Fault error is caused by unforeseen circumstances such as instrument failure, inaccurate calibration or reference drift, equipment leakage, etc. Data correction aims to improve the reliability, accuracy and integrity of data in process production to provide high-quality data for the production and management of process enterprises [2]. Although there are relatively complete commercial product development applications, they are still based on conventional statistical testing and linear analysis methods. It mainly focuses on error finding, data correction and parameter estimation. There are significant defects in its practical application: first, there is no credible reference standard for error detection, which leads to weak recognition and easy-to-miss diagnosis. Second, the process's data correction and parameter estimation are too dependent on the process structure and spatial information, and the process history information is not effectively mined. Third, the algorithm takes too long, making it challenging to realize the real-time correction of the measured data.

In this paper, the problems of data classification, error correction, data correction, and so on are deeply studied, as well as their organic integration with conventional test data correction [3]. In this way, the shortcomings of routine test data correction are solved. This method is suitable for data correction in the manufacturing process of large petrochemical enterprises.

2. Data processing methods.

2.1. Classification of measurement data. The measurement data correction must be based on the redundancy of the process variables. Only the remaining measurement data type and the observation type's non-measurement variables are corrected [4]. Due to many measured variables and constraints in industrial production, it isn't easy to correct and estimate them, so it is necessary to divide the process parameters to reduce the scale of problem-solving. The existing test data division method based on the zero-matrix method is limited in the complex industrial production process. According to the basic theory of graph theory, some scholars have established the sorting method, which does not require matrix operation and saves a lot of storage

*School of Economics and Management, Weinan Normal University, Weinan 714099, China (wnuyangfang@163.com)

space [5]. It is suitable for classifying more complex processes. In the case of no observed data, the equilibrium constraint equation with no observed value is obtained to modify the data directly.

2.2. Fault error detection. The measurement data error can be divided into two categories: random error and negligent error. Before the correction, if the error information cannot be found and excluded, the correction and estimation results will not be able to reflect the actual situation [6]. Therefore, the error correction should be done before the correction of observation data and parameter estimation. In addition, the failure of measuring instruments or pipeline leakage and other factors resulted in human error. Therefore, the conclusion of error discovery can help the operator to maintain the measuring instrument better and can resolve the problem of the instrument operation in time [7]. There are three ways to investigate errors: theoretically analyzing all kinds of data that may cause errors, using various measurement methods to realize the measurement comparison of the same process parameters, and verifying based on the statistical properties of the test data.

The error detection method based on mathematical statistics has a high application value based on the statistical characteristics of measurement data [8]. However, there are significant limitations in practical application. The traditional investigation methods include global inspection, node inspection and measurement data inspection.

Scholars mostly use the MT-NT combined test method to solve the defects of a single test. The idea is to combine the strengths of both. For example, principal component analysis can accurately judge the error orientation but often gives too much fault error. While NT does not spread the error throughout the system, the risk of error is more significant. So, the two can complement each other [9]. In the existing methods of correction of measured data, linear and nonlinear data correction are often treated separately. The standard correction method is linear correction, and nonlinear correction is used for flow rate, temperature and other data. In fact, according to the fundamental needs of data correction, the higher the redundancy of data, the better. If only the linear method is used to correct the data, the number of limiting equations required is limited, so the accuracy of the calculation results is not good. If only the nonlinear iterative method is used for correction, it will consume a lot of iterative operations. In this way, the real-time correction of the measured data of the equipment cannot be realized [10]. Due to the use of a separate limiting equation, no flow data is involved in the correction process, so the result is not accurate enough. This project intends to adopt two methods: linear correction and nonlinear correction. The velocity data after linear correction is used as the initial value, and then nonlinear iterative correction is carried out to minimize the number of iterations and calculation speed. In this way, the data can be corrected in real-time.

3. Time series analysis is applied to data correction. A prerequisite for revising process measurement data is to have some degree of redundancy. There's a lot of redundancy in process systems. This excess information can be divided into two types: one is caused by the presence of connections in the process, called "spatial redundancy," and the data obtained from multiple tests with the same precision instrument at the same measuring point is called "time redundancy." Time limits are domain limits. Equality limits are value limits. The above data correction algorithms are based on spatial redundancy [11]. A real-time database with high reliability is established for petrochemical enterprises using numerical control technology, which can collect and store the data. From the principle of data correction and effective use of information resources, we must consider spatiotemporal redundancy and spatiotemporal redundancy [12]. Therefore, this project intends to use the improved timing analysis method to correct the process data based on the timing characteristics and improve the accuracy of error discovery and data correction.

The improved time series analysis method is a smoothing algorithm based on robust local weighted regression used to analyze time series data. The time series $F = (K, U)$ is divided into three parts: trend component P, periodic component Z, and residual component S. Here $U = \{u_1, \ldots, u_n\}, u_n$ is the n time node; Where $U = \{u_1, \ldots, u_n\}, u_n$ is the data associated with time n.

$$F = P + Z + S$$

A prediction method based on the trend component is proposed. Periodic components can reflect periodic fluctuations in frequency. Residuals are the components that remain after removing the trend and cyclical

components [13]. The improved time series analysis method includes two aspects. It is divided into the outer cycle and the inner cycle. The direction component P and the periodic component Z are obtained by smoothing the timing of F in the inner cycle. The remaining components were collected in the outer cycle section. Assuming that the size of the point (t_l, u_l) in the time series is r_z, then the weight ω_j^l of t_j at any time is calculated in the range with u_i as the core and r_z as the interval:

$$\omega_j^l = \left(1 - \left(\frac{|u_j - u_j|}{u_{\text{farthot}} - u_i}\right)^3\right)^3$$

u_{Farthest} is the point in the region furthest from u_i. Take u_i as a linear regression at any time in this interval and find a smooth curve $f' = \alpha + \beta t$, then the smooth value at time point t_i is f_i. After a given interval length r_z, the timing of F can be decomposed to obtain the corresponding subsequence. After smoothing the subsequence, a periodic subsequence Z' can be obtained, and then the frequency component S is obtained by a low-pass filter, and the periodic component is expressed by $\mathrm{Z} = Z' - S$. And then keep going:

$$P' = F - Z$$

P' is smoothed at intervals of r_q, and a trend component P is obtained. And then, the remaining component is denoted by $S = F - P - Z$. Assuming that the initial data set corresponding to time series $K = \{t_1, \ldots, t_n\}$ is $U = \{u_1, \ldots, u_n\} = \{g(t_1), \ldots, g(t_n)\}$, and assuming that (t_i, u_i) data is missing, then:

$$\begin{gathered}
W(t_i) = g(t_1) + (t_l - t_1)\, g[t_2, t_1] + (t_i - t_1)(t_i - t_2) \\
g[t_3, t_2, t_1] + \ldots + (t_l - t_1)(t_l - t_2)\ldots(t_i - t_{n-1})\, g[t_{n-1}, \ldots, t_2, t_1] \\
S(t_l) = (t_i - t_1)(t_i - t_2)\ldots(t_i - t_n)\, g[t_{n-1}, \ldots, t_2, t_1] \\
W(t_l) = g(t_1) + (t_i - t_1)\, g[t_2, t_1] + (t_i - t_1)(t_i - t_2) \\
g[t_3, t_2, t_1] + \ldots + (t_i - t_1)(t_i - t_2)\ldots(t_i - t_{n-1})\, g[t_{n-1}, \ldots, t_2, t_1]
\end{gathered}$$

$g[t_i, t_j]$ is the first-order differential quotient of $g(t)$ at point t_i, t_j. Where $W(t_i)$ is the Newton interpolation approximation. $S(t_i)$ is a residual function [14]. Define the data set that has been populated with lost values as $U' = \{u_1', \ldots, u_n\}$. The maximum value is $u_{\max}$ and the minimum value is $u_{\min}$. The method of Inormalization is used so that all data values fall within the range of $[0, 1]$. For example:

$$f_i = \frac{u_i^i - u_{\min}}{u_{\max} - u_{\min}}$$

The data set $F = \{f_1, \ldots, f_n\}$ obtained at the end of the pre-processing. The improved time series analysis method eliminates the trend and periodicity components of the sequence. This makes it easier to find outliers and reduce problems such as error alarms caused by outliers [15]. A modified timing method obtained residuals S for the pre-processed data set. S Perform electrostatic protection tests. Here's how it works:

1. Calculate the middle-value M of the residual series data S.
2. Find the deviation of M from the median value.
3. Calculate statistics for each data point in S:

$$R_i = \frac{S_i - \bar{S}}{\text{mad}}$$

   $\bar{S}$ is the sample average.
4. The maximum value in R is statistically treated, and if the value exceeds the critical value ε, it is regarded as an outlier and removed from the time series.
5. Repeat the process (1) to (4).

4. Application of time series analysis method in practice. The improved time series method was tested and evaluated in a refinery's atmospheric and vacuum plant, focusing on the ability of error detection and the influence of time domain value and error size on error correction.

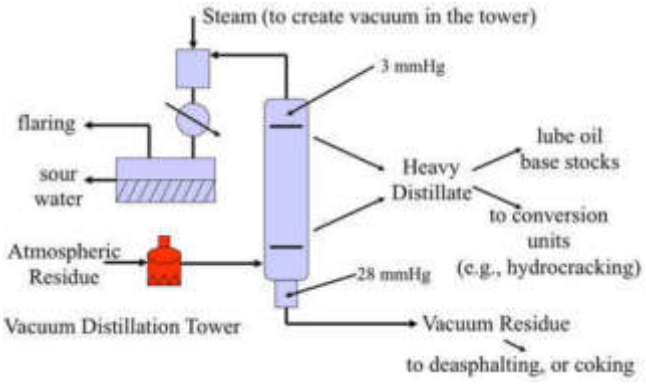


Fig. 4.1: Flow chart of petroleum atmospheric and vacuum device.

4.1. Process Overview. This type of atmospheric and vacuum equipment from a refinery produces oil from an oil depot and sends it to the equipment before heating it through an exchange process and then feeding it to an electric desalination tank. It is desalted and dehydrated in the process. In this process, it is also given a simple flash [16]. The flash-top gas is exchanged with the flash-bottom oil in the atmospheric tower and then heated into the tower for flash. Usually, line, two-line and three-line products are produced from the sideline of the atmospheric tower. The gas is discharged from the standard gas compressor and then through the heat exchange at the bottom of the stabilizing tower to obtain stable gas, liquefied gas, and naphtha. The conventional bottom fraction is divided into four parallel sections and fed into the pressure reducer for heating. It is then fed into a vacuum fractionator for classification. The products of reducing the first, second, third, and fourth lines are produced by the sideline of the decompression tower. The decompression tower extracts the oil at the top of the decompression tower, while the bottom reduction residue is discharged by heat transfer. A simple schematic diagram of the flow structure is obtained through the positioning analysis of each logistic measuring instrument in the equipment (Figure 4.1).

The process data in Fig. 4.1 was classified by sorting rule classification and matrix method. In this way, the simplified process structure of the standard pressure-reducing valve can be obtained [17]. There are 29 flow units and 6 nodes in the whole process.

4.2. Analysis of causes for error discovery. This project intends to establish 500 sets of 29,000 observation samples with the field calibration results of this equipment as actual values and add two 2.5% random deviations of positive and negative values. The random error is added to it to study its detection ability in various cases.

4.2.1. Influence of time domain value on error discovery rate. Select 12-30 errors in the time domain. Add fault errors of 20%, 40%, 60%, 80% and 100% in actual cases. The influence of time domain size on error detection ability is studied [18]. The evolution of the virtual detection rate of fault error over time is shown in Figure 4.2. At 16-20, the improved sequence method has a meager detection rate of fault errors, which can efficiently detect and eliminate fault errors and obtain more reasonable data correction.

4.2.2. Detection of errors of different sizes. As can be seen from the relative deviation between correction values and actual values in Table 4.3, the improved time series method has a good detection of errors of various sizes, and the results are the same as the previous examples.

4.2.3. Error identification in the communication of multiple orders of magnitude. The conventional error detection algorithm cannot locate the error accurately in the minor traffic flow, resulting in a significant deviation between the corrected result and the actual value. The test results show that the improved time series method can accurately detect and exclude. As can be seen from figures 4.4 and 4.5, the correction values are very close to the actual values.

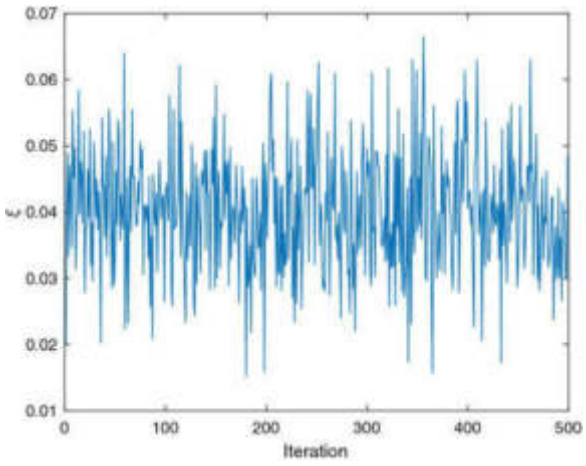


Fig. 4.2: Influence of time domain values on the detection capability of obsolete errors.

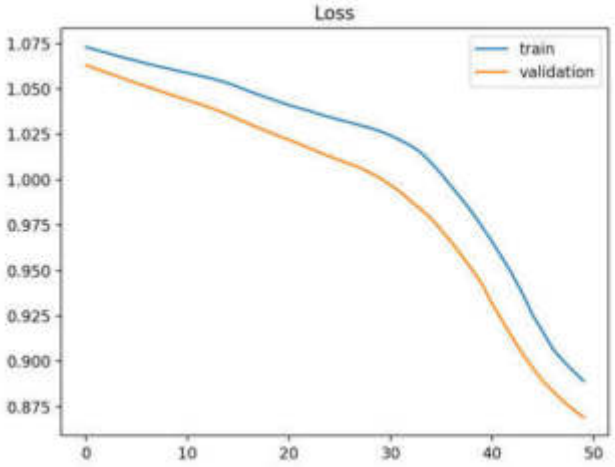


Fig. 4.3: Change of correction accuracy with error size.

4.3. Analysis of results after data modification. The time series method is used to modify 100 measured data. The results of the revised series are shown in Table 4.1.

As shown in Table 4.1, in the test data of the combination, 7,12, and 26 each carry 1 fault error. The average deviation between their correction and actual values is only 1%.

5. Conclusion.

1. According to the accurate division of the chemical production process, the defects of error detection are discussed, and a composite test method with process simulation as the core is established. This allows for better detection of errors.
2. Aiming at the linear and nonlinear problems existing in the system, the joint correction method is studied to increase the redundancy of the observation data. It also improves the speed and accuracy of data correction.
3. The process parameter correction method based on time series analysis is studied to use better the

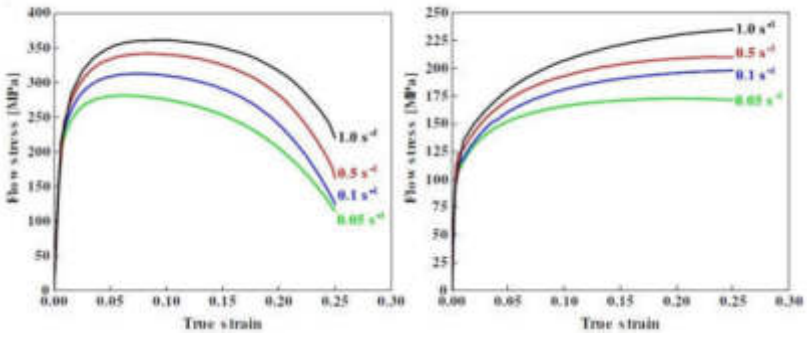


Fig. 4.4: Comparison of the results of a large order of magnitude traffic correction and actual value.

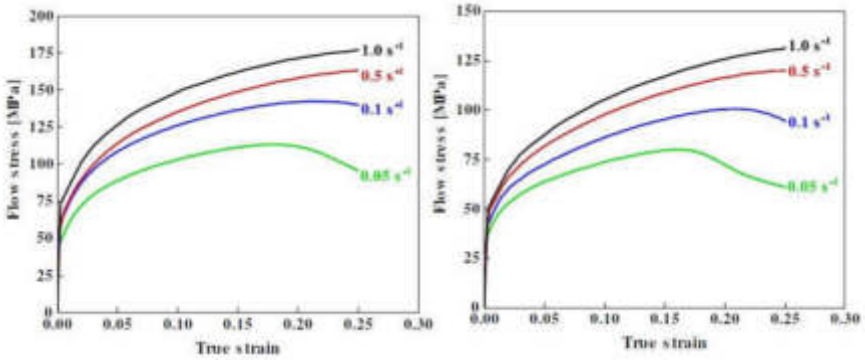


Fig. 4.5: Comparison of the results of a small order of magnitude flow correction and actual value.

massive historical data in the numerical control system. This project intends to use the improved sequence method to analyze observational data in the time domain. This reduces the influence of random error and the threshold of error discovery. The defects of conventional statistical testing methods are solved.

6. Acknowledgement. Shaanxi Provincial Education Department Project (21JK0627): Research on Mechanism and Path of Intellectual Property Protection to Promote Technological Innovation in Equipment Manufacturing Industry in Shaanxi Province.

REFERENCES

[1] Rai, R., Tiwari, M. K., Ivanov, D., & Dolgui, A. (2021). Machine learning in manufacturing and industry 4.0 applications. International Journal of Production Research, 59(16), 4773-4778.

[2] Jin, Z., Zhang, Z., Demir, K., & Gu, G. X. (2020). Machine learning for advanced additive manufacturing. Matter, 3(5), 1541-1556.

[3] Jiang, J., Xiong, Y., Zhang, Z., & Rosen, D. W. (2022). Machine learning integrated design for additive manufacturing. Journal of Intelligent Manufacturing, 33(4), 1073-1086.

[4] Alexopoulos, K., Nikolakis, N., & Chryssolouris, G. (2020). Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. International Journal of Computer Integrated Manufacturing, 33(5), 429-439.

[5] Solke, N. S., Shah, P., Sekhar, R., & Singh, T. P. (2022). Machine learning-based predictive modeling and control of lean manufacturing in automotive parts manufacturing industry. Global Journal of Flexible Systems Management, 23(1), 89-112.

Table 4.1: Results of time series correction.

	Flow rate /(t/h)				Temperature /K			
NO.	Truth value	Measured value	Corrected value	Calibration error /%	Truth value	Measured value	Corrected value	Calibration error /%
1	126.77	126.16	126.66	-0.08	117.13	116.82	117.09	-0.03
2	122.81	124.27	122.28	-0.44	115.01	117.61	114.98	-0.03
3	125.73	125.82	125.67	-0.08	117.11	117.15	117.08	-0.03
4	121.77	121.69	122.08	0.24	115.00	115.90	114.98	-0.03
5	130.42	131.04	130.70	0.20	218.07	217.85	218.15	0.03
6	110.83	111.96	110.43	-0.40	238.88	237.58	238.95	0.03
7	126.35	188.53	126.19	-0.14	209.82	207.42	209.88	0.03
8	129.48	129.47	129.38	-0.07	194.28	193.01	194.33	0.03
9	71.46	71.34	71.44	0.01	202.61	205.34	202.68	0.03
10	105.63	105.77	105.42	-0.20	288.48	288.07	288.57	0.03
11	105.63	105.19	105.71	0.09	288.48	288.58	288.57	0.03
12	108.85	56.23	108.52	-0.30	288.45	286.63	288.54	0.03
13	105.63	106.42	105.60	-0.01	288.48	291.23	288.57	0.03
14	425.63	425.77	425.25	-0.10	382.80	383.23	383.31	0.14
15	76.35	76.47	76.63	0.42	117.21	117.15	117.17	-0.03
16	30.31	30.23	30.33	0.14	164.52	165.52	164.51	0.00
17	76.56	76.42	76.84	0.43	227.79	233.00	227.77	-0.01
18	57.50	56.96	54.61	-5.27	335.84	334.25	335.80	-0.01
19	63.65	63.28	64.10	0.83	374.50	368.61	374.40	-0.03
20	64.58	65.19	65.13	0.83	374.48	369.59	374.36	-0.03
21	63.65	64.44	64.00	0.65	374.50	372.85	374.40	-0.03
22	64.58	65.16	65.03	0.68	374.48	376.29	374.36	-0.03
23	256.46	257.96	258.26	0.75	395.93	399.47	395.54	-0.10
24	15.63	15.50	15.61	0.07	156.27	156.33	156.27	0.00
25	65.00	65.17	65.48	0.83	260.70	257.08	260.75	0.02
26	48.75	71.96	48.90	0.34	316.99	313.24	317.06	0.02
27	0.00	0.00	0.00	0.16	386.46	383.29	386.46	0.00
28	126.15	126.77	127.26	0.95	387.80	391.03	388.07	0.07
29	1.01	1.02	1.01	0.04	83.33	82.96	83.33	0.00

[6] Xia, C., Pan, Z., Polden, J., Li, H., Xu, Y., & Chen, S. (2022). Modelling and prediction of surface roughness in wire arc additive manufacturing using machine learning. Journal of Intelligent Manufacturing, 33(5), 1467-1482.

[7] Fernandes, M., Corchado, J. M., & Marreiros, G. (2022). Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Applied Intelligence, 52(12), 14246-14280.

[8] Ranjan, N., Kumar, R., Kumar, R., Kaur, R., & Singh, S. (2023). Investigation of fused filament fabrication-based manufacturing of ABS-Al composite structures: prediction by machine learning and optimization. Journal of Materials Engineering and Performance, 32(10), 4555-4574.

[9] Jiang, J. (2023). A survey of machine learning in additive manufacturing technologies. International Journal of Computer Integrated Manufacturing, 36(9), 1258-1280.

[10] Putnik, G. D., Manupati, V. K., Pabba, S. K., Varela, L., & Ferreira, F. (2021). Semi-Double-loop machine learning based CPS approach for predictive maintenance in manufacturing system based on machine status indications. CIRP Annals, 70(1), 365-368.

[11] Sing, S. L., Kuo, C. N., Shih, C. T., Ho, C. C., & Chua, C. K. (2021). Perspectives of using machine learning in laser powder bed fusion for metal additive manufacturing. Virtual and Physical Prototyping, 16(3), 372-386.

[12] Chen, L., Yao, X., Xu, P., Moon, S. K., & Bi, G. (2021). Rapid surface defect identification for additive manufacturing with in-situ point cloud processing and machine learning. Virtual and Physical Prototyping, 16(1), 50-67.

[13] Liu, Z., Rolston, N., Flick, A. C., Colburn, T. W., Ren, Z., Dauskardt, R. H., & Buonassisi, T. (2022). Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. Joule, 6(4), 834-849.

[14] Farbiz, F., Habibullah, M. S., Hamadicharef, B., Maszczyk, T., & Aggarwal, S. (2023). Knowledge-embedded machine learning and its applications in smart manufacturing. Journal of Intelligent Manufacturing, 34(7), 2889-2906.

[15] Tercan, H., & Meisen, T. (2022). Machine learning and deep learning based predictive quality in manufacturing: a systematic review. Journal of Intelligent Manufacturing, 33(7), 1879-1905.
[16] Penumuru, D. P., Muthuswamy, S., & Karumbu, P. (2020). Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. Journal of Intelligent Manufacturing, 31(5), 1229-1241.
[17] Barrionuevo, G. O., Sequeira-Almeida, P. M., Ríos, S., Ramos-Grez, J. A., & Williams, S. W. (2022). A machine learning approach for the prediction of melting efficiency in wire arc additive manufacturing. The International Journal of Advanced Manufacturing Technology, 120(5), 3123-3133.
[18] Thakur, V., Kumar, R., Kumar, R., Singh, R., & Kumar, V. (2024). Hybrid additive manufacturing of highly sustainable Polylactic acid-Carbon Fiber-Polylactic acid sandwiched composite structures: Optimization and machine learning. Journal of Thermoplastic Composite Materials, 37(2), 466-492.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 11, 2024
Accepted: Jun 20, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 796–802, DOI 10.12694/scpe.v26i2.3902

APPLICATION RESEARCH OF IMPROVED PARTICLE SWARM COMPUTING INTELLIGENT ALGORITHM IN TRACK AND FIELD TRAINING TARGET OPTIMIZATION

CHAO WANG*

Abstract. This paper studies the influence of the athlete's state and wind direction on the Javelin flight trajectory. The process of plane throwing is analyzed theoretically. An accurate javelin flight mathematical model is established. Using the modified particle swarm optimization method, the paper finds a way to make the Javelin longer distance. This algorithm makes it portable and adaptive. Then, the entropy weight method is used to evaluate the weight of each factor. The experiment shows that this method can reflect the actual Javelin flying condition well, and its prediction effect is in accord with the demand of engineering applications.

Key words: Javelin; Particle swarm optimization; Flat throw motion; Entropy weight method

1. Introduction. Javelin throwing is an ancient sport. In ancient times, it was mainly used for hunting and fighting. Now it's an Olympic sport. In track and field competition, Javelin belongs to the relatively light equipment, but the technology is relatively complex. It can be divided into three stages: javelin holding, run-up and final throw. The most challenging part of this event is the final step after the jump. The quality of its throw will determine the trajectory and final landing of the Javelin. The thrower's motion in the air is determined by the motion form of the thrower and its technical parameters.

Domestic and foreign scholars have done a lot of research on object flight trajectory prediction. Literature [1] simulates surface targets by installing sensors on aircraft. However, this approach is too costly to be popularized. In literature [2], the Elman-NARX neural network is used for pilot track prediction based on Bayesian regularization. However, the application of this method is narrow and not easy to popularize. In reference [3], a method of track deviation compensation, the inverse functional of Lyapunov dynamics, is established. Four kinds of limited variable drag forces are used to simulate the actual situation, and the tracking error is small. The analysis method described in this paper is relatively simple in terms of the influence factors of each factor, so it is more applicable than the above methods.

2. Javelin model construction. The projectile presented in this paper is approximately a square in shape, and its cross-section is always constant. However, the shape of the national standard differs slightly from that described in the title. Firstly, the outline of the Javelin in the national standard is briefly explained (Figure 2.1).

Suppose D_6 is the center of gravity of the Javelin, $D_1 \sim D_9$ is the length of Part I, and the handle position $D_0 \sim D_1$ is Part II. Position $D_4 \sim D_0$ from the handle to the tip of the Javelin is the middle part, marked Part II.

It is necessary to determine the function relation of its external shape to obtain the projectile's area and the projectile center's position. The central axis corresponding to part I is in the range of $[0, 1200]$, while the Javelin gradually thickens as the axis length increases [4]. The long axis of the handle part II is between $[1200, 1800]$ and differs from the image prescribed by the state. The profile radius of this length is constant. In addition, this value includes the length of the long Javelin wrapped with the non-slip rope, which will be considered separately in later calculations. Particle III has a long axis at $[180, 2362)$. It becomes smaller as the radius of the section decreases, and the final tip N at $[2362, 2640]$ is approximately a convex cone. Because the

*Dalian University of Science and Technology, Dalian 116000, China (15040402587@163.com)

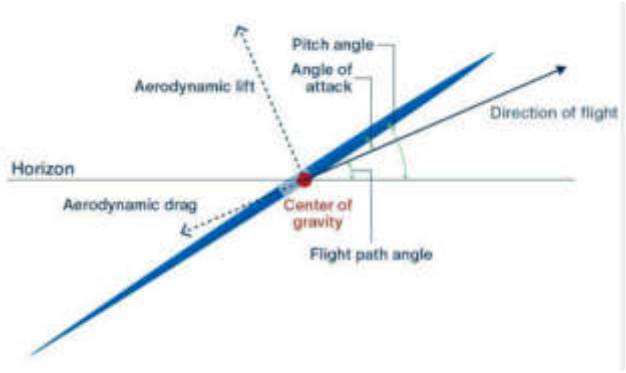


Fig. 2.1: Symbols of each part of the Javelin.

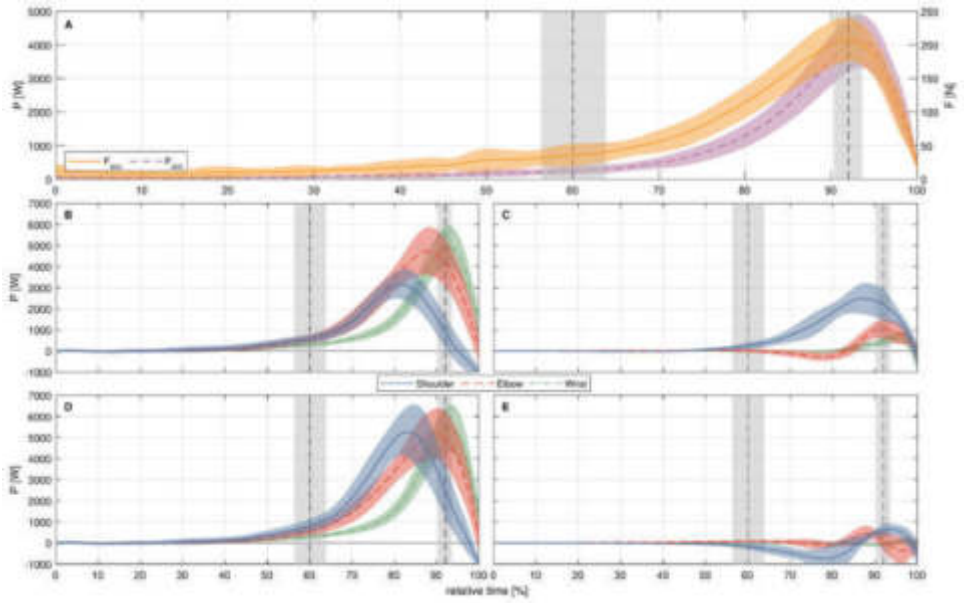


Fig. 2.2: Changes in relative errors of each part of the Javelin after fitting.

corresponding structure of the structure II part is cylindrical, it is not necessary to pay too much attention to it, and the rest of the part can be fitted with nonlinear equations.

Take half the length of each spindle. Use Google's deep network TensorFlow to fit radial parameters. The loss function is determined by the square of the difference between the measured and predicted values [5]. The algorithm uses the stochastic gradient method, sets the learning rate of the network to 0.01 , and conducts 1000 times of repeated training. The final error size obtained in the Tensor Board visual teaching software is shown in Figure 2.2.

When throwing, it is not possible to ensure that the muzzle velocity α of the Javelin is the same angle as the axis of the projectile [6]. If the initial angle of attack β is in the horizontal direction and the angle is more significant than the hand Angle α, the initial angle of attack β is greater than o. If the opposite, the initial angle of attack β is smaller than o. In the case of γ representing the gun Angle, each angle can meet the initial Angle of attack β = gun Angle γ - shot angle α.

This paper assumes that the Javelin will not spin during flight. All the power is concentrated in the center

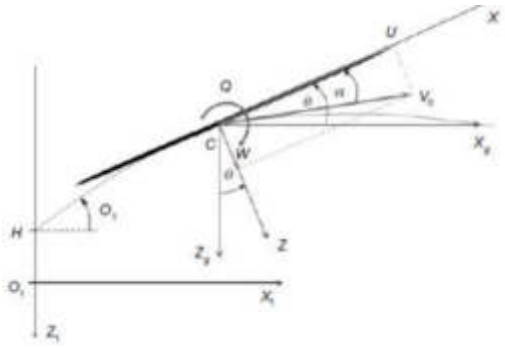


Fig. 2.3: Stress on Javelin during flight.

of the long Javelin. When the Javelin is thrown at the top, it is assumed to be in the horizontal position. At this time, when there is no wind, or the wind direction is horizontal, it is considered that it is not affected by wind resistance [7]. The trajectory of the Javelin has an Angle with the direction of the main shaft of the Javelin. Its dimensions are assumed to match the original conditions for easy calculation. An inverse relationship exists between wind speed and javelin speed under upwind conditions. This value corresponds directly to the squared value of the current rate [8]. In addition, it also relates to the wind coefficient c in the air, the air density p and the force area of the Javelin in the airflow. The forces on the throwing point are shown in Figure 2.3.

In the initial problem, the effect of external wind speed is not considered, so it is only necessary to calculate the drag force caused by the Javelin under the action of the object's relative speed [9]. Its dimensions will also change with the angle β. The β value is the angle of the Javelin's spindle and the rate of fire. During the rising period, its angle continues to decline. The size of β during the decline period can be regarded as constant after it has increased to a certain point.

The basic principles of Newtonian mechanics and momentum are analyzed [9]. The formula to be achieved in the ascending phase is as follows:

$$\begin{cases} F_x = f_x \\ F_y = f_y + G \\ \frac{1}{2} \cdot m \cdot v_1^2 - \frac{1}{2} \cdot m \cdot v_0^2 = -\int_0^h F_x ds - \int_2^h F_y ds \\ f_x = f \cdot \cos\theta \\ f_y = f \cdot \sin\theta \\ f = \frac{1}{2} \cdot \varepsilon \cdot \rho \cdot L \cdot v^2 \\ L = L_1 \cdot \sin\theta \\ \int_2^h F_y ds = \frac{1}{2} \cdot m \cdot v_y^2 \\ v_y = v_0 \cdot \sin\alpha \\ G = mg \end{cases}$$

To go from the highest point to the lowest point during the throw, the formula needed is:

$$\begin{cases} F_x = f_x \\ F_y = G - f_y \\ \frac{1}{2} \cdot m \cdot v_2^2 - \frac{1}{2} \cdot m \cdot v_1^2 = -\int_0^h F_x ds - \int_0^h F_y ds \\ f_x = f \cdot \cos\alpha \\ f_y = f \cdot \sin\alpha \\ f = \frac{1}{2} \cdot \varepsilon \cdot \rho \cdot L \cdot v^2 \\ L = L_1 \cdot \sin\theta \\ \int_0^h F_y ds = \frac{1}{2} \cdot m \cdot v_y^2 \\ v_y = v_2 \cdot \sin\alpha \\ G = mg \end{cases}$$

The action function of the thrower to be constructed in this problem looks like this:

$$\begin{cases} \int_0^{l_1} \left(v_x^2 + v_y^2\right) \cdot \sin 2\theta dv_x \cdot t = \frac{2 \cdot m \cdot v_0^2 \cdot \cos^2 \alpha - 2 \cdot m \cdot v_1^2}{\varepsilon \cdot \rho \cdot L_1} \\ \int_0^{l_2} \left(v_x^2 + v_y^2\right) \cdot \sin \theta \cdot \cos \alpha dv_x \cdot t = \frac{m \cdot v_1^2 - m \cdot v_2^2 - m \cdot v_y^2}{\varepsilon \cdot \rho \cdot L_1} \\ l = l_1 + l_2 \end{cases}$$

In this way, the mathematical models of the following several javelins can be expressed by the above formula.

3. Particle swarm algorithm. Particle swarm is a population-based optimal search method in which each particle in the population represents a possible solution to the problem, and the fit value of the solution is associated with the objective function. During the search, the particle will adjust its speed according to its flight history and the optimal particle position [10]. They can do this by cooperating and competing to find the best particles. The modified equation of motion and positioning mode of the particle swarm optimization method is as follows:

$$v_{id}(t+1) = \lambda v_{id}(t) + \delta_1 \left(\eta_{id}(t) - u_{id}(t)\right) + \delta_2 \left(f_{id}(t) - u_{id}(t)\right)$$
$$u_{id}(t+1) = u_{id}(t) + v_{id}(t+1)$$

where $\delta_1 = z_1$ Rand $1(), \delta_2 = z_2$ Rand $2($ $()$. The constants z_1 and z_2 represent the particle's effect on social and personal cognitive levels, and are generally expressed in the same numerical value. Rand1() And Rand 2() is any number in any range $(0, 1)$. λ is the inertial weight. The concept of the grey correlation degree is used to study the correlation of each

$$U_0 = \left(u_0(1), u_0(2), \ldots, u_0(n)\right),$$

$$U_1 = \left(u_1(1), u_1(2), \ldots, u_1(n)\right), \ldots, .$$
$$U_m = \left(u_m(1), u_m(2), \ldots, u_m(n)\right)$$

The correlation factor $\zeta\left(u_0(r), u_i(r)\right)$ between U_0 and $U_1, U_2, \ldots, U_m$ is:

$$\zeta\left(u_0(r), u_i(r)\right) = \frac{\min_i \min_r \left|u_0(r) - u_i\left(r^r\right)\right| + \xi \max_i \max_r \left|u_0(r) - u_i\left(r^r\right)\right|}{\left|u_0(r) - u_i(r)\right| + \xi \max_i \max_r \left|u_0(r) - u_i(r)\right|}$$
$$r = 1, 2, \ldots, n$$

Here the degree of grey correlation between $\xi \in (0, 1), U_i$ and U_0 is $\zeta\left(U_0, U_i\right) = \frac{1}{n} \sum_{r=1}^{n} \zeta\left(u_0(r), u_i(r)\right)$, where ξ is the differentiation coefficient. Regularization is usually used to improve the learning effect of neural networks.

Regularization is usually used to improve the learning effect of neural networks. The error function of the standard method is expressed by $W = \omega \cdot W_C + \varphi \cdot W_\Lambda$. $W_C = \frac{1}{2} \sum_{r=1}^{K} \sum_{j=1}^{N} \left(c_{jr} - y_{jr}\right)^2, N$ is the sample quantity. K is the number of output nodes of the neural network. c_{jr} is the quantity required. y_{jr} indicates the actual network output. $W_\Lambda = \frac{1}{2} \sum_{i=1}^{\Lambda} \lambda_i^2, W_\Lambda$ is for formal fines. Where λ_i is the weight and threshold for training the network of neurons. Where Λ is the number of connected weights and thresholds. φ, ω is the super parameter. The size of φ, ω directly affects the learning purpose of the network.

4. Model solving. The throw Angle is the angle of the throw point to the ground. The starting angle of attack is the angle between the motion trajectory of the center of the Javelin and the axis of the Javelin during the throwing process [11]. The Javelin has an Angle of attack of 0 during flight. The initial angle of attack is negative when above the vertical axis of the Javelin. The angle of attack is positive when the Javelin's trajectory is down. The same type of Javelin was tested in this study [12]. It is necessary to find the relationship between throwing distance and shooting Angle, initial attack Angle and shooting speed according to the existing experimental data, and then find the trajectory of the Javelin in the air.

First, three lines were constructed with the player's shot Angle, shot speed and initial attack Angle as three independent variables [13]. The internal variables of each line are 24 lines, which are the measured values of

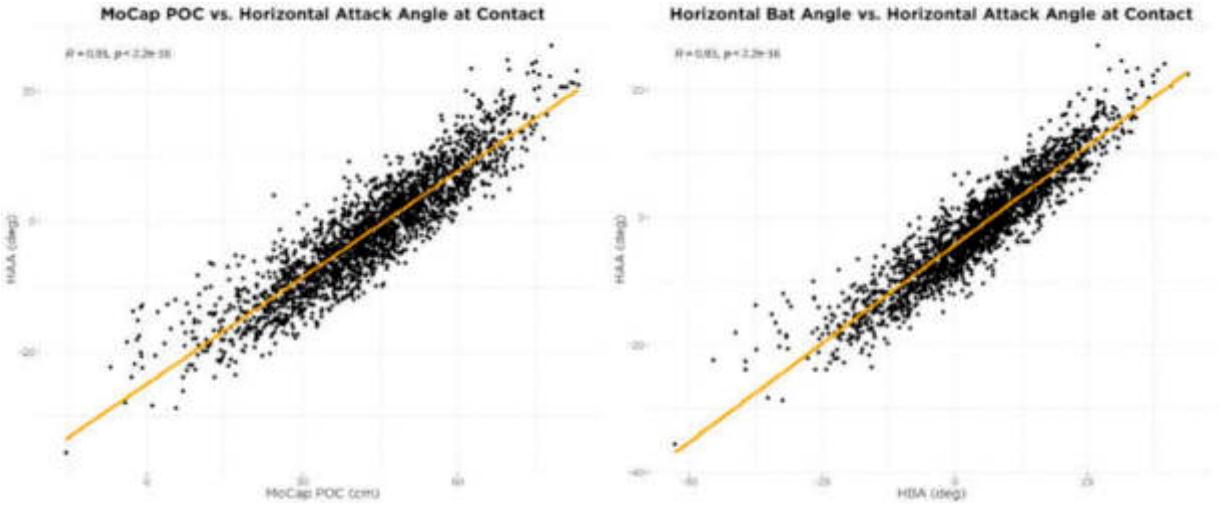


Fig. 4.1: Scatter diagram of the influence of shot Angle on throwing distance.

each independent variable. A line vector containing 24 measured values is created. SPSS statistical analysis tool was used to test the three independent vectors constructed using the throw's distance as the dependent variable. The three independent variables were analyzed statistically and tested by MATLAB. The R square values obtained by the two methods are compared with the regression coefficients obtained [14]. Through the statistics of the experimental results, it is found that factors such as the angle, the speed of the motion and the starting angle of the thrower are related to the motion of the Javelin.

From Figure 4.1, most of the points show a tendency to increase as the angle of the shot increases, but not as significantly as the speed of the shot increases. This is also consistent with the regression coefficient before the release speed is used as the independent variable, which is as high as 0.476. Although still a positive regression coefficient, its value is much smaller than the regression coefficient when the release speed is used as the independent variable. However, as people can see from the graph, there is also a positive correlation between the distance thrown and the angle of the shot [15].

MATLAB software is used to verify the multi-variable regression model. After some error points are eliminated, each data contains zero points, which verifies the accuracy and practicability of the method. When other initial angles of attack remain unchanged, the change of initial angle of attack does not change significantly [16]. The optimal time in the combination of various throwing angles and initial angles of attack is A0=42°, B0=-5°.

The particle swarm optimization method combines the individual optimal results with the population optimization results to solve the local optimization [17]. Positioning constraint [28,44], set the initial population N as 50, repeat 100 times, and randomly generate a 50x1 data matrix from 28-44. Compare the current fitness and the best solution set with the historical record, and update its optimal positioning and solution if it exceeds it.

It can be seen from Figure 4.2 that this method has successfully found the best scheme. The paper gets the best result: A0=42°, B0=-5°. The optimum initial attack Angle of the Javelin varies with the throwing Angle, so the optimum initial attack Angle is not a constant value. There is some correspondence between the two. When the projection Angle is large, the corresponding optimal initial impact Angle is smaller [18]. As the projection Angle decreases, the corresponding optimal initial attack Angle also increases, and in some cases, these two parameters are optimal.

The simulation results show that when other initial conditions are unchanged, whether downwind or upwind, the flight range of the Javelin will gradually become longer with the increase of the throwing Angle in the initial state. It shrinks again when the throwing Angle increases to 42°.

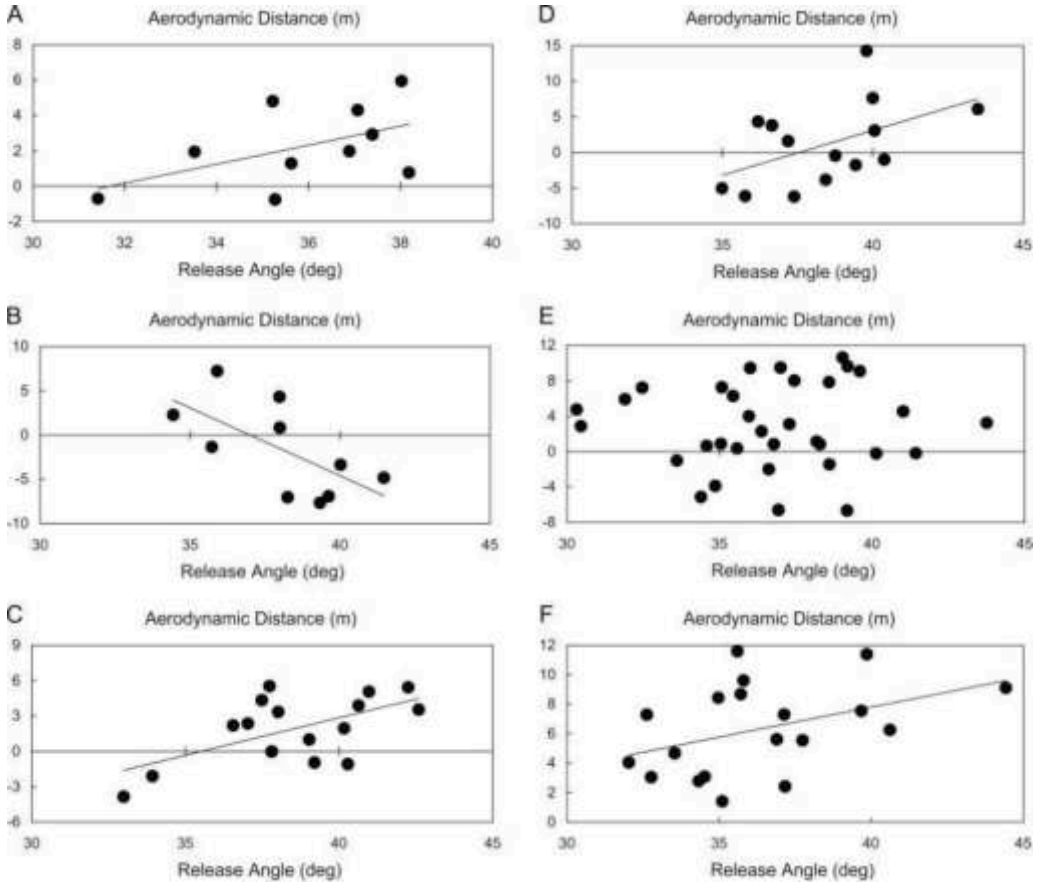


Fig. 4.2: Variation of throwing distance with shot Angle.

5. Conclusion.

1. When other initial values are the same, the particle swarm optimization method is used to conduct the research. The results show that when the throwing Angle is 0°-42°, the throwing amount increases with the increasing angle.
2. Under the premise that other initial conditions are the same and the throwing Angle is fixed, the joint action of the initial attack Angle and the initial inclination angle greatly influence the projectile's range. The optimal throwing quantity of this combination is A0=42°and B0=-5°.

REFERENCES

[1] Pawan, Y. N., Prakash, K. B., Chowdhury, S., & Hu, Y. C. (2022). Particle swarm optimization performance improvement using deep learning techniques. Multimedia Tools and Applications, 81(19), 27949-27968.
[2] Gad, A. G. (2022). Particle swarm optimization algorithm and its applications: a systematic review. Archives of computational methods in engineering, 29(5), 2531-2561.
[3] Cansu, T., Kolemen, E., Karahasan, Ö., Bas, E., & Egrioglu, E. (2023). A new training algorithm for long short-term memory artificial neural network based on particle swarm optimization. Granular Computing, 8(6), 1645-1658.
[4] Liu, X. H., Zhang, D., Zhang, J., Zhang, T., & Zhu, H. (2021). A path planning method based on the particle swarm optimization trained fuzzy neural network algorithm. Cluster Computing, 24(3), 1901-1915.
[5] Berbek, M. I., & Oglah, A. A. (2022). Adaptive neuro-fuzzy controller trained by genetic-particle swarm for active queue management in internet congestion. Indonesian Journal of Electrical Engineering and Computer Science, 26(1), 229-242.
[6] Usmani, U. A., Watada, J., Jaafar, J., Aziz, I. A., & Roy, A. (2021). Particle swarm optimization with deep learning for human action recognition. Int. J. Innovative Comput. Inform. Control, 17(6), 1843-1870.

[7] Yilmaz, A., & Yolcu, U. (2022). Dendritic neuron model neural network trained by modified particle swarm optimization for time-series forecasting. Journal of Forecasting, 41(4), 793-809.
[8] Elsedimy, E. I., AboHashish, S. M., & Algarni, F. (2024). New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. Multimedia Tools and Applications, 83(8), 23901-23928.
[9] Yousef, B. A., Rezk, H., Abdelkareem, M. A., Olabi, A. G., & Nassef, A. M. (2020). Fuzzy modeling and particle swarm optimization for determining the optimal operating parameters to enhance the bio-methanol production from sugar cane bagasse. International Journal of Energy Research, 44(11), 8964-8973.
[10] Ben Seghier, M. E. A., Carvalho, H., Keshtegar, B., Correia, J. A., & Berto, F. (2020). Novel hybridized adaptive neuro-fuzzy inference system models based particle swarm optimization and genetic algorithms for accurate prediction of stress intensity factor. Fatigue & Fracture of Engineering Materials & Structures, 43(11), 2653-2667.
[11] Golafshani, E. M., Behnood, A., & Arashpour, M. (2023). Predicting the compressive strength of eco-friendly and normal concretes using hybridized fuzzy inference system and particle swarm optimization algorithm. Artificial Intelligence Review, 56(8), 7965-7984.
[12] Kumar, G., Singh, U. P., & Jain, S. (2022). An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting. Soft Computing, 26(22), 12115-12135.
[13] Diaz Martinez, D., Trujillo Codorniu, R., Giral, R., & Vazquez Seisdedos, L. (2021). Evaluation of particle swarm optimization techniques applied to maximum power point tracking in photovoltaic systems. International Journal of Circuit Theory and Applications, 49(7), 1849-1867.
[14] Berlin, S. J., & John, M. (2020). Particle swarm optimization with deep learning for human action recognition. Multimedia Tools and Applications, 79(25), 17349-17371.
[15] Sahour, S., Khanbeyki, M., Gholami, V., Sahour, H., Karimi, H., & Mohammadi, M. (2024). Particle swarm and grey wolf optimization: enhancing groundwater quality models through artificial neural networks. Stochastic Environmental Research and Risk Assessment, 38(3), 993-1007.
[16] Nayak, J., Swapnarekha, H., Naik, B., Dhiman, G., & Vimal, S. (2023). 25 years of particle swarm optimization: Flourishing voyage of two decades. Archives of Computational Methods in Engineering, 30(3), 1663-1725.
[17] Balasubramanian, K., Ananthamoorthy, N. P., & Ramya, K. (2022). An approach to classify white blood cells using convolutional neural network optimized by particle swarm optimization algorithm. Neural Computing and Applications, 34(18), 16089-16101.
[18] Saha, S., Saha, A., Roy, B., Sarkar, R., Bhardwaj, D., & Kundu, B. (2022). Integrating the Particle Swarm Optimization (PSO) with machine learning methods for improving the accuracy of the landslide susceptibility model. Earth Science Informatics, 15(4), 2637-2662.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 11, 2024
Accepted: Jun 20, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 803–813, DOI 10.12694/scpe.v26i2.3903

SENSORY STYLING DESIGN OF PHYSIOTHERAPY BEDS BASED ON BP NEURAL NETWORK

CHEN SU*, CHANGJUN LI†, YUPENG JIANG‡, AND XINCAN LI§

Abstract. There is a significant gap between the form of current physiotherapy beds and users' perceptual images. By employing techniques such as computer graphic design and logical operations, analyzing the mathematical relationship between user needs and design elements contributes to enhancing the scientific nature of product design and making the design process more rigorous. This paper, based on an analysis of the product image and design process, employs a comprehensive fuzzy evaluation method to identify representative perceptual vocabulary for physiotherapy beds. The KJ method and Delphi method are utilized to select necessary samples. A morphological analysis matrix is established using the morphological analysis method, and a comprehensive decision-making model for product design is constructed using a BP neural network on the MATLAB platform. Through training the BP neural network on physiotherapy bed products, it becomes possible to predict the perceptual evaluation of product design, achieving a quantification of the design. This provides valuable support for the appearance design of physiotherapy bed products and significantly enhances the efficiency of designers' work.

Key words: BP neural network; Physiotherapy bed; Perceptual image; Modelling design

1. Introduction. In recent years, physiotherapy methods have matured, leading to the active involvement in China about the exploration of physiotherapy products [1]. With increased health awareness and advances in medical technology, the demand for physiotherapy beds has gradually risen. Currently, there is a structural disparity between users' perceptual demands and designs for these physiotherapy products, evolving from rigid requirements to perceptual needs [2]. Beyond functional requirements, users also prioritize inner emotional experiences, necessitating designers to convey product characteristics through specific forms. Therefore, assessing the perceptual cognitive elements in product design holds crucial significance in shortening design cycles and reducing development costs.

As a design approach that translates users' perceptual images into design elements, Kansei Engineering aligns with the trend of constructing mathematical models for quantified research objectives [3]. In past practices, Zhao Yanan et al. constructed a Kansei image prediction model for office chairs based on a BP neural network [4]. Ding Lu et al. applied a BP neural network to optimize the overall design of a programmable paper cutter, yielding improved results in perceptual design [5]. Ma S and Yan X designed an intelligent clothing pattern design system based on BP neural network [6]. Chen DL also combined Kansei Engineering and BP neural networks in the development of a product form design system [7]. However, in the application process, subjective methods often dominate the selection of one or multiple suitable perceptual terms. The utilized BP neural network typically retains a single hidden layer even when dealing with multiple outputs, limiting selection and resulting in lower accuracy. This can lead to non-scientific neural network outputs or suboptimal weight allocation.

Thus, this paper, after obtaining raw data through multiple questionnaire surveys and conducting multivariate analysis and cluster analysis using SPSS, identifies representative perceptual vocabulary for physiotherapy bed design. The morphological analysis method is employed to deeply deconstruct the design elements of physiotherapy beds, resulting in a comprehensive perceptual image assessment matrix. For cases involving multiple outputs, a BP neural network is designed with multiple hidden layers to increase combination possi-

*School of Industrial Design, Hubei University of Technology, Wuhan, Hubei, 430072, China
†School of Industrial Design, Hubei University of Technology, Wuhan, Hubei, 430072, China (Corresponding author, li12717418@163.com)
‡School of Industrial Design, Hubei University of Technology, Wuhan, Hubei, 430072, China
§School of Industrial Design, Hubei University of Technology, Wuhan, Hubei, 430072, China

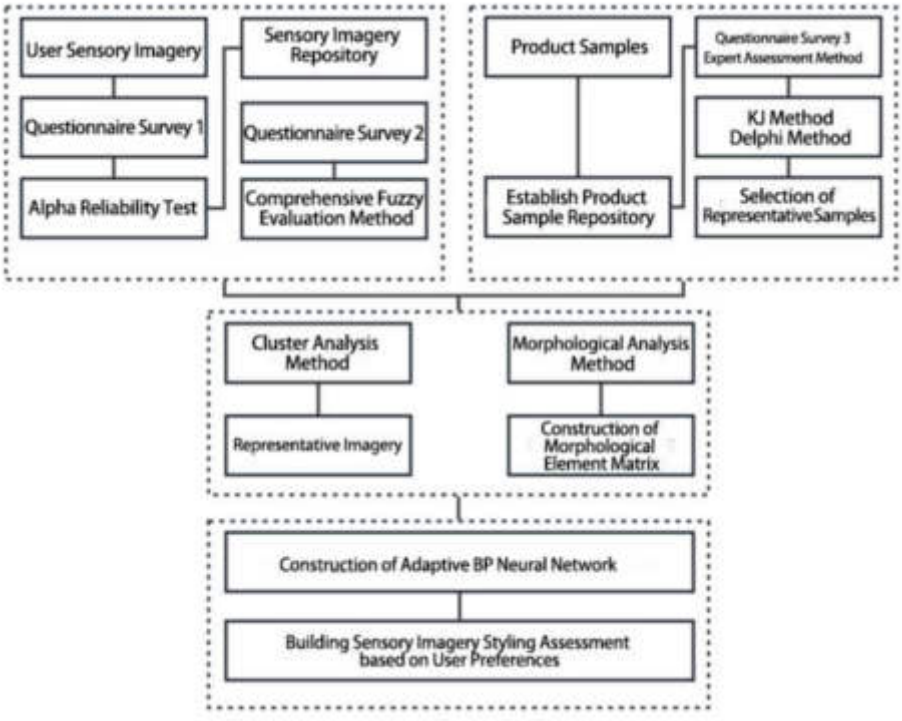


Fig. 2.1: Research Process and Methodology.

bilities, establishing a mapping relationship between the two. Simulated predictions are performed to render the previously experientially driven design process more scientifically grounded.

2. Research Process and Methodology. When considering the conceptual design output, designers need to combine user requirements, design patterns, and bidirectional inference [8]. This involves transforming users' subjective perceptions into comprehensible images, effectively integrating the visual attributes of a product with users' perceptual images. In light of the specific details of the research, this paper establishes the research framework into four main sections, as illustrated in Figure 2.1.

2.1. Collection and Selection of Perceptual Images. In this phase, a diverse range of perceptual vocabulary is gathered to express various potential perceptual images, utilizing these words to embody and elucidate the semantic aspects of the product. This enables the analysis of human perceptions to derive novel content for product development [10]. Based on the Semantic Differential (SD) method questionnaire results, this paper employs comprehensive fuzzy evaluation. As the representative samples are formed in the cognitive space dimension based on the size of the sample population and the complexity of the design form, mathematical techniques such as multivariate analysis and cluster analysis are utilized to establish representative images.

2.2. Sample Filtering and Selection. A crucial step in extracting users' perceptual images involves obtaining samples through methods like interviews or surveys. The sample repository typically encompasses physical samples and virtual samples. Physical samples refer to tangible products available on the market, while virtual samples are designs simulated by various designers. Thus, this paper collects product design samples from sources such as online marketplaces, official company websites, and journals. The representative sample library within these design samples then needs to be selected.

2.3. Construction of Product Design Elements Repository. The repository of product design elements serves as the foundation for constructing perceptual image designs. Products are defined by multiple categories, each containing diverse design elements. Additionally, functional deconstruction of products is necessary to focus on the primary functional units, reducing the influence of other factors. Therefore, in analyzing design elements, categorization is based on design characteristics and functional components.

2.4. Establishment of Product Image Design Model. The construction of image design models often relies on intelligent algorithms in computing, frequently using methods such as BP neural networks, ant colony algorithms, genetic algorithms, and support vector machines [9]. BP neural networks, a powerful algorithm, possess strong nonlinear mapping capabilities. They emulate biological neural processing systems and unique human learning and cognitive patterns [10]. Due to their effectiveness in managing and establishing complex relationships between input and output variables, they find extensive application in product design and related fields [11].

3. Handling User Perceptual Images. The perceptual images obtained from the questionnaire survey serve as raw data. Given the vagueness and complexity of perceptual images, it's necessary to perform quantitative processing on the collected perceptual images.

3.1. Categorization and Fitting of Perceptual Images. The perceptual images collected through user surveys tend to be scattered and lack concentration, making it difficult to identify representative vocabulary or determine the number of categories. Commonly, previous research has relied on subjective judgment for categorization. Therefore, this paper aims to determine the number of categories for perceptual images through distributing questionnaires and processing the data.

In the questionnaire survey, participants are asked to group semantically similar words into the same category. The number of categories ranges from 4 to 10, with varying word counts per group, ensuring that all words are used without repetition. Based on the grouping results, the frequency of pairs of perceptual image words being placed in the same category is calculated, leading to the creation of a matrix of similar frequencies.

Considering the fit analysis of perceptual image words along with the content analysis, the dimensions of the matrix are set to M*M, where M represents the number of initial perceptual images collected. The matrix is represented using $\Delta = (\mathrm{x_{ij}})\,(\mathrm{i,j} = 1, 2 \ldots 40)$, as shown in formula 3.1.

$$\Delta = \begin{pmatrix} \mathrm{X_{1,1}} & \cdots & \mathrm{X_{1,40}} \\ \vdots & \ddots & \vdots \\ \mathrm{X_{40,1}} & \cdots & \mathrm{X_{40,40}} \end{pmatrix}$$

$\mathrm{x_{ij}}$-The frequency of the i-th word and the j-th image appearing in the same category. Based on the matrix of similar frequencies, we can utilize Multidimensional Scaling (MDS) to assess the fit between words, enabling a deeper exploration of their underlying relationships and calculating their coordinates in a lower-dimensional space. MDS constructs a distance matrix $\mathrm{D_{ij}} = (\mathrm{d_{ij}})$ between objects based on the similarity matrix $\mathrm{C_{ij}} = (\mathrm{c_{ij}})$, with the transformation between data represented by formula 3.2.

$$\mathrm{d_{ij}} = \sqrt{(\mathrm{c_{ii}} + \mathrm{c_{jj}} - 2\mathrm{c_{ij}})}$$

By performing k-dimensional fitting, we can generate a grid X of size $\mathrm{n} * \mathrm{k}$, which corresponds to a complex system composed of a distance matrix in the form of $n*k$. This system contains off-diagonal elements. Through arranging these elements and understanding their relationships, effective fitting in K-dimensions can be achieved. To facilitate k-dimensional fitting, a grid X of size $\mathrm{n} * \mathrm{k}$ is constructed, with X corresponding to a distance matrix $\widehat{\mathrm{D}}_{\mathrm{ij}} = \left(\widehat{\mathrm{d}_{\mathrm{j}}}\right)$ that contains information as in equation $D_{ij} = d_{ij}$. Non-diagonal elements from $D_{ij} = d_{ij}$ are selected, sorted in ascending order, and labeled as follows:

$$S^2(\widehat{X}) = \frac{\min \sum_{ipj} \left(d_{ij}^* - \widehat{d}_{ij}\right)}{\sum_{ipj} d_{ij}^2}$$

Table 3.1: Fitting Goodness Empirical Criteria.

Stress Value	Fitting Degree	Stress Value	Fitting Degree
Stress 20%	Poor	Stress 2.5%	Very Good
Stress 10%	Satisfactory	Stress=0	Perfect Match
Stress 5%	Good		

Adjusting $\mathrm{d}_{\mathrm{ij}}^{*}$ to achieve the maximum value of $\mathrm{S}^2(\widehat{\mathrm{X}})$, at this point, $\{\mathrm{d}_{\mathrm{ij}}\}$ corresponding to the extreme minimum value of $S^2(\widehat{\mathrm{X}})$ is referred to as the least squares regression of $\{\mathrm{d}_{\mathrm{ij}}\}$. It is assumed under the premise of invariance, and there exists a $\widehat{\mathrm{X}}$ such that:

$$\mathrm{S}^2(\widehat{\mathrm{X}}) = \min_{\mathrm{X_{n+k}}} \mathrm{S}(\widehat{\mathrm{X}}) = \mathrm{S_k}$$

In this case, $\widehat{X_0}$ is referred to as the best-fitting construction point, while S_k represents the pressure index. Once the perceptual image vocabulary objects in the K-dimensional space are positioned, their original similarity can be reflected by transforming the "distance" coordinates between each word. To assess the fitting effectiveness in the K-dimensional space, the matrix is multidimensionally unfolded, and their composite degree is determined based on the empirical standard of fitting goodness. The criteria for fitting goodness are presented in Table 3.1.

3.2. Selecting Representative Perceptual Vocabulary. After obtaining the number of clusters K for the perceptual image vocabulary, further dimension reduction will be carried out using clustering methods to obtain representative perceptual image vocabulary. Common clustering methods include hierarchical clustering and K-means clustering. In this study, hierarchical clustering will be employed for clustering. By comparing the clustering results, the optimal output solution for representative perceptual image vocabulary clustering will be chosen.

Hierarchical clustering is an approach that organizes data through successive splitting and merging, resulting in a hierarchical tree-like structure. In hierarchical clustering, the Euclidean distance is commonly used as the calculation criterion, and the between-group average linkage method is employed. The distance from sample x to cluster G is defined as shown in formula 3.5.

$$D(x-G) = \frac{1}{n}\sum_{j=1}^{n} D\left(x-g_j\right)$$

n represents the number of samples in the cluster; x represents the number of cases in cluster G; $D\left(c-g_j\right)$ represents the Euclidean distance between a case and another case within cluster G. After conducting hierarchical clustering analysis using SPSS, a dendrogram representing the clustering analysis is obtained. Based on this dendrogram and the predetermined number of clusters K, representative perceptual vocabulary for the product can be identified.

4. BP Neural Network Design. BP neural networks possess the advantage of the error backpropagation algorithm, enabling effective modeling of complex nonlinear functions. This provides an efficient approach to explore the design of physiotherapy beds and its connection with other similar design requirements.

4.1. Structure of the BP Neural Network. A BP neural network consists of an input layer, hidden layers, and an output layer, with varying numbers of interconnected neurons between layers. Signals propagate forward through the input, hidden, and output layers, while errors propagate backward in the opposite direction, iteratively adjusting the weights and biases between layers. Iteration continues until the output error meets the desired precision or a preset number of iterations is reached. The topology of the BP neural network is illustrated in Figure 4.1.

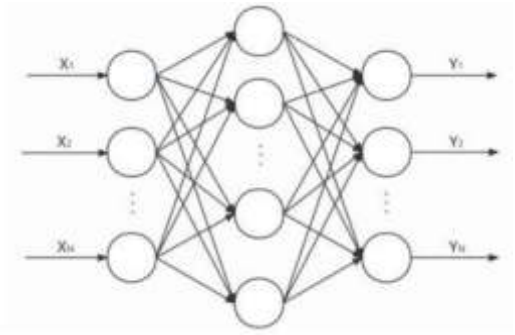


Fig. 4.1: Structure of the BP Neural Network.

4.2. Determining the Number of Hidden Layer Neurons. The number of neurons in the hidden layer plays a role in uncovering underlying patterns in the training samples and converting them into weights. The quantity of hidden layers also impacts the accuracy of the BP neural network. Too few hidden layers can compromise the approximation precision of nonlinear functions, while too many may lead to longer training times and overfitting. Determining the number of neurons in the hidden layer generally relies on empirical formulas to avoid futile attempts and unnecessary calculations. Therefore, an empirical formula is employed here using a trial-and-error method for determination. The preliminary range for estimation is shown in formula 4.1.

$$\mathrm{L} < \sqrt{\mathrm{M} + \mathrm{N}} + \mathrm{A}$$

L - Number of neurons in the hidden layer; M - Number of neurons in the input layer; N - Number of neurons in the output layer; A - Adjustment constant typically taken between 1 and 10

4.3. Selection of Activation Function. In the BP neural network, the transfer function employs a nonlinear transformation function, specifically the Sigmoid function. This function is further divided into unipolar and bipolar types. In the neural network constructed in this paper, the unipolar version will be employed, as shown in formula 4.2, with a value range of $(0, 1)$.

$$\log \mathrm{si}(\mathrm{x}) = \frac{1}{1 + \mathrm{e}^{-\mathrm{x}}}$$

Meanwhile, the output layer will employ the purelin function, as shown in formula 4.3 .

$$y = x$$

4.4. Data Normalization. In order to facilitate rapid convergence of the network and prevent values from being too widely dispersed, which can hinder neural network learning, data normalization is necessary to scale the data to the range $[-1, 1]$. This paper will use formula 4.4 as the normalization formula.

$$\mathrm{x}_{\mathrm{nrom}} = 2 \times \frac{\mathrm{x}^{-\mathrm{x}_{\mathrm{min}}}}{\mathrm{x}_{\mathrm{max}} - \mathrm{x}} - 1$$

5. Example Verification. Using fuzzy analysis and BP neural network f innovative design of physiotherapy bed forms.

5.1. Establishment and Selection of Perceptual Images. After systematic research, a total of 145 perceptual vocabulary words related to physiotherapy beds were collected through various means such as user interviews, journals, and evaluation reports. These vocabulary words encompass research outcomes from multiple fields including perceptual engineering, design studies, and product aesthetics, in order to better represent

Table 5.1: Perceptual Image Vocabulary List after Frequency Screening.

Perceptual Image Vocabulary				
Humanized	Safe	Clean	Technological	Precise
User-friendly	Sturdy	Tidy	Practical	Professional
Stable	Friendly	Reliable	Advanced	Simple
Durable	Approachable	Efficient	Harmonious	Comfortable
Modern	Smooth	Accurate	Neat	Round
Orderly	Strong	Automated	Steady	Diamantine
Steadfast	Pleasant	High-end	Gentle	Expensive
Substantial	Organized	Elegant	Soft	Delicate

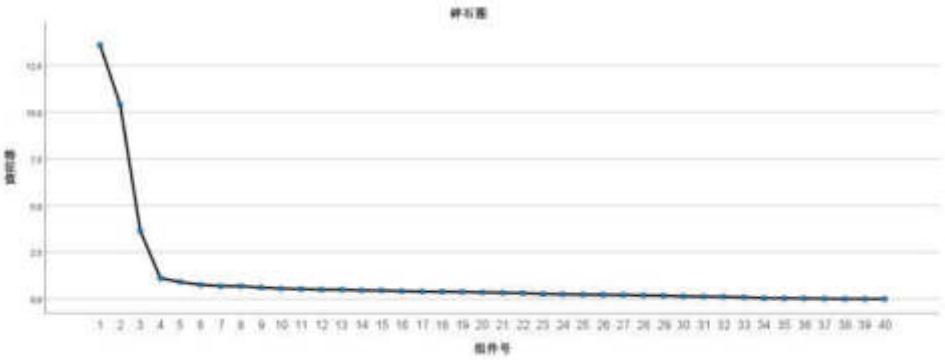


Fig. 5.1: Scree Plot of Perceptual Image.

Table 5.2: Fitting Goodness Evaluation.

Dimension	Stress ( )	Coefficient of Determination (RSQ)	Fitting Degree
4	0.049	0.973	Good
5	0.029	0.982	Good
6	0.019	0.987	Very Good
7	0.014	0.99	Very Good

user perceptual images. After thorough investigation, 152 questionnaires were distributed to engineers and designers associated with physiotherapy beds, yielding 145 valid responses. Following meticulous screening, vocabulary words with frequencies below 5% were excluded, as statistically, frequencies below 5% lack universal representativeness. This process resulted in retaining 70 more representative vocabulary words. After organizing the approximations and inconsistencies of these words, and undergoing selection by an expert group, 40 perceptual vocabulary words were obtained as the initial perceptual image library (See in table 5.1).

After involving 42 individuals with diverse backgrounds, we conducted a grouped questionnaire survey, ultimately receiving 35 completed questionnaires. Upon collecting the questionnaire data, the data was organized into a 40x40 matrix based on formula 1. The collected data was subjected to a correlation analysis of perceptual factors using SPSS. At this point, with the assistance of the scree plot in Figure 5.1, the number of factors can be inferred. When the line passes through 4, the sudden steepness becomes stable, suggesting that the optimal number of reference factors for this matrix is 4 or more groups.

Subsequently, the matrix was positioned in a K-dimensional space. After unfolding the matrix multidimensionally, coordinates for 40 vocabulary words were obtained in the K-dimensional space, as shown in Table 5.2, the fitting goodness was evaluated for dimension numbers of 4, 5, 6, and 7.

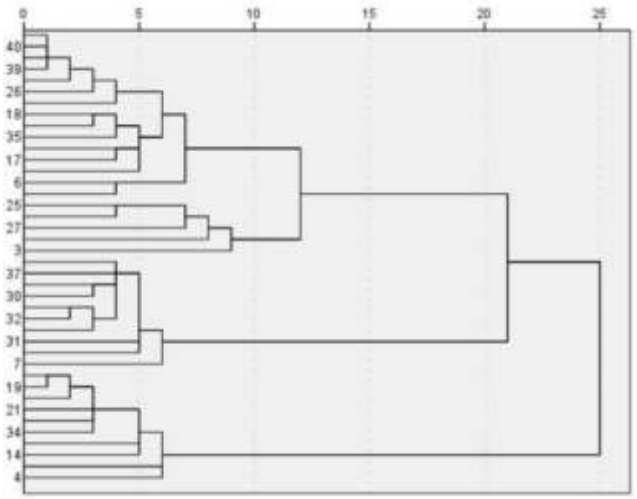


Fig. 5.2: Hierarchical Clustering Dendrogram of Perceptual Image Vocabulary.

Table 5.3: Hierarchical Clustering Grouping Results.

Group	Representative Vocabulary	Words within the Group
Group1	Safe	Technological, Precise, Sturdy, Stable, Reliable, Advanced, Durable, Efficient, Accurate, Neat, Strong, Automated, Steady, Diamantine, Expensive, Substantial, Steadfast, Modern, Professional, Organized, Delicate
Group2	Soft	Comfortable, Gentle
Group3	Friendly	Elegant, Pleasant, Round
Group4	Humanized	Harmonious, Smooth, Approachable, Orderly, High-end, Tidy, User-friendly, Clean, Simple, Practical

Considering the need for simplicity in the BP neural network's output and to avoid setting too many nodes in the output layer, it is more reasonable to set the number of clusters to 4 in the clustering analysis. Moreover, when the dimension is set to 4, with Stress $(S_k) = 0.040 < 0.05$, it indicates a good fitting. Therefore, it is judged that these 40 vocabulary words can be divided into 4 groups, obtaining the number of groups for the perceptual vocabulary in the physiotherapy bed evaluation system.

This paper will employ hierarchical clustering separately to achieve clustering. Through the comparison of clustering results, the optimal output scheme for perceptual image vocabulary selection will be chosen. In this case, 4 representative vocabulary words will be selected as the optimal output. Hierarchical clustering analysis was conducted using SPSS, resulting in a dendrogram as shown in Figure 5.2. Upon analysis, it can be determined that the representative perceptual vocabulary words are "Safe," "Humanized," "Friendly," and "Soft." Further details about the grouping results can be found in Table 5.3.

5.2. Constructing Sample Styling Feature Codes. Through online collection, a total of 68 initial samples were obtained. After undergoing the KJ method and Delphi method screening, 12 samples were selected which shared similar styling functionalities and exhibited representative qualities. Based on the morphological analysis approach, these 12 samples were deconstructed to establish a repository of styling elements. According to the functional aspects and styling characteristics of therapeutic beds, these samples were broken down into 9 distinct styling morphological elements. Each morphological element encompasses design features numbering between 3 and 6, with the aim of ensuring both comprehensiveness and specificity. Furthermore, each element was assigned a unique code, as illustrated in Table 5.4.

Table 5.4: Deconstruction Coding of Styling Design Elements.

Design Element	Type					
Base x1	1 Rectangular	2 Rod	3 Wheels	4 Trapezoidal Angle	5 Pipe-like	6
Support x2	Square	Trapezoidal	Leg	Custom Irregular Shape	Exposed Structure	One Side Suspended
Bed Surface x3	Rectangular	Two Wide, One Narrow	Circular Rectangular	One Side Narrow		
Peripheral Handles x4	None	Expandable Single-Side Handle	Expandable Dual-Side Handles			
Control Panel x5	None	Remote Control	Integrated into Bed Body	Individually Set Console		
Bed Surface Movement Structure x6	Fixed	Sectional Front-Back Extension	Sectional Vertical Movement	Sectional Vertical Tilt	Overall Vertical Tilt	
Bed Surface Material x7	Fabric	Leather	Plastic			
Color x8	Cool Tones	Warm Tones	Black and White Neutral Colors			
Fixing Device x9	None	Additional Structural Fixation	Straps Fixation			

Table 5.5: Deconstructed Design Elements

Sample	Design Element									Sensory Imagination Evaluation Average			
	x1	x2	x3	x4	x5	x6	x7	x8	x9	Safe	Humanized	Friendly	Soft
1	2	4	1	1	2	1	1	1	1	1.82	2.64	2.10	3.12
2	2	3	1	1	1	1	2	2	1	1.54	2.12	1.34	1.56
3	5	2	1	3	2	2	2	3	1	2.02	2.32	2.86	2.48
4	5	3	4	1	1	1	2	3	2	2.74	2.34	1.86	1.42
5	3	5	4	1	3	4	2	3	2	1.98	3.02	1.32	2.22
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
49	2	3	1	3	3	1	1	3	2	2.10	1.84	2.92	2.78
50	1	5	4	1	2	4	2	3	2	1.62	1.58	1.886	1.70

5.3. Evaluation of Sensory Imagination for Physiotherapy Beds. Based on the research questionnaire statistics and the summarized sensory evaluation average values, as well as the deconstruction of the design elements of physiotherapy beds, a sensory evaluation matrix consisting of fifty samples and "nine sensory imageries" was established. The conformity was measured using a 1-5 Likert scale, where a score of 1 indicates non-conformity and a score of 5 indicates complete conformity. After conducting questionnaire surveys with 34 design professionals and 16 non-design professionals, the evaluations were averaged to study the preference of the survey subjects towards the samples. The partial data from the table can be found in Table 5.5.

Based on the aforementioned analysis, the BP neural network is constructed using the MATLAB 2022 software. Given the network's characteristic of multiple inputs and multiple outputs, an additional hidden

Table 5.6: Model Accuracy under Different Hidden Layers.

Hidden Layer Neurons	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Hidden Layer Neurons	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)
5	0.23097	0.10847	0.32935	10	0.16461	0.048648	0.22056
6	0.18153	0.055638	0.23588	11	0.11175	0.019345	0.13909
7	0.18106	0.045277	0.21278	12	0.31526	0.19502	0.44161
8	0.21159	0.061676	0.24835	13	0.21998	0.05842	0.2417
9	0.15716	0.04218	0.20538	14	0.2025	0.06482	0.2546

layer is introduced to determine the optimal number of neurons. Since the network entails 9 design elements, the input layer comprises 9 nodes. Correspondingly, there are 4 sensory imaginations being evaluated, leading to the output layer containing 4 nodes. Computed according to the empirical formula in formula 6 , with M = 9 and N = 4, the value of $L < 5 \sim 15$ is obtained.

For the purpose of accurate performance assessment, the MAE function is utilized. This function is less susceptible to the influence of outliers, allowing for a more precise depiction of data distribution. The specific formula for MAE is as follows:

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|y_i - f\left(x_i\right)\right|$$

y_i-denotes the actual value for the i-th sample. $f(x_i)$-signifies the predicted value by the neural network for the i-th sample. Utilizing the MATLAB software, various parameters were trained, and the corresponding error results for each parameter were presented in Table 5.6.

From Table 5.6, it is evident that when the number of neurons in the hidden layer is 11, both the training model's Mean Squared Error (MSE) and the validation model's MSE are minimized. Therefore, for this neural network architecture, the optimal number of neurons in the hidden layer is chosen to be 11. The BP neural network structure is set as 9*11*11*4, with a learning rate of 0.1 and an expected error of 0.00001. Following the specified criteria, the optimal network is trained and stored as the final quantitative relationship model for the design. The model is established using the purelin transfer function and trained using the gradient descent method.

The correlation coefficient between the training results and the actual evaluation data is R=0.96359, as shown in Figure 5.3. This indicates a high level of prediction accuracy for the model, demonstrating a strong fitting effect on the training dataset.

Model Practical Testing. Once the neural network is established, it can serve as a reference for the design of the therapeutic bed's visual imagery, providing design directions for designers. However, it is necessary to demonstrate its functional rationality. Therefore, different design proposals will be subjected to performance testing using the established BP neural network. To perform cross-validation, five industrial designers will each provide a set of therapeutic bed design proposals. These proposals will be adjusted and modeled in the Rhinoceros software to incorporate the model's design factors, resulting in five sets of therapeutic bed designs. Material and color design will be applied, and the models will be imported into Key shot for rendering. The specific proposals can be seen in Figure 5.4.

These five proposals will undergo expert scoring to obtain actual evaluation values. Subsequently, the five samples to be verified will be encoded according to Table 5.2 and used as input to the constructed BP neural network. The obtained indicator values from the output layer will be compared with the actual evaluation values to determine the relative error in predicted evaluation values.

Based on the data presented in Table 5.7, it is evident that the results generated by the neural network closely align with the evaluations conducted by experts, with a maximum deviation of only 8.5% and a minimum deviation of just 0.8%. This finding demonstrates that the trained neural network for design imagery

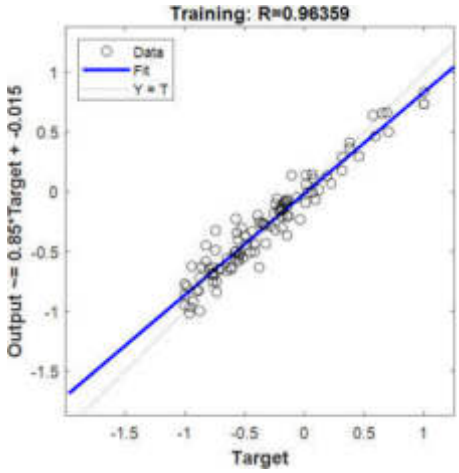


Fig. 5.3: Fitting Goodness of the Training Set.

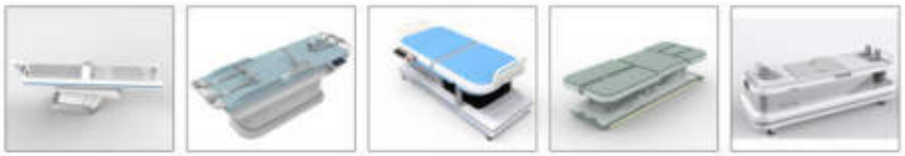


Fig. 5.4: Neural Network Training and Testing Scheme.

Table 5.7: Verification and Comparison Results.

Sensory Vocabulary	Data Error	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
	Predicted Sensory Value	2.1	3.60	1.96	1.64	1.84
Safe	Actual Sensory Mean	2.28	3.80	2.02	1.60	1.82
	Relative Error (%)	8.5%	5.6%	3.1%	2.4%	1.1%
	Predicted Sensory Value	2.1	2.84	1.82	2.19	1.66
Humanized	Actual Sensory Mean	2.06	2.62	1.80	2.10	1.62
	Relative Error (%)	1.9%	7.7%	1.1%	4.1%	2.4%
	Predicted Sensory Value	2.9	3.48	2.42	2.26	2.32
Friendly	Actual Sensory Mean	2.84	3.38	2.36	2.12	2.28
	Relative Error (%)	2.1%	2.9%	2.5%	6.2%	1.7%
	Predicted Sensory Value	3.61	2.67	2.18	2.72	2.32
Soft	Actual Sensory Mean	3.64	2.60	2.00	2.68	2.14
	Relative Error (%)	0.8%	2.6%	8.3%	1.5%	7.7%

can effectively capture the sensory characteristics of the therapeutic bed. Consequently, this network can be employed to comprehensively assess the visual design of therapeutic beds. By employing the neural network model, designers are empowered to move beyond relying solely on experience and sketches to infer the overall appearance of therapeutic beds. This design approach is notably more precise and scientifically grounded.

6. Conclusions. By applying the BP neural network model, we were able to establish a more scientifically grounded connection between the overall visual design elements of the physiotherapy bed and users' sensory imaginations. Additionally, we could enhance the precision of understanding users' emotions, enabling design-

ers to better employ design language to convey the product's meaning. The introduction of mathematical evaluation models helps overcome the limitations of solely relying on designers' subjective experiences in the original design process. This significantly reduces design risks and facilitates more informed decisions for visual design. However, the design of a physiotherapy bed, being a versatile medical rehabilitation equipment, also involves various factors such as specific functional requirements, internal operational structures, material choices, and human-computer interaction. These factors could serve as inputs for the BP neural network or other mathematical models, establishing logical connections between them and users' sensory demands. Furthermore, different neural network architectures or training functions can be utilized to obtain diverse sensory perception evaluation averages, thus balancing the diverse needs of users. With the extensive application of machine deep learning techniques in product sensory image design, more comprehensive and powerful mathematical models will likely facilitate further development in research on design methods based on sensory ergonomics.

REFERENCES

[1] Wang W Y, Zhou H, Wang Y F, (2021). et al. Current policies and measures on the development of traditional Chinese medicine in China. Pharmacological research, 16(3), 105-187.
[2] Feng Yixiong, (2020). Lou Shanhe, Wang Xupeng, Zheng Hao, Gao Yicong, Wang Yun, Tan Jianrong. Research on Performance-oriented Perceptual Image Evaluation Method for Cus Tomized Products. Journal of Mechanical Engineering, 56(9), 181-190.
[3] Lévy P D. (2013). Beyond kansei engineering: The emancipation of kansei design. International journal of design, 7(2), 83-94.
[4] Zhao Yanan,(2021).Study on the Construction of Perceptual Image Prediction Model for Office Chairs Based on BP Neural Network, Furniture & Interior Design, 271(09),118-122.
[5] DING Lu, ZHU Yan. (2020). Application of BP Neural Network in Perceptual Form Design of Programmable Paper Cutting Machine. PACKAGING ENGINEERING, 41(16),135-140.
[6] Ma S, Yan X. (2022). Application of BP Neural Networks in Garment Pattern Design System. Computational Intelligence and Neuroscience,22(6),202-223.
[7] Chen D, (2022). Cheng P. Development of design system for product pattern design based on Kansei engineering and BP neural network. International Journal of Clothing Science and Technology, 34(3), 335-346.
[8] Dong Y, Zhu R, Peng W, (2021). et al. A fuzzy mapping method for Kansei needs interpretation considering the individual Kansei variance. Research in Engineering Design, 32(2), 175-187.
[9] ZHOU Aimin,(2018).Product Multi-Image Prediction Model Based on Aesthetic Measure of Form .JOURNAL OF GRAPHICS, 39(4), 654-660.
[10] Cleophas T J, (2010). Cleophas T F. Artificial intelligence for diagnostic purposes: principles, procedures and limitations. Clinical chemistry and laboratory medicine, 48(2), 159-165.
[11] Wang Y, Zhao Q, Chen J, (2022). et al. Color design decisions for ceramic products based on quantification of perceptual characteristics. Sensors, 22(14), 5415.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 11, 2024
Accepted: Jun 20, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
 Volume 26, Issues 2, pp. 814–820, DOI 10.12694/scpe.v26i2.3931

THE INTELLIGENT COMPUTING AND INFORMATION TECHNOLOGY IN SPORTS PERFORMANCE EVALUATION

YUANYUAN ZHANG*, HUAN LONG†, AND LEI JING‡

Abstract. This paper presents a new method to obtain the training trajectory by using the mean shift method. In this way, the incomplete motion trajectory caused by the rapid movement of the moving target due to the complex background is solved. The human body modeling is regarded as a skeletal model with 51 degrees of freedom and 16 joints, and the motion trajectory is digitally processed. At the same time, the dimension compression of the trajectory is also carried out to reduce the calculation amount. The gradient iteration method based on random distribution is selected to reduce the dependence on environmental parameters. The object color image is the main feature to realize the acquisition of motion trajectory. The experiment proves that the algorithm can reflect the movement state of each part of the athlete well. This method can accurately obtain the training trajectory without any associated parameters.

Key words: Digital learning; Data dimensionality reduction; Average shift algorithm; Gradient iteration method

1. Introduction. MPEG is the expert group on moving images [1]. The algorithm has the advantages of less redundancy and better overall stability, while H.263 algorithm has higher efficiency. In addition, in order to ensure the complete coding of the tested object, the tested object must be separated during video compression [2]. However, the existing compression methods cannot effectively retrieve the video, so there are some difficulties in the effective segmentation of the image. MPEG-4, introduced in 2000, added the ability to retrieve multiple semantic objects based on context and foreground [3]. It can effectively improve the compression efficiency. However, it has poor denoising ability during compression [4]. But a lot of experiments prove that these algorithms cannot accurately describe the pose of the subject and the segmentation effect is not obvious. In this project, the fuzzy clustering method is used to deeply study the motion attitude of the target object, so as to realize the effective use and analysis of the motion video.

2. Motion video fuzzy clustering image segmentation. In order to solve the problems of random posture, unclear change law and difficult to judge the range of motion video, a method of motion video detection and segmentation is proposed. It can distinguish the moving position area and the invalid pixel in the image into the front and back parts. In the process of motion video recording, the method based on fuzzy clustering is used to segment motion video recording [5]. The motion of the object is predicted and compensated by image analysis. Construct multiple images with small intervals into the same background [6]. Then, the boundary extraction technique is used to segment the object's boundary to obtain the background image. Eliminate the foreground area to get the background area. The block diagram of image segmentation is shown in Figure 2.1.

When selecting fuzzy objects, people must first consider what features can accurately depict the moving posture; The second is that the motion posture in the image and the image between the invalid image cannot be completely consistent, and there should be a big difference between the two. After the motion is predicted and compensated, the moving objects in the moving video can be represented by small spaced backgrounds with similar backgrounds [7]. The feature of grayscale can be used when selecting the fuzzy feature. In addition, in view of the problem that a single fuzzy feature is not conducive to segmentation results, this project intends to transform the division of moving position region and invalid pixel region into the processing of non-normal feature distribution pixels [8].

*North China Institute of Science and Technology, Langfang 065201, China (Corresponding author, zyyncist2024@163.com)
†North China Institute of Science and Technology, Langfang 065201, China
‡North China Institute of Science and Technology, Langfang 065201, China

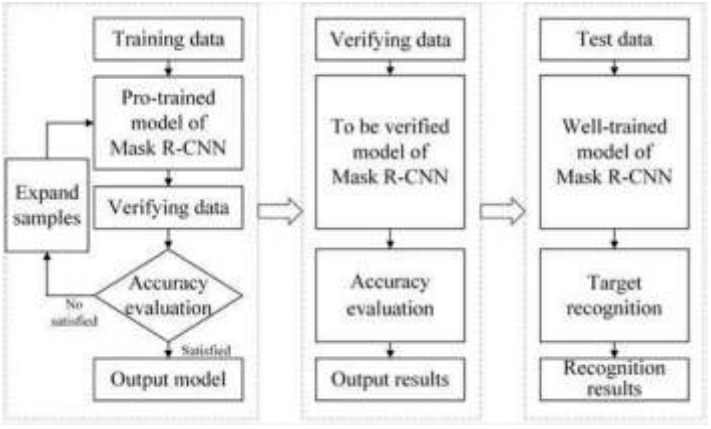


Fig. 2.1: Block diagram of sports image segmentation.

3. Use Mean Shift algorithm to train tracking data capture. Aiming at the fast moving and complex background of objects captured in sports, an average displacement method is proposed to track the trajectory of objects. Measurement is a non-parametric kernel density estimation method based on gradient iteration [9]. It is independent of the associated parameters and has a high convergence rate. The following illustrates the core density estimation of the algorithm: point u in the set is as follows:

$$\hat{g}(u) = \frac{1}{nl^d} \sum_{i=1}^{n} H\left(\frac{u - U_i}{l}\right) \tag{3.1}$$

l represents the window radius, H represents the kernel function, and then the Epanechnikov kernel is used to minimize the mean deviation between the estimated and the true density. Its formula is:

$$H_{P(u)} = \begin{cases} \frac{1}{2} z_d^{-1}(d+2)\left(1 - \|u\|^2\right) & \text{if } \|u\| < 1 \\ 0 & \text{else} \end{cases} \tag{3.2}$$

z_d is the volume of the ball. In general, H is a function of $\|u\|^2$, and $H(u) = h\left(\|u\|^2\right), h : [0, \infty) \to S$ is a shape function of the H kernel. Epanechnikov's shape function can be expressed as:

$$H_{P(u)} = \begin{cases} \frac{1}{2} z_d^{-1}(d+2)(1-u) & \text{if } u < 1 \\ 0 & \text{else} \end{cases} \tag{3.3}$$

From the above formula, it can be seen that h represents inconstant, nonnegative, piecewise continuity, and conforms to $\int_0^\infty h(u)du < \infty$. By replacing the distribution function of formula (3.2) in formula (3.3), we get:

$$\hat{g}_{H(u)} = \frac{1}{nl^d} \sum_{i=1}^{n} H\left(\left\|\frac{u - U_i}{l}\right\|^2\right) \tag{3.4}$$

l^d stands for ball height. If the core function $h(u)$ is derivable in the range $u \in [0, \infty)$, then it is assumed to be $y(u) = -h'(u)$, and at this time $y(u)$ is considered to be typical of the core function, and this core function can also be expressed as $Y(u) = Zy\left(\|u\|^2\right)$, where Z is a normalized constant, if the estimated gradient value is used as the baseline, the estimated probability density gradient can be obtained:

$$\hat{\nabla}_u g_{H(u)} \nabla_u \hat{g}_{H(u)} = \frac{1}{nl^d} \sum_{i=1}^{n} \nabla_u H\left(\frac{u - U_i}{l}\right) \tag{3.5}$$

For $u_1, u_2, \ldots, u_n, \nabla_u$ is a gradient factor. The offset mean vector of the core function Y can be obtained as follows:

$$M_{l,Y}(U) = \left[\frac{\sum_{i=1}^{n} u_i y\left(\left\|\frac{u-U_i}{l}\right\|^2\right)}{\sum_{i=1}^{n} y\left(\left\|\frac{u-U_i}{l}\right\|^2\right)} - u \right] \tag{3.6}$$

The density estimates of u can be obtained from Y :

$$\hat{g}_{Y(u)} = \frac{Z}{nl^d} \sum_{i=1}^{n} H\left(\left\|\frac{u-U_i}{l}\right\|^2\right) \tag{3.7}$$

M is a constant, and (3.4) can be reduced to:

$$\hat{\nabla}_u g_{H(u)} = \hat{g}_{Y(u)} \frac{2/Z}{l^2} M_{l,Y(u)} \tag{3.8}$$

It follows that:

$$M_{hY(u)} = \frac{l^2}{2/Z} \times \frac{\hat{\nabla}_u g_{H(u)}}{\hat{g}Y(u)} \tag{3.9}$$

From formula (3.8), it can be seen that the average-displacement vector obtained from Y coincides with the concentration gradient obtained from H. Therefore, a series of changes at the center point of the core function is taken as $y_j, j = 1, 2, \ldots, ,$ and then:

$$y_{j+1} = \frac{\sum_{i=1}^{n} u_i y\left(\left\|\frac{y_j-u_i}{l}\right\|^2\right)}{\sum_{i=1}^{n} y\left(\left\|\frac{y_j-u_i}{l}\right\|^2\right)} \tag{3.10}$$

y_{i+1} is the weighted average of y_j points calculated using the value of the kernel function, and y_1 is the original positioning [10]. Then the motion characteristics extracted by the average method are used to collect the motion trajectory. The color information remains unchanged when the object is translated, rotated, or deformed. This is one of the most believable image features.

Suppose $U = \{u_1, \ldots, u_N\}$ contains N sampling points and $p(u) = N(u, \delta, N)$ represents the probability distribution, where δ represents the mean vector and Q represents the covariance matrix. If the training object is elliptical, the selected object in this frame is initialized first. u_i is used to represent the specific position of the pixel owned by the entire player, δ_0 is the starting point, and T is the move cycle [11]. The general shape of this goal can then be expressed as:

$$Q_0 = \sum \left(u_i - \delta_0\right)\left(u_i - \delta_0\right)^T \tag{3.11}$$

An image segmentation method based on color histogram is proposed, which divides the color image into M subregion and uses $b\left(u_i\right) : S^2 \to 1, \ldots, M$ function to determine the color degree of each region u_i of the image [12]. The m subinterval is resolved as follows:

$$w_m = \sum_{i=1}^{N_{Q_0}} N\left(u_i; \delta; Q_0\right) \varphi\left[b\left(u_i\right) - m\right] \tag{3.12}$$

Where φ represents the Kronecker function and N represents the Gaussian kernel, the central part of the object is weighted to the maximum, and then N_{Q0} pixels close to the core are used as sub-intervals. Elements other than 2.5σ are not considered during processing to reduce the operational burden [13]. If the object to be captured appears in an image, then the capture finds the captured object in a new image, and then determines

Table 4.1: Statistics of SA values.

Segmentation method	Tier 1	Tier 2
Fuzzy clustering algorithm	0.0776	0.2200
MPEG-4	0.1285	0.2993
Frequency domain segmentation	0.1384	0.4414
Time domain segmentation	0.2728	0.3463

the position of the object in the new sequence by calculating the similarity between the real object and the object to be captured. In this way, the color rendering $s(\delta, Q)$ of the captured object can be determined by the color histogram, then the following values of m color parts can be obtained:

$$s_m(\delta, Q) = \sum_{i=1}^{N_v} N\left(u_i; \delta; Q\right) \varphi\left[b\left(u_i\right) - m\right] \tag{3.13}$$

The Bhattacharyya coefficient is used to judge the following approximation of the two-color subspaces:

$$\zeta[s(\delta, Q), w] = \sum_{m=1}^{M} \sqrt{s_m(\delta, Q)} \sqrt{w_m} \tag{3.14}$$

A Taylor extension of the estimate $s\left(\delta^{(h)}, Q^{(h)}\right)$ is as follows:

$$\zeta[s(\delta, Q), w] \approx z_1 + z_2 \sum_{m=1}^{N_s} \mu_1 N\left(u_i; \delta, Q\right) \tag{3.15}$$

z_1, z_2 is the Lagrange daily number and μ_i is the Bhattacharyya coefficient, which is calculated as follows:

$$\mu_i = \sum_{m=1}^{M} \sqrt{\frac{w_m}{s_m\left(\delta^{(h)}, Q^{(h)}\right)}} \varphi\left[b\left(u_i\right) - m\right] \tag{3.16}$$

In order to achieve the closest similarity between the captured tracking data and the real object, the maximum value in the formula (3.16) must be used.

4. Experimental results and analysis.

4.1. Spatial accuracy. This project takes the 2022 UCLA game video as the research object. Two continuously transformed moving images were selected for morphological filtering and correction respectively [14]. The spatial accuracy of each image was evaluated using the following formula:

$$SA = \frac{\sum_{(u,v)} \phi^{ext}(u, v) \oplus \phi^{ref}(u, v)}{\sum_{(u,v)} \phi^{ref}(u, v)} \tag{4.1}$$

(u, v) represents the pixel, ϕ^{ext} represents the partition range of the image; ϕ^{ref} is the partition of the completed moving object; The operation symbol $\oplus$ represents the additivity of logic. Generally speaking, the segmentation of images with fewer logarithms has higher spatial accuracy [15]. The SA values obtained by various algorithms are shown in Table 4.1. The results show that the proposed fuzzy clustering algorithm has the least SA. It can fill the blank of sports video efficiently. It can accurately segment images with large effects such as background and color. It has high spatial and temporal accuracy.

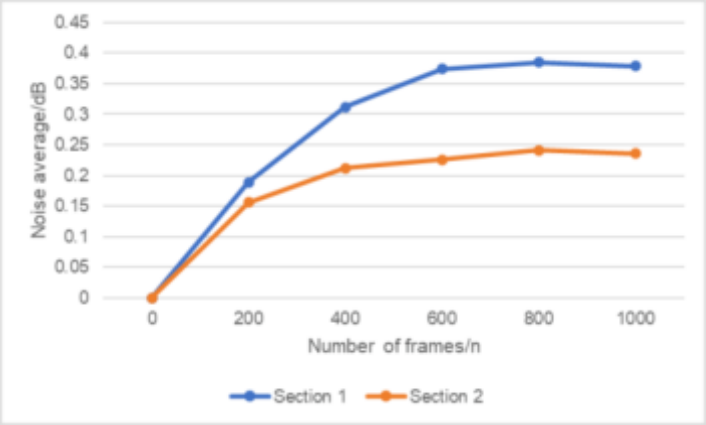


Fig. 4.1: Average image noise curve.

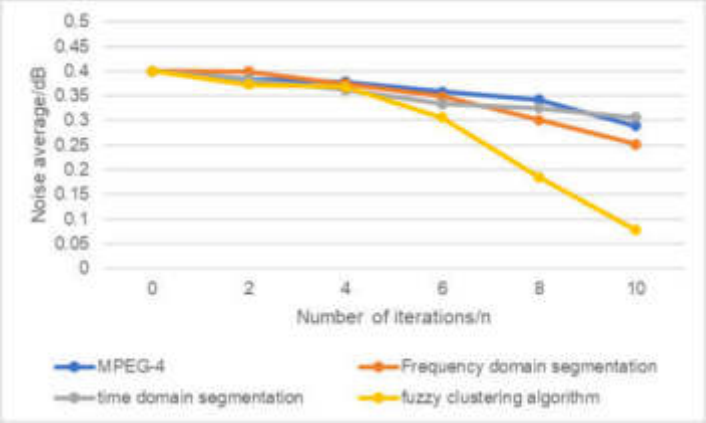


Fig. 4.2: The relationship between the number of iterations in paragraph 1 and the mean noise value.

4.2. Noise iteration performance. For the iterative ability of image segmentation algorithm against noise, the most fundamental requirement is that it can quickly achieve the purpose of self-adaptation. Figure 4.1 shows the average noise curve of two successively changing dynamic images in motion video recording. Fig. 4.2 and Fig. 4.3 shows the changes in the number of iterations for the average noise of various partitioning modes [16]. The overall increase in noise level off. After using this algorithm, the noise in the image is reduced to some extent [17]. The moving image segmentation algorithm based on fuzzy clustering is adopted. Repeat 6 times to reduce the noise of the image to 50%. The algorithm has good noise iteration.

4.3. Spatial distortion rate. Fig. 4.4 and 4.5 show the changes of spatial distortion rates of two dynamic images that change successively when they are processed by various division methods [18]. The spatial distortion rate is determined in the following manner:

$$s_n = \frac{\sum_{(u,v)} P^{ext}(u,v) \oplus P^{ref}(u,v)}{\sum_{(u,v)} P^{ext}(u,v)} \times 100\% \tag{4.2}$$

s_n represents the spatial distortion rate of the NTH frame; Where P^{ext} , P^{ref} represents the value of the divided pixel (u, v), and the value of the complete pixel [19]. The experimental results show that this method can effectively eliminate the| invalid pixels in the moving video, effectively suppress the noise in the moving

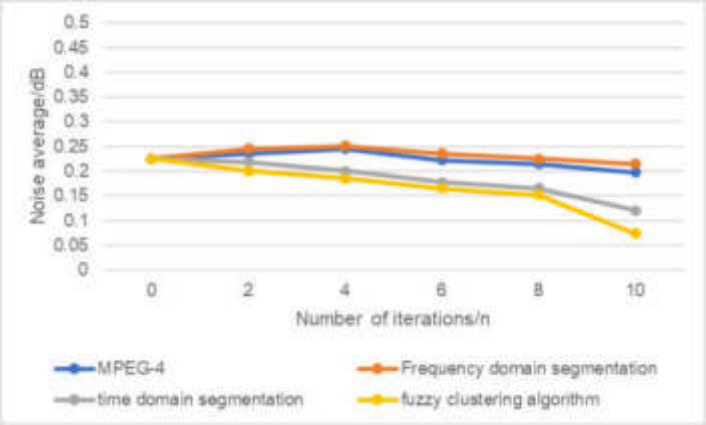


Fig. 4.3: Relation curve between the number of iterations in paragraph 2 and the mean noise value.

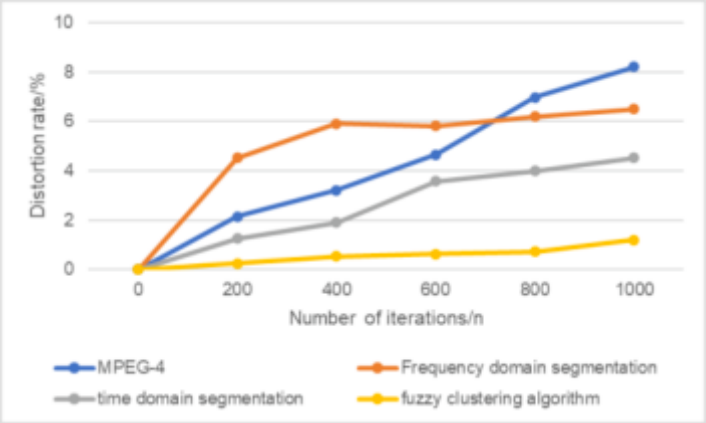


Fig. 4.4: Variation of the spatial distortion rate in paragraph 1.

video and obtain the obvious dynamic change law [20]. However, the other three kinds of algorithms have large spatial distortion, unstable change law, unable to effectively suppress noise points, and difficult to establish background. This leads to the segmentation of the image is not clear and other problems.

5. Conclusion. A moving image processing method based on fuzzy clustering is proposed. Compared with MPEG 4, frequency domain partition and time domain segmentation, the algorithm proposed in this project has higher application value in terms of spatial accuracy, noise iteration ability and spatial distortion rate. Physical education can promote national fitness and promote the development of competitive sports.

REFERENCES

[1] Rangasamy, K., As'ari, M. A., Rahmad, N. A., Ghazali, N. F., & Ismail, S. (2020). Deep learning in sport video analysis: a review. TELKOMNIKA (Telecommunication Computing Electronics and Control), 18(4), 1926-1933.
[2] Liu, Y., & Ji, Y. (2021). Target recognition of sport athletes based on deep learning and convolutional neural network. Journal of Intelligent & Fuzzy Systems, 40(2), 2253-2263.
[3] Zhao, Z., Liu, X., & She, X. (2021). Artificial intelligence based tracking model for functional sports training goals in competitive sports. Journal of Intelligent & Fuzzy Systems, 40(2), 3347-3359.
[4] Deng, Q., Zhang, Z., & Zhong, J. (2020). Image-free real-time 3-D tracking of a fast-moving object using dual-pixel detection. Optics Letters, 45(17), 4734-4737.

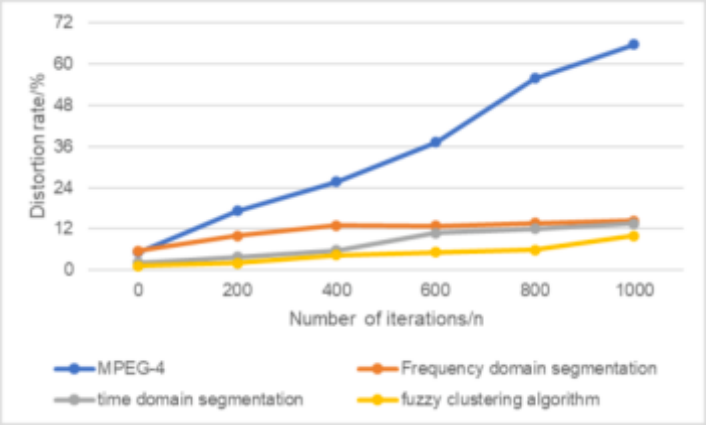


Fig. 4.5: Variation of spatial distortion rate in paragraph 2.

[5] Lu, Y., & An, S. (2020). Research on sports video detection technology motion 3D reconstruction based on hidden Markov model. Cluster Computing, 23(3), 1899-1909.
[6] Yan, L., Cengiz, K., & Sharma, A. (2021). An improved image processing algorithm for automatic defect inspection in TFT-LCD TCON. Nonlinear Engineering, 10(1), 293-303.
[7] Wu, S. (2021). Simulation of classroom student behavior recognition based on PSO-kNN algorithm and emotional image processing. Journal of Intelligent & Fuzzy Systems, 40(4), 7273-7283.
[8] Nakayama, Y., Lu, H., Li, Y., & Kamiya, T. (2020). WideSegNeXt: semantic image segmentation using wide residual network and NeXt dilated unit. IEEE Sensors Journal, 21(10), 11427-11434.
[9] Raabe, D., Nabben, R., & Memmert, D. (2023). Graph representations for the analysis of multi-agent spatiotemporal sports data. Applied Intelligence, 53(4), 3783-3803.
[10] Wang, T. (2022). Exploring intelligent image recognition technology of football robot using omnidirectional vision of internet of things. The Journal of Supercomputing, 78(8), 10501-10520.
[11] Zhou, G., Zhang, Z., Yin, W., Chen, H., Wang, L., Wang, D., & Ma, H. (2024). Surface defect detection of CFRP materials based on infrared thermography and Attention U-Net algorithm. Nondestructive Testing and Evaluation, 39(2), 238-257.
[12] Pasupa, K., Kittiworapanya, P., Hongngern, N., & Woraratpanya, K. (2022). Evaluation of deep learning algorithms for semantic segmentation of car parts. Complex & Intelligent Systems, 8(5), 3613-3625.
[13] Chen, C., Seo, H., Jun, C. H., & Zhao, Y. (2022). Pavement crack detection and classification based on fusion feature of LBP and PCA with SVM. International Journal of Pavement Engineering, 23(9), 3274-3283.
[14] Sun, Z., Xuan, P., Song, Z., Li, H., & Jia, R. (2022). A texture fused superpixel algorithm for coal mine waste rock image segmentation. International Journal of Coal Preparation and Utilization, 42(4), 1222-1233.
[15] Husain, A. A., Maity, T., & Yadav, R. K. (2020). Vehicle detection in intelligent transport system under a hazy environment: a survey. IET Image Processing, 14(1), 1-10.
[16] Liang, C. M., Li, Y. W., Liu, Y. H., Wen, P. F., & Yang, H. (2022). Segmentation and weight prediction of grape ear based on SFNet-ResNet18. Systems Science & Control Engineering, 10(1), 722-732.
[17] Wang, H., Minnema, J., Batenburg, K. J., Forouzanfar, T., Hu, F. J., & Wu, G. (2021). Multiclass CBCT image segmentation for orthodontics with deep learning. Journal of dental research, 100(9), 943-949.
[18] Wang, Y., Lv, H., Deng, R., & Zhuang, S. (2020). A comprehensive survey of optical remote sensing image segmentation methods. Canadian Journal of Remote Sensing, 46(5), 501-531.
[19] Cai, W., Song, Y., Duan, H., Xia, Z., & Wei, Z. (2022). Multi-feature fusion-guided multiscale bidirectional attention networks for logistics pallet segmentation. Computer Modeling in Engineering and Sciences, 131(3), 1539-1555.
[20] He, S., Wang, R., & Luo, H. (2022). Computing metasurfaces for all-optical image processing: a brief review. Nanophotonics, 11(6), 1083-1108.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 14, 2024
Accepted: Jun 17, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 821–828, DOI 10.12694/scpe.v26i2.4010

THE EDGE COMPUTING EXTENSIVE DATA PROCESSING FRAMEWORK AND ALGORITHM FOR THE INTERNET OF THINGS

LUYAO GE*

Abstract. The paper studies a monitoring pedestrian recognition method driven by big data for edge-cloud collaboration. It extends the original centralized computing to edge and cloud collaborative processing. Firstly, the image boundary node N0 is preprocessed, and the extracted image is expressed in multiple levels. Then, the RGB-D multimodal image learning modeling method is applied to the edge nodes of the network using cloud computing. The boundary node uses the existing learning mode to perform action identification and uploads the identified action information to the cloud to form the final action classification. The method of bone surface fitting and dense trajectories is combined to achieve robust, dense human posture feature extraction. The directed principal component histograms of 3D stereo structures are obtained using dense point cloud data. The features of the spatial and temporal neighborhood 3D gradient histogram are extracted from the apparent texture. Through experiments, it is verified that the proposed method can significantly reduce the shortcomings of traditional centralized algorithms in data transmission and cloud storage and improve the pattern recognition accuracy by 2.2% based on edge cloud collaboration.

Key words: Edge cloud collaboration; behavior recognition; multimodal characteristics; internet of things; edge computing

1. Introduction. A traffic video monitoring system is a security system based on computer, image recognition, network communication, information processing and other technologies, including central control equipment, station control equipment, image intake and display, video signal transmission and other equipment, which can provide visual information of monitoring area for dispatchers in the control center, station duty personnel, train drivers, etc. It plays a significant role in ensuring the safe operation of traffic and high-quality services [1]. With the continuous expansion of traffic lines and stations, monitoring scenarios and target types are becoming increasingly complex, with massive scale and wide distribution. This challenges the monitoring system's intelligent, networked, real-time and maintenance management. Road video surveillance systems have experienced analog to digital video surveillance systems, which are the current intelligent video surveillance systems.

The combination of motion feature extraction and video content analysis can effectively solve the current security problems and has important practical significance for constructing a "smart city." The current cloud computing model has been unable to thoroughly and efficiently handle large-scale monitoring video data, so we extend the cloud computing model to edge cloud collaboration to improve monitoring efficiency. Literature [2] researched network-oriented resource allocation and task scheduling in an edge-cloud-oriented collaborative computing environment, conducted integrated forecasting of potential tasks from horizontal and vertical spatiotemporal scales, and optimized their configuration. Literature [3] shows that edge-cloud collaboration has advantages such as bandwidth saving, delay reduction and privacy protection, which is significant for developing power grids. Literature [4] breaks through the traditional single-cloud model and studies the flexible, intelligent manufacturing method of end-edge-cloud collaboration to adapt to SMT production lines' flexible and intelligent production requirements. Literature [5] proposes an optimization algorithm that integrates hardware and software to minimize the weighted delay of each node, aiming at migrating multi-way collaborative tasks in cloud edge collaboration.

This project will study a monitoring pedestrian recognition method driven by big data for edge-cloud collaboration. Extend the original centralized computing to edge and cloud collaborative processing. Then, the RGB-D multimodal image learning modeling method is applied to the edge nodes of the network using cloud computing.

*Dongchang College, Liaocheng University, Liaocheng, Shandong, 252000, China (gly10082024@126.com)

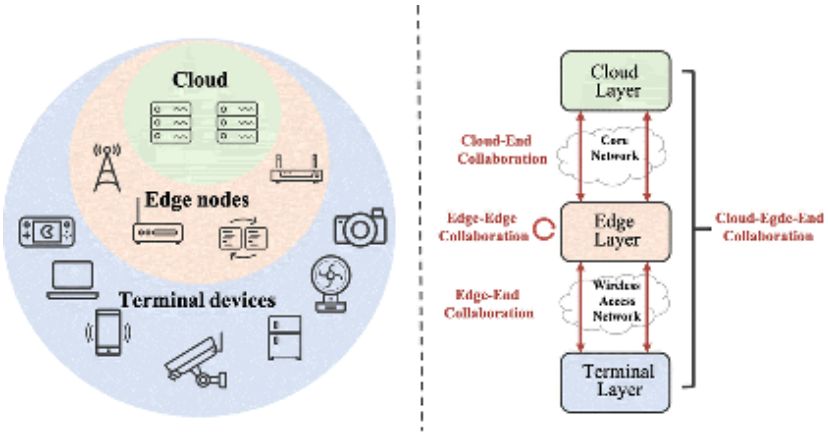


Fig. 2.1: Application architecture of cloud edge collaboration technology for traffic video surveillance system.

2. Surveillance video analysis architecture design based on cloud edge collaboration. This project takes the cloud platform of video surveillance as the research object and researches critical issues such as resource virtualization, video target recognition, surveillance scene classification, and monitoring big data processing of video surveillance, focusing on key technologies such as video image data preprocessing and surveillance scene pre-recognition based on edge computing [6]. By constructing an intelligent collaboration and collaboration framework, the video surveillance cloud and edge network can be fully coordinated to achieve the overall and global performance of video surveillance. The application architecture of cloud-edge collaboration technology for traffic video surveillance systems is shown in Figure 2.1.

The system consists of a data acquisition layer, edge node layer, network transmission layer, cloud computing center layer and service application layer. The line will install cameras in some crucial places according to the needs of the operation and public security departments, which are used to collect and perceive the video images of the original monitoring area, such as passenger flow density, personnel behavior, and on-off order. There are usually several control boxes with AI computing functions in the boundary layer. It includes intelligent hardware, software, and models. It is used for resource scheduling, target identification, scenario analysis, service scheduling, etc. The information transmission is completed in the network transmission layer, and the monitoring devices of the control center, station, yard and other regions are connected through the transmission network. A large amount of video monitoring data is uniformly stored and intelligently analyzed using edge network collaboration, and through collaborative control of the edge network, a full range of cooperation is carried out in resources, data, intelligent analysis, application services and other levels [7]. At the service application level, various road video surveillance system scenarios are intelligently identified, analyzed, and coordinated with other safety subsystems.

3. RGB-D video multimodal feature extraction.

3.1. Feature extraction of binary dense moving human posture. In this paper, the normalized skeletal surfaces are fitted based on the reference of feature-dense trajectories and MP characteristics. Then, it is integrated with moving dense trajectories to obtain dense postural features that can describe human movement [8]. First, the $U_i^{t_0}$ of each frame surface sampling is fitted in the case that the given tracking track length is the frame S. The motion trajectories of several lower layers are extracted to correlate feature P. Then, the average clustering algorithm is used to get the corresponding dictionary C_{MP}. Each path in P is sorted by Euclidean distance to obtain multiple semantic multiple features D consistent with the number of dictionary primitives and other dimensions. The basic tracing features of the authorized MP feature are as follows

$$P = (P_1, P_2, \cdots, P_{M_s}) \tag{3.1}$$

Among

$$\begin{aligned} &M_s = M_{MP}\left\lfloor \frac{M_F}{S} \right\rfloor, P_t = \left(U_i^t, U_i^{l+1}, \cdots, U_i^{i+S-1}\right) \\ &1 \le i \le M_{\sin p\,\rho} \end{aligned} \tag{3.2}$$

$M_{M\rho}$ The number of dictionary primitives extracted from the underlying features; M_F is the number of frames left after the extra time is removed. Where $M_{\sin p}$ is the fitting number of sample points on the surface. The descriptor $U_i^{\sigma_0}$ represents the sampling point of the 3D behavior action at time t_0. The combination of the three-dimensional posture $f_i(t_0)$ and the first and second derivative $\zeta f_i(t_0)$ and $\zeta^2 f_i(t_0)$ of the normalized human skeleton

$$U_i^{\sigma_0} = \left[f_i(t_0), v_j f_i(t_0), \varphi_s^{-2} f_i(t_0)\right] \tag{3.3}$$

v and φ are the associated weights of the velocity and acceleration components of the multipoint motion characteristic descriptor of Dense MP.

3.2. Feature extraction of directionally sparse principal component histogram. This project takes the pedestrian apparent surface as the research object. The apparent surface of the human body is characterized by the main direction of the divergence matrix based on a dense point cloud, and pedestrian apparent features are extracted similarly to plane regular [9]. Firstly, the dispersion of the point cloud in the spacetime neighborhood of the three-dimensional point cloud sequence and its corresponding feature vector are mapped to aspect M, thus obtaining the quantization histogram HOPC as follows:

$$y_t^{\rho *} = \left[x_1^T x_2^T x_3^T\right]^T \in R^{\beta \times M} \tag{3.4}$$

The space adjacent to the skeleton movement is segmented by the space-time pyramid method. The algorithm is integrated with sparse coding technology to obtain the HOPC signal dictionary C, the corresponding coefficient matrix, the HOPC feature of the weighted coefficient and the difference of the visual dictionary [10]. The n_f class adaptive space-time pyramid method is used to decompose the $n_H \times n_W \times 7$ type space-time basis elements and combine them with the weighted average and maximum values of the spatial axis weights of each basis element to produce corresponding feature vectors. Finally, several corresponding flight paths are joined to form SHOPC features (where D is the main direction component of the dispersion matrix). $n_H \times n_w$ is the number of grids in the range of interest and n_f is the number of time cone levels [11]. The weight of the optimal basis function and ELM's structural parameters are obtained using the maximum likelihood and least square method. Combine ELM with the weight of the base unit and add modular and non-negative constraints to the weight vector of the base unit:

$$\begin{aligned} &\min_i \min_{\varphi,\infty} \frac{1}{2}\|\varphi\|_G^2 + \frac{1}{2} z \sum_{i=1}^{M} \|\omega_i\|^2 \\ &\text{st. } \varphi^T \phi(x_i; \delta) = y_i - \omega_i, \forall i \\ &\sum_{f=1}^{M_K} \delta_f = 1\delta_f \ge 0, \forall f \end{aligned} \tag{3.5}$$

Among

$$\begin{aligned} &\varphi^T = [\varphi_1, \varphi_2, \cdots, \varphi_{M_K}] \in R^{\left(\hat{\phi}(-)+++\varphi_k, K(-)\right)_k)} \\ &\varphi_f \in R^{\phi_i(-)}, f = 1, 2, \cdots, M^K \end{aligned} \tag{3.6}$$

$\omega_i = [\omega_{1i}, \omega_{2i}, \cdots, \omega_{Ti}]^T, 1 \le i \le M$ is column i of ω. Where $Y = [y_1, y_2, \cdots, y_M] \in R^{U \times M}$ is the class matrix. In M training samples, according to formula (3.2), the kernel matrix corresponding to the multi-mode visual

Table 4.1: Hardware parameters of edge node and cloud server.

Platform	Hardware	Computing resource	Internal memory	Hard disk
Edge Node	PC	CPU: i5 8400 GPU: RTX2060	16GB	512GB
High in the clouds Server	Rack type Server	CPU: Xeno 4116 GPU: Tesla100	128GB	1TB

features is:

$$\mu(\cdot,;;r) = \mu\left(x_i, x_j; \delta\right) = \phi\left(x_i; \delta\right)^T \phi\left(x_j; \delta\right) = \sum_{f=1}^{M^K} \delta_f K_f\left(x_i, x_j\right) \tag{3.7}$$

The optimal kernel function of the multi-kernel ELM model is combined with the weight vector to obtain the optimal binding kernel function and the optimal binding kernel function in the training sample [12]. Iterate repeatedly to get the best construction parameters v^* and δ^* of the best ELM, and finally get the decision vector for the test sample C:

$$g_c(x) = \sum_{i=1}^{M} v_{ic}^* \sum_{f=1}^{m} \delta_f^* K_f\left(x_i, x\right) \tag{3.8}$$

$v^* \in R^{M \times U}$ is the best subset of Lagrange. Where δ^* is the best core weight vector. The subscript c is the classification mark for the test sample U.

4. Experimental analysis. Simulation experiments were carried out to test the effectiveness of the behavior identification mechanism under edge-cloud collaboration [13]. The test system uses 6 PCS and 1 frame server on the boundary and cloud sides, respectively. The problem of the limited computing power of boundary nodes is simulated using the limitations of VMWare virtual machines on computing and storage. Table 4.1 lists hardware data for border nodes and cloud computing servers.

4.1. Data Set.

4.1.1. TU-RGB+D120 data set. This database has the most extensive collection of bone samples in the world. The sample size was 114,480, and the behavior types were 120. The shooting angles are increased to 155, and 106 targets are tested [14]. The database provides two test criteria for cross-topics and cross-settings. The test criterion based on cross-topic will take 53 individuals as the research objects and take the training set and the test set as the research objects. Number the training set as even and mark the test set as odd in the cross-setting test criteria.

4.1.2. Kinetics Database. This database includes 400 categories of human behavior extracted from YouTube, with at least 300 video clips of each action of approximately 10 seconds. The range of activities is extensive, including the interaction between people and things and between people [15].

4.2. Test the advantages of edge cloud collaborative computing. This article compares using performance in an edge-cloud collaboration environment with the following approach [16]. The system's performance is mainly reflected in three dimensions: network transmission volume, total task time and device energy consumption. In the single boundary operation, the operation is carried out on the boundary without any network transmission. In a single cloud computing environment, when the amount of data increases, the data traffic increases with the increase of the data scale (figure 4.1) [17]. The extracted feature information must be uploaded to the cloud first, then the cloud computing learns and distributes it to the edge nodes, thus reducing the initial network overhead. In the following identification, only the identified data needs to be uploaded to the cloud to reduce the growth of network traffic [18]. With the increased number of tasks, the advantages of edge cloud collaboration are more prominent. Power consumption refers to the proportion of CPU, memory and

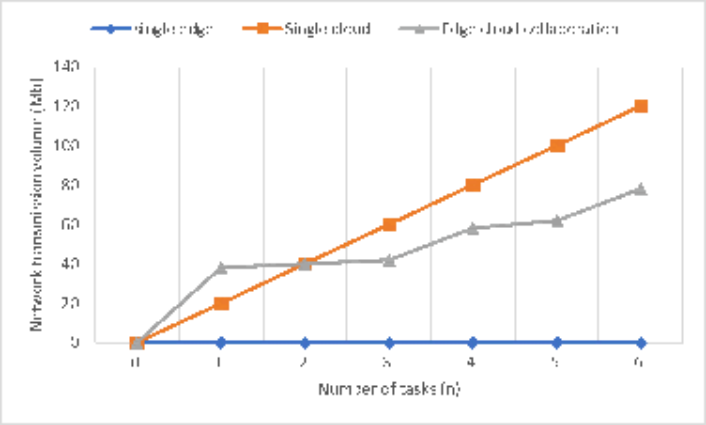


Fig. 4.1: Comparison of data traffic of the three mechanisms under the same number of jobs.

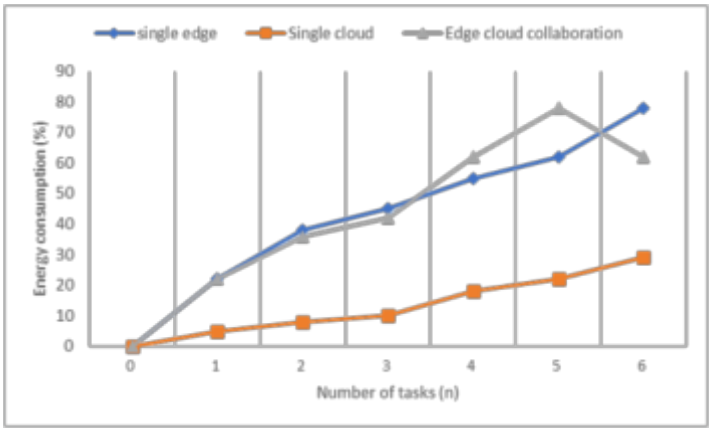


Fig. 4.2: Comparison of energy consumption of the three methods under the same number of operations.

hard disk each machine uses to operate [19]. Under the premise of ensuring the integrity of the identification work, the computing resources of cloud computing and the memory capacity of the edge have been maximized. Figure 4.2 shows the comparison of energy consumption of the three schemes under the same number of tasks.

The results of the comparative test of operation time in the three plans are shown in figure 4.3. Because of the network's limited storage space and CPU resources, single identification takes too long in the single boundary operation mode. And the more tasks there are, the more time it takes. The boundary end indicates that the task is terminated when the capacity limit is reached [20]. More time is needed to learn and deploy the cloud model before the first run under edge cloud collaborative computing. So, the time of first use is more than the time of a single use, but when the number of uses increases, the time of use will be less than the other two methods, which can prove the advantages of edge cloud collaboration.

This article selects a classic case with a resolution of 1280x720, a time length of 2 seconds, and a frame rate of 20 frames [21]. When the number of tasks increases, the more data the video uploads, the more likely it is to cause network congestion [22]. The method in this paper does not need to upload the video to the cloud, but only requires edge cloud collaboration in the range of 10×50x16b=8 kb. This dramatically reduces the network traffic for uploading videos. Edge cloud collaboration shows better performance in realizing human movements considering the impact of all aspects.

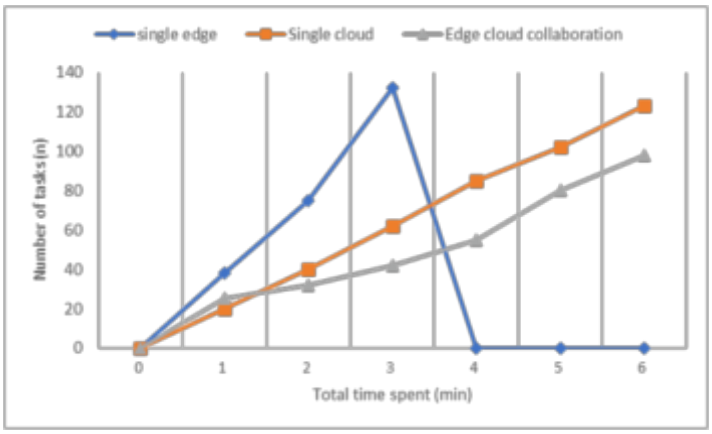


Fig. 4.3: Comparison of the total time of the three plans for the same number of jobs.

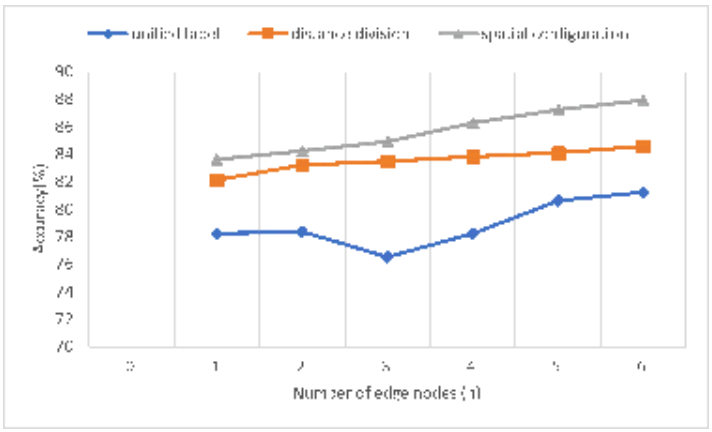


Fig. 4.4: Influence of the number of boundary nodes on identification accuracy.

4.3. Comparison of classification accuracy of RGB-D multi-mode images by three classification methods. The classification accuracy of RGB-D multi-mode images under three classification strategies is studied based on edge-cloud collaboration. The change in identification accuracy can be observed by adjusting the number of boundary nodes in the experiment [23]. The model is classified correctly after using a different number of nodes and division methods. Figure 4.4 shows the influence of the number of edge nodes on the recognition accuracy.

The number of boundary nodes gradually increases under the three partition methods, and the recognition accuracy is gradually improved until all boundary nodes are adopted. However, the segmentation method is based on a single market that lacks local feature differences when processing bone sequences. This leads to a decline in accuracy. Compared with region segmentation, the spatial layout partitioning method can obtain better recognition [24]. Therefore, this project proposes a multi-model modeling method based on RGB-D multimodal images to enhance its practical value in complex environments.

4.4. Cloud fusion effect verification. The identification effect of multiple boundary nodes in cloud computing is tested using two sets of different samples [25]. Firstly, at the edge node N_0, the similar frames in the image are eliminated, and the pose is estimated [26]. Secondly, the action identification is carried out at $N_1 \sim N_m$, and finally, the identification results are uploaded to the cloud computing for data processing. Tables

Table 4.2: Joint identification accuracy of edge nodes and cloud computing based on NTU RGB+D120.

	Cross Subject	Cross View
N1	84.58%	90.73%
N2	86.04%	91.67%
N3	84.06%	90.52%
N4	85.31%	91.04%
N5	85.94%	91.35%
Single cloud	85.52%	91.56%
Merge.	87.40%	93.44%

Table 4.3: Accuracy of boundary node and cloud fusion identification in Kinetics database.

	Top 1	Top 5
N1	86.88%	88.75%
N2	78.54%	89.90%
N3	85.52%	88.23%
N4	84.27%	90.42%
N5	85.10%	88.44%
Single cloud	86.04%	89.06%
Merge.	88.02%	91.88%

4.2 and 4.3 are performance tests conducted on the Kinetics platform in a mix of single-edge, single-cloud and cloud-based modes based on NTU- RGB+D120.

Under the cross-subject test criterion, the accuracy of a single cloud image reached 85.52%, and the accuracy of a single cloud image was 84.06% 86.04%. The mean value is 85.10%. In the cloud computing environment, the experimental effect reaches 87.40%, about 2.29% higher than the average recognition accuracy in a single cloud computing environment. The accuracy of the single cloud model based on cross-view is 91.56%, and the accuracy of the single cloud model is 90.52% to 91.67%. The average was 91.04%, while the cloud computing model reached 93.13%, an increase of 2.40%. It is found that motion recognition accuracy can be improved by 2% based on cloud computing.

The cloud server can perform identity authentication faster and improve accuracy in a single cloud computing environment. However, a large number of video and image uploads will lead to data congestion, which will also cause a significant burden on the performance of the cloud server. There may be some problems in data transmission, such as missing frames, which seriously reduces the system's performance. On the premise that the energy consumption of the edge cloud collaboration algorithm studied in this project is comparable to that of a single cloud algorithm, it not only reduces the data transmission overhead, reduces the cloud server's data storage burden but also effectively plays the computing performance of the edge node and improves the identification accuracy of the system.

5. Conclusion. Firstly, the collaborative computing architecture of edge cloud is studied to realize the effective configuration and efficient use of edge and cloud. Secondly, multi-mode image features based on RGB-D are constructed, and image processing is carried out in the cloud. Finally, the image information of each region is uniformly processed on the cloud platform. The experiment verifies that this project will improve cognitive accuracy and significantly improve the network transmission capacity, equipment energy consumption, total task time and other indicators.

6. Acknowledgements. The Project is supported by Research on Data Governance of Education Enrollment Examination (2023QY038). The project information is 2023 project of the "14th Five-Year Plan" of Education Science in Shandong Province.

REFERENCES

[1] Cui L, Su X, Ming Z, et al. CREAT: Blockchain-assisted compression algorithm of federated learning for content caching in edge computing. IEEE Internet of Things Journal, 2020, 9(16): 14151-14161.
[2] Wang X, Ning Z, Guo S. Multi-agent imitation learning for pervasive edge computing: A decentralized computation offloading algorithm. IEEE Transactions on Parallel and Distributed Systems, 2020, 32(2): 411-425.
[3] Lakhan A, Mohammed M A, Rashid A N, et al. Deadline aware and energy-efficient scheduling algorithm for fine-grained tasks in mobile edge computing. International Journal of Web and Grid Services, 2022, 18(2): 168-193.
[4] Lv Z, Chen D, Wang Q. Diversified technologies in internet of vehicles under intelligent edge computing. IEEE transactions on intelligent transportation systems, 2020, 22(4): 2048-2059.
[5] Tang M, Wong V W S. Deep reinforcement learning for task offloading in mobile edge computing systems. IEEE Transactions on Mobile Computing, 2020, 21(6): 1985-1997.
[6] Xiao H, Zhao J, Pei Q, et al. Vehicle selection and resource optimization for federated learning in vehicular edge computing. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(8): 11073-11087.
[7] Mills J, Hu J, Min G. Multi-task federated learning for personalised deep neural networks in edge computing. IEEE Transactions on Parallel and Distributed Systems, 2021, 33(3): 630-641.
[8] Zhang L, Ansari N. Latency-aware IoT service provisioning in UAV-aided mobile-edge computing networks. IEEE Internet of Things Journal, 2020, 7(10): 10573-10580.
[9] Xiong X, Zheng K, Lei L, et al. Resource allocation based on deep reinforcement learning in IoT edge computing. IEEE Journal on Selected Areas in Communications, 2020, 38(6): 1133-1146.
[10] Dai X, Xiao Z, Jiang H, et al. Task co-offloading for D2D-assisted mobile edge computing in industrial internet of things. IEEE Transactions on Industrial Informatics, 2022, 19(1): 480-490.
[11] Liu Y, Wang S, Zhao Q, et al. Dependency-aware task scheduling in vehicular edge computing. IEEE Internet of Things Journal, 2020, 7(6): 4961-4971.
[12] Ning Z, Dong P, Wang X, et al. Distributed and dynamic service placement in pervasive edge computing networks. IEEE Transactions on Parallel and Distributed Systems, 2020, 32(6): 1277-1292.
[13] Luo Q, Hu S, Li C, et al. Resource scheduling in edge computing: A survey. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2131-2165.
[14] Liu X, Yu J, Feng Z, et al. Multi-agent reinforcement learning for resource allocation in IoT networks with edge computing. China Communications, 2020, 17(9): 220-236.
[15] Yang L, Yao H, Wang J, et al. Multi-UAV-enabled load-balance mobile-edge computing for IoT networks. IEEE Internet of Things Journal, 2020, 7(8): 6898-6908.
[16] Fan G, Chen L, Yu H, et al. Multi-objective optimization of container-based microservice scheduling in edge computing. Computer Science and Information Systems, 2021, 18(1): 23-42.
[17] Zhan C, Hu H, Sui X, et al. Completion time and energy optimization in the UAV-enabled mobile-edge computing system. IEEE Internet of Things Journal, 2020, 7(8): 7808-7822.
[18] Yuan Q, Li J, Zhou H, et al. A joint service migration and mobility optimization approach for vehicular edge computing. IEEE Transactions on Vehicular Technology, 2020, 69(8): 9041-9052.
[19] Wang Y, Lang P, Tian D, et al. A game-based computation offloading method in vehicular multiaccess edge computing networks. IEEE Internet of Things Journal, 2020, 7(6): 4987-4996.
[20] Liu X, Yu J, Wang J, et al. Resource allocation with edge computing in IoT networks via machine learning. IEEE Internet of Things Journal, 2020, 7(4): 3415-3426.
[21] Zeng F, Chen Q, Meng L, et al. Volunteer assisted collaborative offloading and resource allocation in vehicular edge computing. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(6): 3247-3257.
[22] Zhang R, Cheng P, Chen Z, et al. Online learning enabled task offloading for vehicular edge computing. IEEE Wireless Communications Letters, 2020, 9(7): 928-932.
[23] Yuan H, Zhou M C. Profit-maximized collaborative computation offloading and resource allocation in distributed cloud and edge computing systems. IEEE Transactions on Automation Science and Engineering, 2020, 18(3): 1277-1287.
[24] Zhan Y, Guo S, Li P, et al. A deep reinforcement learning based offloading game in edge computing. IEEE Transactions on Computers, 2020, 69(6): 883-893.
[25] Wang L, Wang K, Pan C, et al. Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing. IEEE Transactions on Cognitive Communications and Networking, 2020, 7(1): 73-84.
[26] Bai T, Pan C, Deng Y, et al. Latency minimization for intelligent reflecting surface aided mobile edge computing. IEEE Journal on Selected Areas in Communications, 2020, 38(11): 2666-2682.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 22, 2024
Accepted: Jun 25, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 829–836, DOI 10.12694/scpe.v26i2.4020

THE CASTING PROCESS OPTIMIZATION AND MELT FLOW SIMULATION BASED ON COMPUTER SIMULATION AND DATA MINING

YI DING*, XINGANG SONG†, AND DONGMING ZHANG‡

Abstract. The flow state and velocity of magnesium alloy melt at different temperatures were studied by a contact test method. Using the SOLA-VOF finite difference method, the mathematical model of the three-dimensional flow field and temperature field in the flow field of magnesium alloy twin-roll casting-rolling molding stand plate was established. The liquid magnesium alloy melt takes the form of a convex arc and advances successively on the liquid metal surface. This is mainly due to the mass accumulation of the pyrolytic products in the liquid magnesium alloy melt. The filling time is shortened with the increase of pouring temperature and vacuum degree. In the filling stage, the filling rate showed periodic alternating changes.

Key words: Lost mold casting; Magnesium alloy; Melt flow; Flow velocity; Computer simulation

1. Introduction. Compared with conventional sand casting, evaporative mold filling involves many physical and chemical problems, such as heat transfer, filling flow, chemical reaction and cooling solidification. A specific coupling effect between the above factors leads to the liquid metal filling becoming more complicated. This has a significant impact on the quality of the casting products. In liquid metal filling, evaporation mold vaporization will consume most heat energy, resulting in a sharp drop in its front temperature [1]. The combined effect of the evaporation rate and the return of the pyrolysis product leads to a significant decrease in the flow performance of the melt. This is prone to problems such as cold isolation and insufficient pouring [2]. Achieving a uniform and continuous flow field structure is a critical way to reduce defects such as porosity in castings [3]. Therefore, the research on this problem has been widely concerned by people.

The melting points of magnesium and aluminum are very close, and the casting temperatures are very different [4]. Because the latent heat of crystallization of magnesium alloy is lower than that of aluminum alloy, and it is easy to oxidize, the casting temperature is difficult to be too high. This results in rapid solidification and transient liquid phase retention of the magnesium alloy, resulting in poor workability and increased shrinkage defects related to the plastic deformation of the alloy. In addition, since magnesium alloys have a lighter mass than aluminum alloys, the processing efficiency is also lower [5]. At a certain altitude, the static pressure head weakens the filling, while the negative pressure plays a more significant role. Therefore, completely copying the filling rules of aluminum alloy does not apply to magnesium alloy technology.

In addition to using photographic technology for some exploratory research, there are almost no other research results. The photographic method is a direct measurement method, but because it is impervious to water, its cooling performance differs significantly from loose sand [6]. This results in a significant difference between the measured and measured values. The electric contact method measured the filling process of magnesium alloy liquid in a lost mold [7]. The effect of vacuum degree and pouring temperature on the filling behavior of magnesium alloy melt was studied. This provides a theoretical basis for the further preparation of magnesium alloy melt.

2. Test method and device. The influence of pouring temperature, vacuum degree and gate number on the melt flow state was studied. The rheological properties of polymer foamed plastic mold were studied [8]. The pattern density is 0.020g/cm3. The shape and size of the flat sample and the internal gate are shown in

*College of Mechanical and Electronic Engineering, Huanghe Jiaotong University, Jiaozuo Henan 454950, China (Corresponding author, 15890129126@163.com)
†College of Mechanical and Electronic Engineering, Huanghe Jiaotong University, Jiaozuo Henan 454950, China
‡College of Mechanical and Electronic Engineering, Huanghe Jiaotong University, Jiaozuo Henan 454950, China

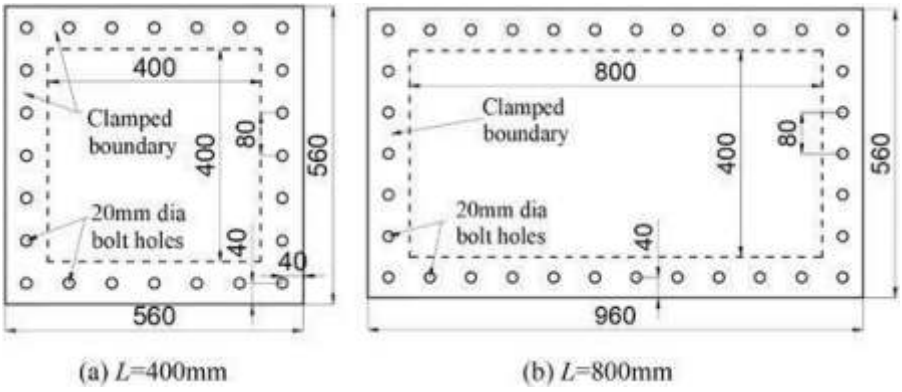


Fig. 2.1: *Shape and size of foam board sample for test.*

Fig. 2.1. The model is divided into single disc and double disc. Both are 10 mm thick. The "+" on the disk indicates electrical contact insertion [9]. The size of the straight gate used in the experiment was 30 mm X30 mm x240 mm, while the size of the cross gate was 30 mm x25 mm x 55 mm.

The influence of different process conditions on the filling morphology of magnesium solution was studied using a contact filling test device. In the test process, the appearance and the pouring parts are bonded by hot melt adhesive, and then the module is painted and baked in the oven [10]. The temperature of the oven should not exceed 50℃. The dried module is placed in the reverse sandbox and filled with water to wash the quartz sand. Vibrate on a three-dimensional vibratory compaction platform and cover with plastic film. Please turn on the vacuum pump so that it is in a stable state. Then, perfusion was performed and lasted for 5 minutes after the end of perfusion [11]. The magnesium alloy ingot is melted in a crucible resistance furnace with a capacity of 5 kg and a power of 5 kW. Flux is used in smelting. When all the charge is melted, the temperature is adjusted to 740-760 degrees, then hexachlorophene is added to remove steam; after refining, stop for 15-20 min, then adjust it to the required temperature and pour.

3. Numerical simulation of flow and temperature field in casting furnace.

3.1. System Assumptions.

1. Magnesium alloy is an exemplary flow material. Magnesium metal liquid is regarded as a continuous incompressible Newtonian fluid during simulation.
2. The apparent viscosity of the liquid magnesium fluid between the liquidus and the solid phase is set as a temperature-dependent model and in a straight-line manner.
3. The numerical calculation is based on the flow diversion mode inside the gate, and the line segment from the liquid outlet end of the gate to the middle of the roll gap is simplified into the overall flow zone with the inner side of the gate [12]. The surrounding materials are all adiabatic, so the temperature change of the liquid steel during the casting and rolling process is ignored, and only the uniformity of the temperature field in the flow zone is concerned.

3.2. Governing Equation. The SOLA-VOF method was used to calculate the difference in magnesium alloy melt flow in the cast, and the following formula was obtained: Continuity equation:

$$\frac{\partial \alpha}{\partial x}+\frac{\partial \beta}{\partial y}+\frac{\partial \gamma}{\partial z}=0 \tag{3.1}$$

Momentum conservation equation:

$$\begin{aligned}
&\frac{\partial \alpha}{\partial t}+\alpha\frac{\partial \alpha}{\partial x}+\beta\frac{\partial \alpha}{\partial y}+\gamma\frac{\partial \alpha}{\partial z}=-\frac{1}{\rho}\frac{\partial p}{\partial x}+g_x+\delta\left(\frac{\partial^2 \alpha}{\partial x^2}+\frac{\partial^2 \alpha}{\partial y^2}+\frac{\partial^2 \alpha}{\partial z^2}\right)\\
&\frac{\partial \beta}{\partial t}+\alpha\frac{\partial \beta}{\partial x}+\beta\frac{\partial \beta}{\partial y}+\gamma\frac{\partial \beta}{\partial z}=-\frac{1}{\rho}\frac{\partial p}{\partial y}+g_y+\delta\left(\frac{\partial^2 \beta}{\partial x^2}+\frac{\partial^2 \beta}{\partial y^2}+\frac{\partial^2 \beta}{\partial z^2}\right)
\end{aligned} \tag{3.2}$$

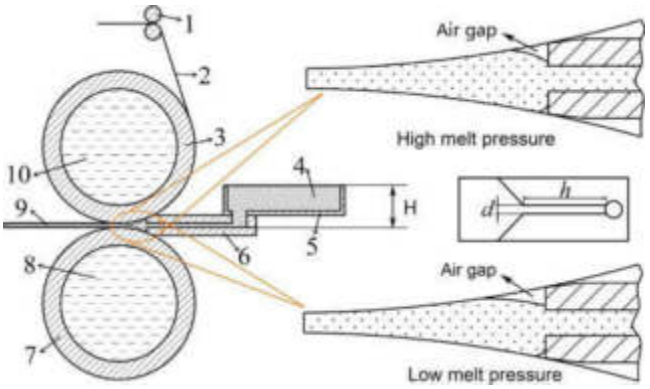


Fig. 3.1: *Schematic diagram of flow field of molten steel from box to gate and roll gap center line during casting and rolling .*

$$\frac{\partial \gamma}{\partial t} + \alpha \frac{\partial \gamma}{\partial x} + \beta \frac{\partial \gamma}{\partial y} + \gamma \frac{\partial \gamma}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial z} + g_z + \delta \left(\frac{\partial^2 \gamma}{\partial x^2} + \frac{\partial^2 \gamma}{\partial y^2} + \frac{\partial^2 \gamma}{\partial z^2} \right) \tag{3.3}$$

α, β, γ is the velocity vector of x, y, z in the coordinate system, and p is the pressure; Where ρ represents the concentration of magnesium alloy liquid, g_x, g_y, g_z represents gravitational acceleration, and δ represents dynamic viscosity. When using the volumetric function method to track the free surface motion, the H-type equation must also be solved:

$$\frac{\partial H}{\partial t} + \alpha \frac{\partial H}{\partial x} + \beta \frac{\partial H}{\partial y} + \gamma \frac{\partial H}{\partial z} = 0 \tag{3.4}$$

Energy equation

$$\frac{\partial E}{\partial t} + \alpha \frac{\partial E}{\partial x} + \beta \frac{\partial E}{\partial y} + \gamma \frac{\partial E}{\partial z} = \zeta \left(\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} \right) + \Delta S \tag{3.5}$$

ζ is the thermal diffusion coefficient, $\zeta = \xi / \rho_{c_p} \cdot c_p$ is the specific heat capacity, ξ is the thermal conductivity and ΔS is the latent heat of condensation.

3.3. Simplification of flow area. The flow field diagram of molten steel from box to gate and roll gap center line during casting and rolling is given (Fig. 3.1). The change of liquid face value in a specific interval is simplified to the constant value of liquid surface by simulation.

After the liquid level in the current tank is fixed, the midpoint of the gap is used as the starting point of the potential field. On the energy surface, the flow of the liquid obeys the Beendeavor formula:

$$p_0 + \rho g l_0 + \frac{1}{2}\rho \beta_0^2 = p_1 + \rho g l_1 + \frac{1}{2}\rho \beta_1^2 \tag{3.6}$$

The liquid metal level is selected at the connection point between the front box and the casting. The pressure p_0 on the liquid surface of the front tank is atmospheric [13]. In this case, when the liquid flows from the flow tank into the front container, the pulsation rate β_0 on the liquid surface is o. In this coordinate system, $\beta_0 = H.P_1$ is the pressure in the flow field and β_1 is the velocity in the flow field. In the numerical simulation of the flow field and temperature field in the twin-roll casting and rolling process, two methods of constant pressure and constant rate can be used for numerical simulation. In the casting and rolling process, when the liquid metal flows into the casting and rolling area from the gate, the speed of the gate inlet is realized by the static pressure formed by the difference in the liquid level of t front groove [14]. Since the liquid metal fills the casting area, the liquid pressure the gate is greater than the air pressure, so the flow rate of the liquid metal at

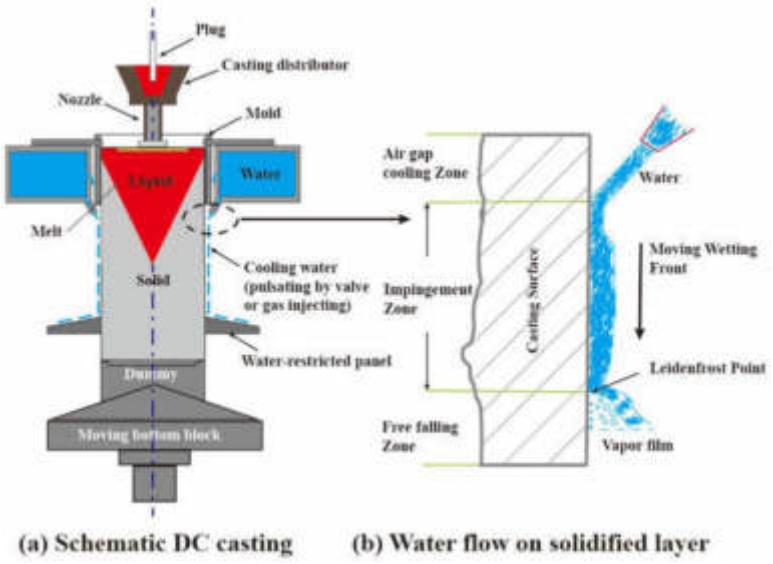


Fig. 3.2: *Schematic diagram of flow area for numerical simulation.*

the gate inlet also changes. The liquid metal in the entire flow field area is driven by constant pressure. In the simulation calculation, the pressure p_1 of the liquid metal at the gate inlet is set as follows:

$$p_1 = p_0 + \rho g l_0 \tag{3.7}$$

When the liquid steel does not slip in the inner surface of the gate and the roll surface, the interface flow rate of the two phases is o. Figure 3.2 shows the digital simulation results of the flow field. In the actual production, the liquid metal flows along the gate to the casting and rolling area, and then through the roll gap's center, the casting and rolling process is realized [15]. The flow and temperature fields of magnesium alloy melt in the continuous casting nozzle and casting and rolling zones were studied.

4. Test results and analysis.

4.1. Rheological properties of magnesium alloys under different vacuum conditions. Fig. 4.1 shows the flow field distribution at the gate plane under different vacuum conditions. When the vacuum degree increases, the rate of pyrolysis products is accelerated, the resistance of liquid propulsion is reduced, the flow rate is accelerated, and the filling time is significantly shortened [16]. In filling a single-track plate, the shape of the flow path shows the central radiation of the inner flow path, and the transition process of mold filling accompanies the transition from the convex arc to the inner flow path. The reason is that the liquid metal will be preferentially filled along the wall of the flow path under negative pressure.

In this process, heat conduction, convection and radiation are carried out between liquid metal and coating, dry sand and non-vaporized foam phase. During gasification mode pyrolysis, a series of physicochemical reactions and material transfer processes occur between molten metal, paint and dry sand. The results show that the Mg-based alloys at different temperatures are dominated by gas phases in different forms at different temperatures. Most gas products are gaseous styrene (about 70%), while small molecules are about 10%. In the process of evaporative molding, the pyrolysis gas formed a significant barrier to the coating layer, the pyrolysis products were discharged under negative pressure, and the pyrolysis products were concentrated in the outer boundary [17]. The accumulation of pyrolysis products at the front end of the flow field inhibits the forward movement of magnesium alloy melt, forming a convex arc shape with the inner gate in the center. With the increase of vacuum degree, the discharge rate of thermal decomposition products at the surface boundary increases significantly, accelerating the filling of the liquid boundary, enhancing the adhesion effect of the wall surface making the flow field's front shape smoother.

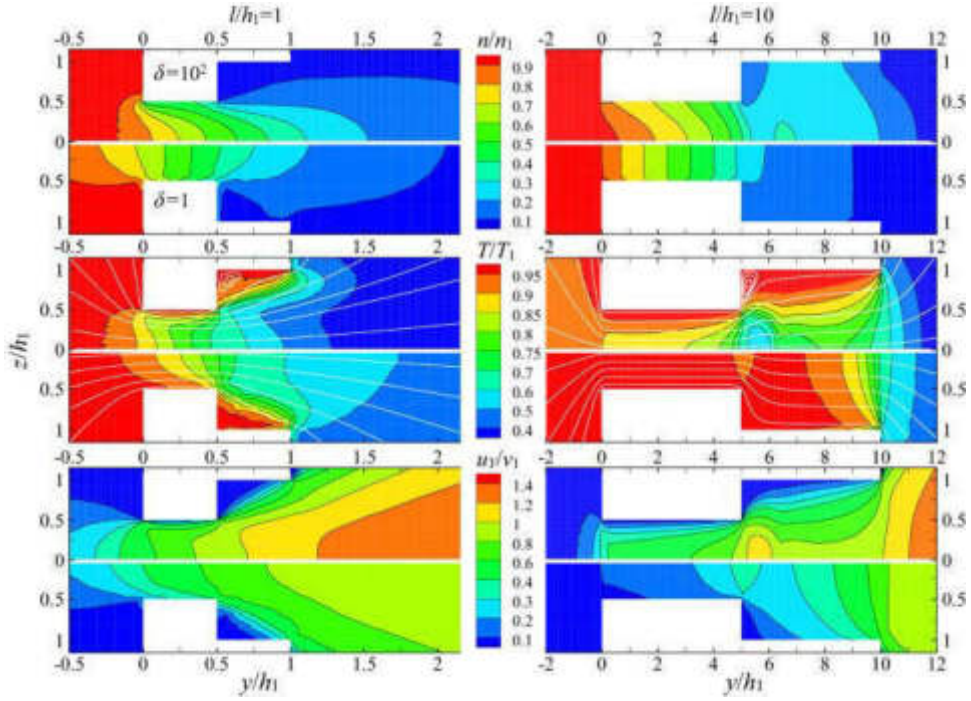


Fig. 4.1: *Flow field distribution of a single plate under different vacuum conditions.*

4.2. Influence of pouring temperature and number of gates on melt structure of MG-base alloy. Fig. 4.2 shows the variation of the shape of the flow field front in a single bottom surface and two channel plates with the pouring temperature when the vacuum degree is 0.02 MPa and the foam density is 0.020 g /cm3, and different pouring temperatures are adopted. The results show that when the liquid metal flows into the sample from the inner gate, the liquid metal forms a raised arc upward, which is roughly symmetric with the gate. When the flow front continues to advance, the liquid metal gradually fills the sample's cross-section, the flow field's curvature decreases during this process, and the movement of the liquid metal tends to be stable [18]. After this, the metal tip progresses in approximately the same pattern until the liquid metal reaches the top of the sample and is filled. The filling shapes of the two materials are not different at different pouring temperatures, but different pouring temperatures greatly influence the filling time, and it decreases with the increase in pouring temperature. With the increase in pouring temperature, the appearance vaporization rate increases, and the escape rate of the pyrolysis product increases under the same vacuum condition, thus reducing the molding time. In the process of filling the mold of the double-layer inner gate, the two liquids are filled with the same curvature, "double peaks," and in the forward process, the two liquids fuse into one in forward process, the bending degree is reduced, and the filling mold is more stable. Compared with a single plate, the double cavity can also reduce the filling time, making the filling process more stable.

4.3. Flow rate of metal melt during casting of lost mold. Fig. 4.3 shows the relationship between the filling rate of magnesium alloy liquid along the flow direction and the flow channel length under different vacuum degrees, pouring temperature and foam concentration [19]. The flow rate is determined by the time difference between the two contact points according to the distance between the two contact points in the flow direction, so the filling rate is the average rate between each contact point. The results show that the flow rate of magnesium increases first and then decreases during the filling process, but over time, the flow rate of magnesium changes continuously. It is necessary to alternate fast and slow to ensure the continuity of filling. A cracking reaction occurs when liquid metals meet under heating conditions, and the cracking product increases the resistance of liquid propulsion, mainly determined by the melt decomposition rate and

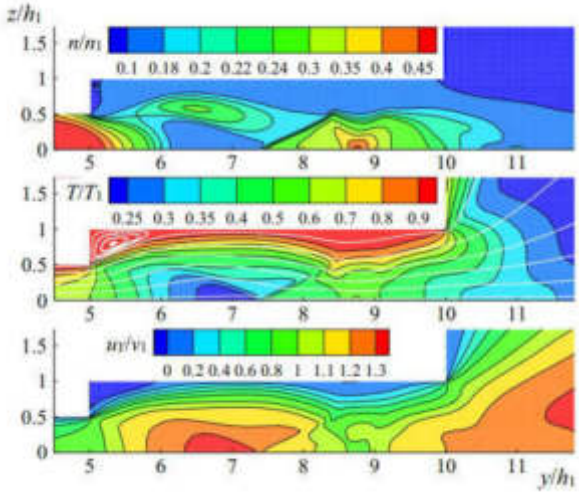


Fig. 4.2: *Flow front morphology of single gate and double gate plate at different pouring temperatures.*

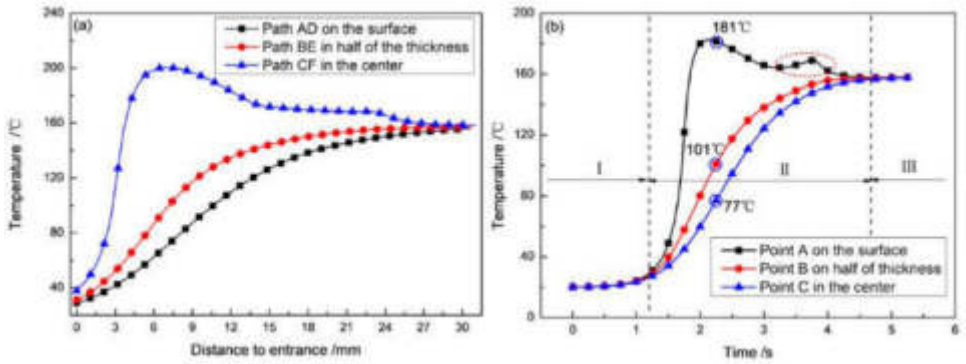


Fig. 4.3: *The relationship between the gate plane sample's filling rate and the flow path length under various processing parameters.*

product discharge range. The thrust force of liquid metal in the filling-blanking mode of liquid magnesium melt is mainly determined by gate pressure and injection volume. The resistance generated during filling is mainly due to the gas and liquid products not being discharged in time. The filling rate of liquid magnesium depends on the relative value of thrust and resistance. The movement rate of liquid magnesium alloy melt in the early casting stage is mainly due to its high gas permeability or negative pressure effect, which reduces the propropulsion resistance of liquid magnesium alloy. At this time, there is a significant gap in the transition layer between the metal melt and the bubble phase. The gap is getting smaller. When the melt flow rate of magnesium alloy is greater than the decomposition rate of the foam phase, the interfacial spacing of melt droplets becomes narrower. The thermal resistance of droplet gas and liquid product increases, and the melt filling rate decreases. At the same time, the heat transfer on the magnesium alloy surface is enhanced with the reduction of the interfacial spacing. At this time, the cracking rate of the morphology is faster than that of the previously reduced liquid melt, so the transition layer increases, the flow resistance decreases, and the filling rate increases again. The velocity of magnesium alloy melt again exceeds the decomposition rate of metal melt. Under different mold filling conditions, the change of void size and resistance in the transition layer can explain

the periodic change in mold filling rate. It is necessary to use different pouring rates to control according to the different flow rates of magnesium alloy melt to ensure the continuous filling of liquid metal. When pouring, the straight gate should be filled. Otherwise, due to the low heat capacity of magnesium alloy wire, it is easy to have the disadvantage of a poor gate. If the injection rate is too high, the appearance of rapid evaporation will occur, and there will be a large amount of gas backflow or lead to magnesium alloy solution leakage from the pouring cup.

5. Conclusion.

1. During vacuum grouting filling, due to the accumulation of products generated by thermal decomposition at the front end of the flow, especially the outer edge, the magnesium alloy solution advances from the inner gate in a convex arc shape. The flow field structure tends to be stable during mold-filling, but wall adhesion will occur when the vacuum degree increases. The filling time is shortened with the increase of pouring temperature and vacuum degree.
2. The filling rate of magnesium alloy is generally from high to low, showing a periodic change during this period.

REFERENCES

[1] Shehata, M. M., El-Hadad, S., Moussa, M. E., & El-Shennawy, M. (2021). The combined effect of cooling slope plate casting and mold vibration on microstructure, hardness and wear behavior of Al–Si alloy (A390). International Journal of Metalcasting, 15(3), 763-779.

[2] Shayan, M., Eghbali, B., & Niroumand, B. (2020). Fabrication of AA2024− TiO2 nanocomposites through stir casting process. Transactions of Nonferrous Metals Society of China, 30(11), 2891-2903.

[3] Javaid, A., & Czerwinski, F. (2021). Progress in twin roll casting of magnesium alloys: A review. Journal of Magnesium and Alloys, 9(2), 362-391.

[4] Wu, G., Wang, C., Sun, M., & Ding, W. (2021). Recent developments and applications on high-performance cast magnesium rare-earth alloys. Journal of Magnesium and Alloys, 9(1), 1-20.

[5] Kumar, A., Rana, R. S., & Purohit, R. (2020). Effect of stirrer design on microstructure of MWCNT and Al alloy by stir casting process. Advances in Materials and Processing Technologies, 6(2), 320-327.

[6] BADIZI, R. M., Parizad, A., Askari-Paykani, M., & Shahverdi, H. R. (2020). Optimization of mechanical properties using D-optimal factorial design of experiment: Electromagnetic stir casting process of A357− SiC nanocomposite. Transactions of Nonferrous Metals Society of China, 30(5), 1183-1194.

[7] Rao, M. S. S., & Kumar, A. (2024). Experimental study and optimization of process parameters for producing semi-solid A392 alloy using vibration-assisted cooling slope process integrated with mould vibration. International Journal of Metalcasting, 18(2), 944-961.

[8] Gurusamy, P., Raj, S. H. K., Bhattacharjee, B., & Bhowmik, A. (2024). Assessment of microstructure and investigation into the mechanical characteristics and machinability of A356 aluminum hybrid composite reinforced with SiCp and MWCNTs fabricated through rotary centrifugal and squeeze casting processes. Silicon, 16(1), 367-382.

[9] Dong, Y. W., Shao, P. F., Guo, X., Xu, B., Yin, C. P., & Tan, Z. Y. (2023). Deformation characterization method of typical double-walled turbine blade structure during casting process. Journal of Iron and Steel Research International, 30(10), 2010-2020.

[10] Nişancı, M. C., & Yurddaş, A. (2020). Compare between the results of the casting simulation and the results of experimental production with calculating the interface heat transfer coefficient of the casting-mold. Celal Bayar University Journal of Science, 16(2), 169-181.

[11] Asano, K., Yamada, H., & Sugimura, S. (2024). Erosion resistance of heat-treated aluminum cast iron to aluminum alloy melt. MATERIALS TRANSACTIONS, 65(5), 534-540.

[12] Kanyo, J. E., Schafföner, S., Uwanyuze, R. S., & Leary, K. S. (2020). An overview of ceramic molds for investment casting of nickel superalloys. Journal of the European Ceramic Society, 40(15), 4955-4973.

[13] Sankhla, A., & Patel, K. M. (2022). Metal matrix composites fabricated by stir casting process–a review. Advances in Materials and Processing Technologies, 8(2), 1270-1291.

[14] Schurmann, D., Glavinić, I., Willers, B., Timmel, K., & Eckert, S. (2020). Impact of the electromagnetic brake position on the flow structure in a slab continuous casting mold: An experimental parameter study. Metallurgical and Materials Transactions B, 51(1), 61-78.

[15] Terwadkar, A. R., Sahasrabuddhe, A. V., & Kulkarni, M. M. (2020). Implementation of Statistical Quality Control Techniques to Minimize the Casting Defects. International Research Journal of Engineering and Technology (IRJET), 7(12), 1533-1544.

[16] Berti, S., Jagus, R. J., Flores, S. K., & González-Martínez, C. (2024). Antimicrobial Edible Starch Films Obtained By Casting and Thermo compression Techniques. Food and Bioprocess Technology, 17(4), 904-916.

[17] Hao, X., Liu, G. H., Wang, Y., Wu, S. P., & Wang, Z. D. (2022). Optimization of investment casting process for K477 superalloy aero-engine turbine nozzle by simulation and experiment. China Foundry, 19(4), 351-358.

[18] Guan, F., Jiang, W., Li, G., Zhu, J., Wang, J., Jie, G., & Fan, Z. (2022). Effect of vibration on interfacial microstructure

and mechanical properties of Mg/Al bimetal prepared by a novel compound casting. Journal of Magnesium and Alloys, 10(8), 2296-2309.
[19] Guo, Z., Jie, J., Liu, S., Liu, J., Yue, S., Zhang, Y., & Li, T. (2020). Solidification characteristics and segregation behavior of Cu-15Ni-8Sn alloy. Metallurgical and Materials Transactions A, 51(3), 1229-1241.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 24, 2024
Accepted: Jul 5, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 837–845, DOI 10.12694/scpe.v26i2.4025

CARBON EMISSION PREDICTION AND SENSITIVITY EVALUATION OF VIRTUAL POWER PLANTS BASED ON BIG DATA AND MULTISCALE ANALYSIS

JIE LI*, ZHOU YANG†, WENQIAN JIANG‡, AND JUNTAO PAN §

Abstract. In order to address the issue of increased prediction errors in the peak carbon emissions of virtual power plants due to various influencing factors of electricity carbon emissions, the authors propose a study on the prediction and sensitivity evaluation of virtual power plant carbon emissions based on big data and multi-scale analysis. Firstly, it analyzes the original data sequence and cumulative sequence, use grey BP neural network to construct a carbon emission peak prediction model, then it analyzes the factors affecting electricity carbon emissions, and use recursive calculation method to calculate electricity carbon emissions. Then, it compress the model coefficients to zero through a penalty function and filter out significant variables. Based on the adjacency characteristics of carbon emission flow, the node carbon potential is calculated through finite recursion, and iterative training is carried out within the allowable error range to solve the model and obtain the predicted peak carbon emissions of electricity. The experimental results indicate that the prediction results of the designed method under three scenarios of benchmark setting, low-carbon, and enhanced low-carbon are 40 million tons, 390 million tons, and 40 million tons, respectively, which are consistent with the actual results, indicating that the prediction error of this method is lower and the prediction results are more accurate. The method studied by the authors can provide technical support for carbon emission control and improve prediction accuracy.

Key words: Grey BP neural network, Virtual power plant carbon emissions, Peak prediction, Penalty function

1. Introduction. With the continuous improvement of the electricity market trading system, the trading models have become more diversified, presenting a mixed trading model of bilateral and joint venture transactions coexisting [1]. However, the existing carbon emission accounting methods are not applicable to all trading models. For bilateral trading models, especially green bilateral trading, the allocation method of "nearby power supply, proportional sharing" does not meet the "bilateral" characteristics, and the low-carbon benefits of clean energy power plants cannot be accurately allocated to users who sign bilateral transactions with them. The power industry is an indispensable basic industry in modern society. With the continuous increase in energy demand, the impact of carbon emissions on the environment is becoming increasingly significant. The global power industry is accelerating its transformation towards clean and low-carbon development [2]. As an important component of the power industry, the electricity market coordinates and manages the generation, sales, transmission, and customer relationships in the power system.

The carbon emissions from virtual power plants are influenced by a variety of dynamic factors. Accurately forecasting these emissions is crucial for developing effective carbon reduction strategies. By predicting future carbon dioxide emissions of power plants, it is possible to optimize the power generation mix and implement specific energy-related carbon reduction measures. At present, there is relatively little research on predicting carbon emissions from power plants, and many existing achievements rely on the analysis of influencing factors to predict carbon emissions. Due to the multiple influencing factors of carbon emissions prediction in power plants, traditional small power plants lack advanced sensor equipment, making it difficult to obtain data on various factors, resulting in significant limitations in predicting influencing factors. The author aims to predict and evaluate the sensitivity of carbon emissions from virtual power plants through big data and multi-scale analysis methods [3]. The development of big data technology provides the energy industry with massive data and rich analytical tools, which can effectively capture and analyze various data generated during the

*Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China.
†Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China (Corresponding author, yangz_gxcsg@163.com)
‡Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China.
§Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China.

operation of virtual power plants. The multi-scale analysis method can deeply explore the dynamic changes and influencing factors of carbon emissions from different time and spatial scales [4,5,6].

2. Literature Review. As the share of renewable energy generation rises and flexible resources like distributed energy storage and electric vehicles are increasingly integrated, the discrepancy between the power system's load demand and the spatiotemporal characteristics of power output has become more pronounced. Consequently, the virtual power plant (VPP) has been developed. VPPs leverage advanced communication technologies to efficiently aggregate distributed power sources, controllable loads, and energy storage systems, allowing them to flexibly respond to scheduling directives and participate in electricity market transactions. Therefore, how to accurately predict the peak carbon emissions, formulate carbon reduction strategies based on the prediction results, and control the peak carbon emissions in a short period of time have become urgent problems in the development process of modern society. Currently, scholars are studying methods for predicting peak carbon emissions from different perspectives and theories. Yang et al. introduced a flexible carbon emission mechanism that coordinates energy between electric hydrogen equipment, SMR factories, and gas turbines. Within this new framework and carbon emission mechanism, the VPP makes purchasing and sales decisions in both day-ahead and real-time markets while optimizing the operational strategies of its internal components [7]. Xuejin, W. et al. developed a virtual power plant scheduling method utilizing a low-carbon multi-objective two-stage optimization algorithm. Initially, this method identifies the production and consumption levels of various energy sources within virtual power plants, such as wind power, thermal power, and hydropower. Subsequently, aiming to minimize costs and reduce carbon emissions, multi-objective optimization algorithms are employed to allocate and schedule the energy resources within the virtual power plant [8]. Wu, Y.et al. introduced virtual power plants (VPP) and power-to-gas (P2G) technologies to enhance energy integration. They first proposed a VPP structure connected to P2G and developed a physical output model. Then, they constructed a multi-objective operational optimization model for VPPs, considering electrical interconnection, with the dual goals of reducing carbon emissions and achieving economic operation. They also proposed a solution method for this model. Finally, they validated the contributions of P2G, demand response (DR), and gas storage technology (GST) through case studies [9].

In response to the existing problems of the above methods, the author proposes a research on virtual power plant carbon emission prediction and sensitivity evaluation based on big data and multi-scale analysis. It can flexibly set parameters according to actual situations, reduce prediction errors, and improve prediction accuracy.

3. Method.

3.1. Design of multi-scale state monitoring methods for big data. The multi-scale state monitoring method proposed by the author for big data is mainly based on the big data provided by the information system, constructing corresponding state monitoring functions that meet the conditions for state parameter reconstruction, and monitoring the corresponding state of industrial equipment [10]. At the same time, multi-scale monitoring of equipment status is carried out through benchmark models and residual fusion of multi-step information. The flowchart is shown in Figure 3.1.

3.1.1. Data preprocessing. For big data in device information systems, there is a lot of noise and incompleteness. In order to monitor devices more accurately, it is necessary to preprocess big data, which can greatly improve the quality of big data and the efficiency of monitoring [11]. Data preprocessing mainly consists of four steps, namely data cleaning, data integration, data transformation, and data reduction [12]. Among them, data cleaning refers to smoothing big data, identifying and removing noisy data, removing isolated points, and processing missing values in big data accordingly; Data integration mainly involves integrating big data from different sources into the same data storage system according to the same rules, facilitating the subsequent use of big data; Data transformation refers to the appropriate transformation of data forms through methods such as data generalization, smooth aggregation, normalization, etc., in order to facilitate subsequent use; Data specification refers to the specification representation of the obtained dataset, mainly used to ensure the integrity of the data. The data preprocessing was completed through the above process, providing data support for the construction of the following benchmark models [13,14].

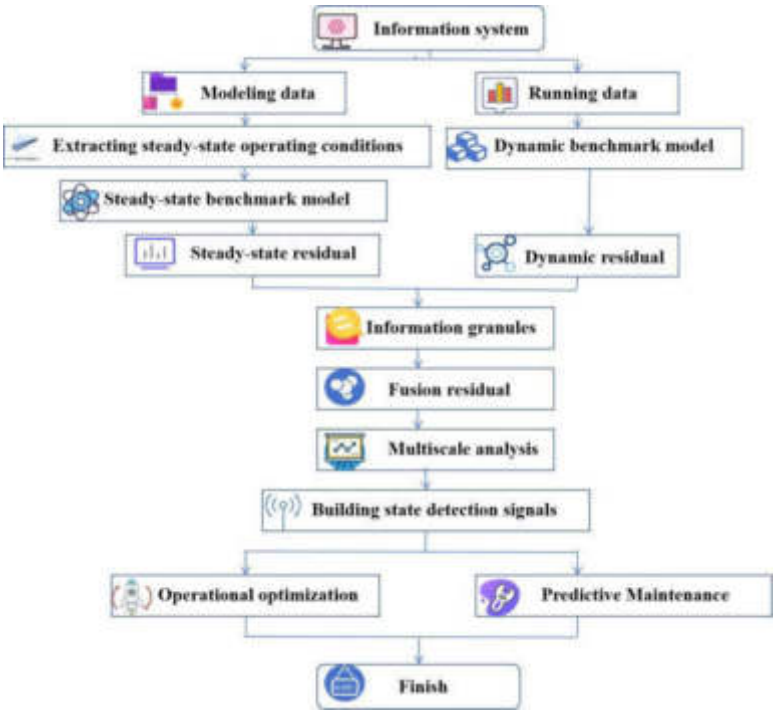


Fig. 3.1: Flow chart of multi-scale state monitoring method for big data

3.1.2. Building benchmark models. Based on the preprocessed data mentioned above, the benchmark model is constructed using deep learning theory. The benchmark model mainly establishes connections between the operating data and design parameters of equipment through certain mechanisms, with the aim of constructing corresponding criteria for establishing industrial equipment parameters and providing reference information for status monitoring. During the operation of the equipment, comparing the reference information provided by the benchmark model with measurable information can construct corresponding residual sequences. If the industrial equipment is working normally, the residual is almost zero; otherwise, it indicates abnormal operation of the equipment. Inconsistencies frequently arise during equipment operation. Consequently, when creating a benchmark model, it is essential to develop both steady-state and dynamic models simultaneously. The steady-state benchmark model is primarily constructed using large datasets of steady-state operating conditions, focusing solely on the parameters during stable equipment operation. In contrast, the dynamic benchmark model is based on extensive data from dynamic operating conditions, reflecting the equipment's performance under varying conditions. During the operation of the equipment, its dynamic working condition data accounts for a large proportion, therefore, this model mainly represents the relationship between this data. To minimize modeling errors resulting from device delays, inertia, and other factors, deep learning techniques are employed to develop the model, thereby enhancing its overall quality [15]. The theoretical model of deep learning is shown in Figure 3.2.

The construction of the benchmark model was completed through the above process, providing reference information for the final state monitoring.

3.2. Construction of prediction model based on grey BP neural network. Carbon emissions are a dynamic process influenced by multiple factors, and artificial neural network models have been proven to have high applicability for predicting carbon emissions. In the construction of artificial neural network carbon emission prediction models, accurately identifying the influencing factors of carbon emissions is the key to improving the accuracy of input layer data in the artificial neural network model, which plays an important role in improving the prediction accuracy of the model. By combining the dynamic variability, nonlinearity, and

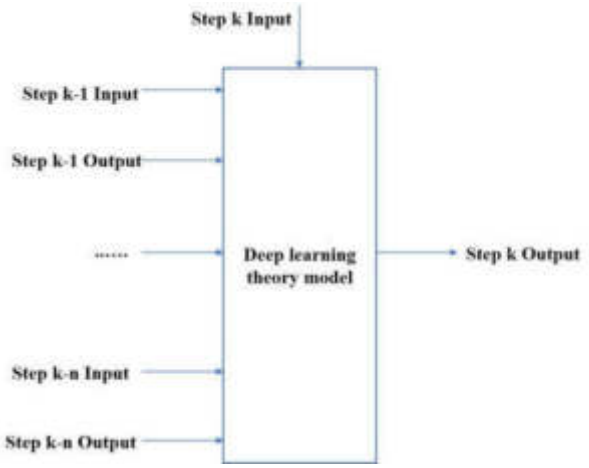


Fig. 3.2: Deep Learning Theory Model

sociality of carbon emissions, optimizing the carbon emission prediction model through parameter optimization, weight initialization, model structure adjustment, learning rate scheduling, and other methods, or constructing a linear nonlinear coupled combination model, the prediction efficiency of the model can be effectively improved. As a data-driven prediction model, the predictive performance of artificial neural network models is affected by initial values, which leads to the network easily falling into local optima and training easily entering overfitting or premature fitting. With the increasing demand for accuracy in carbon emission prediction, further research is needed on the optimization methods of artificial neural network models. The grey BP neural network based on error backpropagation has the advantage of continuously approaching the limit value of the function, which leads to the existence of the optimal solution in the prediction results. In addition, the number of layers, units, and training factors of the model structure can be set according to different environments, making the prediction process more flexible and random, and improving prediction accuracy and efficiency. The grey BP neural network can train a large amount of data and clarify the relationships between the data. By using a fast descent method to continuously adjust the network weights and thresholds through reverse propagation, the error of the network is reduced. The core idea of this method is to introduce a new learning method, which iteratively modifies and trains the network's reverse propagation to ensure that the output vector of the network is consistent with the desired vector [16].

The grey BP neural network model considers random variables as gray variables that vary within a certain interval. After processing, the accumulated sequence shows an exponential growth trend. In this case, the original data sequence and the accumulated sequence are:

$$c^{(0)} = [c^{(0)}(1), c^{(0)}(2), c^{(0)}(3), \cdots, c^{(0)}(n)] \tag{3.1}$$

$$c^{(1)} = [c^{(1)}(1), c^{(1)}(2), c^{(1)}(3), \cdots, c^{(1)}(n)] \tag{3.2}$$

Using the original data sequence and cumulative sequence as input values, a carbon emission peak prediction model based on grey BP neural network is constructed, as shown in Figure 3.3.

Model training is considered complete when the sum of squared errors in the network's output layer falls below a specified threshold. This approach effectively reduces the weighted order deviation within the network [17]. The specific steps of model training are to initialize each node and randomly assign weights and thresholds to each node. After completing the parameter settings, calculate the connection weights and thresholds for the input layer and output layer respectively. Choose the next input method and iterate repeatedly until the network output meets the requirements.

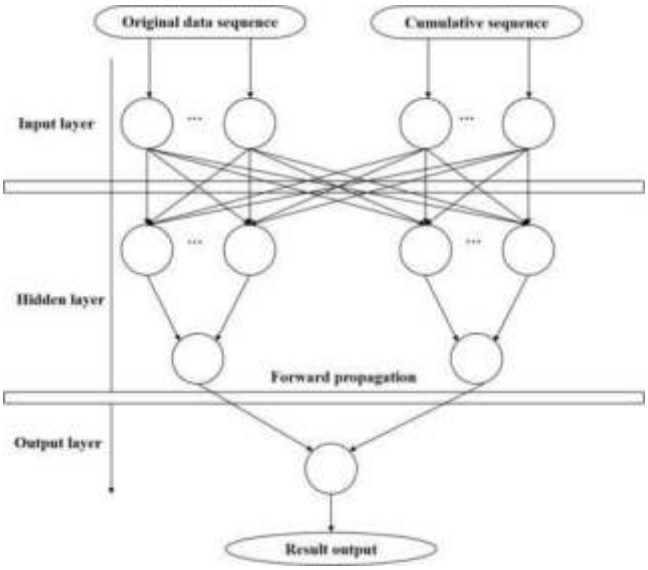


Fig. 3.3: Prediction Model

3.3. Solving the Peak Carbon Emission Prediction Model for Virtual Power Plants. In order to solve the carbon emission peak prediction model for virtual power plants, a recursive algorithm is introduced into the grey BP neural network. This algorithm selects a random carbon emission path, calculates the corresponding impact weights of each factor, and then obtains the corresponding path carbon emission values. The randomly selected new carbon emission path is:

$$L_{i+1} = L_i + \lambda \oplus L'(\xi) \tag{3.3}$$

In the formula, L_i represents the i-th carbon emission path; λ represents the carbon emission coefficient; $L'(\xi)$ represents the carbon emission path extraction function. In the process of calculating the carbon emissions of virtual power plants, the recursive calculation method is used to calculate the carbon emissions of virtual power plants. The formula is:

$$C = \sum_{i=1}^{j} a_i \times f_i \times \lambda_i \tag{3.4}$$

In the formula, a_i represents the i-th electricity consumption mode; f_i represents the i-th power conversion coefficient; λ_i represents the i-th type of electricity carbon emission coefficient; j represents the number of electrical equipment. The recursive algorithm adopts a random selection of carbon emission paths for electricity, mainly by analyzing the influencing factors of carbon emissions on different paths to obtain the maximum and minimum carbon emission cycles. The specific operation process is as follows:

Step 1. Randomly obtain the carbon emissions of each path, and by introducing differential equation parameters and weighting them, a new grey BP neural network model is obtained.

Step 2. Normalization method is used to eliminate non-linear relationships between data and introduce them into the model. The Lasso regression analysis method was used to analyze the impact of various factors on the prediction of peak carbon emissions. Lasso regression analysis compresses the coefficients in the model using a penalty function, turning some factors to 0 to filter out significant variables.

Assuming a as the independent variable and b as the dependent variable, the standard value of the predicted sample obtained after m samples is (a, b), and the kth predicted value of the independent variable a is:

$$x_k = (x_{k1}, x_{k2}, \cdots, x_{km})^T \tag{3.5}$$

In the formula, T represents the prediction period. The regression model of the dependent variable on the independent variable can be expressed as:

$$b_i = \sum_{i=1}^{j} a_i + \epsilon_i \tag{3.6}$$

In the formula, ϵ_i represents a random natural number. If you want to filter out variables that have a significant impact, you need to add a condition to the formula, and the constraint expression is:

$$\begin{cases} arg(t_1, t_2, \cdots, t_t) min||b - ta||^2 \\ s.t. \sum_j \frac{|t|}{\sum t_j^0} \leqslant \phi \end{cases} \tag{3.7}$$

In the formula, t represents the harmonic parameter; ϕ represents the optimal adjustment threshold. Lasso regression is the process of continuously adjusting harmonic parameter values, reducing regression coefficients, compressing variable coefficients until they reach 0, in order to obtain significant variables, namely carbon emission peaks [18].

Step 3. Due to the adjacency of electricity carbon emissions calculation, when calculating the carbon emissions of a node, only the carbon emissions of neighboring nodes need to be obtained, and there is no need to know the carbon emission flow information of that node.

By allocating power to each node, the connections between each node can be obtained. Based on the adjacency of carbon emissions in the power grid, calculate the carbon potential from the initial point to different nodes in sequence. In each iteration, after determining the carbon potential of a certain node, all node carbon potentials can be obtained. Therefore, each iteration can obtain accurate node carbon potential calculation results within any period of time. Finally, the finite recursive method was used to calculate the carbon potential of all nodes in the network. The specific calculation formula is as follows:

$$\sigma_j = \frac{\sum_{i \in \Omega_i} P_i \sigma_i + \sum_{j \in \Omega_j} G_j \sigma_i^G}{\sum_{i \in \Omega_i} P_i + \sum_{j \in \Omega_j} G_j} \tag{3.8}$$

In the formula, P_i represents the active power injected by the node; G_j represents the active power of the power unit branch; Ω_i, Ω_j represents the set of carbon emissions and node injection for the i-th and j-th electricity, respectively. Determine if all nodes have been polled, and if all node carbon potentials have been obtained, complete the recursion.

Step 4. Under the condition of allowable deviation, the prediction model was solved and the peak prediction results of electricity carbon emissions were obtained [19].

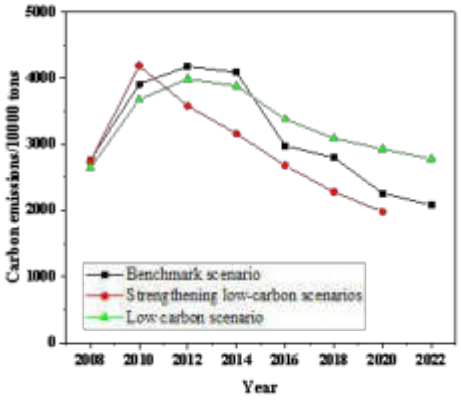
3.4. Experiment. In order to verify the rationality of the author's virtual electric field carbon emission peak prediction method, relevant experiments were designed to verify the feasibility of the method. By analyzing the direction and scale of power exchange between the distribution network and the main network, determine the calculation order of carbon emissions between each major network. If active power is injected into the main grid, it can be used as the main grid power supply, not as the main grid load. If active power is injected into the main grid distribution network, the boundary information between distributed power sources and conventional thermal power plants is used to analyze carbon emissions in the distribution network. Building on this, the carbon potential and injection power of the root node were used as key variables for the main network, and the carbon emissions of the main network were calculated. The distribution of these carbon emissions was then analyzed based on the carbon emission flow within the main network [20].

4. Results and Discussion. The peak of carbon emissions does not mean that the carbon emissions reach their peak in a year, but rather that a stable trend in carbon emissions in a certain region begins from that year. Under the baseline scenario, low-carbon scenario, and enhanced low-carbon scenario, the peak carbon emissions were collected as shown in Table 4.1.

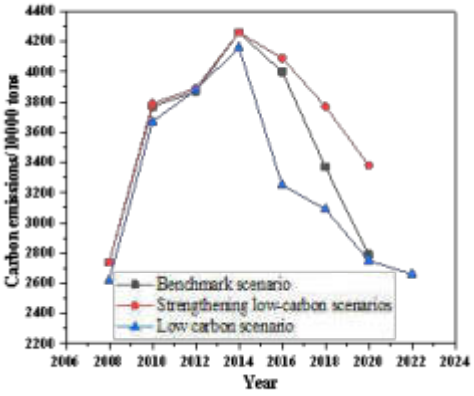
According to Figure 4.1(a), there is a maximum error of 6 million tons between the peak carbon emissions of the STIRPAT model method and the data in Table 4.1 under the baseline scenario; There is a maximum

Table 4.1: Carbon Emission Peak Collection Results

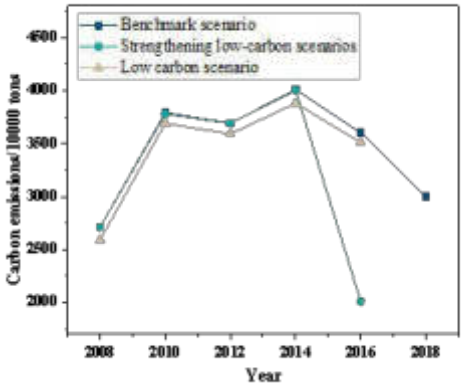
year /year	Peak carbon emissions/10000 tons		
	Benchmark scenario	Low carbon scenario	Strengthening low-carbon scenarios
2008	27001	26001	27001
2010	38001	37001	38001
2012	37001	36001	37001
2014	40001	39001	40001
2016	36001	38001	35001
2018	34001	37001	30001
2020	32001	36001	25001
2022	30001	35001	20001



(a) STIRPAT model method



(b) Double regression prediction model method



(c) Author's research methods

error of 5 million tons between the peak carbon emissions of the STIRPAT model method and the data in Table 4.1 under low-carbon scenarios; There is a maximum error of 4 million tons between the peak carbon emissions of the STIRPAT model method and the data in Table 4.1 under the enhanced low-carbon scenario.

According to Figure 4.1(b), there is a maximum error of 4 million tons between the carbon emission peak of the dual regression prediction model method and the data in Table 1 under the baseline scenario; There is a maximum error of 5 million tons between the carbon emission peak of the dual regression prediction model method in the low-carbon scenario and the data in Table 1; There is a maximum error of 6 million tons between the carbon emission peak of the dual regression prediction model method under the strengthened low-carbon scenario and the data in Table 4.1. According to Figure 4.3(c), it can be seen that the peak carbon emissions under the three scenarios using the research method are consistent with the data in Table 1. Based on the above analysis results, it can be concluded that the peak carbon emissions from electricity using the research method are consistent with actual data, indicating that the prediction results of this method are more accurate.

5. Conclusion. The author proposes a study on virtual power plant carbon emission prediction and sensitivity evaluation based on big data and multi-scale analysis. Due to the non-linear trend of carbon emission changes, the grey BP neural network is used for carbon emission peak prediction. The reason is that the grey BP neural network has good non mapping ability and can flexibly set parameters according to actual situations. The Lasso regression screening method and recursive calculation method are used to solve the model, and the relevant results of electricity carbon emission peak prediction are obtained. The experimental results show that under the baseline scenario, low-carbon scenario, and enhanced low-carbon scenario, the peak carbon emissions of the author's research method are consistent with the actual values, all of which are 40 million tons, 390 million tons, and 40 million tons. The rationality of the author's research method has been verified by the experiment, which can provide technical support for carbon emission control.

6. Acknowledgement. This work was supported by the research project of Guangxi Power Grid Co., Ltd (Project No. GXKJXM20220081)

REFERENCES

[1] Yan, R., Lin, Y., Yu, N., & Wu, Y. (2023). A low-carbon economic dispatch model for electricity market with wind power based on improved ant-lion optimisation algorithm, 8(1), 29-39.
[2] Piao, S., He, Y., Wang, X., & Chen, F. (2022). Estimation of china's terrestrial ecosystem carbon sink: methods, progress and prospects, 65(4), 11.
[3] Sheveleva, G. I. (2022). Corporate governance in generating companies of the russian electric power industry in the context of esg agenda, 5(5), 12.
[4] Leng, K., Li, Z., & Tong, Z. (2022). How will tradable green certificates affect electricity trading markets under renewable portfolio standards?a china perspective, 6(4), 585-598.
[5] Ju, L., Yin, Z., Zhou, Q., Liu, L., Pan, Y., & Tan, Z. (2023). Near-zero carbon stochastic dispatch optimization model for power-to-gas-based virtual power plant considering information gap status theory. International Journal of Climate Change Strategies and Management, 15(2), 105-127.
[6] Su, S., Hu, G., Li, X., Li, X., & Xiong, W. (2023). Electricity-carbon interactive optimal dispatch of multi-virtual power plant considering integrated demand response, 120(10), 2343-2368.
[7] Yang, Z., Li, K., & Chen, J. (2024). Robust scheduling of virtual power plant with power-to-hydrogen considering a flexible carbon emission mechanism. Electric Power Systems Research(Jan.), 6(1), 1-11.
[8] Xuejin, W., Chen, C., Yao, S., & Qiang, C. (2024). Multi-objective two-stage optimization scheduling algorithm for virtual power plants considering low carbon. International Journal of Low-Carbon Technologies, 11(1), 13.
[9] Wu, Y., Wu, J., De, G., & Fan, W. (2022). Research on optimal operation model of virtual electric power plant considering net-zero carbon emission. Sustainability, 47(1), 443-458.
[10] Kuang, Y., Wang, X., Zhao, H., Qian, T., Li, N., & Wang, J., et al. (2023). Model-free demand response scheduling strategy for virtual power plants considering risk attitude of consumers, 9(2), 516-528.
[11] Wang, M., Wang, P., Liang, W. U., Yang, R. P., Feng, X. Z., & Zhao, M. X., et al. (2022). Criteria for assessing carbon emissions peaks at provincial level in china, 13(1), 131-137.
[12] Guo, Q., Xi, X., Yang, S., & Cai, M. (2022). Technology strategies to achieve carbon peak and carbon neutrality for china's metal mines, 29(4), 9.
[13] Liu, E. B., Peng, Y., Peng, S. B., Yu, B., & Chen, Q. K. (2022). Research on low carbon emission optimization operation technology of natural gas pipeline under multi-energy structure, 19(6), 3046-3058.
[14] Abudu, H., Jr, P. K. W., & Lin, B. (2022). Climate pledges versus commitment: are policy actions of middle-east and north african countries consistent with their emissions targets?, 13(4), 612-621.
[15] BAIQinghua, YINXuelian, WANGJing, ZHANGJie, CHUChao, & LIXuejun. (2023). Prediction model of minimum temperature inside solar greenhouse in central hexi corridor based on ridge regression. JOURNAL OF AGRICULTURE, 13(5), 96-100.
[16] FengDONG, ZhaoLI, Ling-HanLI, & Shu-MeiZHANG. (2022). Flow state monitoring of gas-liquid two-phase flow using multiple dynamic kernel principle component analysis. Acta Automatica Sinica, 43(03), 762-773.

[17] Li, J. (2022). Venture financing risk assessment and risk control algorithm for small and medium-sized enterprises in the era of big data. Journal of Intelligent Systems, 31(1), 611-622.
[18] Lee, S. Y., Park, J., & Kim, D. Y. (2023). Context awareness by noise-pattern analysis of a smart factory, 76(8), 1497-1514.
[19] Chaovalitwongse, W. A., Yuan, Y., Zhang, Q., & Liu, J. (2022). Special issue: innovative applications of big data and artificial intelligence, 9(4), 3.
[20] Yuan, X., Deng, H., & Hu, J. (2022). Constructing a ppi network based on deep transfer learning for protein complex detection. IEEJ Transactions on Electrical and Electronic Engineering, 17(3), 436-444.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 24, 2024
Accepted: Jul 8, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 846–853, DOI 10.12694/scpe.v26i2.4045

DESIGN OF A MONITORING AND FEEDBACK SYSTEM FOR ATHLETE TRAINING PROCESS BASED ON MOBILE INTELLIGENT DEVICES

MENGMENG HU *AND XIFENG WU †

Abstract. In order to solve the problem of high probability of information loss in traditional sports training information management systems, the authors propose the design of an athlete training process monitoring and feedback system based on mobile intelligent devices. The system improves the design of sports training information mobile storage servers, mobile information collection sub sites, and intelligent information processors in terms of hardware, and processes users' sports training information in real-time. In terms of software design, classify the sports training information of different users, calculate the similarity of sports training information of different users, and complete intelligent information management. The experimental results show that for traditional information management systems, the average packet loss rate for tablet device users is 0.0534%, and the average packet loss rate for mobile device users is 0.0732%. The data on packet loss rate is relatively high. For the intelligent management system of sports training information based on mobile devices, the average packet loss rate for tablet users is 0.0144%, and the average packet loss rate for mobile device users is 0.0300%. By comparison, it can be seen that the packet loss rate of information intelligent management systems based on mobile devices is lower. The information processing packet loss rate of this system is significantly reduced, which is better than traditional information management systems.

Key words: Sports training, Information management, Mobile devices

1. Introduction. Modern training theory believes that the training process of athletes is a systematic engineering process that combines multiple disciplines and factors to achieve the best combination [1]. Therefore, in order to grasp the process of athletes' competitive career, it is necessary to have a clear understanding of the growth process of athletes and the conditions under which the athlete training system was formed and evolved. This is one of the important prerequisites for athletes to move towards "scientific" long-term systematic training [2].

From the perspective of training theory, the excellent competitive ability of athletes is the foundation and prerequisite for achieving excellent sports results, which has attracted great attention from training management departments and research institutions at all levels to the competitive ability of athletes. Especially in some individual projects, the benefits of the training system can be demonstrated through the individual performance of athletes. Therefore, the training system needs to revolve around the best individual competitive ability of athletes [3]. In addition, for some excellent athletes who have already formed stable competitive abilities, how to maintain their stable development of competitive abilities and delay decline is also a problem that needs to be solved [4]. With the rapid improvement of athletic performance, the increasingly fierce competition in large-scale competitions has put forward higher requirements for the competitive ability of athletes. In order to enable athletes to compete for a place in high-level competitions, it is necessary to pay attention to how the athlete training system can be improved and evolved. Therefore, understanding the mechanism of the evolution of the athlete training system has high theoretical value [5]. The growth process of athletes is a continuous and complete time process, and the ultimate goal of athlete training is to achieve the ideal state of the athlete's organism to ensure the achievement of high-level sports results. Therefore, timely capturing and grasping the characteristics of different stages in the athlete training system and making reasonable diagnoses of the athlete's competitive ability status is of great guiding significance for coaches, athletes, medical supervisors, scientific researchers, administrative managers, and all participants in the training process [6].

*Sports Academy, Guangdong University of Petrochemical Technology, Maoming, 525000, China
†Youth League Committee, Guangdong University of Petrochemical Technology, Maoming, 525000, China (Corresponding author, 135800059640163.com)

2. Literature Review. The core of the athlete training system is to improve competitive ability. The evolution of competitive ability is one of the most significant characteristics reflecting the athlete training system [7]. The process of sports training is a gradual development of an athlete's competitive ability, as well as a process of promoting the gradual evolution of the athlete from the initial state to the target state. Each training stage is a continuation of the previous training stage and also the foundation of the subsequent training stage. The training stages at different levels constitute the entire competitive development process of athletes through a certain sequential arrangement and combination. Mou, Y. et al. mainly studied the application of embedded system technology in sports training and mobile interconnection control systems for bicycle equipment [8-9]. Tan, L. et al. proposed a mobile artificial intelligence terminal technology and developed and designed an athlete training process monitoring system based on C/S mode. GPS is used to obtain real-time location information of athletes and provide real-time guidance for them [10]. Nie et al. proposed a motion perception system using multi-agent technology in motion training. In the study, the relevant concepts and methods of agents were introduced in detail. In order to better simulate the changes in human joints during movement, various comparison methods were adopted. Among them, the experiment of the imaging analysis module was conducted through simulation of the simulation system. The experimental results indicate that the role of multi-agent in simulation systems is enormous [11].

Intelligent management of sports information generated in sports training venues such as schools, gyms, and gyms can further improve the management efficiency of the sports training process. At present, the management of sports training information is mainly manual, supplemented by simple computer management. Although it can reduce workload, it has not fundamentally changed, and there is still a lot of room for improvement. Based on this, in order to reduce the workload of manual management, the author proposes an intelligent management system for sports training information based on mobile devices.

3. Method.

3.1. Sports training information mobile storage server. The management system designed by the author includes two types of servers: One is a real-time information storage and collection server, and the other is a historical information storage and processing server. The author selected the Weifu COP2000 computer as the core platform and arranged it with various unit components in a structural model [12]. When the concurrency is less than 100, the server's built-in RAID card can be used to connect the hard disk; When the concurrency exceeds 100, it is necessary to configure the hard disk as a hardware disk array and use fiber channel storage to achieve high-performance information storage. The structure of the mobile storage server is shown in Figure 3.1.

3.2. Mobile information collection sub sites. The sub site is the middle layer of the information management system, which communicates with the main site terminal to collect and forward telemetry and remote adjustment data. The mobile sub site designed by the author follows the sports training process in real-time, collects sports training information through monitoring components in the sub site, and transmits it to the main site [13]. Choose a CISCO2509 model router and communicate with the main site through an SDH 2M+model router. The number of mobile sub sites should be selected according to the actual needs of users, and should be able to meet the real-time data collection and transmission needs within the jurisdiction of each sub site. If some sports training venues have a large scale and require a lot of real-time recording information, two or more mobile sub sites can be equipped for the same user at the same time to ensure the smooth operation of the site.

3.3. Information intelligent processor. Intelligent processors are mainly used for collecting data and receiving external information. The information resources involved in sports training are very complex, and can only be accessed and applied after integration and processing. In the sports training information management system, the author designs an information intelligent processor with different ports on the outside of the processor, responsible for different functions. The author chooses the information intelligent processor model S3C6410, and its framework is shown in Figure 3.2.

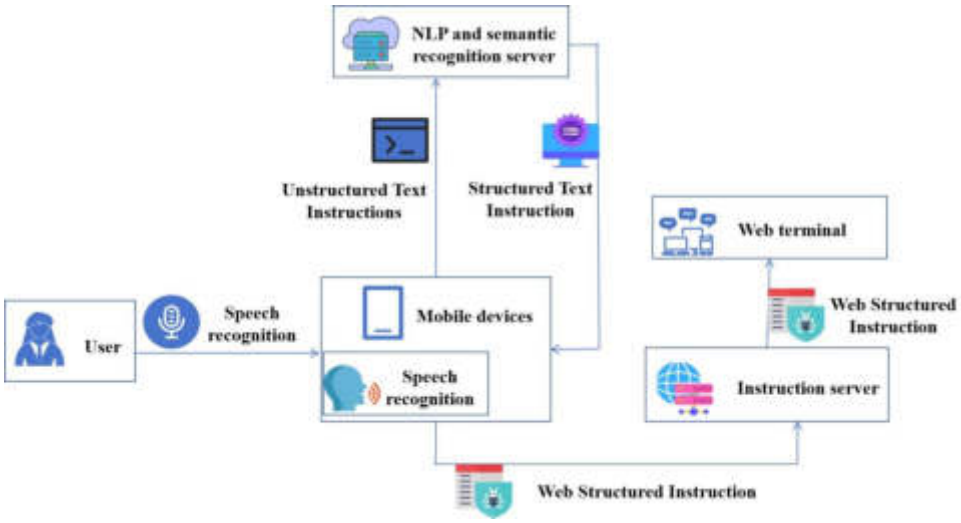


Fig. 3.1: Mobile Storage Server Structure

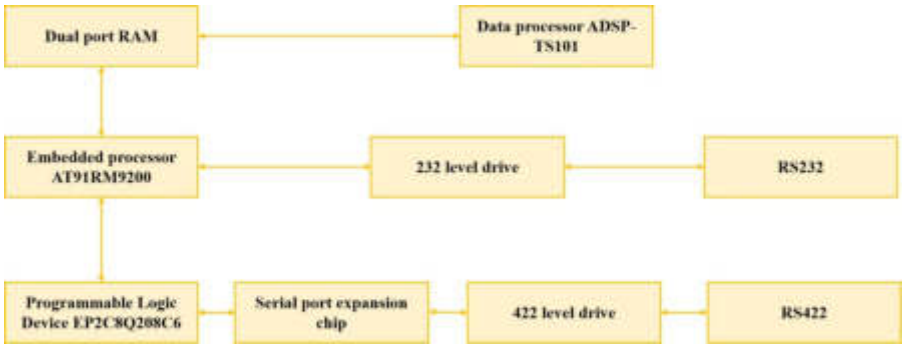


Fig. 3.2: Framework diagram of information intelligent processor composition

3.4. Software design of intelligent management system for sports training information . The receiving end of the software system needs to collect the generated sports training information in real time and classify the information preliminarily based on its characteristics [14]. The calculation formula for information category y is

$$y = \frac{1}{n}\sum_{i=1}^{n}||t - Y|| \tag{3.1}$$

In the formula, n represents the number of sports training information features from different sources; t represents a time variable; Y represents non intersecting features in information data. All information is grouped in the form of user name and ID number, so that the user can continue to register his/her information when he/she performs sports training again. When a user inputs their information again, the system will retrieve all their past information from the background and display key information on the front-end interface [15].

Extract the sports training information stored in the system according to the different users, analyze the content similarity of the sports training information, sort all the sports training information of the user based on the time when the information is recorded, and obtain the corresponding first similar sequence [16]. Integrate the first similar sequence to obtain the second similar sequence of the user, analyze the similarity

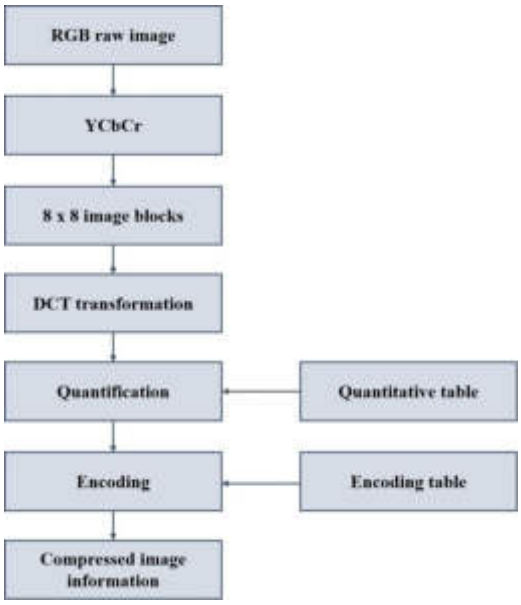


Fig. 3.3: Video monitoring and image processing of the entire process of athlete training

relationship between the first sequence and the second sequence, and each pair of sports training information with co-occurrence relationship is two pieces of information belonging to the same user [17]. Connect two pieces of information and calculate their information similarity. If the similarity between the user and training information is , the process of processing this data information can be expressed as:

$$m_A = \frac{1}{1-x}\sum_{A_m} m_i(A_m) \tag{3.2}$$

In the formula: A represents user identity; x represents the number of times information is processed; Am represents the similarity process quantity of the sports training information; represents an orthogonal set of data information. The calculated result is the final information similarity of the user. By calculating the data information similarity for information processing, intelligent management of sports training information is completed.

3.5. Monitoring video collection throughout the training process. The author combines fuzzy region information fusion and block feature matching to capture video footage of the entire process of athlete training, and adopts the method of image block fusion to capture and fuse images for video monitoring of the entire process of athlete training [18]. As shown in Figure 3.3:

The author used a grouped packet detection method to analyze the ambiguity of the entire training process video monitoring. Based on the regional distributed fusion method, the edge contour feature points of the athlete training process video images were analyzed, and the three-dimensional pixel distribution set for the entire training process monitoring was obtained as follows:

$$T' = \frac{\partial\phi}{\partial t(1-\theta)}\sum_{i=1}(\lambda_i e_i^{LBF} - \mu\lambda_i e_i^{LGF}) \tag{3.3}$$

Among them, ∂, θ, ϕ represents the three-dimensional pixel distribution parameter, λ_i represents the spatial distribution set vector, and e_i^{LHF}, e_i^{LCF} represents the correlation degree of the spatial distribution area in the athlete training video. The expression is as follows:

$$e_i^{LBF} = \int_{\Omega} \lambda_i(y-x)|f(x) - f_i(x)|^2 dx \tag{3.4}$$

$$e_i^{LGF} = \int_{\Omega} \lambda_i(y-x)|f(y) - f_i(x)|^2 dy \tag{3.5}$$

The author combines three-dimensional pixel distribution and corner feature information distribution to perform trajectory tracking and classification detection throughout the training process [19]. Through the collection of three-dimensional information from the video images of the entire athlete training process, assuming a two-dimensional image is I (x, y), the monitoring video acquisition model obtained by using Gaussian search function (DOG) is:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-1} + (x^2 + y^2)/2\sigma^2 \tag{3.6}$$

$$L(x, y, \sigma) = G(x, y, \sigma) \times \frac{I(x, y)}{e^2} \tag{3.7}$$

Among them, σ is the scale space factor, which is used for video image fusion processing throughout the training process.

3.6. Video surveillance feature extraction. With the development of machine vision information recognition technology, the author adopts image vision monitoring recognition method to monitor the entire training process of high-level swimmers. Using optimized image processing technology, a video monitoring and image analysis model for the entire training process under machine vision is established. By sampling video features and extracting information from images throughout the training process, combined with motion video tracking recognition, the ability to monitor and analyze the entire training process of high-level swimmers is improved, which is of great significance for promoting the improvement of swimming training level. Based on the results of video monitoring, collection, and fusion processing obtained from the entire process of athlete training, an RGB feature decomposition model of the video image is established. By calculating the RGB feature space components of the entire training process video, the RGB feature transformation relationship of the athlete training process video is obtained:

$$Y = \sum_{x \neq y} C_x + C_y - \frac{1}{2} \tag{3.8}$$

In the formula, C represents the RGB spatial component, and x and y represent the spatial distribution coordinates. Perform lossless transformation and feature compression on the RGB images of the entire training process video of athletes, and obtain the transformation formula:

$$F(x, y) = \frac{C_x C_y}{8} [\int_{x,y=0} (cos \frac{(2x+1)xy\pi}{16})^2 dx] \tag{3.9}$$

Among them, (x, y) is the image coordinate point.

Assuming that the spatial state feature distribution set of the entire training process video image of athletes follows a normal distribution, that is $n \in N(0, \sigma_n^2)$, where σ_n^2 is the matching feature point of the entire training process video image, the feature variable analysis function of the athlete training video image is obtained [20]. Its expression is:

$$X = \sigma_n^2 \sqrt{\frac{x_{i0} - y_{i0}}{\delta_i} + T' \sum_{i=1} F(x, y) + 1} \tag{3.10}$$

Among them, T' represents the fusion amount of video surveillance throughout the training process, and δ_i represents the entire training process.

The author conducted information monitoring and recognition on the collected video images of the entire training process, and obtained feature variable analysis results through the above analysis. Combined with statistical classification methods, key feature points of high-level athlete training video images were extracted for image classification and recognition.

The process of extracting features from video surveillance images is shown in Figure 3.4.

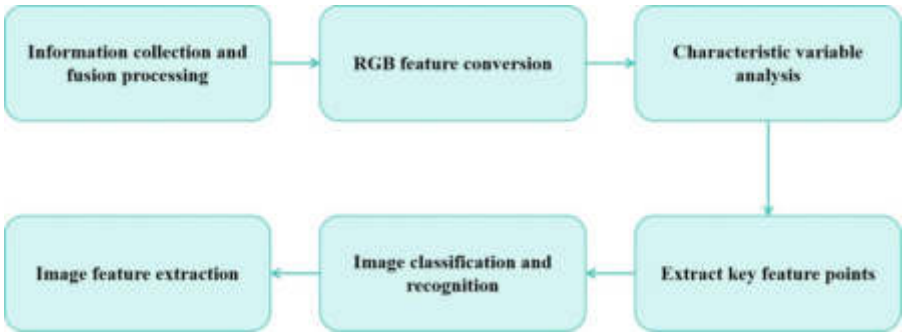


Fig. 3.4: Feature extraction process for video surveillance

3.7. Training 3D Reconstruction. Based on the information recognition of athlete training video action images, combined with the key action feature point extraction model of the entire training process video images, the partition block matching function is obtained as follows:

$$\widetilde{t}(x) = 1 - min_{y\in\Omega(x)}(\frac{I^C(y)}{A^C}) \tag{3.11}$$

Among them, $I^C(y)$ represents the local trajectory of the video monitoring of the entire process of athlete training, and A^c represents the scale information of the video image of the entire process of athlete training.

By using the method of local ambiguity detection, a parameter information analysis model for the entire process of athlete training video images is established, which is expressed as:

$$d(x,y) = \sum_{i=1} d_i(x+1,y) + d_i(x-1,y) + \frac{d_i(x,y+1)}{d_i(x,y-1)} \tag{3.12}$$

In the formula, x and y represent the three-dimensional parameters of image information.

By combining the method of calculating fuzzy frame difference sequences, the three-dimensional decomposition of athlete training action feature points is carried out to obtain the three-dimensional distribution function of the training action trajectory, which is described as:

$$\theta(x,y) = d(x,y) + \sqrt{\frac{L(x-1,y)}{L(x,y+1)} - 1} \tag{3.13}$$

Among them, L represents the fuzzy frame difference sequence coefficients and.

Based on the above analysis, combined with the distribution characteristics of video images, a three-dimensional reconstruction model of athlete training actions is established, which is expressed as:

$$G(x,y,t) = r * \frac{n}{u(x,y,t)} + \sqrt{u(x,y,t)+1} \tag{3.14}$$

Among them, r represents the three-dimensional sequence path ratio of video images, and n represents the distribution feature coefficient of video images.

In summary, a three-dimensional reconstruction model of athlete training actions was obtained, and the training action information was fused based on the three-dimensional reconstruction results.

Based on the three-dimensional reconstruction model of athlete training actions, a fusion matrix of training action information is obtained. Combined with the local fusion degree of video images throughout the training process, a video tracking function is constructed for the entire training process, which is expressed as:

$$H = G(x,y,t) + \sum_{i=1} \frac{L(x,y,t) + u(x,y,t)}{2r} - 1 \tag{3.15}$$

Table 3.1: Parameters of Test Equipment

The name of the parameter	ipad	mobile phone
Memory/GB	4	4
storage/GB	128	128
Running memory/GB	8+128	8+128
operating system	Android	Android

Table 4.1: Comparison Results of Packet Loss Rate

Equipment serial number	The packet loss rate of traditional information management systems/%	Packet loss rate/% of information intelligent management system based on mobile devices
Tablet 1	0.040	0.015
Tablet 2	0.061	0.021
Tablet 3	0.052	0.011
Tablet 4	0.045	0.012
Tablet 5	0.063	0.010
Mobile phone 1	0.074	0.023
Mobile phone 2	0.064	0.031
Mobile phone 3	0.075	0.021
Mobile phone 4	0.081	0.034
Mobile phone 5	0.066	0.033

Based on the above analysis, the trajectory recognition of the entire training process for high-level athletes is:

$$p(x,t) = \lim_{\Delta x \to 0}[\frac{\sigma^{\Delta x}}{\Delta x} * (u + \Delta u)] - p(\omega_i) \tag{3.16}$$

Among them, Δu is the associated pixel point for global threshold segmentation of the entire training process video image, σ is the gray feature of the entire training process video image, and $p(\omega_i)$ is the edge feature distribution set for tracking and recognition of the entire training process video.

3.8. System testing. Prepare 10 mobile devices with the same configuration, 5 tablets and 5 phones each. To ensure the smooth progress of the experiment, a comparison was made between traditional information management methods and the author's proposed intelligent management system for sports training information based on mobile devices. The parameters of the testing equipment are shown in Table 3.1.

4. Results and Discussion. Tablets and mobile phones were tested separately to receive sports training information from 2000 resource packets and record the number of lost packets. The experiment was repeated three times, and the average of the three experiments was taken to calculate the system packet loss rate. The results are shown in Table 4.1. According to the data in Table 4.1, after calculation, it can be concluded that for traditional information management systems, the average packet loss rate for tablet device users is 0.0534%, and the average packet loss rate for mobile device users is 0.0732%. The data on packet loss rate is relatively high; For the intelligent management system of sports training information based on mobile devices, the average packet loss rate for tablet users is 0.0144%, and the average packet loss rate for mobile device users is 0.0300%. By comparison, it can be seen that the packet loss rate of information intelligent management systems based on mobile devices is lower. It can be seen that the intelligent management system for sports training information based on mobile devices is more comprehensive in collecting and processing user data information, and can adapt to different groups of people.

5. Conclusion. The author proposes the design of an athlete training process monitoring and feedback system based on mobile intelligent devices. In response to the problem of incomplete sports training information

management, the author proposes an intelligent sports training information management system based on mobile devices. Compared with traditional information management systems, this system significantly reduces packet loss rate and can comprehensively record user sports training data information, providing scientific guidance for intelligent management of sports training information.

REFERENCES

[1] Nesdoly, A. , Ray, L. , Khurana, R. , & Mchugh, T. L. F. . (2022). Pushing for change: a qualitative study of the experiences of elite athletes during pregnancy. British Journal of Sports Medicine, 56(8), 452-457.

[2] Senner, V. , & Caine, M. . (2023). Physical (mechanical) models for sports equipment research, development and testing:. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 237(1), 3-6.

[3] Zhang, Q. , Zhang, X. , Hu, H. , Li, C. , Lin, Y. , & Ma, R. . (2022). Sports match prediction model for training and exercise using attention-based lstm network, 8(4), 508-515.

[4] Huebner, M. , & Ma, W. . (2022). Health challenges and acute sports injuries restrict weightlifting training of older athletes. BMJ open sport & exercise medicine, 8(2), 1372.

[5] Schick, S. , Chandler, K. , Kasprow, S. , Sisk, M. , Elphingstone, J. , & Wing, J. , et al. (2023). Gender disparities among professional team sports medicine physicians. Clinical Journal of Sport Medicine, 33(6), 648-651.

[6] Yun, Y. , Xiaojuan, H. , Wenping, G. , Jiliang, M. U. , Le, Z. , & Xiangdong, W. , et al. (2022). Human movement monitoring and behavior recognition for intelligent sports using customizable and flexible triboelectric nanogenerator, 65(4), 11.

[7] Szabo, S. W. , & Kennedy, M. D. . (2022). Practitioner perspectives of athlete recovery in paralympic sport:. International Journal of Sports Science & Coaching, 17(2), 274-284.

[8] Biscardi, L. M. , Miller, A. D. , Andre, M. J. , & Stroiney, D. A. . (2024). Self-efficacy, effort, and performance perceptions enhance psychological responses to strength training in national collegiate athletic association division i athletes. Journal of Strength and Conditioning Research, 38(5), 898-905.

[9] Mou, Y. . (2022). Embedded system for mobile interconnection control system of sports training cyclists. International Journal of Embedded Systems(3), 81(1), 211-219.

[10] Tan, L. , & Ran, N. . (2023). Applying artificial intelligence technology to analyze the athletes' training under sports training monitoring system. International Journal of Humanoid Robotics, 20(06), 1309.

[11] Nie, D. , & Liu, J. . (2022). Application of multiagent technology in intelligent distributed sports training simulation system. Wireless Communications and Mobile Computing, 72(244), 493-501.

[12] Gabbett, T. . (2023). Load management: what it is and what it is not!:. Sports Health, 15(4), 478-478.

[13] Cédric YM Morio, Bouten, L. , Duraffourg, S. , & Delattre, N. . (2022). A multidisciplinary approach to the engineering of footwear cushioning: a practical example of gym training shoes:. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 236(1), 5-16.

[14] Terrell, S. L. , & Ficquette, P. . (2023). Exploring training strategies to optimize court performance in older pickleball athletes. Strength and conditioning journal, 45(1), 1-12.

[15] Chai, Y. , Wang, G. , Zhang, C. , Zhu, G. , Wei, F. U. , & Guo, Z. . (2022). Satellite-earth integration intelligent resource management and control architecture of giant sensing constellation. Space-Integrated-Ground Information Networks, 3(3), 13-22.

[16] Fan, J. , Zhang, M. , Sharma, A. , & Kukkar, A. . (2022). Data mining applications in university information management system development. Journal of Intelligent Systems, 31(1), 207-220.

[17] Lu, Z. , Wang, W. , Florez-Arango, J. F. , Seo, J. H. , Hamilton, D. K. , & Wells-Beede, E. . (2023). A framework to develop an immersive virtual reality simulation tool for postpartum hemorrhage management nurse training. Nursing Education Perspectives, 45(3), 189-191.

[18] Darcy, S. , Paramio-Salcines, J. L. , Kitchin, P. J. , & Walters, G. . (2022). Exploring the accessibility of sport stadia for people with disability: towards the development of a?stadium accessibility scale (sas). Sport, Business and Management: An International Journal, 12(1), 93-116.

[19] Meyer, R. , Loncar, L. R. , Jensen, E. , Raja, A. , Tunis, B. , & Moreland, M. L. , et al. (2023). The role of ultrasound in the management of ankle sprains and a clinically relevant geisinger ankle sprain sports ultrasound protocol. Current Sports Medicine Reports, 22(9), 320-327.

[20] Salimi, M. , & Tayebi, M. . (2022). Developing a model for the success of information systems in sports organisations. International Journal of Business Information Systems, 39(1), 76.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 28, 2024
Accepted: Jul 14, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 855–863, DOI 10.12694/scpe.v26i2.4046

INNOVATIVE APPLICATIONS OF MULTIMODAL SENSING TECHNOLOGY IN SPORTS REHABILITATION ASSESSMENT AND TRAINING

CONGCONG MA*, KUN JIANG†, QIAN ZHAO‡, DECAI NI§, AND JIADONG ZHANG¶

Abstract. This study explores an innovative approach to evaluating the training effectiveness of lower limb exoskeleton robots by integrating multiple data types, including electrophysiological signals and kinematic measurements, to assess patients' walking ability quantitatively. Through precisely defined synergistic indicators, this method effectively combines different types of data and dramatically improves the efficiency and accuracy of rehabilitation assessment. First, the patient's lower extremity electro myoelectric activity and movement data were recorded while walking with exoskeleton assistance. Secondly, the key EMG and kinematic features are analyzed and extracted by a collaborative quantization algorithm based on the theory of muscle cooperative work. Then, this information from different levels is integrated to build a feature fusion model, based on which the lower limb motor function score is calculated. The development of multi-channel lower limb exoskeleton human-computer interaction technology for sports training can provide a variety of standardized and standardized auxiliary training for athletes and meet the human body's multi-sensory immersion. A multi-step, multi-degree-of-freedom motion planning algorithm is proposed to reproduce various activities the human body requires. Secondly, the lower extremity-oriented multi-modal human-computer interaction technology is studied to realize the display and guidance of standard movement in information space on the virtual reality competition training simulation platform. Build a motion database to assist and correct basic motion in physical space. The experimental results showed a significant correlation between the extracted myoelectric and kinematic synergistic features and the clinical evaluation tools, with the correlation coefficients reaching 0.832 and 0.859, respectively. The fusion features show a stronger correlation when applying the K-nearest neighbor (KNN) algorithm. This evaluation method cannot only optimize the training strategy of the exoskeleton robot according to the results but also provide the possibility to realize the "man in the ring" mode of evaluation and training simultaneously.

Key words: Rehabilitation assessment; Muscle coordination; Mode fusion; Machine learning; Stroke; Multimodal sensing technology.

1. Introduction. With an aging population and increasing incidence of chronic diseases, lower limb dysfunction has become a severe challenge in the field of global public health. Lower limb rehabilitation is related to individual quality of life and is an essential part of the rational allocation of social medical resources. Traditional rehabilitation assessment and training often rely on the experienced judgment of professionals and a single assessment means, which limits the accuracy and individuation of rehabilitation effects. Therefore, exploring more scientific and efficient lower limb rehabilitation assessment and training methods has become one of the hot spots in rehabilitation engineering.

The development of multimodal sensing technology provides a new opportunity to solve this problem. Multimodal sensing technology refers to integrating a variety of sensors, such as electromyography (EMG), accelerometer, gyroscope, pressure sensor, etc., to capture the electrophysiology, kinematics and dynamics of the human body simultaneously. This technology can provide more prosperous and detailed data in time and space to achieve a more comprehensive and in-depth understanding of the human movement function. In lower limb rehabilitation, applying multi-modal sensing technology can realize the accurate and quantitative assessment of patients' walking ability and provide a scientific basis for developing personalized rehabilitation training plans.

The paper [1] shows that EMG based on a single pathway can automatically evaluate the degree of damage

*Sports Department, Shaanxi Fashion Engineering University, Xi'an 712046, China (Corresponding author, 17754923321@163.com)
†Sports Department, Shaanxi Fashion Engineering University, Xi'an 712046, China
‡Sports Department, Shaanxi Fashion Engineering University, Xi'an 712046, China
§Sports Department, Shaanxi Fashion Engineering University, Xi'an 712046, China
¶Sports Department, Shaanxi Fashion Engineering University, Xi'an 712046, China

in stroke patients and produce quantitative grading indicators corresponding to clinical scores. The researchers developed a portable exercise test bed. A comprehensive evaluation method of upper limb motor function was constructed by using an extreme learning machine. This method is suitable for clinical and home environments. In the paper [2], a neurorehabilitation evaluation method was constructed using the surface EMG signal of limbs. The paper [3] puts forward a new idea of quantitative evaluation of motor function using a mechanical arm's strength and trajectory characteristics. The research results will enhance the effectiveness and credibility of rehabilitation evaluation, expand its scope of application, and promote the broad application of rehabilitation robots. In addition, due to the rapid development of computer science and technology in recent years, machine learning has been paid more and more attention [4]. Current research on action evaluation based on machine learning usually divides behaviors into two categories: right and wrong. The result is only 0 or 1. This method can only be qualitatively described, unable to achieve a quantitative analysis of movement quality at each level and unable to identify incremental changes in patients' imagination. Moreover, the modeling process is relatively complicated, requiring specific training samples.

This project intends to research multi-source sensing data fusion of lower limb exoskeleton systems. In this project, real-time acquisition of lower extremity surface electromyography and movement information was carried out. Then, the patient is comprehensively evaluated from various aspects, such as movement and electrophysiology, to guide individual rehabilitation exercises and carry out guided rehabilitation exercises for the patient. A multi-step, multi-degree-of-freedom motion planning algorithm is proposed to reproduce various activities the human body requires [5]. Secondly, the lower extremity-oriented multi-modal human-computer interaction technology is studied to realize the display and guidance of standard movement in information space on the virtual reality competition training simulation platform. Build a motion database to assist and correct basic motion in physical space. Study the relationship between evaluation criteria and scale [6]. Finally, the score data are imported into the data fusion model to obtain the score of the machine test. The research results of this project will lay a foundation for applying the lower limb exoskeleton robot in stroke.

2. Lower limb exoskeleton multi-modal interaction system.

2.1. System Architecture. This project draws on the next generation of information-physical fusion technology (HCPS) to study the multi-modal interaction technology for the human body. In this process, the communication between man and machine and the control of the path is required, and it is required to accurately understand and implement the person's intention [7]. At the same time, it can meet the purpose of efficient information transmission and feedback. Under the background of the rapid development of the new generation of information technology, the digital intelligent technology represented by virtual reality and extended reality has laid the foundation for optimizing man-machine collaboration. Compared with the traditional CPS, the new generation of urban health management is a human-centered ternary system [8]. Its goal is to achieve cooperation and co-prosperity and guide the improvement of work efficiency. Figure 2.1 depicts the hierarchical structure of HPCS (image cited in A Triple Human-Digital Twin Architecture for Cyber-Physical Systems). It controls the entire physical system by sensing the user's presence, views and behaviors. The information system is added to the human-computer interaction process through the interactive cycle of pedestrian, information, and entity models. In this way, the profound combination of the trinity of "man-machine-thing" is gradually achieved.

2.2. System Functions. This project will study human-machine integration technology for lower limbs. It consists of three modules: a virtual reality simulator, a motion database, and a lower limb exoskeleton (Figure 2.2). Establish the information interaction model between the human endoskeleton and VR. The combination of reality and virtual provides more realistic movement training guidance for the human body to improve the training effect. The general applicability of the system is improved by collecting 4 motion nodes and transforming them into the limb movement of the training object [9]. The lower limb exoskeleton robot proposed in this project has the advantages of a simple structure, good adaptability and a flexible response mechanism. It can provide a variety of force feedback methods for all kinds of moving objects. The virtual body controller of the upper computer and the actual model of the lower limb exoskeleton robot arm of the slave foot constitute the whole control system. It mainly includes a virtual reality competitive training simulator and action data set design. Through the detection of virtual objects, the trajectory information of virtual objects

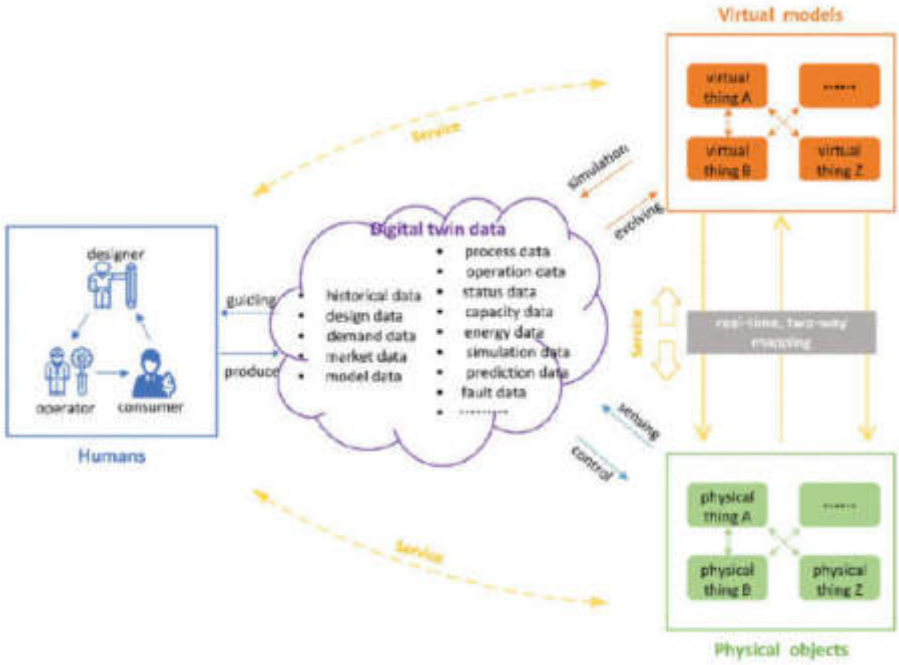


Fig. 2.1: *Human-information-physical system structure.*

is fed back to the actuator [10]. Then, the lower limb exoskeleton robot is supported by force perception, and VR simulators enhance the vision and hearing so students can obtain standardized and immersive exercise experiences in movement training.

Students wear an exoskeleton robot, virtual reality headset and motion capture tracker, and follow the instructor to do basic movements in virtual reality. Students can complete the standard sports training according to their own wishes through pre-programmed programming in the virtual simulation system. The lower limb exoskeleton robot arm adopts a force feedback device to realize the dynamic perception of the body [11]. The system ensures the standardization of the body and improves the training effect. The system helps students carry out basic competitive sports training using multi-sensory information such as vision, hearing and touch.

2.3. System Hardware. The system hardware completed the pressure feedback, including the motor drive and vibration two-force feedback. The hardware control scheme is given in Figure 2.3. The system comprises an STM32 development board, HC-05 Bluetooth module, MPU6050 module, vibration motor, EPOS drive, MAXON-Re35 motor, etc.

The athletes' hearing and force perception are enhanced by VR technology and lower limb exoskeleton technology. A method based on motor drive and vibration feedback is proposed to realize multi-mode human-computer interaction [12]. The first is to set the relevant parameters, and then put on the robotic arm and helmet on the leg. The Angle marker sensor is used to track the movement of the current trainer. The current motion orientation of the trainer and the changed reference motion orientation are determined. Figure 2.4 shows the multi-modal interactive execution process of the lower extremity exoskeleton.

3. Quantitative method of lower limb movement based on muscle synergy theory.

3.1. Muscle collaborative extraction. Muscle coordination, as an optimal central regulation mode, is clinically significant for recovery after a stroke. Studies have shown a high degree of consistency in healthy people's coordination of body movements [13]. This project seeks the mechanism of muscular coordination as universal neuromuscular coordination and to provide the temporal characteristics of the active-induced sEMG quantified muscle coordination. The existing decomposition methods mainly include factor analysis,

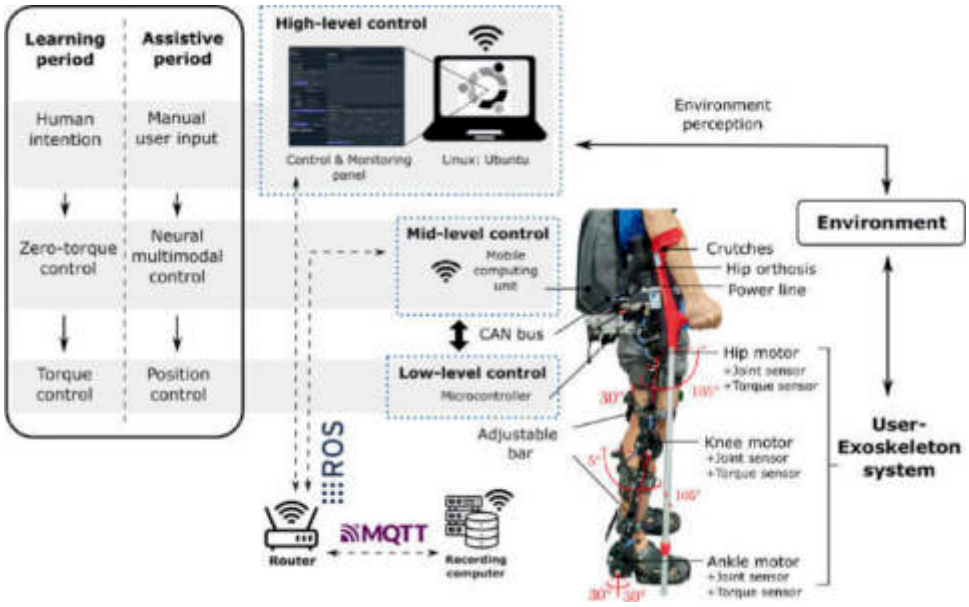


Fig. 2.2: *Framework of multi-modal interaction system for lower extremity exoskeleton.*

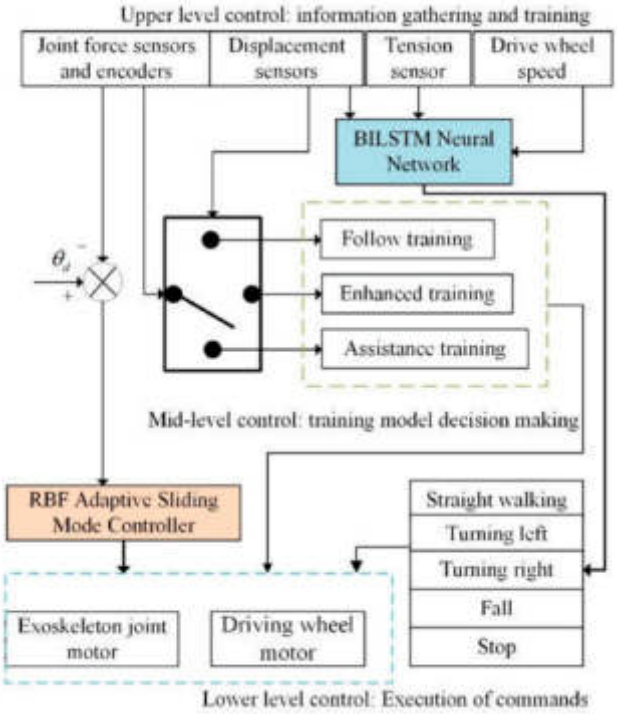


Fig. 2.3: *Hardware control block diagram of lower extremity exoskeleton multi-mode interaction system.*

non-negative matrix decomposition, principal component analysis, etc., which decompose multi-channel sEMG data and obtain fewer samples with high characterization ability. The basic mathematical model of code

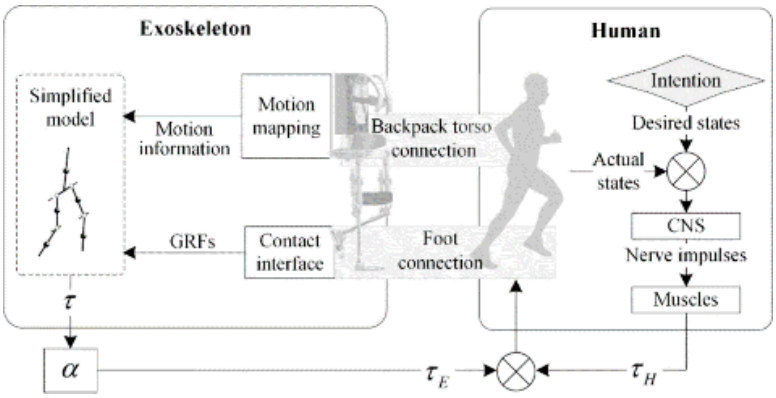


Fig. 2.4: *Multi-modal interactive execution flow.*

composition is expressed in formula (3.1):

$$Y_{n\cdot\tau} = R_{n\cdot\lambda} \times J_{\lambda\cdot\tau} = [R_1, R_2, \cdots, R_\lambda] \times [J_1, J_2, \cdots, J_\lambda]^T = \sum_{i=1}^{\lambda} R_i J_i + D \tag{3.1}$$

It is known that the matrix $Y_{n\cdot\tau}$ can be divided into two parts: one is the coordination unit $R_{n\cdot\lambda}$, and the other is the excitation factor $J_{\lambda\cdot\tau}$. Where D is the error matrix that can be ignored. Where n represents the number of channels, τ represents the number of samples, λ represents the number of cooperative elements, T represents the exchange number, and i represents the number of matrices.

The traditional joint extraction method uses matrix decomposition to map the original sample to the lower dimension while maintaining the characteristic. Principal component analysis (PCA) is a typical feature extraction method based on probability distribution, which can effectively deal with samples containing redundant information [14]. It has significant application value in pattern recognition, machine vision, etc. By solving A linear orthogonal transformation of R matrix, the obtained data U is converted into an implicit low-dimensional matrix C :

$$C = R^T U \tag{3.2}$$

In equation (3.3), the method of obtaining each component of matrix C is expressed:

$$c_{ij} = r_i^T u_j \tag{3.3}$$

where c_{ij} is the element of a low-dimensional matrix C, r_i is its orthogonal vector, u_j is the eigenvector, and T is the commutation sign. The loss function F is represented by the maximum value of the converted difference to obtain the orthogonal vector r_i :

$$\begin{aligned} F &= \max \frac{1}{N} \sum_{i=1}^{s} \sum_{j=1}^{N} \left(r_i^T u_j - r_j^T \bar{u} \right)^2 = \max \sum_{i=1}^{s} r_i^T B r_i \\ &\text{s.t. } r_i^T r_i = 1 \end{aligned} \tag{3.4}$$

where π represents the mean of u, and B represents the covariance matrix of U. The eigenvector refers to the local maximum eigenvalue on r_i. The corresponding eigenvalue δ_i and corresponding eigenvector v_i are obtained by the HSVD method, and then the orthogonal matrix $\tau \cdot \lambda$ is obtained. Here, U is the trained $\tau \cdot n$ dimension EMG signal (τ is 1 sample number of gait cycles, n is 8 channels), and R is the number of muscle coordination units in the human body (λ is the number of muscle coordination units). C is data converted to dimension $\lambda \cdot y$. The eigenvalues and eigenvectors after singular value decomposition are used to measure the similarity. When walking, the muscle group of the leg can be divided into 5 cooperative modes, so the cooperative unit λ is represented by 5 .

3.2. Collaborative Quantification. $Y = \{y_1, y_2, y_3, \cdots, y_n\}$ is the action data, where n is the number of extracted steps $y_i (i = 1, 2, \cdots, n)$, and is the data of a step. Through principal component analysis, the convector v_i, u_i is obtained, corresponding to the eigenvector δ_i, ξ_i. When two objects have similar motions in higher dimensions, their orbits should be similar. So, the sequence must also be the same when the corresponding action is similar. A joint decomposition method based on principal component analysis is proposed to realize the joint analysis of motion parameters [15]. In this way, the retrieval and classification of motion parameters are realized.

$$\mathcal{S}_i = \max_{\mathrm{din}(Z)} = \min_{a \in z} \frac{\|Bu\|_2}{\|u\|_2} \tag{3.5}$$

If two actions are similar, their conformal vectors v_1 and u_1 should be roughly parallel. There is $|v_1 \cdot u_1| = |v_1|\|u_1\|\cos(\alpha) \approx v_1 \| u_1| = 1.\alpha$ is the Angle of the two common vectors v_1 and u_1. The sEMG, and action sequences of different training segments were analyzed by weight similarity measure to evaluate patients' recovery status quantitatively.

$$\chi(A, Q) = \frac{1}{2} \sum_{i=1}^{n} \left(\left(\frac{\mathcal{S}_i}{\sum_{i=1}^{n} \mathcal{S}_i} + \frac{\eta_i}{\sum_{i=1}^{\pi} \eta_i} \right) |v_i \cdot u_i| \right) \tag{3.6}$$

Where n is the eigenvalue number, A is the baseline data set by the reference control group, and Q is the patient data. $\mathcal{S}_i, \eta_i$ is the i eigenvalue of the reference and experimental data. It corresponds to the i-cooperation vector v_e, u_i. In this way, only the convectors of the two action matrices and their corresponding eigenvalues are obtained without the influence of other action information [16]. The algorithm can capture the similarity between λ conformal vectors and use corresponding eigenvalues for weight calculation. χ is in the range from o to 1 . The closer it is to 1 , the better the patient fits the control group.

3.3. Modal fusion evaluation model. The most significant difference in each stage is the cooperation mode and the movement mode of each independent part. It focuses on gradually transforming the overall rehabilitation process to a standard and complex movement mode to achieve the brain reassembly effect [17]. This method is challenging to detect if a single sensor is used for detection, so additional characteristics are needed to classify each specific action more deeply. A joint coordination vector G_i is constructed to verify the practicality of this collaboration feature, expressed by the following formula (3.7)

$$G_i = [D_i, \cdots, D_n, \Lambda_i, \cdots, \Lambda_n] \tag{3.7}$$

Where D_i is the neurophysiological characteristic of cooperation, Λ_i is the motor coordination characteristic, and n is the number of cooperative characteristics. A predictive score Z_i for the patient's lower limb recovery can be obtained using a guided machine-learning model:

$$Z_i = I\left(G_i\right) \tag{3.8}$$

The KNN method is used to train 5 neighbors. The eigenvalue space U is defined as an n-dimensional vector space of real numbers $(u_i, u_j) \in U, u_i = \left(u_i^{(1)}, u_i^{(2)}, \cdots, u_i^{(n)}\right), u_j = \left(u_j^{(1)}, u_j^{(2)}, \cdots, u_j^{(n)}\right)$ .H is expressed as a function of the distance of (u_i, u_j). Using the formula (3.9) in Euclidean geometry:

$$H\left(u_i, u_j\right) = \left(\sum_{i=1}^{a} \left| u_i^{(i)} - u_j^{(n)} \right|^2 \right)^{\frac{1}{2}} \tag{3.9}$$

Here l is the dimension of the eigenspace U. The least squares support vector machine is regarded as a multilayer feedforward neural network whose training model comprises 130 neurons. The quasi-Newton algorithm is used to optimize the weights and deviations, and the convergence efficiency of the algorithm is improved. Set the maximum depth of the RF to 5 . The radiation basis function is the core of the algorithm. Where Z_{i} is the result of the pattern, and does the evaluation value correspond to the result of the pattern, it is more detailed than the conventional category evaluation. From the 6 people, 60 cases were selected as experimental samples, and the other 4 cases were used as test data.

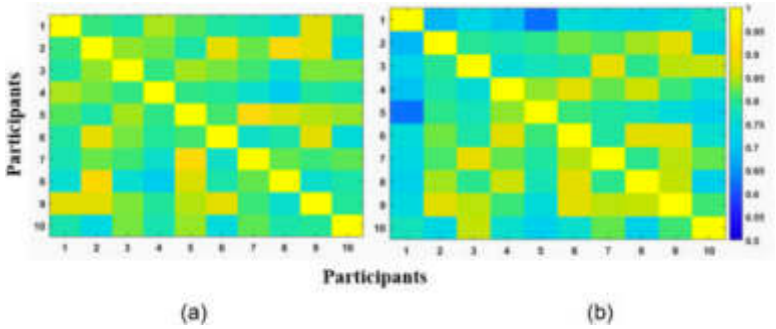


Fig. 4.1: *Weight comparison of synergies between healthy subjects and patients.*

4. Test results and analysis.

4.1. Comparison between healthy subjects and patients. The collaborative element weights of the healthy and patient sides were shown in Figure 4.1 for the BRS_3 stage subjects. This index ranges from o to 1 , indicating the strength of muscle activity from large to small. Due to the contraction of local muscles after stroke, the local myoelectric activity is reduced, so the weight of the affected limb is generally lower than that of the healthy side [18]. However, in the BRS_3 stage, the activation weight of the affected area of the left ventricle was significantly higher than that of the healthy area. This is due to increased limb strength caused by muscle spasms in the left ventricle.

The moving sample size tables BRS6 and BRS3 taken from the patient are shown in Figure 4.2. Patients in the BRS6 stage can walk independently, and their indicators are comparable to those of ordinary people. However, BRS3 stage patients often have instability due to limb pain and decreased ability to control movement. In particular, weakness of the hip flexors can significantly reduce the range of motion of the hip joint. The affected knee cannot fully flex and extend during both the upright and the rocking phases due to the weak flexor muscles of the knee flexor tendon, which prevents it from achieving normal hip flexion during both the upright and the rocking phases. Under normal circumstances, the maximum bending of the patient's foot occurs at the beginning of walking, but the degree of bending of the patient's knees is small, resulting in the patient's feet not being lifted from the ground. In addition, because the patient cannot fully lift the front ball of the affected side during the rocking phase, the initial contact appears to be on the flat ball of the foot rather than the heel.

The flexion and extension motion of the affected upper limb decreased significantly in the swaying period, and due to the lack of lower limb muscle strength, it could not independently complete its weight when standing. By studying sEMG and joint Angle, the cause of the limited range of limb movement after stroke. Grade 3 patients, due to their limited joint flexibility, will produce muscle convulsions and other symptoms, so their movements often end prematurely or deviate. The above two points can more directly reflect the shortcomings of patients in lower limb movement.

4.2. Research on the characteristics of the cooperation index.. The paper divided 10 patients into 4 different BRS stages. In Figure 4.3, there are significant differences in the characteristics of the joint index among patients with various BRS stages. Among them, the collaborative characteristics of SEMG appear more in Stage 6 and Stage 3. The S7 muscle twitching was more severe in BRS5 patients, so the cooperative properties of sEMG were very different from normal controls, with two abnormal values. During the same period, the data of the movement index were relatively stable. In terms of movement parameters, the data distribution of patients with the BRS3 stage is very different, and the variation is considerable compared with the average population, mainly because the symptom of patient S6 is that he is unable to speak in the right limb, and because his disease has only been over one month, there are a small number of movement samples. BRS3 stage patients have severe sports injury and spasmodic muscle tonia, and their motor and electromyographic coordination index is lower than usual. However, since the standard deviation of BRS6 stage patients was significantly lower than that of 3-5 stage patients, both sEMG and motor coordination index were relatively

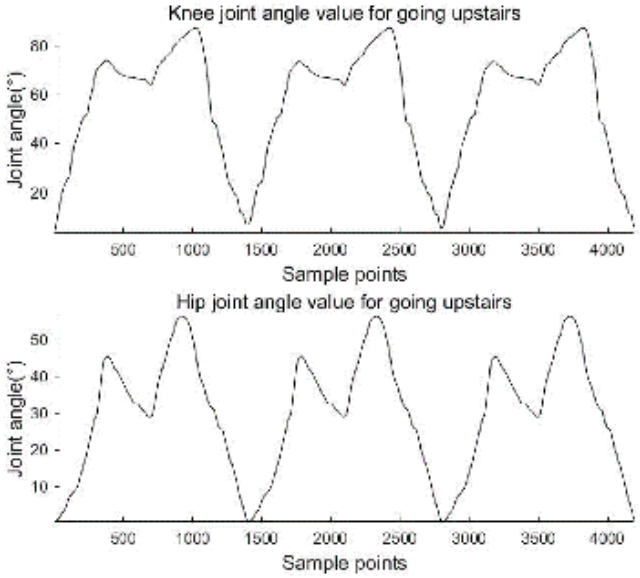


Fig. 4.2: *Comparison of subjects' lower extremity angles.*

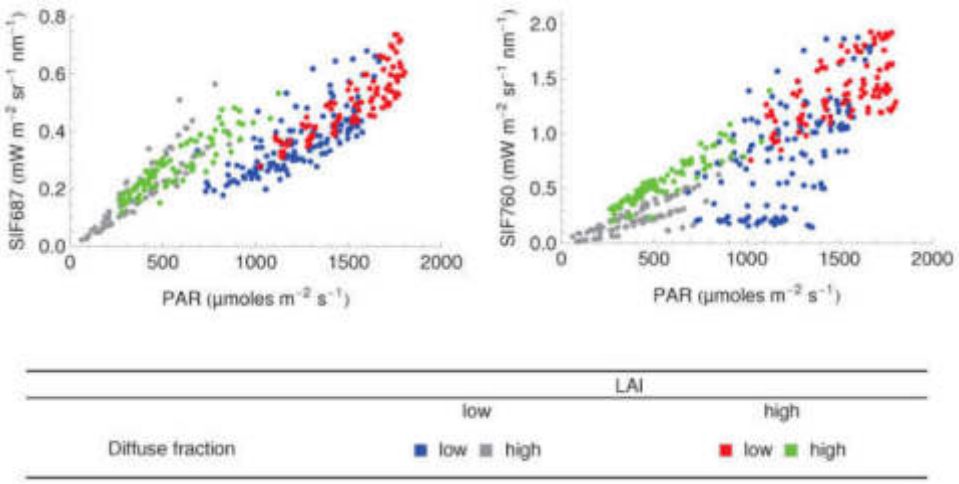


Fig. 4.3: *Data distribution of BRS staging coordination index.*

stable, suggesting that the muscle control function and Angle curve of BRS6 stage patients were relatively stable when walking.

5. Conclusion. This paper intends to construct a set of collaborative feature-based walking function assessment methods for stroke patients and combine this assessment method with BRS scale grading to verify the correctness of this method. Studies have shown that this scale is highly correlated with the conventional BRS scale. The research results can improve the accuracy of clinical evaluation and help update the motion trajectory of the lower limb exoskeleton in real time based on evaluation data. From the clinical level, the evaluation system proposed in this study can better meet the needs of newly admitted patients, provide evaluation indicators for newly admitted patients, and shorten the evaluation cycle of rehabilitation doctors. The research results of this paper provide a basis for rehabilitation doctors to make a reasonable rehabilitation plan.

Acknowledgement. University-level project of Shaanxi Fashion Engineering University , project name: Innovative application research of multi-mode Sensing Technology in fitness course – Taking Shaanxi Institute of Fashion Engineering as an example, project number: 2024JG082.

REFERENCES

[1] Sherry, N. S., Fazio-Sumrok, V., Sufrinko, A., Collins, M. W., & Kontos, A. P. (2021). Multimodal assessment of sport-related concussion. Clinical journal of sport medicine, 31(3), 244-249.
[2] Moral-Munoz, J. A., Zhang, W., Cobo, M. J., Herrera-Viedma, E., & Kaber, D. B. (2021). Smartphone-based systems for physical rehabilitation applications: A systematic review. Assistive Technology, 33(4), 223-236.
[3] Bhardwaj, S., Khan, A. A., & Muzammil, M. (2021). Lower limb rehabilitation using multimodal measurement of sit-to-stand and stand-to-sit task. Disability and Rehabilitation: Assistive Technology, 16(5), 438-445.
[4] Debnath, B., O'brien, M., Yamaguchi, M., & Behera, A. (2022). A review of computer vision-based approaches for physical rehabilitation and assessment. Multimedia Systems, 28(1), 209-239.
[5] Powell, M. O., Elor, A., Teodorescu, M., & Kurniawan, S. (2020). Openbutterfly: Multimodal rehabilitation analysis of immersive virtual reality for physical therapy. American Journal of Sports Science and Medicine, 8(1), 23-35.
[6] Samyoun, S., Islam, M. M., Iqbal, T., & Stankovic, J. (2022). M3sense: Affect-agnostic multitask representation learning using multimodal wearable sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(2), 1-32.
[7] Yang, L., Amin, O., & Shihada, B. (2024). Intelligent wearable systems: Opportunities and challenges in health and sports. ACM Computing Surveys, 56(7), 1-42.
[8] Capecci, M., Ceravolo, M. G., Ferracuti, F., Iarlori, S., Monteriu, A., Romeo, L., & Verdini, F. (2019). The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(7), 1436-1448.
[9] Liao, Y., Vakanski, A., & Xian, M. (2020). A deep learning framework for assessing physical rehabilitation exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(2), 468-477.
[10] Hysing, E. B., Smith, L., Thulin, M., Karlsten, R., Bothelius, K., & Gordh, T. (2019). Detection of systemic inflammation in severely impaired chronic pain patients and effects of a multimodal pain rehabilitation program. Scandinavian Journal of Pain, 19(2), 235-244.
[11] Seshadri, D. R., Magliato, S., Voos, J. E., & Drummond, C. (2019). Clinical translation of biomedical sensors for sports medicine. Journal of medical engineering & technology, 43(1), 66-81.
[12] Vieluf, S., Scheer, V., Hasija, T., Schreier, P. J., & Reinsberger, C. (2020). Multimodal approach towards understanding the changes in the autonomic nervous system induced by an ultramarathon. Research in Sports Medicine, 28(2), 231-240.
[13] Roche, J., De-Silva, V., Hook, J., Moencks, M., & Kondoz, A. (2021). A multimodal data processing system for LiDAR-based human activity recognition. IEEE Transactions on Cybernetics, 52(10), 10027-10040.
[14] Cortell-Tormo, J. M., García-Jaén, M., Ruiz-Fernandez, D., & Fuster-Lloret, V. (2019). Lumbatex: a wearable monitoring system based on inertial sensors to measure and control the lumbar spine motion. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(8), 1644-1653.
[15] Zhai, B., Perez-Pozuelo, I., Clifton, E. A., Palotti, J., & Guan, Y. (2020). Making sense of sleep: Multimodal sleep stage classification in a large, diverse population using movement and cardiac sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(2), 1-33.
[16] Ding, Y., Cao, Y., Duffy, V. G., Wang, Y., & Zhang, X. (2020). Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. Ergonomics, 63(7), 896-908.
[17] Hutabarat, Y., Owaki, D., & Hayashibe, M. (2020). Quantitative gait assessment with feature-rich diversity using two IMU sensors. IEEE Transactions on Medical Robotics and Bionics, 2(4), 639-648.
[18] Yunas, S. U., & Ozanyan, K. B. (2020). Gait activity classification from feature-level sensor fusion of multi-modality systems. IEEE Sensors Journal, 21(4), 4801-4810.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 28, 2024
Accepted: Jul 18, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 864–870, DOI 10.12694/scpe.v26i2.4047

INTELLIGENT OPTIMIZATION AND RECOMMENDATION SYSTEM DESIGN FOR PERSONALIZED TRAINING PROGRAMS FOR MARATHON ATHLETES BASED ON MACHINE LEARNING

LUXIA GUO*

Abstract. This research focuses on developing an innovative machine learning-based intelligent optimization and recommendation system for marathon runners' personalized training schemes. The system aims to provide accurate and dynamically adjusted training guidance for athletes through real-time monitoring, training effect evaluation and intelligent recommendation. First, the system uses advanced wearable technology to achieve real-time monitoring of multiple physiological and athletic data during athlete training, including but not limited to heart rate variability, lactate threshold, and gait analysis. This data forms the basis of a personalized training program. Secondly, the support vector machine (SVM) algorithm is used to evaluate the training effect of the collected data. Finally, the system combines individual characteristics and the historical performance of athletes and generates personalized training plans through optimization strategies such as support vector machines. This process not only considers the short-term training goals but also considers the long-term sports career planning. Through algorithm modeling and computer simulation, it is found that the system can realize the continuous optimization of the training scheme in the process of continuous iteration. The intelligent system significantly improves athletes' training efficiency and competition performance compared to traditional training methods. This study provides a new perspective and practice path for intellectualization in sports training.

Key words: Machine learning; Marathon; Personalized training; Training effect; Support vector machine.

1. Introduction. With the rapid development of science and technology, the application of machine learning in sports training is increasing daily, especially in optimizing and recommending personalized training programs for marathon runners. Marathon training not only requires athletes to have good physical strength and endurance but also needs scientific training methods and strategies to improve the training effect and competition performance. Traditional training methods often lack individualized adjustment for individual differences, and the introduction of machine learning technology makes the training program more accurately adapt to the specific situation of each athlete.

Real-time monitoring system plays a crucial role in marathon training. Through wearable devices such as GPS watches and heart rate monitors, coaches and athletes can obtain real-time training data, including key indicators such as movement trajectory, speed and heart rate [1]. These data help coaches understand the athletes' training status in time, adjust the training intensity, and prevent overtraining and sports injuries. Machine learning algorithms can process and analyze training data, extract useful information, and provide a scientific basis for athletes' training [2]. For example, by analyzing an athlete's historical training data, a machine learning model can predict an athlete's optimal training load and recovery time to optimize training schedules. Support vector machine (SVM) is a commonly used machine learning algorithm suitable for small samples of nonlinear and high-dimensional pattern recognition problems. In the marathon training effect evaluation, SVM can be used to distinguish the training data of elite athletes and ordinary athletes, identify the key factors affecting the training effect, and provide personalized training feedback for athletes.

There is a close relationship between marathon training real-time monitoring systems, machine learning and support vector machines [3]. Real-time monitoring systems provide the data source, machine learning algorithms process and analyze the data, and support vector machines are used to evaluate the training effects [4]. This combination can improve the science and effectiveness of training and realize the real-time monitoring and instant feedback of athletes' training status to achieve personalized training.

*School of Physical Education, Shanxi Vocational University of Engineering Science and Technology, Jinzhong 030600, China (Corresponding author, jtxytyz@163.com)

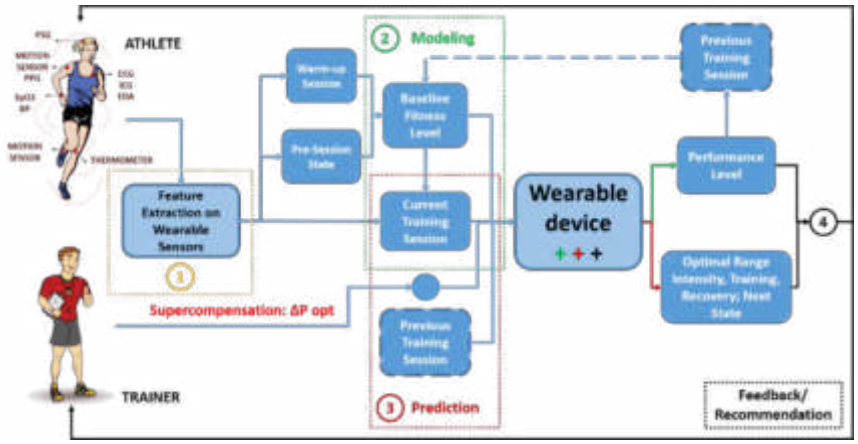


Fig. 2.1: *Architecture of real-time monitoring and guidance system for marathon runners during training.*

This paper aims to design an intelligent optimization and recommendation system based on machine learning for marathon runners' personalized training schemes [5]. The system will integrate the functions of real-time monitoring, data analysis and effect evaluation and verify the effectiveness and practicability of the system by establishing algorithm models and conducting simulation tests [6]. It is expected that the application of this system can significantly improve the training efficiency and competition performance of marathon runners while reducing the risk of sports injuries.

2. Design and establishment of a real-time monitoring system for marathon training.

2.1. Methods and basis for real-time monitoring of athletes' heart rate information. Heart rate is a physiological parameter often used in sports activities and a sensitive metabolic state of the body. Therefore, this topic takes heart rate as a "window" to conduct in-depth research on the body's metabolic state [7]. The study of the causal relationship between exercise load and the change in body function found that the heart rate significantly correlates with exercise intensity, oxygen uptake and energy metabolism. With the increase in load intensity, the body's oxygen consumption and oxygen uptake increase, and the heartbeat also increases, so when the exercise intensity changes, the heartbeat will also change [8]. If the athlete's heartbeat can be monitored in real-time and the heartbeat signal generated by the body for the exercise load can be transmitted to the athlete, then the intensity of the exercise can be adjusted according to their actual situation. This way, you get the best results. This paper now chooses a portable heartbeat monitor to realize wireless heartbeat monitoring. In this way, the athlete's heartbeat is fully integrated into the exercise state.

2.2. Realization of real-time monitoring of positioning during movement. In training for marathon and cross-country running, to accurately grasp the running position of the runners at a certain point in time and the geographical environment, the usual practice is to be driven by the coach. This makes it easy for vehicles to run during training, but it does not ensure the correct guidance of each athlete. There are no new monitoring methods in the training for marathon and cross-country running in our country. The use of GPS for real-time positioning, the specific position of the players displayed on the monitoring screen can facilitate the coach to understand the status of the players [9]. The system designed in this paper presents the player's position directly on the background of the electronic map and regards the player himself as a geographical target. By measuring, analyzing and simulating the spatial information that the players rely on, the training state of the players can be obtained indirectly. Figure 2.1 shows the basic framework for on-site monitoring of a marathon runner's training flow.

2.3. Methods of information exchange between athletes and coaches during training. The development trend of domestic competitive sports is how to realize instant communication with athletes during physical education [10]. According to the actual conditions of the competition, it determines whether the

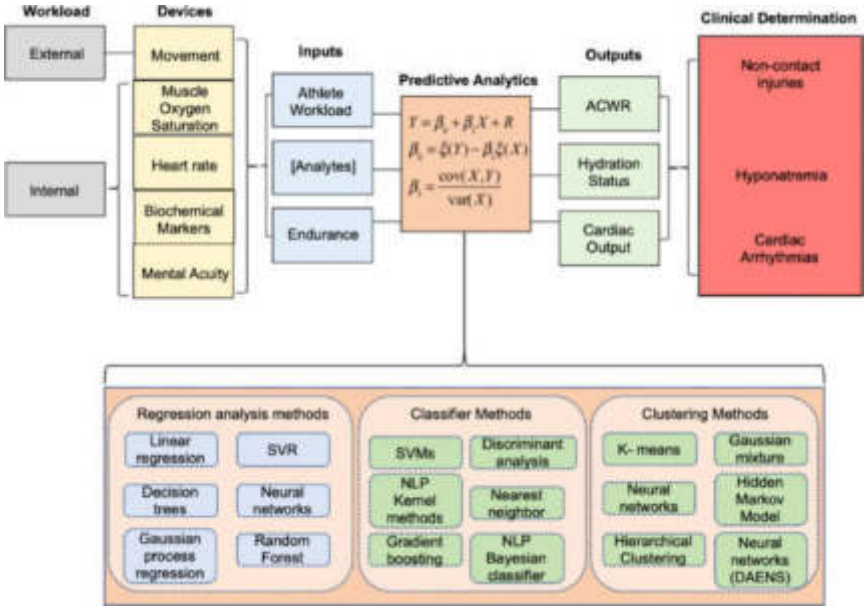


Fig. 2.2: *Sensor architecture of athlete physiological information monitoring system.*

wireless data transmission can be carried out efficiently, whether the communication coverage area and the function area can meet the actual needs, and whether the investment and operation cost of the communication equipment can be borne. The current mainstream mobile communication technology includes analog cluster, digital cluster, CDPD, GPRS, CDMA, and synchronous satellite communication. Trunking digital mobile communication is this system's most suitable communication mode, so this paper decides to use this method to realize the information exchange in physical exercise. By wearing small headsets, athletes can listen to the coach's instructions to achieve better exercise results and adapt to the game's needs.

2.4. Information Integration System Architecture. The athletes' heart rate, position, surrounding environment and other information are transmitted to the training monitoring center through the wireless network using the method of multi-sensor information integration and information fusion [11]. Using GIS or mobile geographic information system, mobile phone mapping, etc., the player's heart rate information, location information, traffic conditions and other information through the road classification method in an image to display on the monitoring terminal. The speed, heart rate and other parameters of the athletes were calculated. Figure 2.2 shows the sensor architecture of the athlete's physiological information monitoring system (see Wearable sensors for monitoring the physiological and biochemical profile of the athlete). In the training process of marathon runners, the information is integrated, the computer information analysis system is built, and the basic information of athletes is managed by database technology. At the same time, the athlete's initial state can be scientifically guided and portable monitoring devices can be used [12]. The real-time exchange of information between athletes and coaches is realized to achieve timely and efficient training guidance for coaches.

3. Evaluation of athlete training effect of machine learning algorithm.

3.1. Machine learning algorithm. If $G(c)$ is a probability metric located in space c, and the set of functions is $P(c, e)$ and $e \in \Theta$, then the least risk functional is the purpose of machine learning. Its expression is:

$$S(e) = \int P(c, e) dG(c) \tag{3.1}$$

The random variable $G(c)$ is uncertain, but there is a fixed random sample. Suppose u is the output of the training system, which can be o or 1. Suppose $g(v,e)$, where $e \in \Theta$ represents a set of indicator functions, and the expression to obtain the loss function is:

$$H(u, g(v,e)) = \begin{cases} 0, & u = g(v,e) \\ 1, & u \neq g(v,e) \end{cases} \tag{3.2}$$

The risk function is used to determine the index function $g(v,e)$ and the output probability of the trainer, and the function with the minor identification error is obtained from the given sample set. An evaluation problem is a nonlinear problem transformed into a high-dimensional feature space based on a given nonlinear function [13]. Then the optimal classification hyperplane is constructed.

By using $Z(v_i \cdot v_j)$ kernel function satisfying Mercer's condition, the nonlinear classification problem on the best classification plane can be transformed into a linear classification problem.

$$Q(\varphi) = \frac{1}{2} \sum_{i,j=1}^{I} \varphi_i \varphi_j u_i u_j Z(v_i \cdot u_j) - \sum_{i=1}^{I} \varphi_i \tag{3.3}$$

The optimal classification function is as follows:

$$g(v) = \operatorname{sgn} \left\{ \sum_{i=1}^{l} \varphi_i u_i Z(v_i \cdot v) + \varepsilon^* \right\} \tag{3.4}$$

A multivariable support vector machine method based on difference is proposed. A base-based method is selected to evaluate the athletes' training performance with physical kernel function quality. Its formula is:

$$Z(v, v_i) = \exp\left(-\frac{\|v - v_i\|^2}{2\sigma^2} \right) \tag{3.5}$$

3.2. Evaluation of athletes' training effect. $V = \{v_1, v_2, \cdots, v_n\}$ and $U = \{u_1, u_2, \cdots, u_n\}$ are a group of physiological parameters that can reflect the effectiveness of an athlete's training, including heart rate, oxygen uptake, hemoglobin and sarcosine. For the metric index set U, the matrix $E = (e_{ij})_{n \times m}$ represents the exponential matrix of the athlete training sample set V. For the measured index value u_j, the athlete training sample v_i is expressed as $e_{ij} = u_j(v_i)\,(i = 1, 2, \cdots, n; j = 1, 2, \cdots, m)$. The normalized exponential matrix E to $S = (S_{xj})_{n \times m}$ can effectively overcome the difference between the evaluation results of each dimension. Take the measured value vector $(s_{i1}, s_{i2}, \cdots, s_{mm})$ of the athlete's training sample v_i under the physiological metric index u_j as the input. The evaluation result z_i of the athlete's exercise effect v_i as the output unit, then there is a nonlinear correspondence between the normalized matrix S and the evaluation result, and it is denoted asG:

$$z_i = G(s_y) \tag{3.6}$$

A sample $v_i = (s_{il}, s_{i2}, \cdots, s_{im})$ of sports performance is taken as the input vector of the support vector machine [14]. The learning sample set is constructed by taking the training effectiveness sampling evaluation value as a regression index. The following regression function is obtained in $F = \{(v_i, z_i)\}_i^n$:

$$z = \sum_{k=1}^{s} (\varphi_k - \varphi_k^*) Z\left(v, v^k\right) + \varepsilon \tag{3.7}$$

where v^k and s are the number of support vectors and the number of support vectors, and φ_k is a Lagrange multiplier, $v^k = (s_{k1}, s_{k2}, \cdots, s_{kn}), k = 1, 2, \cdots, s$. Under the measured physiological index u_j, the index value $(s_{11}, s_{12}, \cdots, s_{mm})$ of the athlete training sample v_i and the evaluation value G of the training effect are nonlinear mapped [15]. Svm-based learning strategies were introduced into physical education teaching, and the quality of physical education was evaluated (Fig. 3.1). The steps to evaluate the effectiveness of an athlete's training are:

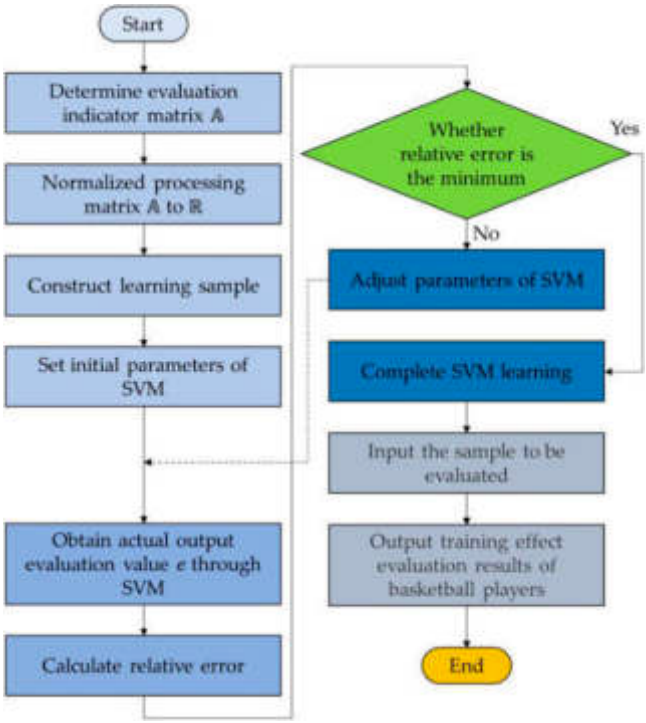


Fig. 3.1: *Structure of athletes' training effect evaluation.*

1. The evaluation index matrix E was established according to the physiological parameters related to physical education teaching effectiveness;
2. The index matrix E, which evaluates the training effectiveness of athletes, is transformed into a normalized matrix S;
3. Based on athletes' training effect sampling $v_i = (s_{i1}, s_{i2}, \cdots, s_{im})$ and evaluation index z_i, a learning sampling set C was constructed, and then SVM was trained and tested.
4. The corresponding regression equation is obtained using a radial basis as the kernel of the support vector machine. The selection index standard for evaluating athletes' exercise effectiveness is set up. These include:

$$\begin{cases} MAE = \frac{1}{h}\sum_{i=1}^{h}\left|z_p^i - z_{SVM}^i\right| \\ MSE = \frac{1}{h}\sum_{i=1}^{h}\left(z_p^i - z_{SVM}^i\right)^2 \end{cases} \tag{3.8}$$

   MAE and MSE are the mean absolute error and mean-variance of the test sample, z_p^i and z_{SVM}^h are the expert evaluation value of the test sample and the SVM calculation result of the confirmed sample, and h represents the total number of samples to be evaluated.
5. The learning process of the support vector machine terminates after the desired parameters are obtained [16]. The trained SVM was used to evaluate the training effect. By analyzing the physiological index vector $(s_{i1}, s_{i2}, \cdots, s_{im})$ of training effect sampling v_i of the evaluated athletes, the training effect evaluation result z_{STM}^i of SVM was obtained.

4. Case analysis. In this study, 10 marathon runners were selected as subjects. Let players participate in football, basketball, volleyball, swimming, running, 5 sports training [17]. Heart rate and lung capacity were collected during each exercise. Morning venous blood was collected at each stage through 10000 test data, including 2000 training samples and 8000 training samples. This program can be well applied to Windows and complete SVM classification and regression problems. Finally, the SVM was divided into 10 groups, each

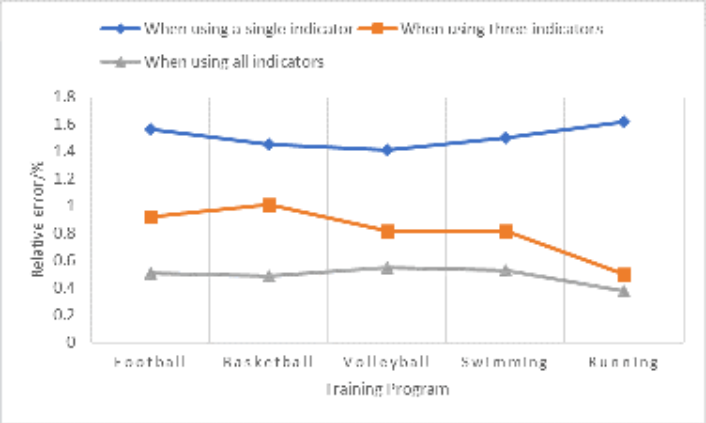


Fig. 4.1: *Relative error for different index quantities.*

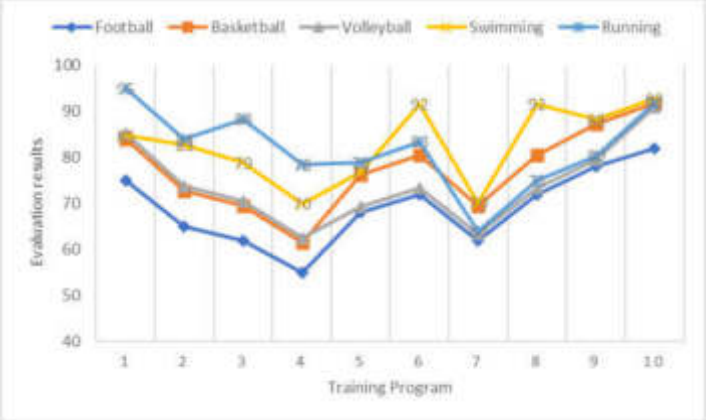


Fig. 4.2: *Evaluation results of athletes' training effect.*

group of 200, and finally, the parameters of the SVM with C=150 and $\sigma 2$=0.016 were obtained. It was used to train SVM, and the number of support vectors and parameters b=30 and -0.218 of the regression equation were obtained. Heart rate, maximal oxygen uptake, hemoglobin, creatine kinase and blood lactic acid were selected as physiological test items to evaluate the athletic performance of athletes [18]. When only heart rate was used as the physiological test index, 3 physiological test indexes, such as heart rate, maximal oxygen uptake and hemoglobin, were used, and the relative error of 5 physiological test indexes was applied to the evaluation of the training effect of each item was statistically treated. According to the experimental data in Figure 4.1, introducing more physiological parameters to evaluate training effects can improve the diversity of evaluation and thus improve the accuracy of evaluation results.

The evaluation method proposed in this paper is applied to the comprehensive evaluation of 10 athletes in 5 events. The results are shown in Figure 4.2.

The evaluation criteria for evaluating the training effect of athletes are selected. The relative deviation of the evaluation results is less than 1%. The relative deviation of physical fitness and professional skill methods in evaluating athletes' various sports skills is less than 1%. The comparison shows that the method proposed in this paper has a good performance in evaluating the athlete's performance. The method in this paper can help the coach to guide the training of athletes better to provide a reference for the training practice in the future.

5. Conclusion. The training effect is directly related to the final result of the competition, and the competition performance of the players is affected by many factors such as training mode and personal quality, and its performance has excellent fluctuations. The physical function of athletes is taken as the evaluation object, and the optimal SVM model is used to evaluate sports. The proposed evaluation model is empirically studied, taking marathon runners as an example. This provides a theoretical basis for improving the training level of athletes.

REFERENCES

[1] De Leeuw, A. W., van der Zwaard, S., van Baar, R., & Knobbe, A. (2022). Personalized machine learning approach to injury monitoring in elite volleyball players. European journal of sport science, 22(4), 511-520.
[2] Qiu, S., Hao, Z., Wang, Z., Liu, L., Liu, J., Zhao, H., & Fortino, G. (2021). Sensor combination selection strategy for kayak cycle phase segmentation based on body sensor networks. IEEE Internet of Things Journal, 9(6), 4190-4201.
[3] Rigamonti, L., Albrecht, U. V., Lutter, C., Tempel, M., Wolfarth, B., & Back, D. A. (2020). Potentials of digitalization in sports medicine: a narrative review. Current sports medicine reports, 19(4), 157-163.
[4] Nath, R. K., Thapliyal, H., Caban-Holt, A., & Mohanty, S. P. (2020). Machine learning based solutions for real-time stress monitoring. IEEE Consumer Electronics Magazine, 9(5), 34-41.
[5] Husnain, A., Hussain, H. K., Shahroz, H. M., Ali, M., & Hayat, Y. (2024). A Precision Health Initiative for Chronic Conditions: Design and Cohort Study Utilizing Wearable Technology, Machine Learning, and Deep Learning. International Journal of Advanced Engineering Technologies and Innovations, 1(2), 118-139.
[6] Jowitt, H. K., Durussel, J., Brandon, R., & King, M. (2020). Auto detecting deliveries in elite cricket fast bowlers using microsensors and machine learning. Journal of sports sciences, 38(7), 767-772.
[7] Rana, M., & Mittal, V. (2020). Wearable sensors for real-time kinematics analysis in sports: A review. IEEE Sensors Journal, 21(2), 1187-1207.
[8] Li, B., & Xu, X. (2021). Application of artificial intelligence in basketball sport. Journal of Education, Health and Sport, 11(7), 54-67.
[9] Nwachukwu, B. U., Beck, E. C., Lee, E. K., Cancienne, J. M., Waterman, B. R., Paul, K., & Nho, S. J. (2020). Application of machine learning for predicting clinically meaningful outcome after arthroscopic femoroacetabular impingement surgery. The American Journal of Sports Medicine, 48(2), 415-423.
[10] Worsey, M. T., Espinosa, H. G., Shepherd, J. B., & Thiel, D. V. (2020). An evaluation of wearable inertial sensor configuration and supervised machine learning models for automatic punch classification in boxing. IoT, 1(2), 360-381.
[11] Johnson, W. R., Mian, A., Robinson, M. A., Verheul, J., Lloyd, D. G., & Alderson, J. A. (2020). Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. IEEE Transactions on Biomedical Engineering, 68(1), 289-297.
[12] Giles, B., Kovalchik, S., & Reid, M. (2020). A machine learning approach for automatic detection and classification of changes of direction from player tracking data in professional tennis. Journal of sports sciences, 38(1), 106-113.
[13] O'Driscoll, R., Turicchi, J., Hopkins, M., Horgan, G. W., Finlayson, G., & Stubbs, J. R. (2020). Improving energy expenditure estimates from wearable devices: A machine learning approach. Journal of Sports Sciences, 38(13), 1496-1505.
[14] Wilkens, S. (2021). Sports prediction and betting models in the machine learning age: The case of tennis. Journal of Sports Analytics, 7(2), 99-117.
[15] Incel, O. D., & Bursa, S. Ö. (2023). On-device deep learning for mobile and wearable sensing applications: A review. IEEE Sensors Journal, 23(6), 5501-5512.
[16] Ramkumar, P. N., Luu, B. C., Haeberle, H. S., Karnuta, J. M., Nwachukwu, B. U., & Williams, R. J. (2022). Sports medicine and artificial intelligence: a primer. The American Journal of Sports Medicine, 50(4), 1166-1174.
[17] Taj, I., & Zaman, N. (2022). Towards industrial revolution 5.0 and explainable artificial intelligence: Challenges and opportunities. International Journal of Computing and Digital Systems, 12(1), 295-320.
[18] Alexopoulos, K., Nikolakis, N., & Chryssolouris, G. (2020). Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. International Journal of Computer Integrated Manufacturing, 33(5), 429-439.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 28, 2024
Accepted: Jul 18, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 871–879, DOI 10.12694/scpe.v26i2.4048

ASSET MANAGEMENT OF SMART GRID USING DIGITAL TWIN TECHNOLOGY AND MACHINE LEARNING ALGORITHMS

XIAOTAO DENG*

Abstract. In order to solve the problem of data insecurity in the current smart grid asset monitoring and management platform, the author proposes a research on smart grid asset management using digital twin technology and machine learning algorithms. This study achieves data security performance verification and monitoring through the application of data consensus mechanism. When verifying data in data nodes, data regulation verification information can be published in the form of a set of data packets to the data chain. The best node in the node is used as the data packet in the blockchain to build a platform structure. Management is achieved through data encryption, data monitoring, and device status classification. The experimental results show that the data theft rate simulated using this method has been reduced by 25.48%, and the security performance of the scheme is higher. The use of digital twin technology and machine learning algorithms for smart grid asset management has improved the security issues existing in traditional solutions and has enormous potential for application.

Key words: Smart grid, Asset monitoring, Management platform, Blockchain, Equipment status classification

1. Introduction. With the strengthening of power regulation in various countries around the world, the aging of power equipment, and fierce competition in the market environment, power companies are forced to optimize investment and reduce costs. At the same time, the development of the social economy continuously increases the reliability requirements of the power system. The market and business models of power companies are gradually changing, and there is an urgent need to further optimize management processes and improve asset management levels. Therefore, as an advanced management concept, asset lifecycle management has developed rapidly in foreign power companies. The goal of asset lifecycle management in foreign power grids is mostly to pursue overall economic optimization while meeting certain constraints (such as reliability constraints, environment, user needs, etc.) [1].

With the continuous deepening of smart grid construction, the deployment and application of power grid dispatch automation systems, intelligent operation and inspection systems, distribution automation systems, and electricity consumption information collection systems, etc [2]. On the one hand, it has improved the ability to operate, control, and maintain equipment, and on the other hand, it has greatly improved the ability to collect data, and the system has accumulated more and more data. How to manage and make good use of the increasingly large amount of data is an urgent problem that power grid enterprises need to solve. Smart grid big data processing requires the adoption of new technologies and means to solve data processing problems [3]. Data asset management is a management model that combines traditional asset management methods with the characteristics of data [4]. It is a variety of management activities adopted by enterprises or organizations, which is a function of planning, controlling, and supplying relevant data and information to ensure the security and integrity of data assets, improve the economic benefits brought by data assets through the discovery of big data value, and ensure and promote the development of various businesses of enterprises [5]. Through effective data management and big data analysis and mining, power grid data assets will play an increasingly important role in areas such as power grid scheduling, intelligent operation and maintenance, and intelligent power distribution. This will reduce labor costs, improve equipment utilization, and ensure the safe and stable operation of the power grid [6].

2. Literature Review. In recent years, in the production and operation of the power grid, data analysis can effectively improve work efficiency and effectiveness. For example, by integrating power grid operation data and external environmental data, big data analysis can predict weak points in the power grid in advance,

*Wuhan Railway Vocational College of Technology, Wuhan, Hubei, 430205, China (deng001125@sina.com)

identify problems in a timely manner, and handle them, which can greatly improve the reliability of the power grid and reduce economic losses for enterprises. The traditional smart grid asset monitoring and management platform is mainly realized by the distributed business model of the Internet, so it is necessary to establish a centralized structural system [7]. Completing data exchange, equipment maintenance, and user information resource processing in the operating system of the smart grid, the centralized structural system, due to the existence of a large number of transaction management systems, will cause interference and security risks to the numerous data of the smart grid. Tiwari, S. et al. used various machine learning based algorithms to estimate power grid stability to avoid fault situations. Using the Bagging classifier algorithm to accurately predict electricity demand will help avoid power grid failures, thereby improving the stability and robustness of the power grid [8]. Hudani, D. et al. proposed a data-centric approach (DaC) to generate such invariants. The entire study was conducted using operational data from a functional smart grid, which is also a live laboratory [9]. Wang, R. et al. proposed an anti electricity theft prediction method based on power big data. This method constructs electricity theft data samples based on anomaly rules and introduces constraints on the growth rate of line loss rate [10].

In order to solve the problems in traditional solutions, the author conducted research on the asset monitoring and management platform of smart grids based on the use of digital twin technology and machine learning algorithms.

3. Method.

3.1. Smart Grid with Digital Twin Technology. The smart grid under the digital twin technology includes both the physical level physical grid and the virtual grid in the digital world. The two contents that exist in different dimensions are essentially coexisting and intertwined. The virtual grid will fully restore the various conditions and related parameters of the physical grid, thereby obtaining data and case support for the control of the physical grid through observation and adjustment of the virtual grid [11]. In this process, other digital information technologies can be applied to assist, enabling the smart grid to form deep learning capabilities in the process of continuous autonomous optimization, thereby enabling the smart grid to derive development models that can self detect, improve, and optimize, gradually establishing the ability to operate and manage itself [12].

3.2. Structure Construction of Smart Grid Asset Monitoring and Management Platform. The author applies decentralized technology to monitor and manage electrical equipment assets in the smart grid, and sets corresponding monitoring and management system structures based on different data and equipment characteristics for block management [13-14]. Firstly, a decentralized ledger management structure for electrical equipment should be established, which includes statistics on the model, operating status, installation date, purchase date, and coverage range of electrical equipment. It is also necessary to backup complete information records on the maintenance and inspection frequency of electrical equipment, and scientifically build a safety database that can be extracted and retrieved in a timely manner. Electrical equipment inevitably incurs certain self losses during operation, and it is necessary to manage potential vulnerabilities in the equipment. The loss and vulnerability information of electrical equipment should be recorded in different regions and promptly reported to relevant power grid departments. The main task is to promptly investigate and manage vulnerabilities in electrical equipment. The monitoring of the status of electrical equipment is one of the very important tasks in the operation of the smart grid [15]. A complete decentralized database needs to be established for the daily maintenance data of equipment, equipment failure status, equipment operation time, and corresponding maintenance plans, ensuring the normal operation of the smart grid and the monitoring and management of electrical equipment assets in the smart grid [16].

In order to prevent the occurrence of information leakage caused by exhausted data verification of nodes, the author constructed a decentralized data chain structure. The optimal nodes are divided according to time periods, and the data is selected for data security verification according to different time periods [17]. When parsing data packets in nodes, it is necessary to obtain the data content in the packets through the data key, ensuring the one-way data security of the smart grid. The author's structural framework is shown in Figure 3.1.

The smart meter in the structure is mainly installed in the customer's electrical equipment, facilitating the collection of multidimensional data by the smart meter [18]. Smart meters are customized based on the

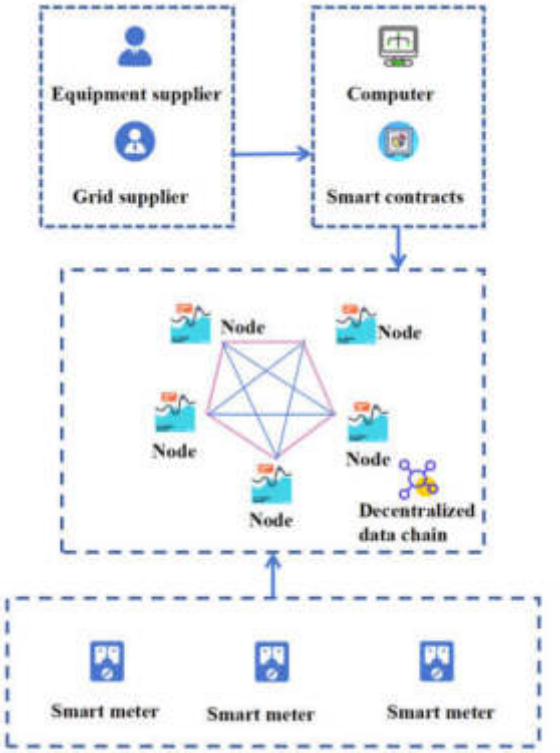


Fig. 3.1: Architecture diagram of intelligent grid asset monitoring and management platform

user's electricity consumption characteristics and the characteristics of the electrical equipment, and require the user's real name login to connect with the data chain in the block [19]. Smart meters will directly send users' multidimensional data to the data nodes of the smart grid, achieving flexibility and accuracy in data application, while also reducing communication load in the smart grid [20]. After receiving data, the operating unit of the smart grid formulates different electricity feedback policies based on different types of customers, and monitors and manages the set electricity equipment in real-time in the form of contracts. When the contract in the smart grid meets the set conditions, the content of the contract will be automatically implemented, and users will cooperate with the conditions in the contract to provide feedback on electricity consumption information and adjust the electricity equipment. This not only improves the power supply efficiency of the smart grid, but also reduces the power supply burden of the smart grid. The efficiency calculation function for smart grid power supply in the contract is:

$$\partial = x \times J(N) \tag{3.1}$$

In the formula: ∂ represents the working efficiency of the smart grid; x is the channel pressure coefficient in the contract; J(N) is the time parameter for multi-dimensional data feedback.

The smart grid contains a control center for asset monitoring and management platforms, which utilizes the blockization and decentralization of data communication links to achieve the independence and permanence of data storage, management, and detection. The implementation of this function does not require human involvement and is achieved through the automatic data allocation technology of the intelligent system, avoiding data account theft caused by personnel participation. In addition, the data management structure of the control center is also equipped with power grid operation chips, which monitor the data information of power grid assets for faults and violations, and issue timely warnings to management personnel.

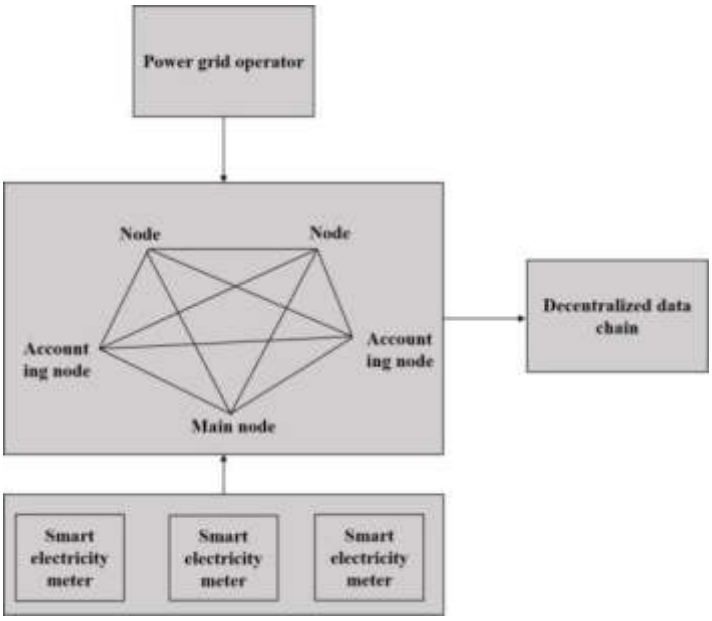


Fig. 3.2: Implementation flowchart of aggregate encryption

3.3. Workflow Analysis of Smart Grid Asset Monitoring and Management Platform. The workflow of the smart grid asset monitoring and management platform is mainly reflected in the analysis of multidimensional data. From obtaining multidimensional data to managing multidimensional data, it needs to be coordinated and allocated through a decentralized data chain structure. The smart meter first collects user's smart grid asset data and publishes it to decentralized grid nodes, which then further transmit the data to the corresponding management platform through blockchain technology. In order to ensure the security of data, the transmission of multidimensional data is mainly carried out in the form of encrypted files. The encryption mechanism of multidimensional data is implemented through aggregation encryption. The encryption mechanism function of multidimensional data is:

$$U = \frac{K-1}{N} \tag{3.2}$$

The encryption public key N is defined as:

$$N = g^N \times modN^2 \tag{3.3}$$

In the formula: U represents the aggregated encryption result; K is the encryption security factor for multidimensional data; N is the public key code. Figure 3.2 shows the implementation flowchart of aggregate encryption.

After the successful data acquisition of the smart grid, it is necessary to monitor the key data, check whether the anti theft function of the data is complete, and prevent illegal users from tampering with the data in the high-voltage grid and stealing high-voltage electricity. By monitoring the data, three-phase voltage data, current preprocessing data, and data dimension can be obtained. The monitoring data needs to be sent to the data chain, and the data from the main node should be transmitted to the blockchain for permanent storage to ensure the stable operation of the smart grid. Data monitoring can be achieved through diagnostic networks in power equipment. Figure 3.3 shows the data monitoring structure diagram.

Smart grid operators also need to effectively control and manage real-time data to ensure stable electricity supply within the coverage area of the smart grid and avoid insufficient power supply during peak periods, based on the situation where users obtain electricity. The implementation of this technology is mainly achieved

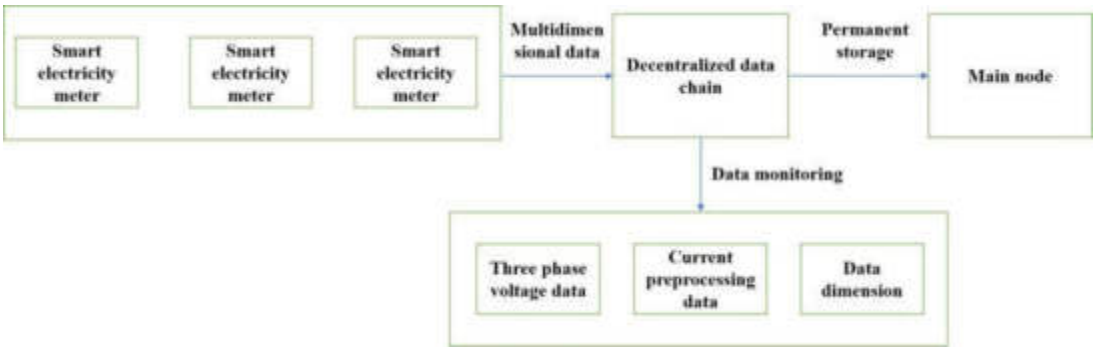


Fig. 3.3: Data monitoring structure diagram

through the contract mechanism between operators and users, and the smart contract contains two operation functions: user electricity consumption and grid electricity supply. When the data in the contract changes, an alert will be issued, and a decoding threshold value needs to be set to limit the continuous power supply of the smart grid. The threshold limit function in the contract is shown in equation 3.4.

$$e(c,1) = e[k, H(N)] \tag{3.4}$$

In the formula: (c,l) represents the threshold value result; K is the multidimensional data electricity consumption coefficient; H(N) is the user's electricity consumption. When the user's electricity consumption exceeds the power supply cost, the smart contract will also reduce the continuous power supply to the user. In severe cases of user arrears, the power supply will be stopped. The data related to equipment assets in the smart grid also needs to be monitored and managed. Equipment suppliers first provide certain factory data and process the safe operation status of the equipment in a decentralized data chain structure. This structure divides the operation status of power grid equipment into normal operation status, alert status, emergency status, and shutdown status. When a device malfunctions, it is necessary to immediately transmit the fault information to the management personnel. The management personnel should make corresponding adjustment plans based on the factory information. The equipment supplier should also transmit the relevant data of the device, Figure 3.4, and the schematic diagram of the device status classification to the blockchain, forming a decentralized permanent database. Figure 3.4 is a schematic diagram of the device's state classification.

3.4. Experimental preparation. The implementation of this scheme first requires the installation of smart meters within the conventional coverage range of the smart grid, the collection of asset data of the smart grid, the introduction of communication contracts in the smart meters, and the removal of centralization in the content data of the contracts, so that the data communication of the smart terminal is connected in real-time with the bus of the smart grid. In order to ensure stability during the test process, the model of the smart meters used in the test is DB2750 three-phase smart meters. The contract in smart meters mainly transmits data to the outside in duplex communication mode, and the working mode of the communication contract is determined based on different data types. The most common communication frame rate in the contract is 48H, and 8 bytes are a separate communication area. During data transmission, the bytes are transferred to form a barcode, which has encryption function and requires a key from the management personnel to parse the data content, enhancing the security and stability of the experiment.

In order to ensure that there is no interference from other power grid data in the experimental environment, a data signing process is established in the smart grid data environment. A key system is required for the transmission and exchange of user related data and smart grid related data, and the data receiving port can have an independent key. When attackers of smart grid assets install stolen files in the data channel, the eavesdropping files cannot be effective until users publish data information to a decentralized data chain structure, and the data content needs to be decrypted through a signed password. The ciphertext in channel transmission is in a sealed state, and even if the attack file obtains the data information in the channel, it cannot

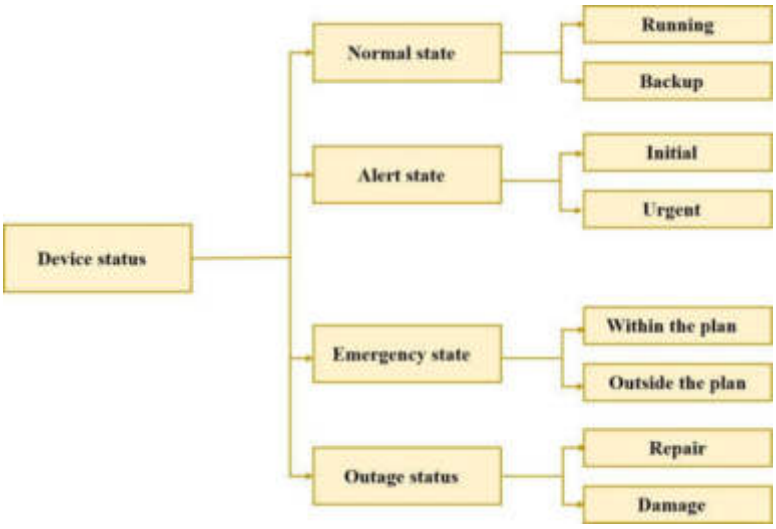


Fig. 3.4: Schematic diagram of device status classification

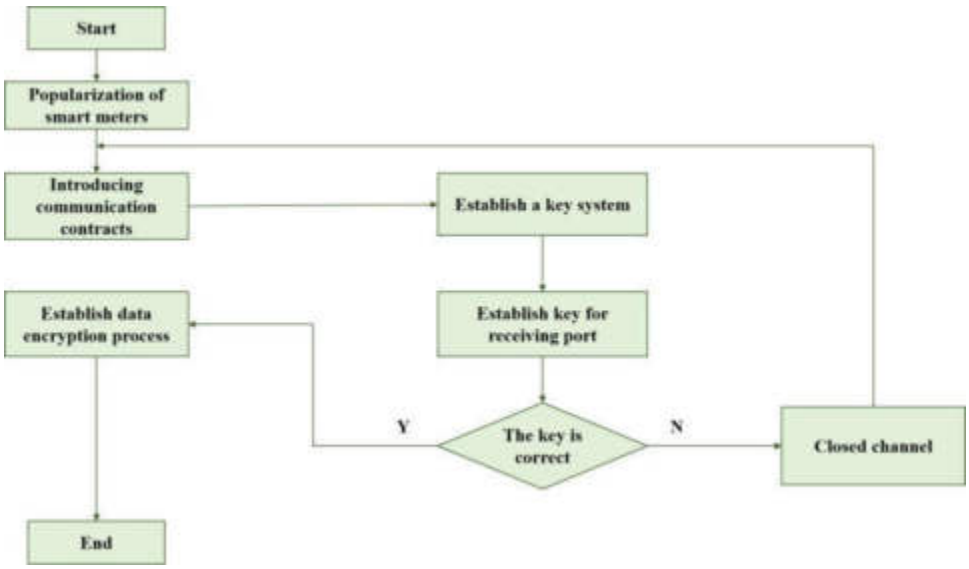


Fig. 3.5: Experimental Environment Construction Flow Chart

crack the encrypted file. The permanent and decentralized nature of multi-dimensional data in electricity can adapt to the confidential characteristics of blockchain data blocks. Figure 3.5 shows the flowchart for constructing the experimental environment.

4. Results and Discussion. The asset information and management platform of the smart grid in the experiment require multidimensional data content. The reception and processing of the management platform should cooperate with the regulation of the smart grid, and the decentralized data chain structure has permanence. The data stored by the management platform cannot be corrected again. The comparison results of the operational effectiveness and performance of the asset monitoring and management platform in the face of data attacks in the same smart grid by three schemes based on the content of the smart treaty are shown in

Table 4.1: Performance Comparison of Three Schemes

Performance	Internet distributed business model	Multidimensional data aggregation feedback mechanism	Author's proposal
Confidentiality	good	good	excellent
Data integrity	excellent	good	excellent
Dimension collection	good	good	excellent
data management	good	excellent	excellent
Permanent	excellent	good	excellent

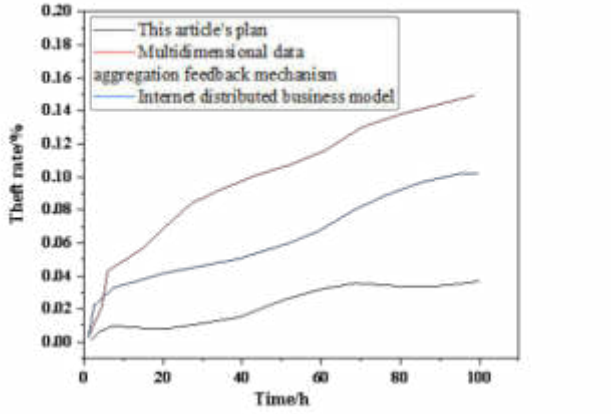


Fig. 4.1: Comparison of Security Processing Capabilities of Three Schemes for the Same Stolen Files

Table 4.1.

The author's plan has a wide coverage of the implementation of the intelligent grid asset monitoring and management platform, including the monitoring and management of intelligent grid electrification equipment assets and user data information assets. For the asset management of electrification equipment, this article establishes a decentralized ledger management structure to achieve monitoring and management of electrification equipment resources in multiple states such as operation, maintenance, and repair. In the experiment, the safety performance of electrical equipment asset monitoring and management in the smart grid was first compared. The same electrical equipment data information was monitored in the same experimental environment, and the same data theft files were installed in three different schemes. Figure 4.1 compares the security processing capabilities of three schemes for the same stolen files.

It can be seen from Figure 4.1 that the security performance of the author's scheme is the highest. After 100h, the data theft rate reaches the highest (about 0.04%), followed by the Internet distribution business model, which reaches about 0.08%, and the data multi-dimensional aggregation feedback mechanism is the worst, which reaches about 0.15%. In addition, the security processing capacity of the author's scheme is growing steadily within 100h, while the traditional method rises to 0.03% to 0.04% within 10h. The security stability is poor. The multidimensional data source for asset electrification data in the smart grid in this plan was initially from equipment suppliers. This plan is based on the data provided by suppliers as the foundation of a multi-dimensional data chain structure, and monitors whether there will be data failures and leaks during the operation of electrical equipment based on relevant safety performance data. It scientifically manages vulnerabilities in the electrical equipment channel, establishes real-time contact with the regulatory authorities of the smart grid, and enhances the stability of security performance. The author has also completed decen-

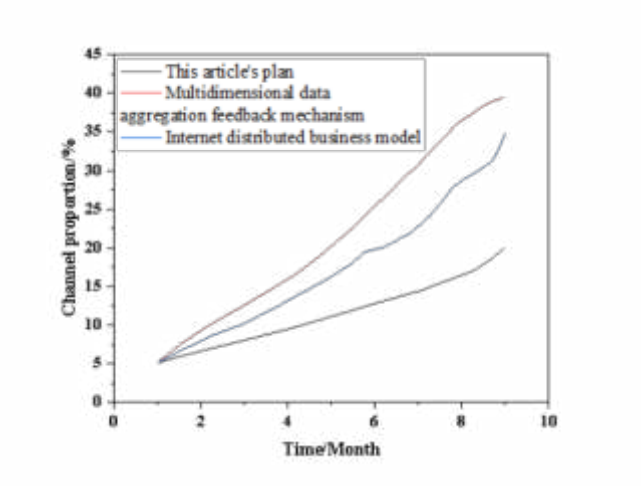


Fig. 4.2: Comparison of Communication Channel Pressure

tralization in the overall structure, making the monitoring and management of intelligent grid electrification equipment assets an independent operating system, and ensuring higher safety performance. Asset monitoring and management in smart grids require communication to be completed through data nodes, and the accuracy of data transmission between nodes is also one of the comparison objects of this experiment. The same type of data is obtained separately and transmitted between nodes, and the accuracy of data transmission is monitored in the channel.

The intelligent grid asset monitoring and management of this scheme has a higher degree of accuracy, achieving 100% accurate data transmission, while the other two traditional schemes have a minimum accuracy of nearly 30%. In the same area, this scheme applies smart meters to collect power supply data from users. The collected data has real name characteristics, and the transmission of nodes requires security verification to be transmitted in real time, thereby ensuring the accuracy of the data transmission process. The application of smart contracts as communication protocols in the asset monitoring and management communication channel of this scheme can timely feedback inappropriate data, reducing the overall channel pressure of the structure. The decentralized node blockchain adopts a phased working form, which also reduces the communication pressure of the channel. Figure 4.2 shows a comparison of communication channel pressures for intelligent grid asset monitoring and management under three different schemes in one year.

From Figure 4.2, it can be seen that the channel pressure under this scheme is the smallest, and the channel proportion of the author's scheme is controlled within 20%, while the channel proportion of the traditional scheme has reached 30%. The author has mainly improved the data transmission source and transmission medium of the channel, enhancing the smoothness of data transmission in asset monitoring and management, and thereby enhancing the overall efficiency of the structure.

5. Conclusion. The author proposes the study of using digital twin technology and machine learning algorithms for smart grid asset management. The smart grid asset monitoring and management platform is the basic guarantee for efficient and stable power supply in the smart grid. The traditional implementation scheme of the smart grid asset monitoring and management platform mainly operates in a centralized mode, which is easily affected by interference from other power grid data. In response to the security, accuracy, and efficiency of the intelligent grid asset monitoring and management platform, the author establishes a decentralized structure for the intelligent grid asset monitoring and management platform, and applies technologies such as channel smart contracts, node data transmission, and key processing to improve the problems existing in traditional solutions.

REFERENCES

[1] Zhou, C., Wang, L., & Wang, L. (2022). Lattice-based provable data possession in the standard model for cloud-based smart grid data management systems:. International Journal of Distributed Sensor Networks, 18(4), 137-147.
[2] Hamada, D., Sugimoto, M., Yamazaki, Y., Nagata, N., Takahara, H., & Takeko, R., et al. (2023). Construction of kansei evaluation model for protective films for smartphones. Transactions of Japan Society of Kansei Engineering, 22(2), 207-216.
[3] Devi, N. M., & Kasireddy, S. R. (2022). Querying and retrieving data from graphs stored in distributed architecture (qrda). International Journal of Smart Grid and Green Communications, 2(2), 139.
[4] Davis, M., & Sébastien Lleo. (2024). Jump-diffusion risk-sensitive benchmarked asset management with traditional and alternative data. Annals of Operations Research, 336(1-2), 661-689.
[5] Tang, J., & Qu, W. (2023). The application mode of blockchain in university asset management, 11(1), 12-25.
[6] Monteiro, J., Pedro, André, & Silva, António Joo. (2022). A gray code model for the encoding of grid cells in the entorhinal cortex. Neural Computing and Applications, 34(3), 2287-2306.
[7] Li, D., Xiao, K., Wang, X., Guo, P., & Chen, Y. (2023). Towards sparse matrix operations:graph database approach for power grid computation, 6(1), 50-63.
[8] Tiwari, S., Jain, A., Yadav, K., & Ramadan, R. (2022). Machine learning-based model for prediction of power consumption in smart grid. Int. Arab J. Inf. Technol., 19, 323-329.
[9] Hudani, D., Haseeb, M., Taufiq, M., Umer, M. A., & Kandasamy, N. K. (2022). A data-centric approach to generate invariants for a smart grid using machine learning, 45(2), 114-119.
[10] Wang, R., Duan, Z., Ge, X., & Zhang, J. (2022). Application research of computer machine learning algorithm in anti-theft analysis under smart grid. IEEE, 16(3), 1513-1535.
[11] Jiang, X., Sun, A., Sun, Y., Luo, H., & Guizani, M. (2023). A trust-based hierarchical consensus mechanism for consortium blockchain in smart grid. Tsinghua Science and Technology, 28(1), 69-81.
[12] Yang, Y., Du, S. Q., & Chen, Y. (2023). Real-time pricing method for smart grid based on social welfare maximization model. Journal of Industrial and Management Optimization, 19(3), 2206-2225.
[13] Perumal, S., & Rajendiran, S. (2022). Low power multiplier based long short-term memory hardware architecture for smart grid energy management. International Journal of System Assurance Engineering and Management, 13(5), 2531-2539.
[14] Darzi, S., Akhbari, B., & Khodaiemehr, H. (2022). Lpm2da: a lattice-based privacy-preserving multi-functional and multi-dimensional data aggregation scheme for smart grid. Cluster Computing, 25(1), 263-278.
[15] Dou, C., Wu, D., Yue, D., Jin, B., & Xu, S. (2022). A hybrid method for false data injection attack detection in smart grid based on variational mode decomposition and os-elm. CSEE Journal of Power and Energy Systems, 8(6), 1697-1707.
[16] Zhou, C., Wang, L., & Wang, L. (2022). Lattice-based provable data possession in the standard model for cloud-based smart grid data management systems:. International Journal of Distributed Sensor Networks, 18(4), 137-147.
[17] Jogunola, O., Adebisi, B., Anoh, K., Ikpehai, A., Hammoudeh, M., & Harris, G. (2022). Multi-commodity optimization of peer-to-peer energy trading resources in smart grid, 10(1), 11.
[18] Oula, F. Z. E., Eddine, C. S., Julian, B. G., & Hamidou, T. (2022). Hierarchical mean-field type control of price dynamics for electricity in smart grid, 35(1), 17.
[19] Chen, M., Sharma, A., Bhola, J., Nguyen, T. V. T., & Truong, C. V. (2022). Multi-agent task planning and resource apportionment in a smart grid. International Journal of System Assurance Engineering and Management, 13(1), 444-455.
[20] Li, J., Gu, C., Xiang, Y., & Li, F. (2022). Edge-cloud computing systems for smart grid: state-of-the-art, architecture, and applications, 10(4), 13.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 28, 2024
Accepted: Jul 24, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 880–887, DOI 10.12694/scpe.v26i2.4053

THE MEDICAL TESTING EQUIPMENT MANAGEMENT SYSTEM BASED ON ARTIFICIAL INTELLIGENCE

HONGLI PEI*, LEI SUN†, AND WENTAO GUO‡

Abstract. This study focuses on developing a medical testing equipment management system based on artificial intelligence. The system integrates advanced sensor technology to monitor patient's physiological characteristics data in real-time, such as heart rate, blood pressure, body temperature, etc. and processes the data through a differential entropy analysis algorithm to extract key health indicators. Then, this study constructed a deep learning neural network model to predict the changing trend of patient health status and optimized the configuration and use of medical detection equipment accordingly. This paper proposes a feature extraction method based on neural network model, which can effectively identify abnormal patterns in physiological signals and provide high-quality input data for subsequent prediction models. Simulation results show that the proposed neural network model has high accuracy and practicability in predicting patients' health status. The model can assist healthcare workers in identifying potential health risks in time to improve treatment results and patients' quality of life.

Key words: Artificial intelligence; Medical testing equipment management; Sensor; Physiological characteristics data; Differential entropy; Neural network model; Algorithm design; Model simulation.

1. Introduction. In today's rapidly developing medical science and technology field, the application of artificial intelligence (AI) is gradually changing the traditional medical service model. In particular, as an essential part of medical services, the medical testing equipment management system's intelligent upgrade is significant for improving medical efficiency, reducing costs, and enhancing patient experience. Researching medical testing equipment management systems has always been a hot spot in medical engineering. In the research field of medical testing equipment management systems, literature [1] proposed a real-time monitoring system based on the Internet of Things (IoT), which realized remote monitoring and status assessment of medical equipment by integrating a variety of sensors, effectively improving the efficiency and response speed of equipment management. However, the data processing capabilities of the system still need to be strengthened, especially for real-time analysis and decision support in the face of large-scale data flows. Literature [2] focuses on analyzing physiological feature data and proposes a feature extraction method based on differential entropy, which can extract essential information reflecting heart health status from electrocardiogram (ECG) signals. Studies have shown that the differential entropy algorithm has high sensitivity and specificity in identifying early signs of heart disease. However, the scalability of this method in processing multi-channel or multi-type physiological signals has not been fully verified. The application of neural network models in medical data analysis is also increasing. A convolutional neural network (CNN) model was constructed in reference [3] to identify and classify diseased areas in medical images automatically. The model performs well on multiple public datasets, but the size and diversity of the datasets limit its generalization ability. Literature [4] proposes a system combining sensor technology and recurrent neural networks (RNN) to monitor patients' daily activities and physiological states with chronic diseases. The system can analyze patients' movement patterns and vital signs in real time, which provides a basis for formulating personalized treatment plans. However, the long-term stability of the system and its adaptability to different patient populations still require further study. Literature [5] discusses the application of differential entropy in electroencephalogram (EEG) signal analysis and proposes a new feature extraction process that can effectively distinguish between normal and abnormal brain wave patterns. The results show that the differential entropy algorithm has potential application value in

*Department of Equipment Management and Maintenance, Liaocheng People's Hospital, Liaocheng, 252000, China
†Department of Equipment Management and Maintenance, Liaocheng People's Hospital, Liaocheng, 252000, China
‡Department of Equipment Management and Maintenance, Liaocheng People's Hospital, Liaocheng, 252000, China (Corresponding author, guown2024@163.com)

EEG signal processing, but its computational complexity and real-time processing ability are urgent problems to be solved. Literature [6] proposes a fault prediction model for medical testing equipment based on deep learning, which predicts potential failure modes by analyzing the historical operation data of the equipment. The experimental results show that the model is efficacious in improving the initiative of equipment maintenance and reducing unplanned downtime. However, the model's interpretability and universality to different medical device types still need improvement.

Current research has made positive progress in managing medical detection equipment, physiological characteristics data analysis, and the application of neural network models, but it still faces challenges in data processing speed, model generalization ability, and system stability. Future research should focus on developing more efficient and intelligent algorithms to meet the demand for precision and personalized services in the medical field. Firstly, this paper designs a comprehensive medical testing equipment management system that can collect and analyze the physiological characteristics data from various medical testing equipment in real-time [7]. Secondly, the differential entropy algorithm is introduced, which can extract the key health indicators from the complex physiological signals. Finally, this paper constructs a prediction model based on a neural network, which can predict the change in patients' health status according to the extracted feature data and provide decision support for the configuration and use of medical detection equipment [8]. The results show that the proposed AI-based medical testing equipment management system can significantly improve the quality and efficiency of medical services and provide patients with more personalized and accurate medical services.

2. Design of testing equipment management system.

2.1. System Structure Design. The system consists of four modules: the medical equipment in use status management module, the daily maintenance management module, the medical equipment annual inspection information management module and the query and data processing module [9]. Among them, query and data processing are standard components that realize information query, report printing and output. This paper presents a hospital testing equipment management system based on SSH2+ MVC. The specific system structure is shown in Figure 2.1 (the picture is quoted in the Journal of Healthcare Engineering, 2021, 2021(1): 6685456.).

The system realizes the interaction of each function module through the DAO interface and database. The primary data table consists of the permission data table, equipment basic information table, equipment annual inspection data table, routine maintenance test data table, and failed equipment information table [10]. These data tables are associated with specified values. Authorization data forms are used to set user rights. The primary instrument data form records the basic instrument data during the inspection. The annual inspection data of instruments is the statistical analysis of the annual inspection of hospital instruments. The routine maintenance check data table results from a routine maintenance check on the device. The Retired Equipment Information form is used to record the retired equipment information.

2.2. System module design. The whole system has the characteristic of modularity. This paper gives a detailed analysis of each functional module (Figure 2.2).

2.2.1. Medical equipment in use management module. The status quo and fundamental data of hospital equipment were analyzed and divided into four parts: new, deactivated, scrapped, and transferred [11]. This component data is stored in the primary data table of the device. There are hospital numbers, equipment names, use departments, situations, annual inspection cycles, test units, etc. The system automatically generates new equipment with a unique number to ensure its unified management. Remove the obsolete equipment data from the equipment basic data form and enter it into the equipment information form to ensure it does not have relevant data during routine maintenance and annual inspections. So as not to cause administrative confusion [12]. The transfer of equipment can be achieved by changing the department of use in the primary data table to the department currently used. The basic information of the equipment can be displayed in the report by inputting the hospital code of the hospital equipment. You can also perform fuzzy queries by device name and department. The basic information about the equipment will be presented in the report.

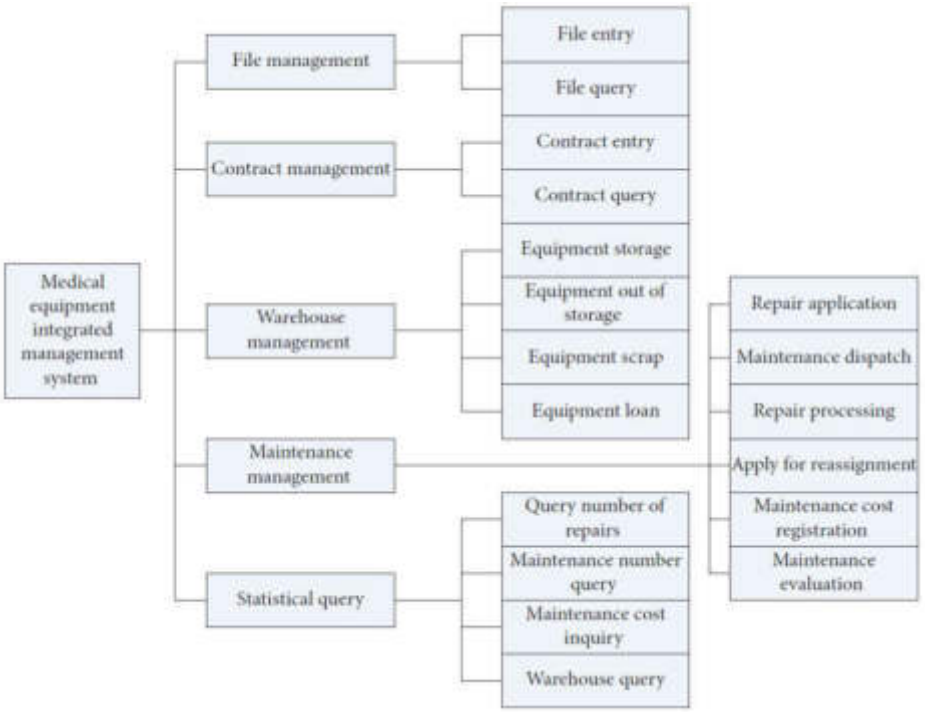


Fig. 2.1: *Overall architecture of hospital testing equipment management system.*

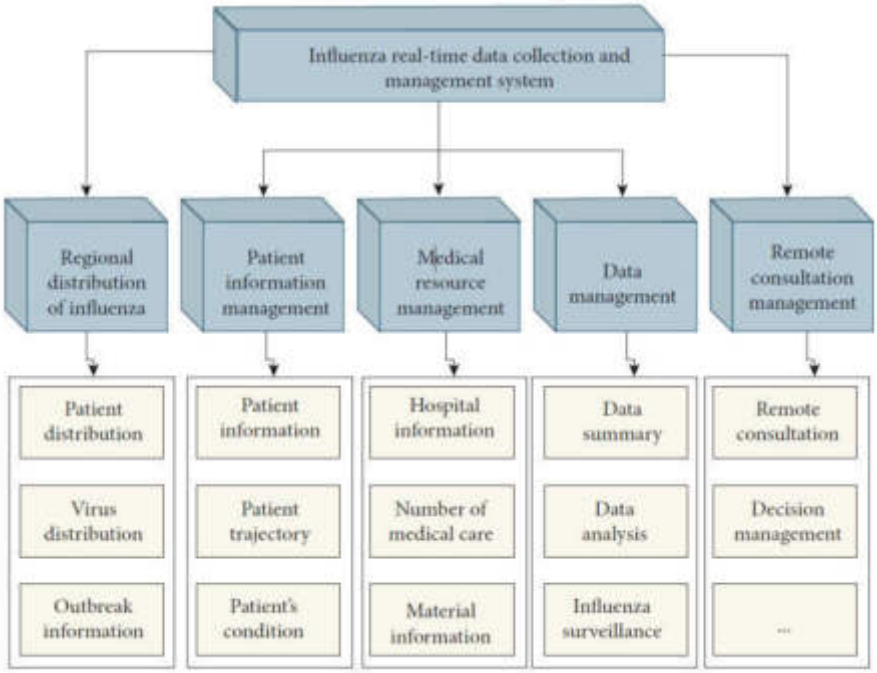


Fig. 2.2: *Functional module architecture of medical testing equipment system.*

2.2.2. Routine repair and maintenance module of health equipment. This module's data is mainly based on daily maintenance test data, including hospital number, maintenance status, post-maintenance inspection, daily inspection, inspection time, inspection personnel, maintenance methods, etc. The data table is linked to the data in the device basic data table by particular keywords, namely the hospital number. When equipment needs maintenance, a distribution work order is generated [13]—to record maintenance and inspection status after maintenance. When the inspection time approaches, the system will pop up an "inspection prompt," prompting maintenance personnel to inspect the relevant facilities regularly.

2.2.3. Annual inspection information module of hospital inspection instruments. The data of this system is based on the annual inspection data of the hospital, which includes the number of the hospital, the annual inspection time, the certificate number, the validity period and so on. This module can complete the retrieval of annual inspection records and original inspection certificates.

2.2.4. Information retrieval and information processing module. This module is a public module that realizes data query, report printing and output. By using departments, equipment names and annual inspection time, the equipment's basic information, annual inspection information and maintenance information can be retrieved and output in the form of reports [14]. In this system, the instruments that must be checked can be checked. The staff can print and output the report of the instrument.

3. Medical detection equipment system based on artificial intelligence algorithm. In the Internet of Things hardware network composed of multiple sensors, how to accurately and effectively identify the required information from the massive data is crucial to the operation of the medical inspection system. This paper presents a data mining method based on a neural network. Data exploration is a processing method to extract hidden information from extensive data [15]. It requires not only a specific data structure but also according to a specific calculation method to construct the processing process. This paper uses the artificial neural network method to construct the mining process. Firstly, feature extraction is carried out on the hardware of the sensor system [16]. Finally, the training samples are imported into the artificial neural network to achieve accurate medical information identification.

3.1. Feature Extraction. Medical information usually has the characteristics of differential entropy, rational asymmetry, energy spectrum, difference asymmetry, etc. Differential entropy is used to describe the random uncertainty of each band quantitatively. That is, the logarithm of the energy spectrum of a particular band is the difference entropy. ES is the average value of signals in 5 frequency bands, so this paper defines differential entropy as the characteristics of 5 frequency bands [17]. Asymmetric structures include rational asymmetry and differential asymmetry. DASM is defined as:

$$DASM = DE\left(B_{\text{left}}\right) - DE\left(B_{\text{right}}\right) \tag{3.1}$$

RASM is defined as:

$$RASM = DE\left(B_{\text{left}}\right) / DE\left(B_{\text{right}}\right) \tag{3.2}$$

DASM is used to represent the entropy difference of 28 groups of data. RASM represents the scaling property. These characteristics are treated with gradient dimension reduction [18]. The model is calculated according to the following formula:

$$f\left(a_n / a_{n-1}\right) = G\left(a_n \mid W a_{n-1}, \Delta\right) \tag{3.3}$$

$$f\left(c_n / a_n\right) = G\left(c_n \mid \Lambda a_n, \Theta\right) \tag{3.4}$$

where c_n represents the observed variable of the initial characteristics of the computer. a_n stands for implicit variable. W stands for implicit variable transformation matrix. Where Δ, Θ is the parameter of the neural network. Λ is the transmission matrix of the neural network.

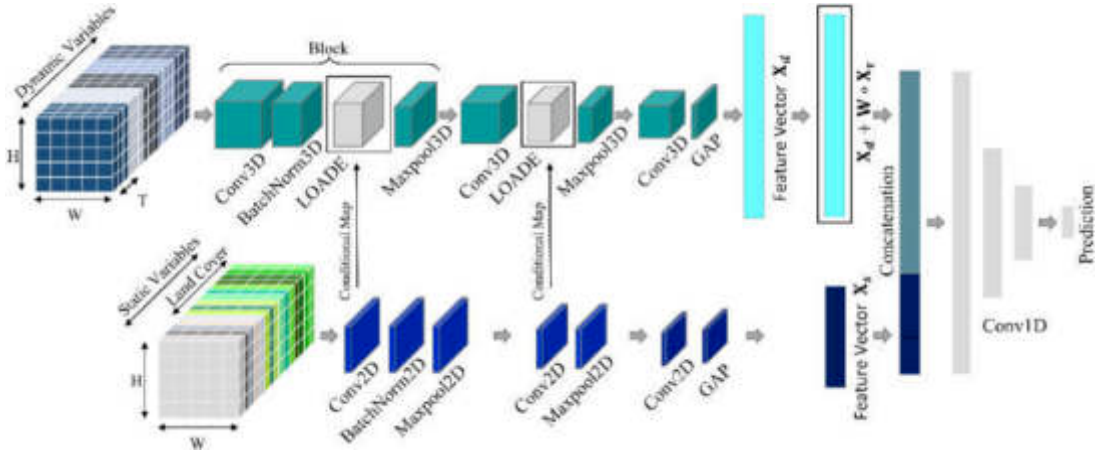


Fig. 3.1: *Structure diagram of adaptive neural network.*

3.2. Neural network model. The neural network structure designed in this paper is shown in Figure 3.1.

It can be seen from Figure 3.1 that the weight sum of the ANN network on excitation output at k_n is:

$$F_{j,t}(k_n) = \sum_{i \in G_{k-1}} \sum_{k=1}^{K_{jit,-1}} \chi_{jik,t-1} a_{i,t-1}(k_n - \gamma_{jik,t-1}) \tag{3.5}$$

In layer j Part l, the output of the excitation is:

$$a_{j,t}(k_n) = h[F_{j,t}(k_n)] \tag{3.6}$$

$h(a)$ is an excitation function, which is widely used in neural networks [19]. All Mses of the neural network at time k_n are:

$$S(k_n) = \frac{1}{2} \sum_{j \in G_2} [s_j(k_n) - a_{j,2}(k_n)]^2 \tag{3.7}$$

$s_j(k_n)$ is the expected output of the i neuronal network at time k_n. In most cases, the gradient descent algorithm is used to reduce error $O(k_n)$:

$$\Delta\chi_{jk,t-1} = -\gamma_1 \frac{\partial S(k_n)}{\partial \chi_{jkt,-1}}, t = 1, 2 \tag{3.8}$$

$$\Delta\gamma_{jk,t-1} = -\gamma_2 \frac{\partial S(k_n)}{\partial \gamma_{jkt,t-1}}, t = 1, 2 \tag{3.9}$$

The chain method is used to adjust the weight of the neural network:

$$\frac{\partial S(k_n)}{\partial \chi_{jk,t-1}} = \frac{\partial S(k_n)}{\partial F_{j,t}} * \frac{\partial F_{jL}(k_n)}{\partial \chi_{jk,t-1}} \tag{3.10}$$

and because:

$$\frac{\partial F_{j,t}(k_n)}{\partial \chi_{jik,t-1}} = a_{il-1}(k_n - \gamma_{jk,t,-1}) \tag{3.11}$$

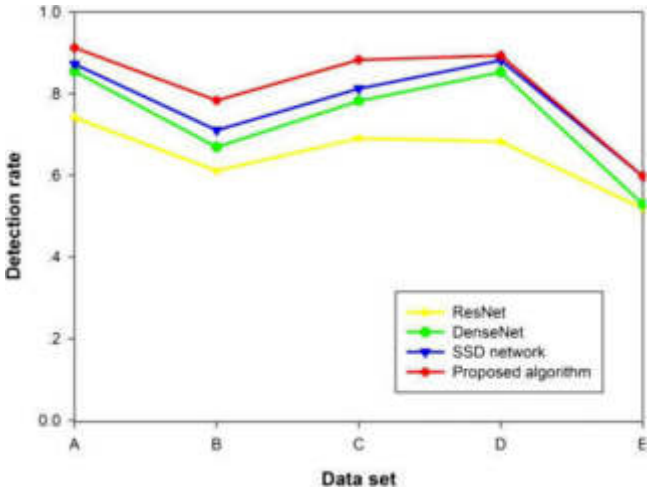


Fig. 4.1: *Efficiency diagram of the data processing algorithm.*

Table 4.1: *Success rate of system experiment.*

Number of experiments	Success rate (%)
100	98.96
200	96.88
300	95.83
400	93.75
500	92.71

There is:

$$\Delta\chi_{jkk,t-1} = \gamma_1 \xi_{j/t}\left(k_n\right) a_{L,t-1}\left(k_n - \gamma_{jk,t-1}\right), t = 1,2 \tag{3.12}$$

The correction of the delay parameters is done in the same way:

$$\Delta\gamma_{j\pi t,-1} = \gamma_2 \xi_{j,t}\left(k_n\right) \chi_{jd,t-1} a_{i,t-1} \cdot \left(k_n - \gamma_{jk,t-1}\right), t = 1,2 \tag{3.13}$$

To reduce the error in the artificial neural network identification process, the parameters obtained are more accurate by adjusting the weight and the lag time [20].

4. Simulation experiment. Finally, the management system of the medical testing instrument is tested and verified. Firstly, the data processing time consumption of the designed artificial intelligence algorithm and the literature method are compared. The experimental results show that the method has good performance [21]. As shown in Figure 4.1, when the amount of data required is small, the neural network algorithm designed takes the same time as other algorithms. However, with the increase in data size, the calculation time of these methods will change significantly, and the calculation time of the method proposed by Resnet will increase the most [22]. The results show that the proposed AI algorithm has more obvious advantages and can meet real-time processing needs. Finally, the comprehensive performance of the method is tested. The results are shown in Table 4.1.

As shown in Table 4.1, with a small number of tests, the method's success rate reached 98.96%. When the number of tests increases, the accuracy of the test will decrease, but the overall performance can meet the accuracy requirements. From the point of view of real-time and accuracy, the medical testing instrument management system with artificial intelligence technology proposed in this paper has a better working effect. It can be well adapted to the needs of the medical industry.

5. Conclusion. This study realized real-time monitoring and acquisition of patients' physiological characteristics data by integrating advanced sensor technology. The differential entropy algorithm is used to analyze the data deeply, and the critical information reflecting the health status of patients is extracted. The neural network model is constructed. This model can predict the health trends of patients based on historical data and provide decision support for the scheduling and optimization of medical detection equipment. The results show that the proposed system improves the management efficiency of medical testing equipment and significantly enhances the accuracy and foresight of medical services. Through the application of differential entropy algorithm, the system can effectively identify abnormal patterns in physiological signals, which provides the possibility for early diagnosis of diseases. The predictive function of the neural network model provides valuable early warning information for the medical staff and helps to adjust the treatment plan in time.

REFERENCES

[1] KHodadadi, V., Bakrani, A., & Vafaie, M. H. (2021). Factors affecting medical equipment management in the COVID-19 pandemic crisis: A mixed qualitative and quantitative study. Hospital Practices and Research, 6(1), 23-28.
[2] Hanifah, N. A., Kairudin, W. N. H. W., Mohd, E., & Ismail, M. N. (2021). Medical Equipment Maintenance Management System with Fingerprint Authentication. Evolution in Electrical and Electronic Engineering, 2(1), 236-245.
[3] Belhouideg, S. (2020). Impact of 3D printed medical equipment on the management of the Covid19 pandemic. The International journal of health planning and management, 35(5), 1014-1022.
[4] Kudratillaev, M., & Yakhshiboyev, R. E. (2023). The Role of Indigenous Medical Devices and Equipment Development in Medicine. Economics and Innovative Technologies, 11(5), 327-341.
[5] Chaminda, J. L. P., Dharmagunawardene, D., Rohde, A., Kularatna, S., & Hinchcliff, R. (2023). Implementation of a Multicomponent Program to Improve Effective Use and Maintenance of Medical Equipment in Sri Lankan Hospitals. WHO South-East Asia Journal of Public Health, 12(2), 85-92.
[6] Liu, Q., Zhang, Q., Guo, W., & Wang, T. (2022). Application Effect Analysis of RCA on Improving the Management of Medical Equipment-Related Near-Miss Events in ICU of Hospital. MEDS Clinical Medicine, 3(4), 115-120.
[7] Jain, A., & Rayal, S. (2023). Managing medical equipment capacity with early spread of infection in a region. Production and operations management, 32(5), 1415-1432.
[8] Aunión-Villa, J., Gómez-Chaparro, M., & García Sanz-Calcedo, J. (2020). Assessment of the maintenance costs of electromedical equipment in Spanish hospitals. Expert Review of Medical Devices, 17(8), 855-865.
[9] Khinvasara, T., Ness, S., & Tzenios, N. (2023). Risk Management in Medical Device Industry. J. Eng. Res. Rep, 25(8), 130-140.
[10] Daki, A. M., Mugop, N. S., Berinyuy, F. C., & Mbinkong, S. R. (2020). An assessment of measures medical imaging personnel use to avoid the breakdown of diagnostic radiologic equipment in some health facilities in Northwest and Southwest regions of Cameroon. Int J Innov Sci Res Technol, 5(8), 36-42.
[11] Suarez, O. M. D., Mármol, A. L., & Camperos, J. A. G. (2023). RCM-PMBOK hybrid methodology for managing the maintenance of critical biomedical equipment in IPS of medium complexity. Journal of Applied Engineering Science, 21(3), 795-809.
[12] Salami, S. R., Bamdadsoofi, J., Khatami Firouz Abadi, S. M. A., & Shafiee, M. (2021). Formation and Development of Technological Innovation Capability in New Technology Base Firms in the Medical Equipment Sector of Fars Province. Journal of Improvement Management, 15(1), 69-94.
[13] Ma, C., & Wen, L. (2020). Discussion on standardization management mode of medical equipment based on whole process quality control. Zhongguo yi liao qi xie za zhi= Chinese journal of medical instrumentation, 44(3), 270-275.
[14] Ali, H. (2024). Utilizing Predictive Analytics for Lifecycle Management and Maintenance of Medical Equipment: Utilizes machine learning algorithms to predict maintenance needs for medical equipment, reducing downtime and improving operational efficiency in healthcare facilities. Journal of AI in Healthcare and Medicine, 4(1), 57-70.
[15] Rahoma, M., Hwisa, S. A., Jira, M. A., Faraj, M. M., Ramih, E. A., Almesawey, S. A., & Godid, F. A. S. (2023). Evaluation of medical equipment of the infertility treatment centers in North western of Libya. Libyan Journal of Medical Research, 17(2), 154-167.
[16] Sarfraz, M., & Pamucar, D. (2024). A parametric similarity measure for spherical fuzzy sets and its applications in medical equipment selection. J. Eng. Manag. Syst. Eng, 3(1), 38-52.
[17] Li, K., Zhao, Y., Yang, Y., & Qian, Y. (2023). Introduction to Quality control guidelines for cleaning and sterilizing equipment. Sterile Supply, 1(3), 81-85.
[18] Singh, N., Tang, Y., & Ogunseitan, O. A. (2020). Environmentally sustainable management of used personal protective equipment. Environmental science & technology, 54(14), 8500-8502.
[19] Sutejo, J., Setiawan, S., & Syam, B. (2022). The experience of nurse managers implementing a nursing management system in COVID-19 wards: A descriptive phenomenology study. Jurnal Keperawatan Padjadjaran, 10(1), 73-82.
[20] Wang, X., Zhang, X., & He, J. (2020). Challenges to the system of reserve medical supplies for public health emergencies: reflections on the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic in China. Bioscience trends, 14(1), 3-8.
[21] Singh, N., Ogunseitan, O. A., & Tang, Y. (2022). Medical waste: Current challenges and future opportunities for sustainable

management. Critical Reviews in Environmental Science and Technology, 52(11), 2000-2022.
[22] Kumar, M., Fatma, A., & Bharti, N. (2022). Access to Medicines and Medical Equipment during COVID-19: Searching Compatibility between the WTO and the WHO. India Quarterly, 78(1), 68-87.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: JMay 28, 2024
Accepted: Aug 3, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 888–898, DOI 10.12694/scpe.v26i2.4057

IMPROVING NODE LOCALIZATION ACCURACY IN WIRELESS SENSOR NETWORKS BASED ON COMPUTER VISION AND DEEP LEARNING OPTIMIZATION

LIANJUN YI *

Abstract. In order to solve the problem of angular effects and reduced positioning accuracy caused by rapid speed changes in position tracking and positioning methods in wireless sensor networks, as well as the difficulty of improving positioning accuracy with a single solution, the author proposes a research on improving node positioning accuracy in wireless sensor networks based on computer vision and deep learning optimization. The author proposes a tracking and localization method using Kalman filtering (KF) and visual assistance on the TI CC2431 ZPS platform. On the basis of normalized cross-correlation, visual assistance calibration technology is used to extract the position of reference nodes as landmarks using visual assistance methods. Then, the KF method is used to calibrate the position estimation, which randomly generates virtual nodes for neural network training. Then, the priority positioning node is located and used as the anchor node for the next positioning, and the wireless loop is used for positioning calculation. The experimental results show that both the TI ZPS method and the KF based method have an estimated position error distance of over 55%, which is less than 2.2m and 1.8m, respectively, The proposed tracking and positioning method has an estimated position error distance of over 55% less than 1.4m. The method proposed by the author effectively avoids uncertainty caused by system errors in actual dynamic environments, reduces angular effects in position estimation systems, and improves positioning accuracy.

Key words: Wireless sensor network, Tracking and positioning, Kalman filtering, Visual assistance, ZPS platform

1. Introduction. With the development of information technology, data is growing at a massive speed, and more and more data is being discovered and utilized. Data that was once unattainable or unused can be collected under the background of new technologies, thus forming the era of big data [1]. Sensors, as an important source of information collection in the physical world, are widely used in various industries. Different sensors have different functions, and a large number of sensors use the same protocol to form a network called Wireless Sensor Networks (WSNs) [2,3]. Wireless sensor networks can achieve functions such as data perception, data collection, and data transmission. Usually, wireless sensors use hardware devices that collect information as wireless sensor nodes, such as smoke alarms, water level monitoring sensors, water quality monitoring sensors, video monitors, etc. These sensors are mainly used to collect certain data. The broad definition of wireless sensors refers to all devices that can provide data, which expands the range of wireless sensors from both ends [4]. In traditional application industries, data collection, data transmission, and other functions can be achieved by installing embedded devices. For example, in traditional washing machines, adding power starters and wireless receiving modules can achieve semi-automatic control. The washing machine can be operated through mobile phones, and shared bicycles can be shared in real-time by adding Bluetooth locks or 4G modules [5]. The emergence of wireless sensor technology has upgraded many functions in traditional industries, and traditional devices are empowered with new functions through new technologies. In the context of new technologies, with the development of 5G and cloud computing, a large number of new products come with the function of implanting wireless sensors. For example, autonomous vehicles are a new device filled with wireless sensors, such as SLAM photodetectors, steering wheel controllers, brake systems, path imaging, etc., all of which collect data through sensors and provide real-time feedback through data transmission. Various wireless sensors are no longer simple single individuals, but they become a collaborative whole. After being hard IPed by intelligent algorithms, sensors can even become execution agencies [6]. Wireless sensors with unlimited types, functions, manufacturers, and regions can form a wireless sensor network within a certain range using the same communication protocol. For example, ZigBee New communication methods such as Lora, NB IoT,

*School of Information Engineering, Gannan University of Science and Technology, Ganzhou, 341000, Jiangxi, China (Corresponding author, 9320240018@gnust.edu.cn)

Sigfox, and WiFi can be embedded in devices through communication modules for self-organizing networks.

2. Literature Review. Numerical optimization is an accurate iterative optimization method, and the quality of the solutions obtained greatly depends on the selection of the initial state [7]. When applied to node localization problems, improvements need to be made based on the inherent properties of the problem itself. For example, Mobile Forest Protection Network (FPWSN) has shown significant advantages in reducing forestry economic losses and improving fire prevention efficiency, Mobile Forest Protection Wireless Sensor Networks (FPWSN) have been widely used in the forest conservation sector. Nevertheless, at the current stage of development, the impact of sensor nodes on the system is frequently overlooked in forest fire prevention and control efforts. Xie, J. et al. devised a novel clustering method tailored for mobile FPWSN termed Boltzmann Adaptive Chaotic Salp Swarm Optimization Clustering (BACSSOC). This innovative approach aims to significantly extend the lifespan of the system, minimize energy usage, and diminish system latency [8]. Jin, Z. et al. explored strategies for deploying sensor nodes to minimize deployment costs while guaranteeing continuous target coverage. They utilized the Confidence Information Coverage (CIC) model to formulate the Minimum Deployment Cost Target Permanent Coverage (CICMTP) problem, with the objective of minimizing the sensor node count needed for deployment [9]. Yu, Q. et al. introduced a framework that integrates wireless sensor networks with the Internet of Things (WSN IoT SEC) for advanced environmental monitoring. The study leverages SEM technology to partition the analysis, exploring individual targets with the aid of detectors, machine learning models, and classifiers. A thorough investigation was undertaken, drawing insights from evaluation results and identified patterns, with a focus on highlighting crucial recommendations and underscoring the importance of SEM analysis [10].

The author proposes a position tracking system combined with visual assistance methods in the WSN environment, with the aim of exploring how to improve the corner effect caused by the tracking system. Compared to traditional position estimation, based on normalized cross correlation, the visual assistance method of the NCC method detects landmark positions as calibration, alleviating the angular effects caused by filtering and tracking techniques. More accurate position estimation was obtained in a static WSN environment.

3. Research Methods.

3.1. Basic knowledge.

3.1.1. State and observation equations. In a state-space representation of a dynamic system, if the system is described by a probability density, the state equation of the system measured by MT at time k is given by the following mathematical model 3.1:

$$x_{\lambda+1} = fun_x(x_k, u_K) \leftrightarrow f(x_{k+1}|x_k) \tag{3.1}$$

The observation equation is as follows 3.2:

$$z_k = fun_z(x_k, \epsilon_k) \leftrightarrow h(z_k|x_k) \tag{3.2}$$

Among them,x_k,$fun_k()$,u_k,z_k,$fun_z()$ and ϵ_k are respectively state vectors, transformation equations, process noise with a known distribution, observation vectors, observation equations, and observation noise with a known distribution. The mathematical model and measurement of linear dynamic systems are represented by the following equations 3.3 and 3.4:

$$x_{k+1} = \Phi_k x_k + u_k, u_k \sim N(0, Q_k) \tag{3.3}$$

$$E\{u_n, u_k^T\} = \begin{cases} Q_k, n = k \\ 0, n \neq k \end{cases} = \delta(k-n)Q_k \tag{3.4}$$

The observation equation is as follows 3.5,3.6:

$$z_k = H_k x_k + \epsilon_k, \epsilon \sim N(0, R_k) \tag{3.5}$$

$$E\{\epsilon_n, \epsilon_k^T\} = \begin{cases} R_k, n = k \\ 0, n \neq k \end{cases} = \delta(k-n)R_k \tag{3.6}$$

Among them, x_k,Φ_k,$u_k$and Q_k are the state matrix, state transition matrix, model noise matrix, and model noise covariance matrix, respectively [11,12]. u_k and ϵ_k is a Gaussian vector that is independent of the zero mean of the corresponding covariance matrices Q_k and R_k.

3.1.2. Kalman filtering. Assuming that vector $x = [x_1, \cdots, x_n]^T$ is composed of independent components, $i = 1, \cdots, n$. The PDF of x consists of independent PDFs of $x_1, \cdots, x_n$. $N(x; m, P)$ is The Gaussian density of n-dimensional vectors, the n-dimensional Gaussian density function is defined as follows 3.7:

$$N(x; m, P) \triangleq |2\pi P|^{-1/2} exp\{(-\frac{1}{2}(x-m)^T P^{-1}(x-m))\} \tag{3.7}$$

Among them, x, m, P are parameters, mean, and covariance, respectively. At time $t = t_k$, the vector $x(t)$ value is x_k. At time $t = t_k$, vector $x(t)$ gives the estimated value $x_{k|j}$ at time $t = t_j$ is represented by double subscripts [13].

(1) Prediction stage (time update stage). From From k to k+1, the state prediction and prediction error covariance are calculated using the following equations 3.8, 3.9, and 3.10:

$$\widetilde{x_{k+1}} = \phi_k \hat{x_k} \tag{3.8}$$

$$\widetilde{P_{k+1}} = \Phi_k \hat{P_k} \Phi_k^T + Q_k \tag{3.9}$$

$$\hat{P_k} = \{\hat{e_k} \hat{e_k}\}, \widetilde{P_k} = E\{\tilde{e_k} \tilde{e_k}^T\} \tag{3.10}$$

Among them, $e_{kj} \triangleq x_{k|j} - x_k, e_{k|k} \triangleq \widehat{x_k} - x_k = \hat{e_k}, e_{k|k-1} \triangleq \widetilde{x_k} - x_k = \tilde{e_k} \circ \widehat{x_k}$, $\widetilde{x_k}, e_{k|j}, \hat{e_k}, \tilde{e_k}$ are state estimation matrix, the status forecast matrix, the status error matrix, the estimated error matrix, and the forecast error matrix are respectively[14].

(2) Innovation stage. The calculation formula for the innovation stage is as follows 3.11,3.12:

$$zz_k = z_k - H_k \tilde{x_k} \tag{3.11}$$

$$K_k = \tilde{P_k} H_k^T [H_k, \tilde{P} H_k^T + R_k]^{-1} \tag{3.12}$$

Among them, zz_k and K_k are the innovation matrix and the Kalman gain matrix, respectively.

(3) Correction phase. The state estimation and the update of the estimation error covariance are as follows 3.13,3.14:

$$\hat{x_k} = \tilde{x_k} + K_k zz_k \tag{3.13}$$

$$\hat{P_k} = [I - K_k H_k] \tilde{P_k} \tag{3.14}$$

Among them, I represents the identity matrix.

3.1.3. Normalized cross-correlation. To determine a reference node as a route marker (RN) the author combined a pattern recognition method with RN (ID) for estimating position calibration [15]. It is well-known that NCC is one of the most basic area-based matching techniques. The NCC scheme is extensively utilized in image processing applications, as it can mitigate brightness discrepancies between images and templates resulting from varying lighting conditions. The NCC equations used by the author are 3.15, 3.16, 3.17, 3.18, 3.19, and 3.20:

$$G_T = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} G_T(x_i, y_j)}{n \cdot m} \tag{3.15}$$

$$G_S = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} G_S(x_i, y_j)}{n \cdot m} \tag{3.16}$$

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(G_T(x_i, y_j) - G_T)^2}{n \cdot m - 1}} \tag{3.17}$$

$$\sigma_S = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(G_S(x_i, y_j) - G_S)^2}{n \cdot m - 1}} \tag{3.18}$$

$$\sigma_T S = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}[(G_T(x_i, y_j) - G_T) \cdot (G_S(x_i, y_j) - G_S)]}{n \cdot m - 1} \tag{3.19}$$

$$r = \frac{\sigma_{TS}}{\sigma_T \sigma_S} \tag{3.20}$$

Among them, $G_T(x, y)$ and $G_S(x, y)$ are grayscale images of the target window and search window, respectively; G_T and G_S are the grayscale values of the target window and search window, respectively; m and n are the number of rows and columns, respectively.

3.1.4. ZigBee positioning system. The ZigBee network, which is IEEE 802.15.4, is designed to facilitate the communication of information at a low, low-power and cost-free time while maintaining the route. In Figure 1, a fundamental location system utilizing a ZigBee network is shown, wherein the Received Signal Strength Indicator (RSSI) is used to measure the power of a received wireless signal. The ZigBee Positioning System (ZPS) includes three kinds of nodes, namely, an index, a RN, and a BN. The coordinator is attached directly to the computer as shown in Figure 1. The RNs, which are located in a known position, send to the BN their equipment ID and its coordinates. Each stationary RN is located at a predefined location. A BN node collects a signal from an RN in accordance with an RSSI value for an RN. When a BN receives a signal from an RN, it identifies a distance between a BN and an individual RN by means of the RSSI sample collected together with a route loss model. The location of the BN is then computed according to the RNN coordinates. Eventually, the BN's estimated location is transferred through a WSN to a Location-Based Services (LBS) app.

3.2. Proposed algorithm. The method for measuring uncertainty in WSN utilizes the widely-used commercial system ZPS, which is based on the CC2431 positioning engine developed by Texas Instruments (TI). Detailed technical information about the TI CC2431 can be found at http://www.ti.com/product/cc2431. The TI CC2431 is a hardware localization engine specifically designed for low-power local area network (ZigBee) applications within WSN [16-18]. In Ad Hoc wireless networks, it is used to evaluate BN position estimation based on RSSI values (from RN, obtained by centralized computing methods). According to the RSSI value, The TI CC2431 positioning engine independently outputs (X , Y) Coordinates. The author focuses on having independent (X , Y) The localization and tracking scheme for the group.

3.2.1. Formulaization of problems. To enhance the accuracy of position tracking, the tracking method can be reformulated as a filtering problem. This involves integrating Kalman Filter (KF) based techniques into position estimation systems using tracking algorithms to enhance the accuracy of the positioning system. While the state and measurement models typically operate within two-dimensional linear Gaussian systems, extending the approach to three-dimensional models is straightforward within the framework of filtering methods. The two-dimensional model vector $x_k = [x_{1,k}, x_{2,k}, \dot{x}_{1,k}, \dot{x}_{2,k}]^T$ represents the state of MT at time k, where, $x_{1,k}$ and $x_{2,k}$ respectively represent X , Y coordinate. $\dot{x}_{1,k} = s_{1,k}$ and $\dot{x}_{2,k} = s_{2,k}$ are represent the velocities in the X and Y directions, respectively [19]. The speed model for the MT is founded on speed noise. The two-dimensional model of the MT characterizes its movement and observed position by incorporating random elements, and the

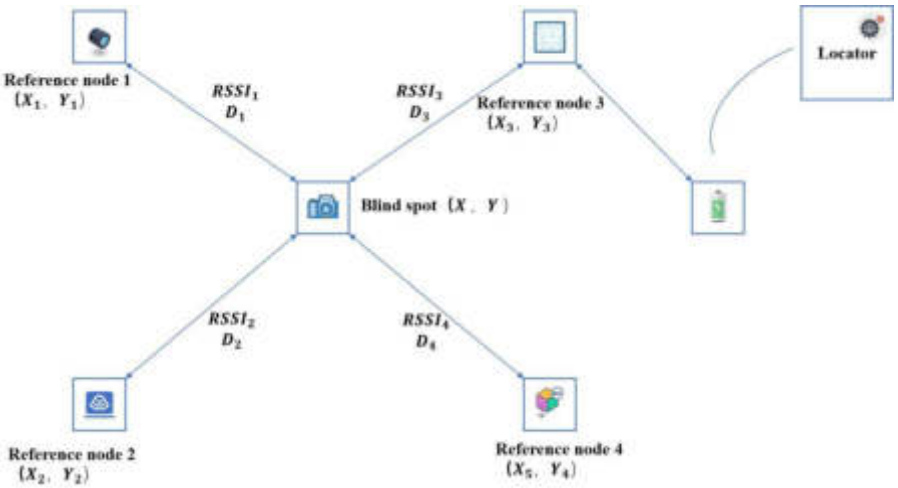


Fig. 3.1: Simple positioning system based on ZigBee network

formula for calculation is provided in equations 3.21 and 3.22.

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \dot{x}_{1,k+1} \\ \dot{x}_{2,k+1} \end{bmatrix} = \begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ s_{1,k+1} \\ s_{2,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta_k & 0 \\ 0 & 1 & 0 & \Delta_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \dot{x}_{1,k+1} \\ \dot{x}_{2,k+1} \end{bmatrix} = \begin{bmatrix} u_{1,k+1} \\ u_{2,k+1} \\ u_{3,k+1} \\ u_{4,k+1} \end{bmatrix} \tag{3.21}$$

$$\begin{bmatrix} z_{1,k} \\ z_{2,k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \dot{x}_{1,k+1} \\ \dot{x}_{2,k+1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,k} \\ \epsilon_{2,k} \end{bmatrix} \tag{3.22}$$

Among them, Δ_k is the measurement period between k and k+1. The difference between equations 3.3 and 3.5 and equations 3.21 and 3.22 can be represented by $u_k = [u_{1,k}, u_{2,}, u_{3,k}, u_{4,k}]^T$ for process noise; $z_k = [z_{1,k}, z_{2,k}]$ and $\epsilon_k = [\epsilon_{1,k}, \epsilon_{2,k}]$ represents respectively Observation information and measurement noise of MT at time k [20].

3.2.2. Visual assisted position estimation. In the WSN environment, signals are affected by reflection, diffraction, scattering, and attenuation effects during propagation, and the instability of RSSI values can also affect positioning accuracy. The signal feedback of the TI CC2431 positioning engine is greatly affected by RSSI information, which can reduce positioning accuracy [21]. To enhance positioning accuracy, a technique leveraging auxiliary markers is employed to counteract angular distortions caused by filtering and tracking methods in dynamic, changing environmental conditions. This method facilitates calibration and positioning based on sensor landmark positions. RN landmarks are extracted from video features using the NCC method for precise position estimation. Moreover, to conserve energy and extend the lifespan of intelligent mobile terminals utilizing the NCC method for position estimation, this approach enables the simultaneous operation of two modes based on a shared threshold: Sleep mode and activity mode. The threshold, derived from RSSI and ZPS testing platforms, determines whether the visual assistance algorithm remains in active mode based on the detected RSSI level [22]. As the BN nears the RN, the video camera of the intelligent mobile terminal initiates path recording and subsequently identifies landmarks using the NCC method. Conversely, when the RSSI surpasses the threshold, the visual assistance algorithm transitions into sleep mode, halting path recording by the video camera. Figure 3.2 illustrates an instance of the visual assistance scheme utilizing the NCC method. Experimental findings demonstrate the efficacy of the NCC method in accurately extracting RN from video features as the correct landmark.

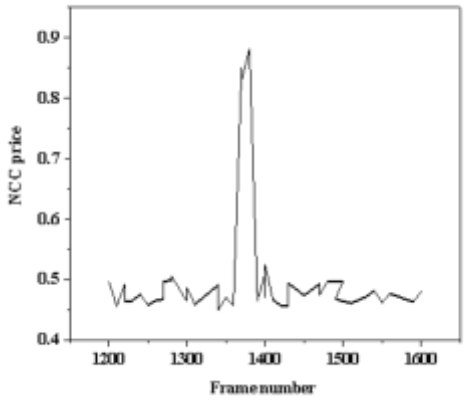


Fig. 3.2: Example of extracting RN as a landmark

3.3. Analysis of RBF-HOP localization algorithm.

3.3.1. The process of positioning algorithm. The positioning algorithm mode adopted by the author is centralized. In this model, the key point of this model is to transmit the hop number of the WSN to one of the WSN's non-known nodes and its anchor points, and then analyzes and determines the specific positions of each unknown node through this processing center. This approach doesn't demand high hardware specifications for wireless sensors and concurrently lowers the energy consumption of wireless sensor node processing during positioning operations. However, this computational technique necessitates awareness of the minimum hop count value between nodes within the wireless sensor network. The positioning calculation process mainly includes the following points: Firstly, adjust the hop count of each anchor node to 0, that is, adjust it to the initialization position, so that further calculations can be carried out; Secondly, once an anchor node has been identified, it will be able to transmit a particular position message to the adjacent nodes; Thirdly, for each neighboring node that receives information packets, the hop count should be updated to the anchor node's hop count, and then an additional 1 should be added to this hop count. After that, the information packets should be propagated to the next adjacent node; Upon receiving hop count information from the preceding node, the subsequent node adjusts its hop count to ensure that each node's hop count reaches the minimum value. Then, the next hop count information is propagated to other nodes; The fifth step is to repeatedly update the node hop count operation until the node hop count information no longer changes [23].

For instance, if the sensor has if the whole WSN has N nodes, and there are M anchor nodes ($m \leq$), then Among m anchor nodes may be determined to be n, the planar position coordinates of the t-th node can be represented as $m_t = (x_t, y_t)$. In the process of calculating the minimum hop count of an unknown node, the hop count information is directly sent to the service processing center, and the information of that node is recorded and processed. Randomly generate several virtual nodes in a wireless sensor network, and number them in order, counting them as 1, 2, 3, etc. The node plane coordinates are marked according to the numbers. Then, using the known radiation radius of the sensor nodes, combined with the hop calculation process, calculate the specific hop count value from each virtual node to the anchor node. Record their hop count in a matrix format, and input the matrix hop count into the input layer of the RBF neural network for training, after the training is completed, its corresponding expected output value is the actual coordinates calculated by the node [24]. Moreover, it is also possible to take the hop count of the anchor point as a training input in RBF net and compute the coordinate of the anchor point.

3.3.2. Localization of Unknown Nodes. The RBF neural network has memory after being trained with input from nodes, and can use its memory function to simulate the hop characteristics and position relationships in wireless sensor networks. There is a matrix hop count between each unknown node and the

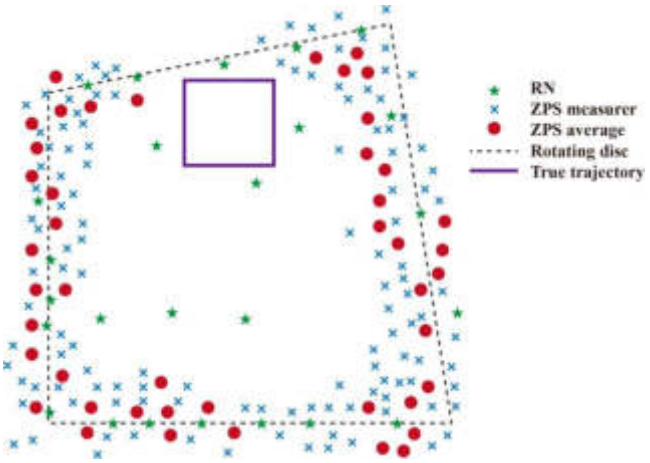


Fig. 4.1: Position estimation results of MT (BN) moving along the test path in the TI ZPS platform

anchor node, so in the process of selecting priority positioning nodes, the hop count matrix can also be selected. For example, when the hop matrix of the anchor node is $nm_t = [h_{t1}h_{t2}h_{t3}\cdots h_{tm}]$ and the hop matrix is input into the trained RBF neural network, there are: Nodes the coordinates of n_t are (x_t, y_t). The localization algorithm process for its unknown nodes is as follows: firstly, randomly generate several unknown nodes and anchor nodes, assuming that the number of unknown nodes is n and the number of anchor nodes is m; The second is to randomly generate several virtual nodes, assuming that the number of virtual nodes is denoted as v; Thirdly, utilizing virtual nodes to train RBF neural networks; The fourth is to select the optimal localization node from several unknown nodes, and then input it into the network for localization; The fifth is to use a series of nodes with known positions as anchor nodes; The sixth step is to locate the loop nodes until all unknown nodes are accurately located.

4. Result analysis.

4.1. Experimental setup. The experimental platform is located on the top floor of the Remote Sensing Research Center, with a sampling distance of 1 meter, RNs (1-26) are widely distributed on rooftops, with 15 RNs distributed in the closed-loop path of sampling locations. In addition, in order to obtain the optimal and precise position of the sampling points, the isotropic radiation characteristics of the antenna are enhanced in the positioning system. Hence, to enhance the accuracy of RSSI measurements, a straightforward antenna radiation mapping approach was employed. This involved positioning the BN on a turntable capable of rotating to ensure that the RN's radiation antenna consistently faced north, west, south, and east. RSSI information was then recorded at various distances between the RN and BN (ranging from 1 to 10 meters) in each direction. This data was subsequently input into the TI CC2431 system to estimate the location. Nonetheless, within wireless network systems, positional errors depend on the information environment RN deployment mode and density. Usually, the more RNs available within the same given area, the more accurate the localization.

4.2. ZigBee positioning system. In wireless network systems, ZPS uses the RSSI value of the TI CC2431 positioning engine for localization. This experiment used a visual assistance scheme combined with TI CC2431 ZPS platform for position estimation and trajectory tracking. As the MT traverses the designated testing path, the experimental outcomes of position estimation based on the TIZPS platform are depicted in Figure 4.1. In this representation, the symbol "x" denotes the estimated position, representing the ZPS measurements. To validate the proposed scheme's experimental findings, Figure 4.1 showcases the corresponding positional parameters associated with the estimated results. It is assumed, without loss of generality, that the MT in equation 3.21 maintains a stable state, with the measurement period (sample time) between k and k+1 set to 1 second. Additionally, Figure 4.1 illustrates the visual position as defined in equation 3.22.

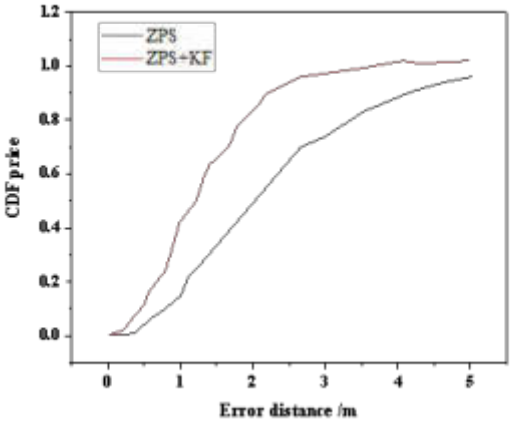


Fig. 4.2: CDPF of error distance for ZPS (observation) and KF based tracking methods

4.3. Position estimation based on Kalman filtering. Figure 4.2 shows the results of the simulation for the location estimation in a tracking system. The results show that the location precision obtained by this method is superior to that of nontracking. In particular, Kalman Filter (KF), being based on the optimum tracking of a Gauss model, usually produces the best linear estimation in the mean square error. Therefore, it is possible to consider the location precision in the KF tracing scheme as an upper limit of the CDF (CDF) for location estimation under Gauss circumstance [25]. The results show that the TI ZPS scheme has an estimated position error distance of over 56% less than 2.2 meters; With the KF-based approach, over 55% of estimated position error distances fall below 1.8 meters. The precision of KF tracing is better than that of ZPS nontracking, as demonstrated by the experiments in Figure 4.3. Notably, the KF method's hallmark is its recursive minimum mean square error state algorithm. Consequently, the position error observed in the KF tracking algorithm can serve as the upper bound for the cumulative distribution function (CDF) across various position estimation methods.

4.4. Visual assisted position estimation. Based on the experience values provided by TI company, The TI CC2431 positioning engine sets the threshold for RSSI in the ZPS platform to 60. When the Received Signal Strength Indicator (RSSI) exceeds 60, the visual assistance algorithm shifts into active mode, initiating landmark detection via the NCC method. Figure 4.3 showcases the experimental outcomes of a visual assistance scheme utilizing the NCC method. The 'r' value denotes the most probable video frame when encountering landmarks. Drawing insights from the experimental findings depicted in Figure 4.3, When the r value is around 0.6, most high peaks can be clearly distinguished, and the optimal value is estimated to be 0.62. Therefore, in the NCC method The threshold for r value is set to 0.62, at this point, RN can accurately detect landmarks and achieve good results. The results indicate that over 55% of estimated position error distances are less than 1.4 meters. The experimental results in Figure 4.3 show that the proposed scheme has better positioning accuracy than non visual assisted methods. Visual assistance methods can provide high-precision positioning estimation and trajectory tracking. The combination of NCC based visual technology and KF based algorithms shows that the tracking scheme proposed by the author results in close proximity to MT's tracking and positioning. Therefore, the visual assistance method based on NCC method can effectively solve and overcome the impact of dynamic time changing environments on the path. Due to the suitability of this position tracking platform for various practical applications.

In scenarios where GPS signals are unavailable, such as indoors or in outdoor environments with signal blockages, continuous real-time position tracking becomes challenging. Hence, integrating various strategies into position estimation systems becomes crucial for enhancing positioning accuracy. Combining visual assistance

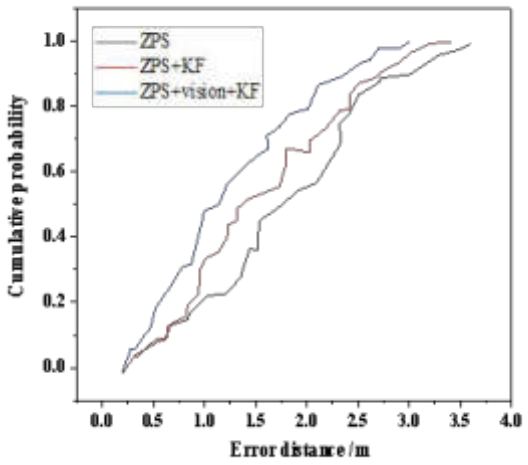


Fig. 4.3: Comparison of ZPS (observation), KF based and visually assisted KF trajectory tracking methods

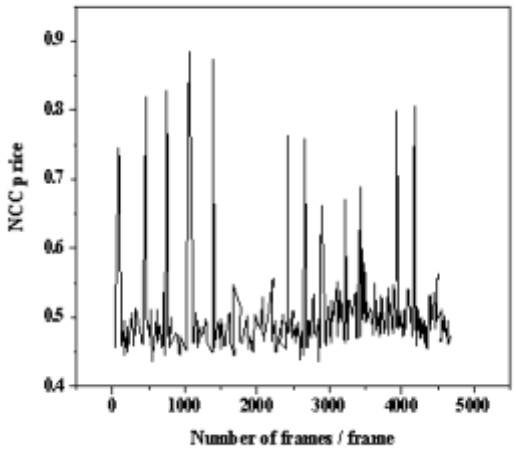


Fig. 4.4: RN image matching results based on NCC method when MT (BN) moves along the test path

methods with position tracking technology forms a pivotal component, acting as a fuse within multi-sensor positioning systems. This fusion approach significantly improves the accuracy of MT position estimation in Location-Based Services (LBS) applications.

5. Conclusion. The author proposes a visual assisted method for position estimation and trajectory tracking based on the WSN environment. In the ZPS platform, based on the KF method, MT can accurately track position changes and improve positioning accuracy. In addition, The NCC process matches RN images and extracts RN as landmarks from video features. In a stable operational setting, the KF tracking scheme leverages RN information to refine position estimation, effectively mitigating angular distortions. Experimental findings underscore the superior accuracy of the proposed scheme over non-tracking and non-visual assistance

methods. Notably, more than 55% of estimated positions exhibit error distances of less than 1.4 meters. By integrating visual assistance and KF tracking methodologies within the ZPS platform of WSN, the author's proposed positioning and tracking platform offers significant enhancements compared to standalone approaches, which has strong advantages for various LBS applications. The author only considered static WSN, mainly to facilitate formal description of the problem. In the future, more challenging dynamic WSN tracking and localization problems will be considered.

6. Acknowledgement. Science and Technology Project of Jiangxi Provincial Department of Education (No.GJJ2204405)

REFERENCES

[1] Kim, T., Vecchietti, L. F., Choi, K., Lee, S., & Har, D. (2020). Machine learning for advanced wireless sensor networks: A review. IEEE Sensors Journal, 21(11), 12379-12397.
[2] Kumar, M., Mukherjee, P., Verma, K., Verma, S., & Rawat, D. B. (2021). Improved deep convolutional neural network based malicious node detection and energy-efficient data transmission in wireless sensor networks. IEEE Transactions on Network Science and Engineering, 9(5), 3272-3281.
[3] Nurlan, Z., Zhukabayeva, T., Othman, M., Adamova, A., & Zhakiyev, N. (2021). Wireless sensor network as a mesh: Vision and challenges. IEEE Access, 10, 46-67.
[4] Lin, C., Han, G., Qi, X., Du, J., Xu, T., & Martínez-García, M. (2020). Energy-optimal data collection for unmanned aerial vehicle-aided industrial wireless sensor network-based agricultural monitoring system: A clustering compressed sampling approach. IEEE Transactions on Industrial Informatics, 17(6), 4411-4420.
[5] Ouyang, A., Lu, Y., Liu, Y., Wu, M., & Peng, X. (2021). An improved adaptive genetic algorithm based on DV-Hop for locating nodes in wireless sensor networks. Neurocomputing, 458, 500-510.
[6] Safaldin, M., Otair, M., & Abualigah, L. (2021). Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks. Journal of ambient intelligence and humanized computing, 12, 1559-1576.
[7] Dao, T. K., Nguyen, T. D., & Nguyen, V. T. (2023). An improved honey badger algorithm for coverage optimization in wireless sensor network. Journal of Internet Technology, 24(2), 363-377.
[8] Xie, J., Zhang, M., Jin, B., Zhai, J., Wang, Z., & Xiao, J., et al. (2023). Bacssoc: a novel clustering method for mobile forest protection using wireless sensor network with lower energy consumption and lower latency. Simulation modelling practice and theory: International journal of the Federation of European Simulation Societies, 114(3), 2017-2042.
[9] Jin, Z., Geng, Y., Zhu, C., Xia, Y., Deng, X., & Yi, L., et al. (2024). Deployment optimization for target perpetual coverage in energy harvesting wireless sensor network. Digital Communications and Networks, 10(2), 498-508.
[10] Yu, Q., Xiong, F., & Wang, Y. (2022). Integration of wireless sensor network and iot for smart environment monitoring system. Journal of Interconnection Networks, 22(Supp02), 11975-12023.
[11] Cui, Y., Zhang, L., Hou, Y., & Tian, G. (2021). Design of intelligent home pension service platform based on machine learning and wireless sensor network. Journal of Intelligent & Fuzzy Systems, 40(2), 2529-2540.
[12] Kunhoth, J., Karkar, A., Al-Maadeed, S., & Al-Ali, A. (2020). Indoor positioning and wayfinding systems: a survey. Human-centric Computing and Information Sciences, 10, 1-41.
[13] Liu, F., Liu, J., Yin, Y., Wang, W., Hu, D., Chen, P., & Niu, Q. (2020). Survey on WiFi-based indoor positioning techniques. IET communications, 14(9), 1372-1383.
[14] Farahsari, P. S., Farahzadi, A., Rezazadeh, J., & Bagheri, A. (2022). A survey on indoor positioning systems for IoT-based applications. IEEE Internet of Things Journal, 9(10), 7680-7699.
[15] Holmqvist, K., Örbom, S. L., Hooge, I. T., Niehorster, D. C., Alexander, R. G., Andersson, R., ... & Hessels, R. S. (2023). RETRACTED ARTICLE: Eye tracking: empirical foundations for a minimal reporting guideline. Behavior research methods, 55(1), 364-416.
[16] Liu, S., Liu, D., Muhammad, K., & Ding, W. (2021). Effective template update mechanism in visual tracking with background clutter. Neurocomputing, 458, 615-625.
[17] Chen, X., Li, Z., Yang, Y., Qi, L., & Ke, R. (2020). High-resolution vehicle trajectory extraction and denoising from aerial videos. IEEE Transactions on Intelligent Transportation Systems, 22(5), 3190-3202.
[18] Li, Y., Zhuang, Y., Hu, X., Gao, Z., Hu, J., Chen, L., ... & El-Sheimy, N. (2020). Toward location-enabled IoT (LE-IoT): IoT positioning techniques, error sources, and error mitigation. IEEE Internet of Things Journal, 8(6), 4035-4062.
[19] Kuvondikovna, K. S., & Hakima, B. (2024). Designing Visual Aid. European Journal of Higher Education and Academic Advancement, 1(2), 165-167.
[20] Mbanda, N., Dada, S., Bastable, K., & Ingalill, G. B. (2021). A sco** review of the use of visual aids in health education materials for persons with low-literacy levels. Patient education and counseling, 104(5), 998-1017.
[21] Khan, M. A., Paul, P., Rashid, M., Hossain, M., & Ahad, M. A. R. (2020). An AI-based visual aid with integrated reading assistant for the completely blind. IEEE Transactions on Human-Machine Systems, 50(6), 507-517.
[22] Al Aqad, M. H., Al-Saggaf, M. A., & Muthmainnah, M. (2021). The impact of audio-visual aids on learning English among MSU third-year students. ENGLISH FRANCA: Academic Journal of English Language and Education, 5(2), 201-214.
[23] Martiniello, N., Eisenbarth, W., Lehane, C., Johnson, A., & Wittich, W. (2022). Exploring the use of smartphones and tablets

among people with visual impairments: Are mainstream devices replacing the use of traditional visual aids?. Assistive Technology, 34(1), 34-45.
[24] Sokhiba, R. (2024). EFFECTIVENESS OF VISUAL AIDS IN TEACHING LANGUAGE. Journal of new century innovations, 51(1), 93-96.
[25] Chiekezie, P. N., & Inyang, M. J. P. (2021). The significance of audio-visual aids in teaching of English vocabulary. GNOSI: An Interdisciplinary Journal of Human Theory and Praxis, 4(2), 54-70.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 29, 2024
Accepted: Jul 13, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 899–906, DOI 10.12694/scpe.v26i2.4066

A PRECISE HEALTH FOLLOW-UP MANAGEMENT INFORMATION SYSTEM FOR COMMUNITY CHRONIC DISEASES BASED ON BIG DATA ANALYSIS

QINGTIAN MIAO*

Abstract. To enhance the efficiency of medical big data utilization, the author introduces a study on a precise health follow-up management information system tailored for community chronic diseases, leveraging advanced big data analytics. The system includes a meticulously designed chronic disease health record management framework, optimized for efficient ETL (Extract, Transform, Load) processes of medical data. This approach aims to minimize indexing overhead, enhance query execution speed and search capabilities, and maximize the use of aggregated computing resources. The system can be decomposed into four modules: ELT, data block creation, index creation, and querying. As an intermediate layer between users and distributed data management systems, this system can provide data upload, query execution mechanisms, and provide indexes to facilitate data search operations. The experimental results show that when the simulation step size is 50t, the maximum amount of data can reach 10 × 104. This method has a much higher retrieval efficiency than heuristic algorithms when processing massive data, further verifying the effectiveness and practicality of the chronic disease health record management system.

Key words: Big data, Chronic diseases, Medical health, Archive management, data retrieval

1. Introduction. Chronic diseases, due to their long incubation period and slow onset, are often overlooked by people. However, with the increasing desire and pursuit for a better life, people's emphasis on health is becoming more and more important. In the medical field, it's widely recognized that the majority of chronic disease patients are low-risk individuals. Traditional approaches to chronic disease management typically emphasize patient self-care to prevent disease progression. However, the development of a chronic disease health management system utilizing big data offers promising prospects for enhancing the overall management and care standards for these patients [1,2,3].

Big data has revolutionized the current situation of disease and health management, gradually shifting from a traditional doctor centered management model to a patient-centered model. This transformation truly realizes patient-centered health management. In the context of big data, patients in the chronic health management system upload their health sign information data to the cloud platform through smart devices they wear. Health care organizations and community health service institutions can link to the cloud platform server of health management data to obtain various physical health data of designated patients. The disease and health management network built with cloud servers as the core reduces the difficulty of managing diseases with large amounts of data. Through the constructed chronic disease health management system, medical institutions can carry out risk assessment on patients' health information at any time without the regional restrictions of time and space, and for patients who need medical guidance, they can achieve point-to-point health guidance for patients who need medical guidance by means of Internet social platform, telephone, SMS and other communication methods [4]. We provide comprehensive intelligent health management and medical services for patients with long incubation periods and high risk randomization rates through long-term tracking, testing, and risk warning. Big data has promoted the integration of medical institutions at all levels, reducing management costs and communication costs for patients. The implementation of a big data-driven chronic disease health management system has facilitated seamless communication across all levels of medical institutions. It has enabled automated and intelligent connectivity among healthcare providers, significantly enhancing the management and care coordination for chronic disease patients. When patients with chronic diseases change diagnosis and treatment institutions, the current reception medical structure can directly obtain a series of

*Institute of Population Health, Faculty of Health & Life Sciences, University of Liverpool, Liverpool, UK (miaoqingtian0528@163.com)

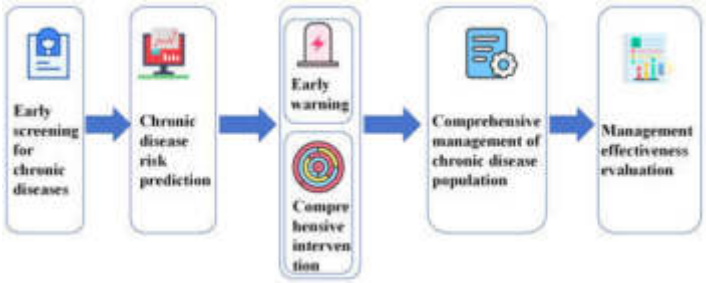


Fig. 3.1: Schematic diagram of chronic disease management

health signs and treatment data from the cloud server database, greatly reducing the medical cost of patients, eliminating repetitive inspections and laboratory processes, and greatly improving treatment efficiency [5,6].

2. Literature Review. As medical information technology advances, the growth rate and scope of medical data are rapidly increasing. The advent of the big data era is unlocking fresh opportunities in health and healthcare sectors, paving the way for transformative advancements. The concept of big data itself is not new in the field of healthcare. Healthcare providers not only improve the quality and details of handling large amounts of medical records, especially chronic diseases, but also continuously increase their scale due to technological advancements. Medical big data primarily originates from sources such as medical records, MRI scans, CT scans, health monitoring data, and genomic data. However, these datasets can often be incomplete, biased, and contaminated with noise. Factors like fuzzy information, redundancy, noise, and high dimensionality significantly hinder the effective utilization of medical data [7]. Li, Q. and colleagues aim to gather extensive data from the sports industry and employ data mining techniques along with neural network methods. Their objective is to comprehensively analyze and predict correlations within sports economic data, offering valuable management insights for companies in the sports industry [8]. Son, J. et al. have devised an analytical model for crafting an optimal alert strategy in asthma management. Their research findings offer actionable insights that can benefit patients, healthcare providers, and healthcare companies specializing in technical support [9]. Kiryu, Y. et al. introduced a domain driven design and development example of an artificial intelligence analysis system for clinical practitioners of traditional Chinese medicine to effectively interpret computational data and minimize noise, with the aim of further utilizing medical big data and artificial intelligence analysis [10].

However, traditional indexing techniques often underperform when applied to medical big data. Hence, the author's focus is on studying ETL management of medical data, aiming to reduce index overhead, enhance query execution, improve search performance, and meet the processing demands of medical and health record management systems while optimizing the utilization of computing resources. This research considers both the creation time and size of indexes to minimize overhead. Additionally, a simulation case study evaluates index traversal time and data retrieval time to assess query performance and search efficiency.

3. Method.

3.1. Definition of Chronic Disease Management. Chronic disease management involves a systematic approach of regular testing, continuous monitoring, evaluation, and comprehensive intervention aimed at managing chronic non-communicable diseases and their risk factors. It encompasses early screening and risk prediction, proactive early warning systems, comprehensive interventions, population-wide management, and effectiveness evaluations. Figure 3.1 illustrates this process. The goal of chronic disease management is to provide holistic, continuous, and proactive care for patients, promoting healthy lifestyles, delaying disease progression, lowering disability rates, improving quality of life, and reducing healthcare costs.

3.2. Characteristics of chronic disease management.

(1) Long term. The development of chronic diseases is mainly caused by exposure to occupational and environmental factors, lifestyle and behavioral patterns. The course of chronic diseases is long, and as the

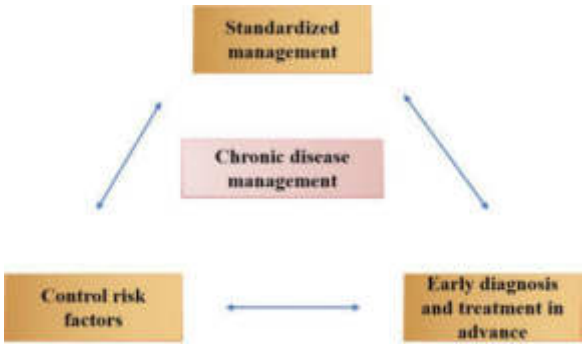


Fig. 3.2: Key Points of Chronic Disease Management

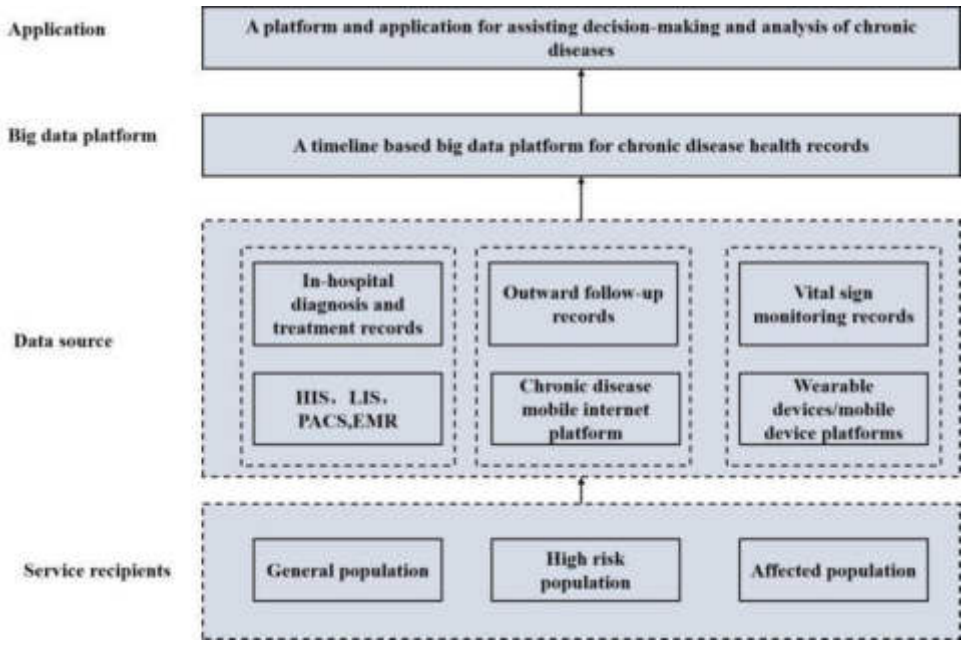


Fig. 3.3: Conceptual model of chronic disease big data

disease progresses, it manifests as progressive functional impairment or disability, causing serious damage to health.

(2) Normativeness. The prevention and treatment of chronic diseases is a continuous process, and chronic disease intervention needs to focus on controlling risk factors, early diagnosis and treatment, and standardized management. See Figure 3.2. Among them, standardized management is the key to the management and treatment of chronic diseases.

3.3. Conceptual model of chronic disease big data . From the perspective of chronic disease management, chronic diseases mainly target three types of population: General population, high-risk population, and diseased population, achieving management of pre hospital, in hospital, and post hospital processes. The conceptual model of chronic disease big data is shown in Figure 3.3.

The management and service targets for chronic diseases include three categories of population: general population, high-risk population, and diseased population. Different intervention measures can be taken to achieve the benefits of chronic disease intervention [11,12].

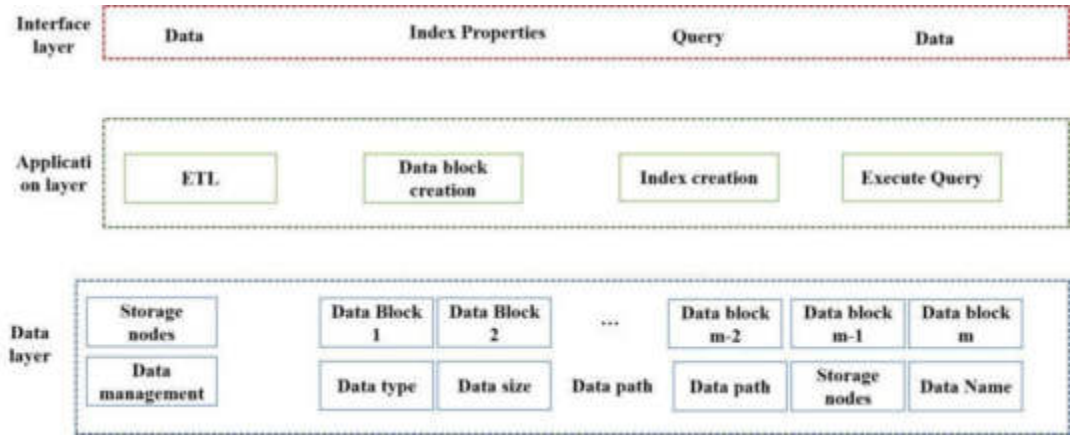


Fig. 3.4: Architecture of Health Record Management System

The main sources of big data on chronic diseases come from hospital information systems (HIS, LIS, PACS, EMR, etc.) within the hospital, as well as monitoring data collected by mobile internet platforms for chronic diseases outside the hospital and various wearable devices. It mainly includes structured, semi-structured, and unstructured multimodal data [13].

The chronic disease health big data platform mainly consists of in-hospital diagnosis and treatment health records, out of hospital disease follow-up records, and long-term monitoring data platforms for vital signs.

- Internal chronic disease diagnosis and treatment health records: mainly include patient diagnosis and treatment information, medication information, medical record information, etc., forming patient diagnosis and treatment health records. The data is sourced from information systems such as HIS, LIS, PACS, and EMR within the hospital.
- Outward follow-up health records: Mainly for the population under chronic disease management, including medication use, chronic disease development, and recovery. The data is mainly collected digitally and automatically through the Internet plus chronic disease platform [14].
- Vital sign monitoring files: mainly for the daily vital sign monitoring data of chronic disease management population, including commonly used indicators such as blood pressure, blood oxygen, respiration, pulse, etc., to achieve continuous and regular collection of vital sign data.

The application of chronic disease assisted decision-making analysis mainly integrates in hospital, out of hospital, and daily monitoring data of chronic disease management population to form patient health record data based on timeline, intelligently realizing the auxiliary diagnosis and treatment decision-making of chronic disease management.

3.4. Architecture. The Health Record Management System is a universal framework for patient health record management and big data indexing, which can be implemented on any distributed system. As an intermediate layer between users and distributed data management systems, this system can provide data upload, query execution mechanisms, and provide indexes to facilitate data search operations. The system architecture is shown in Figure 3.4.

It consists of three layers: (1) User Interface (UI) layer; (2) Application layer; (3) Data layer. Users initiate data upload and index creation operations through the interface layer, while the results of queries and index searches are returned from the application layer to the interface layer for users to browse. The application layer receives data upload and indexing instructions from the interface layer, and calls data block creation to store data in the data layer, while calling index creation to create indexes in the data warehouse. At the application layer, queries are received through the interface and processed by initiating index searches on data stored within the file system. The retrieved data is then delivered back to the user via the interface. Meanwhile, at the data layer, the system manages data blocks and ensures redundancy by storing designated replicas across available

storage resources, thereby safeguarding data integrity and security [15].

3.5. System composition and functions. According to the system architecture diagram, the health record management system can be decomposed into four modules: ETL, data block creation, index creation, and querying. The ETL module will extract data from multiple sources, such as chronic disease and health monitoring data sources, and clean, customize, and insert it into the data warehouse; The second module focuses on efficiently organizing data into blocks to enhance storage and retrieval efficiency. The index creation module employs a B-tree structure to store key-value pairs derived from these data blocks. Lastly, the query execution module retrieves the desired data, showcasing enhanced search performance especially beneficial for handling larger datasets. Below is a detailed introduction to the system composition and functions.

ETL (Extract, Transform, Load) is a fundamental database operation process within medical data warehouses. It plays a crucial role in ensuring data accuracy and reliability, as inaccuracies can potentially impact medical decision-making. Data for these processes is sourced from diverse origins, including various operational databases for chronic disease management and health monitoring across different organizational departments, as well as external suppliers. These sources often present data of varying quality, utilizing inconsistent representations, codes, and formats.

Contemporary big data processing systems offer distributed storage solutions that enhance data reliability through replication. Furthermore, each system incorporates its own data segmentation approach, defining the size and placement of data blocks within the data warehouse. When data is split into fixed size data blocks, the last record will face the threat of interruption, resulting in data corruption. Therefore, accessing multiple sites to retrieve corrupt records will increase the overall data loading time. Thus, to minimize access time for result records, it's essential to retrieve each record in its entirety from a single location. That is to say, the pattern of introducing data blocks ensures that the last record in each block is never segmented [16].

During the data block creation process, records are sequentially read and stored until the block reaches its storage capacity. The author suggests adjusting the block size based on the typical record size in the dataset or the default block size of the ETL system. Let's denote the dataset as D containing x records, as described by equation 3.1.

$$D = \sum_{c=1}^{x} record_c \tag{3.1}$$

Among them, $record_c$ represents the c-th record. Furthermore, Algorithm 1 provides the process of creating data blocks. In a distributed data system, the creation of data blocks occurs before data upload and divides the data into smaller blocks. Then load each block into the data warehouse with an adjustable replication factor.

Algorithm 4 Creating Data Blocks

Input: block_l=D_, which means the data block capacity limit is D_s; flag_c=true, The identifier for whether the data block capacity is full; block_n=0, indicates the number of initialized data blocks;
While reading data do
If flagtrue and block
Add data to a data block
Else
Load data block, return block,block
blockblock
flagtrue
End if
End while
Load data block, return block,block_n

The index creation process occurs after the data is loaded into the data warehouse, and creating an index can shorten the data retrieval time. Furthermore, it is crucial to optimize the index creation process to minimize both the delay between data upload and query execution and the additional space required by the

index. Utilizing the B-Tree structure for indexing helps in managing space and time overhead efficiently. During index creation, each attribute record is assigned as a key, with its position serving as the corresponding value in the index structure.

The process of creating an index is shown in Algorithm 2.

Algorithm 5 Create Index

Input: Index attribute index_attr, data block block_id; Data block content
block_con;
For index_attr do
Create an empty B-Tree structure
End for
Valueblock_id
While reading data do
for index_attr do block_con
Add<key, Value>to B-Tree
End for
End while
Store all established B-Tree structures

As mentioned earlier, indexes play an important role in big data processing and can lead to increased system overhead. Hence, the search efficiency gained from the index must outweigh the overhead incurred during index creation. This section will showcase the functionality of the system's query execution module. This module conducts index searches to efficiently retrieve both indexed and non-indexed attributes. The reliability and availability of data blocks are ensured by the underlying data layer. As long as the data blocks and indexes stored in the underlying data warehouse remain accessible, the system's query execution module will operate seamlessly. In addition, using this module to complete the query execution and data retrieval process depends on the choice of predicates in the query. The query execution module does not provide services for queries that use non indexed attributes as selection predicates [17-18].

The process of using indexes to execute queries is shown in Algorithm 3.

Algorithm 6 Execution Query

Input: Error message err msg; Query; Index; Target name; Target attribute attr; Block position blockreloc; B-Tree <key,Value>; Data block
If query error then
Return err_msg
End if
Get name, attr, block_loc from query
If attr has index do
Key
Valueblock loc
For block do
Get the index of all blocks
Get the key in the middle index
End for
End if

Initially, the system analyzes incoming queries to validate syntax and parameters, promptly notifying users of any errors like typos or syntax issues. Queries that don't match existing files in the file system are automatically discarded. By accurately parsing the query string, the system ensures the availability of query indexes, resorting to full-scan operations only when necessary. It loads and navigates through indexes to pinpoint record locations, facilitating direct access to data from the file based on the specified query. The subsequent section will present findings highlighting the efficiency of search operations [19].

Table 3.1: System Dataset

Data	Size	Number of items	Number of attributes	Number of data blocks
Medical records	77.2	13362	10	3
MRI	405	121850	15	7
CT	1500	33133	15	25
Health monitoring	6450	2298707	15	103
genome	16110	19291856	36	250

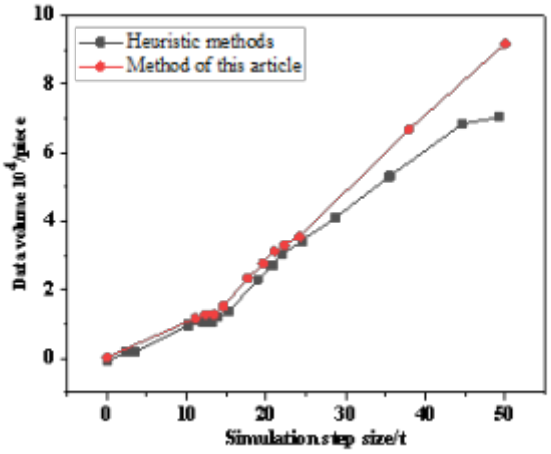


Fig. 4.1: Comparison of Data Classification and Retrieval Efficiency of Different Algorithms

3.6. Simulation analysis. In order to verify the performance of the architecture proposed by the author, a simulation analysis will be conducted using a case study. All modules and related algorithm development environments proposed by the author were developed using Python under the Ubuntu system. The system dataset for evaluating the monitoring and management system is shown in Table 3.1.

The dataset used for system validation includes medical record data, MRI data, CT data, health monitoring data, and genomic data. These datasets are of varying sizes, formats, and contain varying amounts of information. These characteristics of the dataset can affect data loading overhead, indexing overhead, and ultimately search performance.

4. Results and Discussion. Furthermore, compare the retrieval algorithm proposed by the author with the results of heuristic search. The efficiency comparison results of different algorithms for data classification and retrieval are shown in Figure 4.1.

As shown in Figure 4.1, when the amount of data is small, the performance of the two algorithms is almost the same. But as the amount of data continues to increase, the retrieval efficiency of this method continues to improve, and it is far higher than heuristic algorithms. Especially when there are a large number of indexed files, the retrieval rate of heuristic algorithms becomes increasingly low and overwhelming. Therefore, this method is more suitable for the retrieval of massive data. When the simulation step size is 50t, the maximum amount of data can reach 10×104 [20].

5. Conclusion. The author proposes a research on a precise health follow-up management information system for community chronic diseases based on big data analysis. The author has conducted research on medical data preprocessing, data modeling, security, data retrieval, etc., and proposed a health record management

system with data management and efficient retrieval. The system provides the ability to create multiple indexes on a dataset with minimal index overhead, fast creation and traversal, and less space occupation, thereby improving data management and search performance.

REFERENCES

[1] Qianqian, C., & Lijuan, T. (2023). Study on the management of chronic diseases in american and british community pharmacy, 18(2), 157-164.
[2] Lin, X., Lei, Y., Chen, J., Xing, Z., Yang, T., & Wang, Q., et al. (2023). A case-finding clinical decision support system to identify subjects with chronic obstructive pulmonary disease based on public health data. Tsinghua Science and Technology, 28(3), 525-540.
[3] Sisodia, A., & Jindal, R. (2022). An effective model for healthcare to process chronic kidney disease using big data processing. Journal of Ambient Intelligence and Humanized Computing, 14(10), 1-17.
[4] Zhou, X., Lee, E. W. J., Wang, X., Lin, L., Xuan, Z., & Wu, D., et al. (2022). Infectious diseases prevention and control using an integrated health big data system in china. BMC Infectious Diseases, 22(1), 1-9.
[5] Ed-Daoudy, A., Maalmi, K., & Ouaazizi, A. E. (2023). A scalable and real-time system for disease prediction using big data processing. Multimedia Tools and Applications, 82(20), 30405-30434.
[6] Pakhale, S., Visentin, C., Tariq, S., Kaur, T., Florence, K., & Bignell, T., et al. (2022). Lung disease burden assessment by oscillometry in a systematically disadvantaged urban population experiencing homelessness or at-risk for homelessness in ottawa, canada from a prospective observational study. BMC Pulmonary Medicine, 22(1), 1-9.
[7] Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. Journal of Big Data, 9(1), 1-19.
[8] Barbanti, P., Egeo, G., Aurilia, C., Fiorentini, G., Proietti, S., & Tomino, C., et al. (2022). The first report of the italian migraine registry (i-graine). Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, 43(9), 5725-5728.
[9] Li, Q., & Pan, W. T. (2022). Application of multisource big data mining technology in sports economic management analysis. Mathematical Problems in Engineering: Theory, Methods and Applications(Pt.16), 2022.
[10] Son, J., Kim, Y., & Zhou, S. (2022). Alerting patients via health information system considering trust-dependent patient adherence. Information technology & management, 18(7), 719-730.
[11] Kiryu, Y. (2023). Development of a medical big data analysis system utilizing artificial intelligence analytics in clinical pharmacy. YAKUGAKU ZASSHI, 143(6), 501-505.
[12] Shafqat, S., Majeed, H., Javaid, Q., & Ahmad, H. F. (2022). Standard ner tagging scheme for big data healthcare analytics built on unified medical corpora, 2(4), 152-157.
[13] Hulsen, T., Friedecky, D., Renz, H., Melis, E., Vermeersch, P., & Fernandez-Calle, P. (2023). From big data to better patient outcomes. Clinical Chemistry and Laboratory Medicine CCLM , 61(4), 580-586.
[14] Zhou, Y., & Varzaneh, M. G. (2022). Efficient and scalable patients clustering based on medical big data in cloud platform. Journal of Cloud Computing, 11(1), 1-10.
[15] Senhao, C., Yingnan, C., Jingran, G., & Yuwen, C. (2023). Analysis and enlightenment of big data platform for adverse drug reaction supervision in china and the united states, 18(3), 213-220.
[16] Fazel-Najafabadi, A., Abbasi, M., Attar, H. H., Amer, A., Taherkordi, A., & Shokrollahi, A., et al. (2024). High-performance flow classification of big data using hybrid cpu-gpu clusters of cloud environments. Tsinghua Science and Technology, 29(4), 1118-1137.
[17] Kunnumakkara, A. B., Hegde, M., Parama, D., Girisa, S., Kumar, A., & Daimary, U. D., et al. (2023). Role of turmeric and curcumin in prevention and treatment of chronic diseases: lessons learned from clinical trials. ACS Pharmacology And Translational Science, 6(4), 447-518.
[18] Chimezie, R. O. (2023). Health awareness: a significant factor in chronic diseases prevention and access to care. Journal of Biosciences and Medicines, 11(2), 16.
[19] Filippo, A. D., Perna, S., Pierantozzi, A., Milozzi, F., Fortinguerra, F., & Caranci, N., et al. (2022). Socio-economic inequalities in the use of drugs for the treatment of chronic diseases in italy. International journal for equity in health, 21(1), 157.
[20] Sisodia, A., & Jindal, R. (2022). An effective model for healthcare to process chronic kidney disease using big data processing. Journal of Ambient Intelligence and Humanized Computing, 14(10), 1-17.
[21] Lin, X., Lei, Y., Chen, J., Xing, Z., Yang, T., & Wang, Q., et al. (2023). A case-finding clinical decision support system to identify subjects with chronic obstructive pulmonary disease based on public health data. Tsinghua Science and Technology, 28(3), 525-540.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 30, 2024
Accepted: Jul 27, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 907–915, DOI 10.12694/scpe.v26i2.4088

THE OPERATION AND MAINTENANCE STRATEGY OF SMART GRID BASED ON INTELLIGENT PERCEPTION AND OPTIMIZATION ALGORITHM

DEXIONG LI*, JUNYI HUO†, YU WANG‡, JING LI§, AND HUICHAO JIN¶

Abstract. Due to the existence of a sensor model, the state perception of the distribution network can obtain higher RMSE. Because of this situation, this topic intends to use artificial intelligence technology to realize the embedded sensing system of stable operation of distribution network: Front-end sensor and wireless inlet design. According to the stable operation characteristics of the distribution network, a stable data collection system is established. Various algorithms based on data unification and identification are proposed to sense calculation parameters. An adaptive dynamic stability detection method is designed based on a deep neural network. Experiments show that an RMSE of 0.031 can be obtained by this method. This method can realize the accurate perception of the running state of the distribution network.

Key words: Artificial intelligence; Distribution network; Stable operation; Embedded system; Behavior perception; Calculation parameter

1. Introduction. With the large-scale entry of new energy into the distribution network, the distribution network has gradually developed in the direction of active power supply. Active distribution network operation status monitoring is a vital link to improve. With the increasing complexity of the distribution network, the traditional static sensing method cannot meet future distribution network development requirements. The research shows that the stable state detection method of a dehumanized intelligent distribution network based on 3D Lidar is significant. Literature [1] proposes to adopt a combination of static and dynamic sensing modes to realize real-time monitoring of distribution network characteristics in real scenarios based on obtaining a complete panoramic distribution network modeling. Literature [2] obtains global status awareness based on feature monitoring data to improve maintenance and management efficiency. However, the detection effectiveness of this method is highly variable. Literature [3] studies different sensing methods according to different physical parameters to evaluate the proper degree of sensing results. In this way, the optimal sensing physical quantity is selected. Literature [4] intends to adopt sensing technology and integrate big data with 3D inspection to realize real-time monitoring of the environment. However, the scalability of this algorithm is not robust. The research shows that the new sensing technology of distribution networks based on real-time phase information has important theoretical significance and application value. In this paper, the observation equipment is set up in the distribution system, and the influence factors of its state are extracted. Then the collected data is analyzed by support vector machine algorithm. Finally, the data are analyzed using the extended and short-time memory network. Finally, a distributed sensor network based on artificial intelligence is constructed, and its operating state in the distribution network is studied.

2. Application of embedded sensing system in stable operation of distribution network.

2.1. System hardware design.

2.1.1. Front-end perceptron. CC2530 is used as the front-end sensor to consider network energy consumption and signal acceptance sensitivity [5]. Then, it realizes data interaction with the MCU I/O interface

*Department of Electrical Engineering, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050041, China
†Department of Electrical Engineering, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050041, China
‡Department of Electrical Engineering, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050041, China
§Department of Electrical Engineering, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050041, China
¶Department of Electrical Engineering, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050041, China (Corresponding author, jhcmail@163.com)

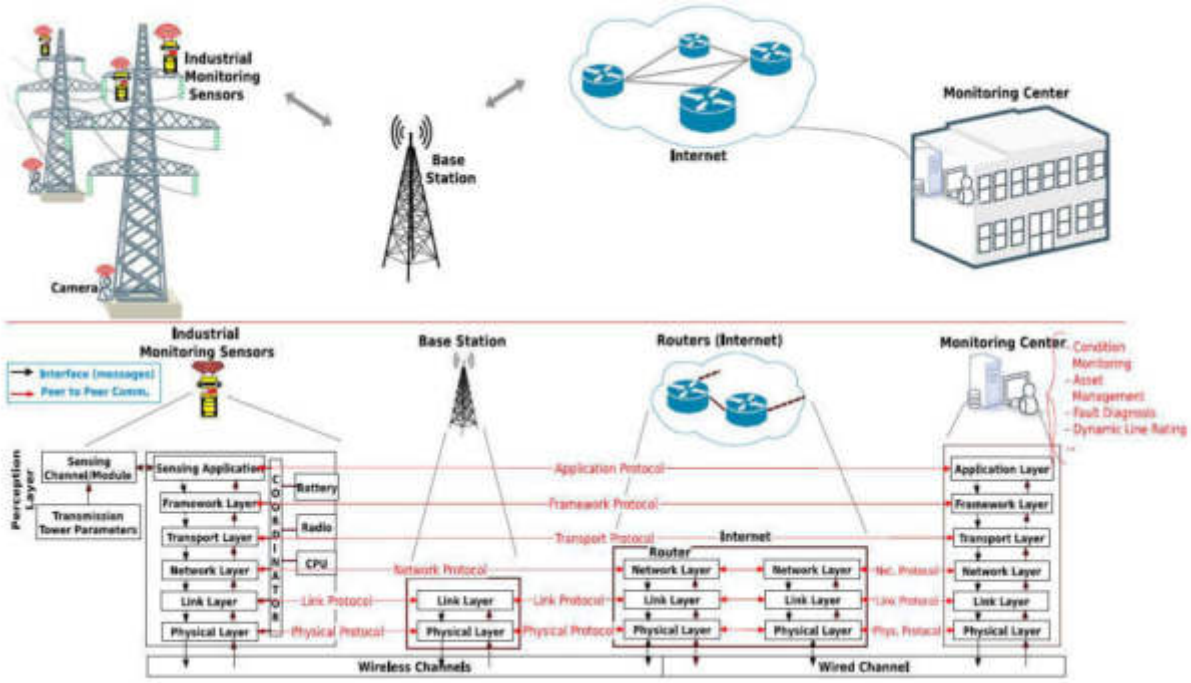


Fig. 2.1: *Hardware architecture of embedded sensing system gateway.*

through the CC2530 interface. Finally, the sensor, the wireless transceiver module, and the clock module constitute the entire system's front-end sensing module.

2.1.2. Implementation of Wireless Gateway. The implementation of a wireless gateway includes the functions of data transmission, encapsulation and analysis. The analysis of the S3C2440 microcontroller shows that it works at more than 400 MHz, which can fully meet the needs of the sensing field [6]. The module is combined with the TFT-LCD display, remote control button and other modules to realize the overall architecture of the wireless gateway (Figure 2.1). Install the LM25965-5.0 switching regulator on the gated power supply to enhance the stability of the gated application.

2.2. Software Design.

2.2.1. Establish a stable power network monitoring system. The stable state monitoring of the distribution network cannot be separated from a lot of data support [7]. This paper presents a data collection system for the operating state of the distribution network (Figure 2.2 cited in Water 2019, 11(3), 562). R is a data collector and Z is an encoder. H_1, H_2 is the length of the channel. This paper determines the information collected by each collector to ensure the completeness of data collection under a steady state [8]. The information collected is obtained using the coding function, and the input signal is generated. Select a point in time in the data acquisition construct in Figure 2.2. The data acquisition channel specification expression is:

$$A_v = B_{jv} + C_v, C_v \in N \tag{2.1}$$

B is an input signal. A is an output signal. v is a kind of acquisition time. m is the steady-state operating performance data collected for a distribution network. j is a specific data collection point. C is a disturbing noise. N is the variance. Set the critical capabilities of the channel to transmit information consistent with the minimum code transfer requirements [9]. The formula for calculating the upper bound of the channel is as

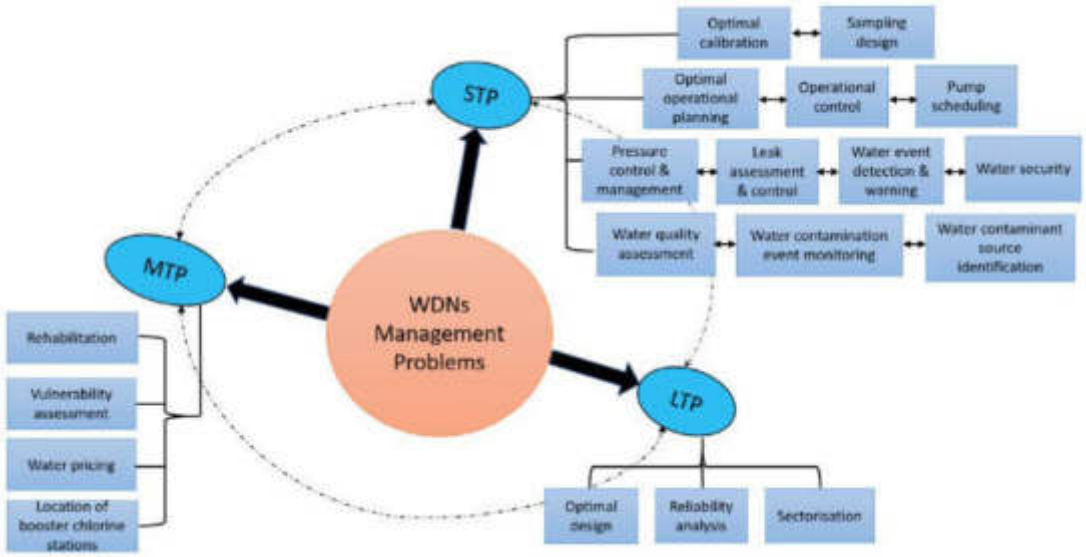


Fig. 2.2: *Distribution network operating status information collection architecture.*

follows:

$$W = H_1 \sum_{j}^{\pi} \log\left(1 + \frac{\phi_j}{N_j}\right) + H_2 \left[\log\left(1 + \frac{\phi_g}{N_f}\right) + \log\left(1 + \frac{\phi_c}{N_c}\right)\right] \tag{2.2}$$

W is the upper bound of the channel. ϕ_j, ϕ_g, ϕ_c is the average of the noise in the information transmission. N_j, N_g, N_c is the variance of noise in information transmission. For the analysis of each average noise power, the power limit of each information transmission stage can be obtained:

$$\begin{cases} \phi_j \geq \frac{1}{n} \sum_{i=1}^{n} \left[B_{j1}(\lambda, v)\right]^2 \\ \phi_g \geq \frac{1}{n} \sum_{i}^{n} \left(B_{12}, \cdots, B_{m2}, i\right)^2 \\ \phi_c \geq \frac{1}{n} \sum_{i}^{\pi} C_i^2 \end{cases} \tag{2.3}$$

λ is the number of facilities in the distribution network [10]. According to the restriction given in equation (2.3), the sampling interval of the front end is set under steady state, and the global acquisition of steady-state data is completed.

2.2.2. Extraction of calculation parameters of intelligent sensing. Data normalization and recognition methods are adopted based on the analysis of steady-state data collection of distribution networks [11]. Because the collected data are spatiotemporal dependent, a multi-level data consistency modeling method is proposed. For obtaining steady-state data, record each characteristic quantity and generate the following matrix:

$$S = \begin{bmatrix} s_{11} & \cdots & s_{1\pi} \\ \vdots & \ddots & \vdots \\ s_{71} & \cdots & s_{2\pi} \end{bmatrix} \tag{2.4}$$

S represents the acquisition matrix, and 5 represents a single property of the collected data. A is to get the number of columns and rows in the matrix.

$$\tilde{\lambda}_{\pi} = (\zeta_1, \zeta_2, \cdots, \zeta_{\pi}) \tag{2.5}$$

$\tilde{\lambda}$ is a stable data vector while ζ represents a matrix column vector. Because of the different static sampling frequencies, some sampling data is missing. This paper proposes a dynamic optimization algorithm based on time series to estimate the similarity of discrete sequences and then extend and compress them [12]. This ensures consistency of sequence size. Select a random column vector in the formula (2.5) as the reference vector. Euclidean distance operations are performed on other column vectors, resulting in several distance matrices:

$$P_i = \begin{bmatrix} U_{11} & \cdots & U_{1,} \\ \vdots & \ddots & \vdots \\ U_{71} & \cdots & U_{\pi} \end{bmatrix} \tag{2.6}$$

Where k is the column vector, P_k is the distance matrix, and U is the European distance. The distance matrix is deduced, generating several distance loss matrices [13]. In this way, the approximate calculation of the column vector is achieved:

$$\begin{gathered} \sigma = \begin{bmatrix} \varphi_{11} & \cdots & \varphi_{17} \\ \vdots & \ddots & \vdots \\ \varphi_{71} & \cdots & \varphi_{\pi} \end{bmatrix} \\ W = \{W_1, W_2, \cdots, W_7\} \end{gathered} \tag{2.7}$$

σ is a group of distance loss matrices, φ is a group of loss degrees, W is a group of optimally regulated sequences, and it is also a group of shortest methods [14]. The minimum vector spacing is ensured by adjusting the vector spacing in the steady state using the dynamic rules. PC.A method was used to evaluate the steady-state data and eliminate the redundant duplication. First, the spacing of the adjustment vectors is normalized to obtain a normalized matrix of the form:

$$L = \delta - \frac{\delta}{\gamma} \tag{2.8}$$

L stands for standardized matrix. $\mathcal{S}$ is the parameter data after spacing adjustment based on formula (2.9). The expression for covariance and uniformization is now obtained:

$$\begin{gathered} Y = \frac{1}{\gamma} L \\ \mathrm{sid}(Y)[G, Z, K] \end{gathered} \tag{2.9}$$

Y is the covariance matrix. svd is SVD. G, Z, K represents the matrix formed after decomposition. G is a dimensionality reduction matrix. The SVM algorithm is used to solve the intelligent sensing parameters of a single sampling point, and the likelihood function is derived by referring to the independent relationship between the observation points in the sampling point [15]. The intelligent detection and analysis of sensor data are realized based on the above content.

2.2.3. Construct the AI adaptive perception model. The embedded sensing system studied in this paper is based on artificial intelligence. A sensor network with independent intellectual property rights is established, combining the theory and method of machine learning [16]. The input and output of dynamic behavior are calculated, and the weights in the model are adjusted according to the obtained state information. The system consists of four levels. From the principle of sensing mode in Figure 2.3, it can be seen that the system consists of the current stable state of the distribution network and the period when the failure occurs. The input vector is represented as:

$$\omega(t) = (\omega_1(t), \omega_2(t), \cdots, \omega_2(t)) = \{\pi(t), \pi(t-1), \cdots, \pi[t-(\partial-1)\tau]\} \tag{2.10}$$

The adaptive perceptual model transmits input-level information to the hidden layer of the algorithm. Then, it is calculated by several hider nodes to obtain:

$$s(t) = \frac{1}{1+\alpha'} \tag{2.11}$$

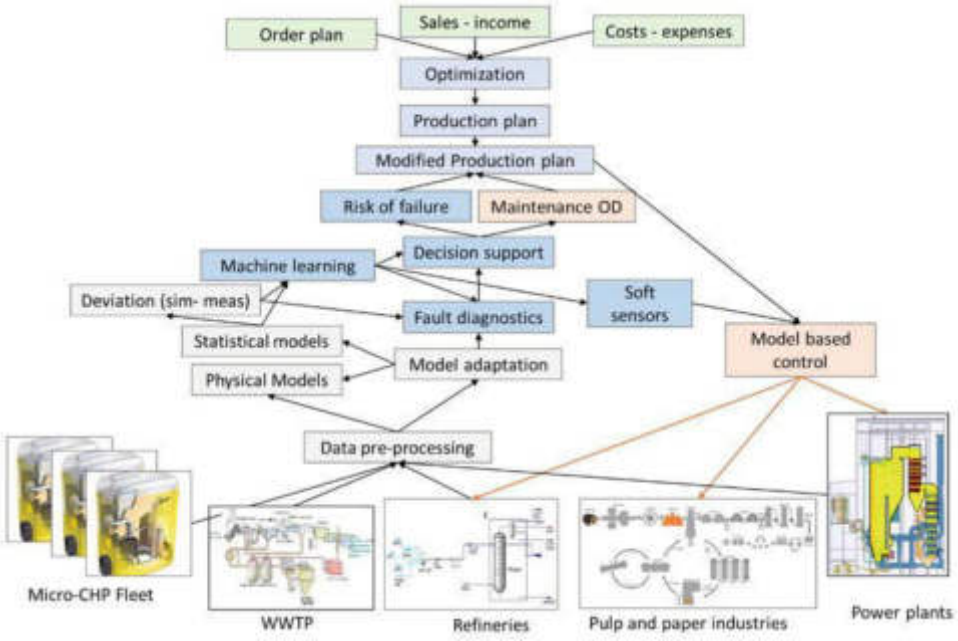


Fig. 2.3: *Adaptive perception model based on artificial intelligence.*

t is the output of the hidden layer, α is the constant, and r is the parameter weight. The Gaussian distribution characteristics of steady-state data are obtained by applying hidden layer measurements on a random layer [17]. This paper uses it to characterize the output characteristics of the system. The random layer's output result is expressed by solving each hidden node's Gaussian distribution.

$$\theta\left[\imath(t), r_0\right] = \frac{1}{1+\alpha^{\imath(t) r_0}} \tag{2.12}$$

Where θ is the output value of any layer, and r_0 is the weight of the implicit node parameter.

The reinforcement learning algorithm is introduced. The output of the random layer is studied in depth. It's expressed in one-dimensional Gaussian form [18]. Then, an adaptive learning method based on random stratification errors is adopted to realize state monitoring under dynamic conditions (FIG. 2.3).

2.3. Main Functional Components.

2.3.1. Data capture module. The information collected in today's intelligent distribution network includes current, voltage, frequency, temperature and other multi-dimensional information. Various sensors are added to the device, such as a phase measuring device, infrared sensor, ultrasonic sensor, etc. This new type of sensor can not only realize the online monitoring of the operation of the distribution network but also determine its position according to the specific accident or abnormal situation. A more complete view of power network operation can be obtained by fusing the signals collected by multiple sensors. In addition, real-time data is particularly critical in security situation awareness, so it needs to be quickly and accurately positioned. The solution uses high-speed technologies such as 5G, optical fiber communication or broadband power lines to ensure low delay and highly reliable data transmission. In addition, data synchronization is essential for multi-sensor or multi-location observation data. High-precision clock synchronization represented by GPS, IEEE1588, etc., can ensure the consistency of all data in time. This lays a good foundation for future multi-source information fusion and analysis.

2.3.2. Data processing and storage. The information processing module has become the system's core to ensure accurate operation. This project introduces Apache Spark technology into the distributed network to

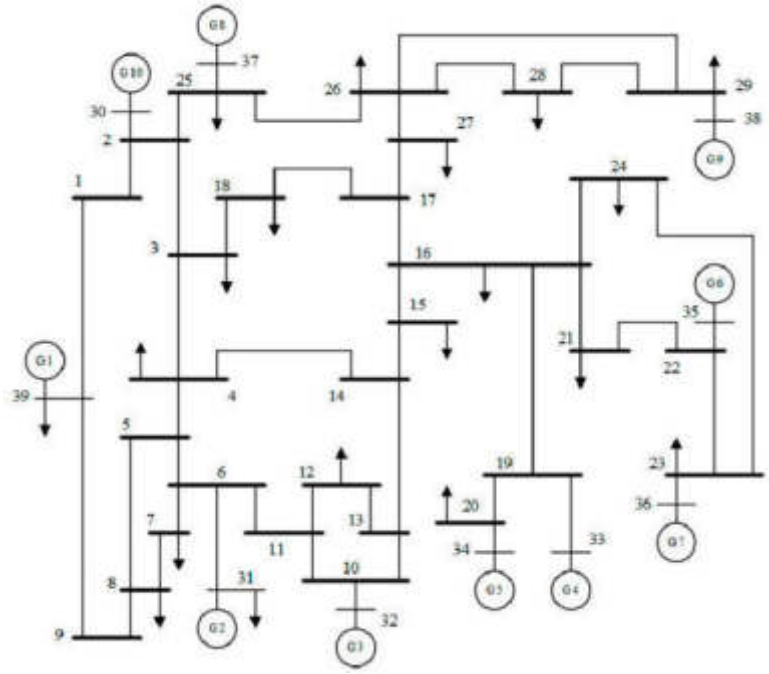


Fig. 3.1: *IEEE39 node system structure.*

realize efficient power scheduling for the massive real-time distribution network. The signal pre-processing stage mainly includes noise removal, normalization and anomaly extraction [19]. The Kalman filter and other efficient filtering methods can suppress the noise in the signal. The standardization process ensures the consistency of data. Advanced outlier detection methods, such as isolated forests, can ensure accurate identification and elimination of possible error information. Principal component analysis, self-coding and other methods can effectively extract important information related to security situation cognition from massive videos. Because of the system's demand for persistent storage, Hadoop lays a good foundation for the efficient storage of massive data. In addition, to realize real-time data retrieval, people also developed Cassandra, MongoDB and other databases based on NoSQL to realize high-speed data retrieval. Methods such as disk array and error correction code ensure that the required data can be restored entirely after some nodes fail to ensure data integrity and consistency. In addition, the system also encrypts the data at multiple levels to ensure the confidentiality of the data. In addition, the intelligent sensing grid also contains the concept of flexible computing and storage. When the data is increased, the system can be expanded horizontally to cope with future data processing and storage problems.

3. System test.

3.1. Establishment of the test environment. The Linux Ubuntu19.04 operating system and JDK1.8 programming components are used. The test platform was built using 7 virtual institutions. The four VMSs serve as the secondary nodes of the data node. The system then consists of two hosts and one management node. There are two configurations for Hadoop users from the SSH protocol. First, the SSH protocol is installed on each virtual machine, creating a post-SSH folder. This facilitates the subsequent system guidance and instruction operation. A non-key pair encryption scheme based on SSH is proposed. The address configuration from node to host is realized through the configuration of Core components such as core-site.xml and MapReduce. The main result of this experiment is 4IEEE39 nodes. It consists of 10 units, 46 lines and 19 load nodes. The detailed topology is shown in Figure 3.1.

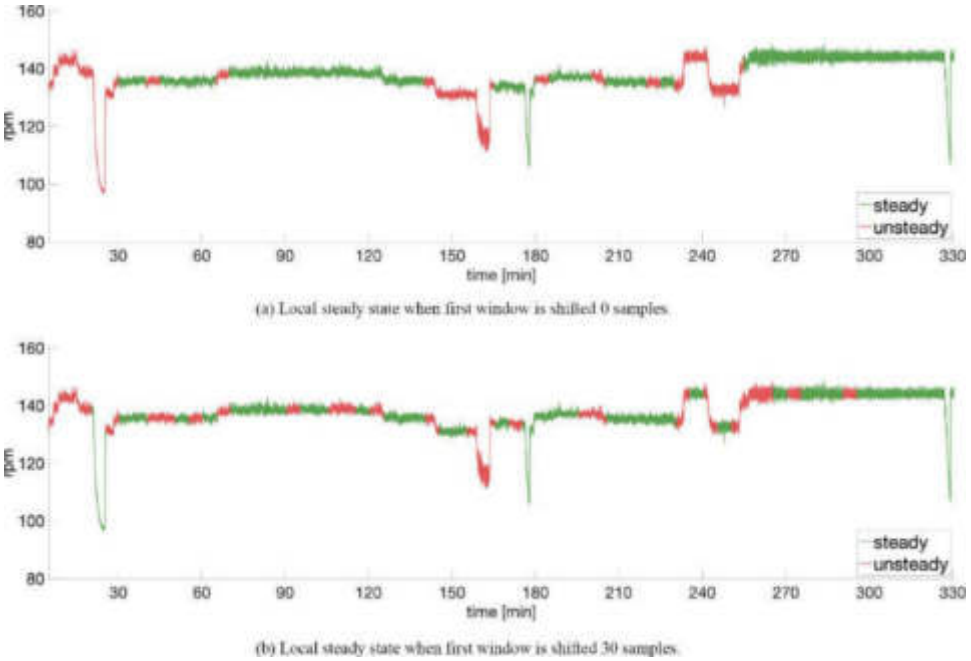


Fig. 3.2: *Time series data of steady-state action behavior of distribution network.*

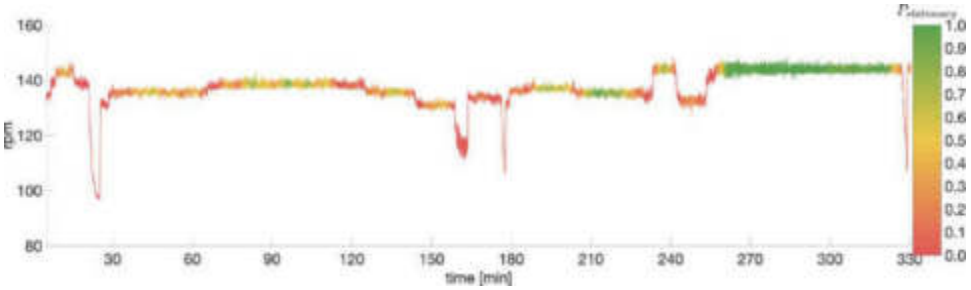


Fig. 3.3: *Line chart of perceptual results.*

3.2. Setting the parameters of the perception model. The system is correctly set before execution to improve the detection accuracy further. Nessus software was used to analyze the running state characteristics, and 200 situation time series data, as shown in Figure 3.2, were collected. The steady-state time series data is regarded as a nonlinear sequence, and the nonlinear relationship is established in the output space of each dimension to realize the perception of the steady-state operation of the distribution network. By analyzing 200 steady-state time series data and setting the dimension of the input vector to 3 and 5, 197 test samples were obtained. The data set collected above is used in the learning of neural networks. The final parameter of the algorithm is determined by comparing the deviation of each parameter with its output. Through analysis, it is found that the number of hidden layer nodes of this method can reach 20 when 5 input vector dimensions are set so that the stable operation state of the distribution network at the next moment has high accuracy.

3.3. Analysis of test results.. This method monitors the whole-day performance of the IEEE39-node system in real time and is integrated with the measured data. This results in the running effect curve shown in Figure 3.3. By comparing the measured data with the measured results, the performance of the proposed system in practical application is illustrated. The stable state values obtained by the detection system proposed

in this paper mostly agree with the actual situation. Reversals occur only at 10 and 15 o 'clock. A method based on RMSE was proposed to measure the method's accuracy to make the method more vividly reflect the use efficiency. The larger the RMSE value is, the higher the detection accuracy of the method is.

RMSE refers to the square root error of the mean, n is the amount of behavior data in A stable state, i is some stable data sample, x_i is the real situation value, and $\underline{x}_i$ is the perceived situation value. According to equation (3.1), the minimum mean square error of the system in this paper is o.o31. This method can accurately detect the dynamic characteristics of the distribution network.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|x_i - x_j|^2} = 0.031 \tag{3.1}$$

4. Conclusion. This project intends to adopt an adaptive sensing model based on artificial intelligence. Then the model is introduced into the reinforcement learning of each hidden node to realize real-time adjustment of the weights of each hidden node. In this way, the average value of each hidden node is less than 3.1%, which can meet the stable operation requirements of the distribution network. However, due to the limited development time, the interface of the perceptual effect display is relatively simple, which requires some beautiful design on the interface so that the user has a better experience.

REFERENCES

[1] Reebadiya, D., Rathod, T., Gupta, R., Tanwar, S., & Kumar, N. (2021). Blockchain-based secure and intelligent sensing scheme for autonomous vehicles activity tracking beyond 5g networks. Peer-to-Peer Networking and Applications, 14(5), 2757-2774.

[2] Wang, W., Lou, B., Li, X., Lou, X., Jin, N., & Yan, K. (2020). Intelligent maintenance frameworks of large-scale grid using genetic algorithm and k-mediods clustering methods. World Wide Web, 23(2), 1177-1195.

[3] Doghri, W., Saddoud, A., & Chaari Fourati, L. (2022). Cyber-physical systems for structural health monitoring: sensing technologies and intelligent computing. The Journal of Supercomputing, 78(1), 766-809.

[4] Sharma, A., & Chauhan, S. (2020). A distributed reinforcement learning based sensor node scheduling algorithm for coverage and connectivity maintenance in wireless sensor network. Wireless Networks, 26(6), 4411-4429.

[5] Kingsley-Amaehule, M., Uhunmwangho, R., Nwazor, N., & Okedu, K. E. (2022). Smart Intelligent Monitoring and Maintenance Management of Photo-voltaic Systems. International Journal of Smart Grid, 6(4), 110-122.

[6] Yu, X., Ergun, K., Cherkasova, L., & Rosing, T. Š. (2020). Optimizing sensor deployment and maintenance costs for large-scale environmental monitoring. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 39(11), 3918-3930.

[7] Mufana, M. W., & Ibrahim, A. (2022). Monitoring with Communication Technologies of the Smart Grid. IDOSR Journal of Applied Sciences, 7(1), 102-112.

[8] Ren, Y., Wang, T., Zhang, S., & Zhang, J. (2023). An intelligent big data collection technology based on micro mobile data centers for crowdsensing vehicular sensor network. Personal and ubiquitous computing, 27(3), 563-579.

[9] Ochuba, N. A., Usman, F. O., Okafor, E. S., Akinrinola, O., & Amoo, O. O. (2024). Predictive analytics in the maintenance and reliability of satellite telecommunications infrastructure: a conceptual review of strategies and technological advancements. Engineering Science & Technology Journal, 5(3), 704-715.

[10] Lima, A. L. D. C. D., Aranha, V. M., Carvalho, C. J. D. L., & Nascimento, E. G. S. (2021). Smart predictive maintenance for high-performance computing systems: a literature review. The Journal of Supercomputing, 77(11), 13494-13513.

[11] Serradilla, O., Zugasti, E., Rodriguez, J., & Zurutuza, U. (2022). Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Applied Intelligence, 52(10), 10934-10964.

[12] Li, J., Gu, C., Xiang, Y., & Li, F. (2022). Edge-cloud computing systems for smart grid: state-of-the-art, architecture, and applications. Journal of Modern Power Systems and Clean Energy, 10(4), 805-817.

[13] Osunsanmi, T. O., Aigbavboa, C. O., Oke, A., & Onyia, M. E. (2022). Making a case for smart buildings in preventing corona-virus: focus on maintenance management challenges. International journal of construction management, 22(16), 3109-3118.

[14] Shcherbakov, M., & Sai, C. (2022). A hybrid deep learning framework for intelligent predictive maintenance of cyber-physical systems. ACM Transactions on Cyber-Physical Systems (TCPS), 6(2), 1-22.

[15] Yan, Y., Liu, Y., Fang, J., Lu, Y., & Jiang, X. (2021). Application status and development trends for intelligent perception of distribution network. High Voltage, 6(6), 938-954.

[16] Adekanbi, M. L. (2021). Optimization and digitization of wind farms using internet of things: A review. International Journal of Energy Research, 45(11), 15832-15838.

[17] Wang, J., Wang, X., Ma, C., & Kou, L. (2021). A survey on the development status and application prospects of knowledge graph in smart grids. IET Generation, Transmission & Distribution, 15(3), 383-407.

[18] Kong, Q., Lu, R., Yin, F., & Cui, S. (2020). Privacy-preserving continuous data collection for predictive maintenance in vehicular fog-cloud. IEEE Transactions on Intelligent Transportation Systems, 22(8), 5060-5070.

[19] Wang, F. Y., Guo, J., Bu, G., & Zhang, J. J. (2022). Mutually trustworthy human-machine knowledge automation and hybrid augmented intelligence: mechanisms and applications of cognition, management, and control for complex systems. Frontiers of Information Technology & Electronic Engineering, 23(8), 1142-1157.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 31, 2024
Accepted: Jul 13, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 916–923, DOI 10.12694/scpe.v26i2.4079

THE ENVIRONMENT PERCEPTION AND PATH PLANNING ALGORITHM FOR DRIVERLESS CARS BASED ON COMPUTER PROCESSING AND MULTI-SENSOR FUSION

YOUJIN ZHAO*

Abstract. This paper selects suitable sensors and equipment based on the characteristics of urban road traffic. Then, a multi-sensor information fusion system based on millimeter-wave radar/camera is established. In this way, the road traffic safety can be effectively controlled. Then, by improving D-S evidential reasoning, a "goal-decision" two-level information fusion method is established. Then, the multi-layer fusion experiment of multi-source sensing data under a tunnel environment is carried out. Experiments show that the ROI region correlation using camera and millimeter wave radar can improve the detection accuracy by 9.66%, effectively solving the limitations of single-sensor detection for automatic driving under complex conditions. The D-S evidential reasoning method processes the automatic driving sensor data. This reduces the false report rate of the sensor by 2.75%.

Key words: Tunnel environment; Driverless; Multi-sensing technology; D-S Principle of evidence; Goal - Determine the Layer 2 integration strategy.

1. Introduction. In recent years, due to the rapid development of artificial intelligence, its role in car driving is becoming increasingly apparent. This makes the car's operation on the road more efficient and safe. Driverless perception is the "eyes" of the car. At present, unmanned vehicles take the road surface as the main body, and the accuracy of their perception can meet most of the traffic monitoring requirements [1]. But most highways in China have underground tunnels at present. Its driving conditions are bad, noise interference is serious, and it is easy to cause the phenomenon of "missing detection" and "false detection" in driving. This causes traffic accidents inside the tunnel. However, the current domestic and foreign researches lack sufficient attention to the unmanned vehicles in the tunnel environment, and it is urgent to break through the problem of safety and reliability perception under this specific condition.

Some scholars use the D-S evidence principle and multi-sensor data for system fault diagnosis [2]. Some scholars have compared and studied the primary probability distribution from the perspectives of neural networks, Bayes and D-S. Some researchers have used SVM to solve fundamental probability assignment problems in D-S evidential reasoning and have applied it to practical fault diagnosis experiments. Some scholars add the unknown variable m(x) to the D-S evidence theory to reduce the contradiction between the evidence [3]. Some studies have used box-line and average substitution methods to detect and repair abnormal points and then used the adaptive weight fusion method for multilevel fusion of multiple sensor points of the same type. Finally, the D-S evidence information fusion method is adopted to solve the multi-objective coordination problem in complex scenes. Some scholars have studied improved D-S evidence reasoning methods to overcome the severe contradictions in the existing methods [4]. This project intends to use the D-S evidence theory to optimize it. Then, the multi-dimensional information fusion method is designed. This project will effectively integrate multi-source sensor data at the target layer and the decision layer, which can obtain accurate information and lay a foundation for the application of vehicle control decision-making, autonomous obstacle avoidance, path planning and other aspects.

2. Design of automatic driving tunnel environment awareness system. Given the poor lighting conditions and instability of unmanned driving in underground traffic, it is necessary to carry out a unique design of its hardware [5]. Considering the tunnel's specific working conditions and the sensor's hardware characteristics, a collaborative sensing and identification method based on millimeter-wave radar and camera is

*Department of Automobile Engineering, Anhui Automobile Vocational and Technical College, Hefei 230601, China (ahqczyjsxy@163.com)

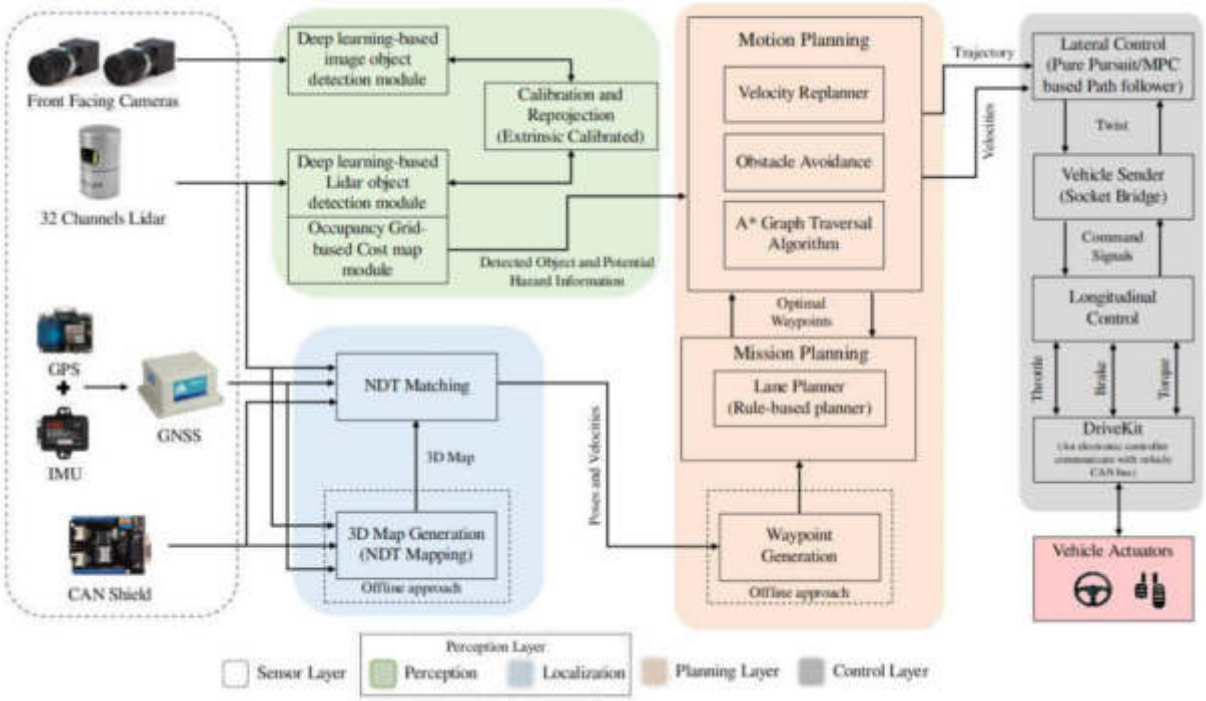


Fig. 2.1: Architecture of unmanned multi-sensor fusion sensing system.

proposed [6]. In addition, multiple ultrasonic radars can be set around the vehicle body to achieve the purpose of short-range collision avoidance. The layout scheme of the multi-sensor sensing System for unmanned vehicles is shown in Figure 2.1 (the picture is quoted in System, Design and Experimental Validation of Autonomous Vehicle in an Unconstrained Environment).

2.1. Goal-decision two-level information fusion strategy. For the specific driving scene of the tunnel, selecting the appropriate sensor and layout is the fundamental guarantee of obtaining efficient detection information. An efficient calculation method for sensing information is designed to achieve accurate detection [7]. Then, post-processing based on the data collected by multiple sensors is the key to improving detection accuracy.

2.1.1. Two-layer information fusion induction method. A reasonable information fusion method using multiple sensors is proposed. First, the ultrasonic sensor has some problems, such as small range and difficulty locating obstacles, and the fusion of multi-sensor information is directly related to the accuracy of recognition results. Secondly, the diversity of the detection environment leads to the scattered arrangement of ultrasonic sensors, and there is no compelling data fusion [8]. Therefore, the design of this paper only plays the role of collision avoidance alarm.

This paper proposes a fusion algorithm based on the target layer. That is the feature and decision fusion algorithms [9]. Because of the light conditions in the tunnel, the information obtained is more complicated than the road conditions on the ground. Therefore, data loss should be reduced as much as possible to ensure data quality. This project intends to study the information fusion method at two ”target layer - decision layer” levels for unmanned vehicles in tunnels. Its data Fusion process is shown in Figure 2.2 (image cited in Sensor Fusion for Radar Detection).

2.1.2. Information fusion model of two-layer perceptron. This paper proposes a two-layer information fusion model based on the ”target layer - decision layer.” This project uses millimeter-wave radar and a camera as the hardware platform for integration [10]. The fusion of target layer and decision layer and the final

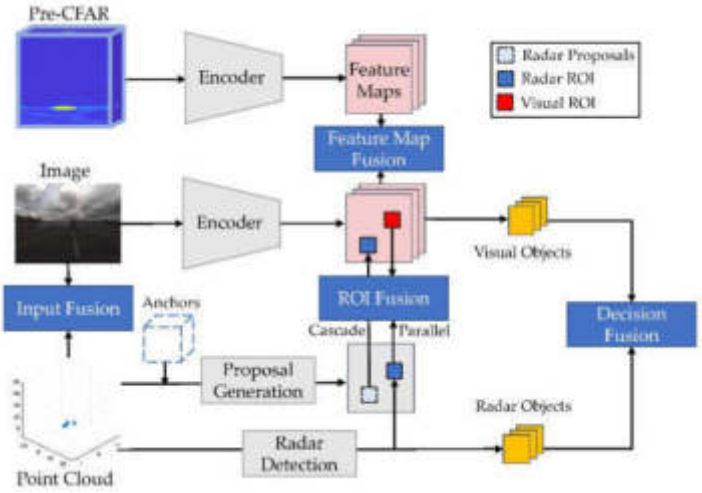


Fig. 2.2: Two-layer fusion strategy based on millimeter wave and camera.

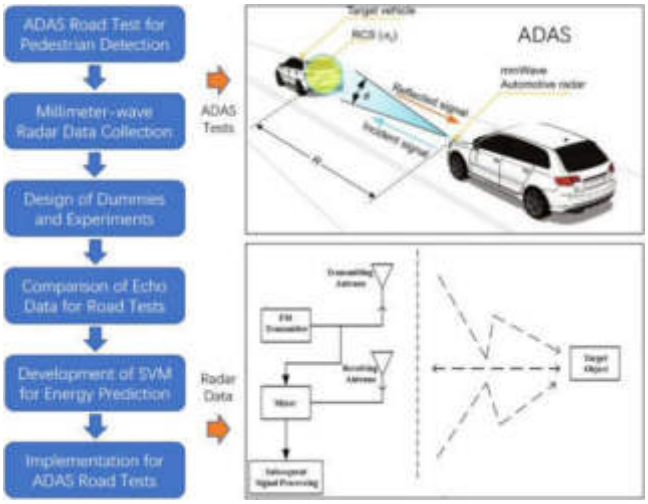


Fig. 2.3: Two-level information fusion model of millimeter wave radar and camera.

result output are realized through the acquisition and processing of multiple sensors. First, the camera detects objects, forming two types of sensing areas. Secondly, the correlation analysis of the two categories of interest regions is carried out and compared with the threshold. In this way, the Adaboost quadratic discriminant method is adopted. Finally, through the D-S theory of evidence, Obtain the motion state of an unmanned vehicle (Figure 2.3 cited in Using millimeter-wave radar to evaluate the performance of dummy models for advanced driving) assistance systems test).

D-S evidence theory has lower prior requirements. It has significant advantages in dealing with uncertain information such as combination and decision [11]. The Adaboost method uses weighted changes to train and learn the primary classifier and then uses different sampling weights to complete the kernel of multiple classifiers (Figure 2.4). This is to reduce the weight of the correct and wrong samples so that the following classification will focus on identifying the problematic samples. It improves classification accuracy by reclassifying new training samples with existing weights. It combines all classifiers according to some criteria to form a new

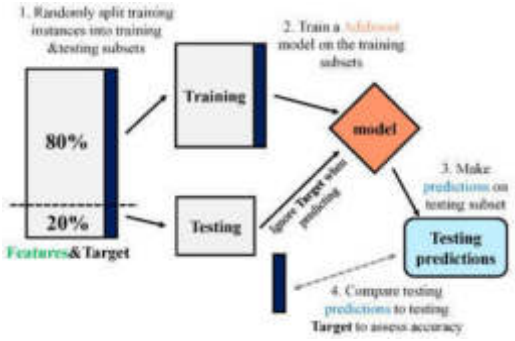


Fig. 2.4: Overall flow chart of Adaboost algorithm.

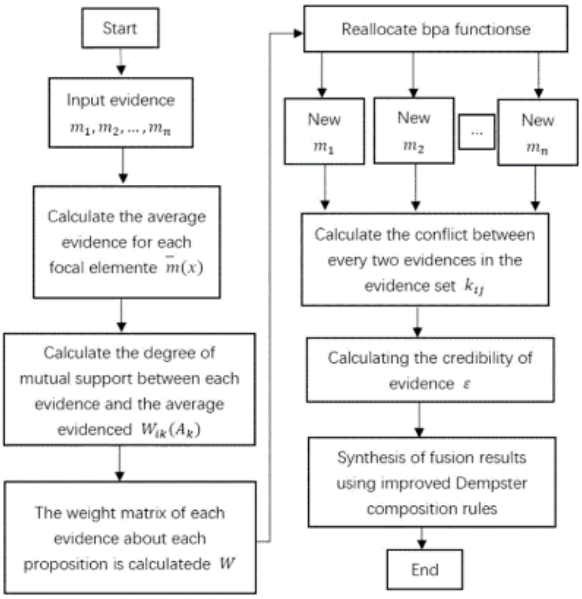


Fig. 2.5: D-S evidence fusion process diagram.

robust classifier, the final decision classifier.

2.2. The basic probability is reconfigured by constructing an evidence vector. The principal diagram of the D-S evidence fusion process based on multi-source information is shown in Figure 2.5. It fully uses the advantages of various sensors to detect the distance and Angle of obstacles ahead through vision and millimeter wave radar [12]. The D-S evidence theory comprehensive criterion is used to fuse multi-source information and judge the final fusion effect.

Define the identity frame as $F = \{B_1, B_2, B_3, \ldots, B_n\}$. Let's call $H_i = [h_i(B_1), h_i(B_2), \ldots, h_i(B_n)]^T$ the evidence vector. The element $h_i(B_j)$ in the evidence vector H_i represents the confidence of class j provided by sensor i. Here $i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$; So, the evidence set is $HH = \{H_1, H_2, H_3, \ldots, H_m\}$. Construct the distance A between the evidence vectors. The distance of the vectors is the degree of directional deviation between the two vectors [13]. This paper uses the Pearson correlation coefficient to characterize the distance between two evidence vectors. Compared with cosine similarity, the Pearson correlation coefficient can prevent error interference when the vector only shifts. It can measure the difference between two spatial vector individuals well.

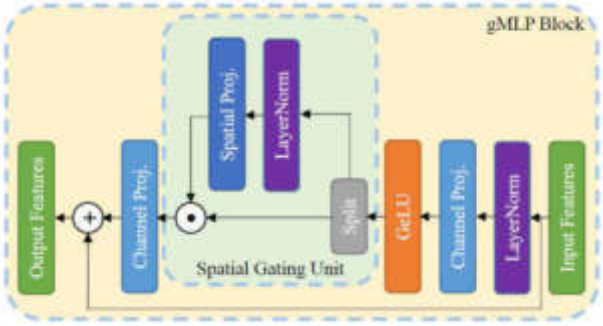


Fig. 3.1: Radar wave detection road surface model.

The distance $s_{\alpha\beta}$ between the two vectors is constructed, which is the directional deviation between the two vectors. This paper uses the Pearson correlation coefficient to characterize the distance between two evidence vectors. This method overcomes the error of the traditional feature extraction method based on wavelet transform. It's an excellent way to measure the difference between two vectors. Under the identification framework $F = \{B_1, B_2, B_3, \ldots, B_n\}$, the distance $s_{\alpha\beta}$ between evidence H_α and H_β is determined by equation (2.1), where $H_\alpha, H_\beta \in HH$:

$$\begin{aligned} s_{\alpha\beta} &= e^{1-\varrho(H_\alpha, H_\beta)} \\ Q\left(H_\alpha, H_\beta\right) &= \frac{\operatorname{Cov}\left(H_\alpha, H_\beta\right)}{\sqrt{\operatorname{Var}\left[H_\alpha\right] \operatorname{Var}\left[H_\beta\right]}} \\ &= \frac{h \sum_{r=1}^{m} h_\alpha\left(B_r\right) h_\beta\left(B_r\right) - \sum_{r=1}^{m} h_\alpha\left(B_r\right) h_\beta\left(B_r\right)}{\sqrt{\sum_{r=1}^{m} h_\alpha^2\left(B_r\right) - \left(\sum_{r=1}^{m} h_\alpha\left(B_r\right)\right)^2} \sqrt{\sum_{r=1}^{m} h_\beta^2\left(B_r\right) - \left(\sum_{r=1}^{m} h_\beta\left(B_r\right)\right)^2}} \end{aligned} \tag{2.1}$$

The evidence weighting factor $\eta\left(H_\alpha\right)$ was determined. The model is used to characterize the importance of the corresponding evidence information H_α in the image. The weighted assignment factor $\eta\left(H_\alpha\right)$ during the final fusion processing is:

$$\eta\left(H_\alpha\right) = \left[1 - in\left(H_\alpha\right)\right] e^{-in\left(H_\alpha\right)} \tag{2.2}$$

3. Low-resolution tunnel identification under multi-sensor information fusion.

3.1. Uncertainty in tunnel detection by radar wave. Millimeter wave radar has the advantages of strong anti-interference, good penetration and high detection resolution. It is suitable for monitoring the road environment throughout the day and is a powerful aid to road perception and detection [14]. Through the analysis of the damaged road surface pits, the concrete steps of detecting the road surface with radar waves are described (Figure 3.1).

When the radar sends out multiple beams to detect the road, the middle component wave in Figure 3.1 deflects outward from the bottom of the pit to form an effective wave. Bounce off the ground and back onto the radar. Multiple pulse signals emitted by radar are processed. The damaged road surface topography can be detected using the residual signal and the actual signal [15]. When the distance between two rays on the boundary of the damaged road is L, and the Angle between one ray and the ground is ζ, which can be calculated by the phase difference between the echo signals received by the radar, and the depth is h, then the value of 1 can be measured and calculated by using the lost waste wave. Then, the irradiation relation shown in Figure 3.1 can be obtained:

$$\begin{cases} h = \frac{l+\Delta\tau}{\sin(\zeta \pm \Delta\omega)} \\ h > \frac{l+\Delta\tau}{\cos(\zeta \pm \Delta\omega)} \end{cases} \tag{3.1}$$

$\Delta\omega$ represents the radar's sweep Angle deviation, ω represents the measurement error, and h represents the size of the crater on the damaged road surface. Its value can be calculated according to the minimum values measured [16]. It can construct a road recognition system based on radar waves.

3.2. Sensor information fusion strategy. Assuming that the boundary detected by vision and radar wave detectors is U_{cam} and u_{rad} , the overlapping area of the two boundaries is R_j, and the area of the two boundaries is R_h, then the intersection ratio of the IOU is:

$$IOU = \frac{R_j}{R_h} \tag{3.2}$$

If the IOU in Formula 3.2 is 0.6 to 1 , it can be determined that both the camera and the radar wave have detected the road, and the measurement data of both are verified and consistent. It has high credibility. If A suitable number is not found, then the global neighbor algorithm is used to conduct data correlation, set the radar and camera failed to match the target point m, n, their distance to the specified point l_r, l_c, according to the distance between the position and the coordinate point, according to this number to arrange the coordinate, you can obtain the vision and radar target point matrix T, U. The residual amount of the radar and camera detection points at the time u is denoting $\xi_{ij}(u)$. It can be expressed by the formula (3.3):

$$\xi_{ij}(u) = \Xi_j(u) - HR_i(u) \tag{3.3}$$

H stands for state transfer matrix, The standardized distance from the camera detection point j to the radar detection point. The following formula can express me

$${l^2}_{ij} = {\Xi^T}_{ij}(u) R_{ij}^{-1}(u) \xi_{ij}(u) \tag{3.4}$$

$R_{ij}^{-1}(u)$ represents the covariance matrix of $\xi_{ij}(u)$.F_i is the threshold of radar detection point i. If the normalized distance is not greater than the threshold, it indicates that the camera detection point is in the coverage area of the radar detection point. They can work together [17]. Each radar detection point corresponds to the camera detection point. The cost distribution in the GNN algorithm is as follows

$$\Phi_{ij} = \begin{cases} l_{ij}^2, l_{ij}^2 \leq F_i \\ F_i, l_{lj}^2 > F_i \end{cases} \tag{3.5}$$

The docking and matching point cloud data on the unpaved road are realized by solving the cost equation. Then, the radar and camera perception data are organically combined.

4. Vehicle fusion perception test in a tunnel environment.

4.1. Identification parameter setting. First, set up the buffer queue system. Because the camera's frame rate selected in this paper is 30 FPS, and the maximum detection range is 60 meters. Setting a buffer time of 30 frames at a time can also ensure driving safety. When an accident occurs during data fusion, the ultrasonic radar's proximity collision avoidance command can be immediately transmitted to the industrial control computer, and an emergency stop command can be executed [18]. It guarantees the safety of the vehicle. A basic probability distribution model based on D-S evidence theory is proposed. Then, in groups of 20 frames, each level of data is gradually entered into the cache. The D-S evidential reasoning method is used to combine the criteria to achieve the purpose of enhancing the detection accuracy. Finally, data such as target type, relative position and relative speed are obtained.

4.2. Test results and analysis. This project selects 20 indicators found in the actual road operation as the research object to test the application effect of the two-layer information fusion method of "target layer + decision layer" in the actual traffic scene. The fusion results of environmental perception in two stages of tunnel testing are shown in Table 4.1. By studying the "target layer and decision layer" fusion, the information fusion between millimeter wave and image can be accomplished efficiently under practical test conditions. In this way, objects in the line of sight can be accurately detected and obtain status information.

Table 4.1: Results of a two-stage perception test in a tunnel.

Target ID	Longitudinal coordinates (m)	Lateral coordinates (m)	Longitudinal relative velocity (m/s)	Lateral relative velocity (m/s)	Target level fusion results
1	24.146	-0.729	6.281	0.010	car/0.80
2	11.208	0.542	-8.552	0.000	person/0.87
3	16.073	-0.708	5.333	0.000	car/0.88
4	11.021	-0.719	1.219	0.000	car/0.87
5	49.927	-0.729	8.115	0.010	car/0.38
6	18.417	-0.646	4.469	0.010	car/0.77
7	32.521	-0.490	5.646	0.000	truck/0.44
8	32.156	-0.552	4.802	0.021	car/0.77
9	36.740	-0.302	3.490	0.010	car/0.88
10	19.875	-0.583	6.229	0.019	car/0.77
11	28.479	0.500	-8.510	0.000	car/0.41
12	73.708	-0.500	3.271	0.010	car/0.85
13	45.469	-0.521	1.854	0.010	car/0.84
14	6.396	-0.490	3.760	0.021	car/0.73
15	19.198	-0.438	6.135	0.010	car/0.74
16	53.125	0.021	10.125	0.000	sign/0.47
17	19.490	-0.677	5.396	0.010	car/0.78
18	21.760	-0.417	6.979	0.010	car/0.77
19	14.917	-0.500	6.406	0.000	car/0.57
20	38.729	-0.448	8.125	0.000	car/0.77

5. Conclusion. According to the characteristics of the current tunnel unmanned vehicle environment, the approach to improve the accuracy of intelligent perception of unmanned driving is studied from the perspectives of hardware selection, layout and information fusion. The experimental platform of the experimental vehicle was constructed, and the underground tunnel with suitable lighting was selected as the experimental site for experimental testing. Experimental results show that this method can improve the image detection accuracy by 9.66%. It can solve the problem that the traditional camera has poor visual perception ability in the roadway. The image information fusion method based on the decision layer proposed in this project will effectively reduce the false reports of images. So, the detection ability of tunnel images is improved effectively.

REFERENCES

[1] Li, Q., Queralta, J. P., Gia, T. N., Zou, Z., & Westerlund, T. (2020). Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments. Unmanned Systems, 8(3), 229-237.
[2] Natan, O., & Miura, J. (2022). Towards compact autonomous driving perception with balanced learning and multi-sensor fusion. IEEE Transactions on Intelligent Transportation Systems, 23(9), 16249-16266.
[3] Yang, P., Duan, D., Chen, C., Cheng, X., & Yang, L. (2020). Multi-sensor multi-vehicle (MSMV) localization and mobility tracking for autonomous driving. IEEE Transactions on Vehicular Technology, 69(12), 14355-14364.
[4] Yi, C., Zhang, K., & Peng, N. (2019). A multi-sensor fusion and object tracking algorithm for self-driving vehicles. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of automobile engineering, 233(9), 2293-2300.
[5] Rick, M., Clemens, J., Sommer, L., Folkers, A., Schill, K., & Büskens, C. (2019). Autonomous driving based on nonlinear model predictive control and multi-sensor fusion. IFAC-PapersOnLine, 52(8), 182-187.
[6] Wang, K., Cao, C., Ma, S., & Ren, F. (2021). An optimization-based multi-sensor fusion approach towards global drift-free motion estimation. IEEE Sensors Journal, 21(10), 12228-12235.
[7] Kovacova, M., Oláh, J., Popp, J., & Nica, E. (2022). The algorithmic governance of autonomous driving behaviors: Multi-sensor data fusion, spatial computing technologies, and movement tracking tools. Contemporary Readings in Law and Social Justice, 14(2), 27-45.
[8] Lian, H., Pei, X., & Guo, X. (2020). A local environment model based on multi-sensor perception for intelligent vehicles. IEEE Sensors Journal, 21(14), 15427-15436.
[9] Tao, X., Zhu, B., Xuan, S., Zhao, J., Jiang, H., Du, J., & Deng, W. (2021). A multi-sensor fusion positioning strategy for intelligent vehicles using global pose graph optimization. IEEE transactions on vehicular technology, 71(3), 2614-2627.

[10] Osman, M., Mehrez, M. W., Daoud, M. A., Hussein, A., Jeon, S., & Melek, W. (2021). A generic multi-sensor fusion scheme for localization of autonomous platforms using moving horizon estimation. Transactions of the Institute of Measurement and Control, 43(15), 3413-3427.
[11] Zhang, S., Zhao, X., Lei, W., Yu, Q., & Wang, Y. (2020). Front vehicle detection based on multi-sensor fusion for autonomous vehicle. Journal of Intelligent & Fuzzy Systems, 38(1), 365-377.
[12] Weon, I. S., & Lee, S. G. (2019). Environment recognition based on multi-sensor fusion for autonomous driving vehicles. Journal of Institute of Control, Robotics and Systems, 25(2), 125-131.
[13] Liu, Z., Xiao, G., Liu, H., & Wei, H. (2022). Multi-sensor measurement and data fusion. IEEE Instrumentation & Measurement Magazine, 25(1), 28-36.
[14] Jung, R., & Weiss, S. (2021). Modular multi-sensor fusion: A collaborative state estimation perspective. IEEE Robotics and Automation Letters, 6(4), 6891-6898.
[15] Yin, R., Cheng, Y., Wu, H., Song, Y., Yu, B., & Niu, R. (2020). Fusionlane: Multi-sensor fusion for lane marking semantic segmentation using deep neural networks. IEEE Transactions on Intelligent Transportation Systems, 23(2), 1543-1553.
[16] Cui, Y., Chen, R., Chu, W., Chen, L., Tian, D., Li, Y., & Cao, D. (2021). Deep learning for image and point cloud fusion in autonomous driving: A review. IEEE Transactions on Intelligent Transportation Systems, 23(2), 722-739.
[17] Li, A., Zheng, B., & Li, L. (2020). Intelligent transportation application and analysis for multi-sensor information fusion of Internet of Things. IEEE Sensors Journal, 21(22), 25035-25042.
[18] Ravindran, R., Santora, M. J., & Jamali, M. M. (2020). Multi-object detection and tracking, based on DNN, for autonomous vehicles: A review. IEEE Sensors Journal, 21(5), 5668-5677.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 31, 2024
Accepted: Jul 5, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 924–939, DOI 10.12694/scpe.v26i2.3774

GENOCARE PROGNOSTICATOR MODEL: HOST GENETICS PREDICT SEVERITY OF INFECTIOUS DISEASE

SHIVENDRA DUBEY*, SHWETA SINGH†, DINESH KUMAR VERMA‡, SUDHEER KUMAR LODHI§, AND SAKSHI DUBEY¶

Abstract. Scientific community understanding of the variance in severity of infectious disease like COVID-19 across patients is an important area of focus. The article presents an innovative voting ensemble GenoCare Prognosticator (GCP) model that incorporates XGBoost and Random Forest classifiers, two cutting-edge machine learning approaches. A large dataset that incorporates medical covariates like gender and age along with biological WES (Whole Exome Sequencing) data was used to train these models. Five-fold stratified cross-validation was used to process the dataset in order to improve model stability and avoid overfitting. Two medical covariates and sixteen recognized candidate gene variants were among the eighteen major features on which our GCP model had been verified using data from earlier studies. Specific post-hoc clarification of the model's predictions was provided by ExplainerDashboard, a Python open-source library, to improve interpretability. Furthermore, we utilized OpenTarget and Enrichr, two bioinformatic resources, to establish connections between the discovered variations in genetics and pertinent ontologies, biological pathways, and possible drug/disease relationships. Unsupervised clustering of SHAP key feature values was included in the analysis, which revealed intricate genetic interactions that affect the severity of the disease. Our results show that although gender and age are the main factors influencing the severity of COVID-19, complex genetic interactions cause severe symptoms in a specific subset of patients. This work contributes to our comprehension of the biological variables influencing the severity of COVID-19 and offers a reliable, comprehensible model that can help recognize patients at high risk and guide individualized treatment plans.

Key words: GCP model, COVID-19 severity, ExplainerDashboard, Genetic, Random Forest, XGBoost.

1. Introduction. Since the initial breakout around late 2019, the SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) remains an imminent danger to humanity. The illness has an extensive range of clinical severity, which indicates that individuals have a complicated and extremely variable response from the host [1]. Genetic variations in the human (host) relation to the severity of infection or susceptibility may offer hints to the best places to focus on developing therapies or possibly preventative efforts to assist in developing drugs and vaccines to combat the infection with SARS-CoV-2 [2]. The field of science is most interested in these discoveries because they might offer crucial hints about how to use already available medications to treat the deadly COVID-19 sickness and SARS-CoV-2 infection [3]. Additionally, we could likely identify populations of people that are either innately resistant to the infection caused by SARS-CoV-2 or who may be in abnormally elevated danger and require safety in the general community [4–7]. Uncommon abnormalities in genes that can lead healthy people to develop a dangerous reaction to COVID-19 disease are another way that the SARSCoV-2 genomic severity and vulnerability might present themselves [6].

Numerous medical studies have found considerable correlations between patients' and comorbidities' illness susceptibility and severity, including hepatitis, diabetes, kidney-related issues, HIV, gender and age [7, 8]. A few hosts become more prone to contracting a serious illness, most likely due to the regulated impact caused by

*Department of Artificial Intelligence and Machine Learning, Manipal University, Jaipur, Rajasthan, India, 303007 (shivendrashivay@gmail.com).
†School of Engineering and Technology, Jagran Lakecity University, Bhopal, Madhya Pradesh, India, 462042 (drshweta.singh2011@gmail.com).
‡Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, Guna, Madhya Pradesh, India, 473226 (dinesh.hpp@gmail.com).
§Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, Vadodara, Gujarat, India, 391760 (sudheer.lodhi30@gmail.com).
¶Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, Vadodara, Gujarat, India, 391760 (sakshi.pateriya27@gmail.com).

genetics, environment, and associated risks. Unanswered questions surround why individual patients respond to COVID-19 infections so differently. Numerous research linking the investigation of human genomics to illnesses have identified certain relationships between particular patient subgroups' severity of disease and these associations [9]. For instance, after being exposed to this sickness, fit, young individuals without any pre-existing health issues had serious symptoms, while others even passed away due to the ailment. Recent evidence demonstrates that asymptomatic individuals' antibody responses against the COVID-19 virus are less robust [10–12].

The variation in COVID-19 severity susceptibility and consequences within individuals can be explained in part by certain intricate genetic connections in the illness on the host side [13]. Their DNA may include significant data about how the illness varies dramatically among individuals. In addition to revealing trends inside human immune systems, the expression of genes could be a crucial factor in regulating whether the host's immune system responds to the virus. To better understand those intricate interactions, which are essential for shedding greater clarity on the genetics of the illness, choose substances for repurposing - and understand those who are particularly susceptible or offering certain kinds of defense toward infection [14, 15], it's now needed to examine the genetic makeup of individuals that demonstrate a severe reaction to COVID-19. The majority of academic investigations have thus far used the GWS (Geno-Wide Association studies) method to pinpoint the genetic variations and marker chromosome sites [16, 17, 18]. Such investigations have proven significant in identifying important gene variations connected to the illness. In patients, their investigations revealed an enhancement of these genetics, leading them to conclude that genes may influence an infection's clinical outcome. The GWS strategy, however, often uses stricter criteria to weed out variations in genes linked to an illness. For instance, to discover extremely significant gene variations associated with a specific disease [19, 20], the acceptable level of the tiny p-value is typically decreased. Furthermore, no scientific model explains how genetic elements in patients' COVID-19 susceptibility and potential illness severity can work together. For those who work with WES (Whole Exome Sequence) data sets associated with forecasting genetic severity and offering clarification in the association of discovered genetic variations related to the SARSCoV-2 severity between individuals, ML techniques are still not readily accessible at the moment. Specialists in all areas of healthcare science might profit from the biological knowledge and analysis provided by understandable machine learning (ML) methods which connect recognized inherited variation indicators to more fully comprehend the complicated interactions between genes that could promote or prevent methods of therapy on the way towards customized medicine [21–23].

Many research investigations have used model-interpretation methodologies to find novel information and hypothesis that can be tested [24–26]. Thus, the genetic variants revealed by a prior investigation and associated with COVID-19's severity among individuals. Employing 16 detected variations of genes and medical variables (gender as well as age) as of an earlier investigation employing a 2000 sample of individuals' WES dataset [20], the present research interprets a post-hoc model. The ExplainerDashboard post-hoc evaluation is based on the GCP prediction models. Researchers in social science may employ the ExplainerDashboard, a freely available Python tool, to clarify the findings related to the model's local and global predictions. Considering the help of individuals' medical data and likely complicated genomic relationships, we use this method to assess how the degree of severity of the illness's outcomes would affect the host's health [20]. To increase the model's predictive strength, we used a five-fold cross-validation split analytical technique over the initial issue data to combine three trained models based on decision trees (XGBoost and Random Forest classifiers). Using an enriched dataset (Pathways, Transcription, Drugs/Diseases, Ontologies, and Cells Types), they subsequently performed domain interpretation of the genetic variations related to the disease's severity [27]. To achieve these goals, this research is designed to employ the GCP model created from a prior study of COVID-19 severities on 2000 cohorts.

1.1. Background Information.

Information about COVID-19 and WES. We discuss the importance of WES (Whole Exome Sequencing) and the potential influence of genetic variations on the severity of COVID-19. Explain the importance of combining the analysis of clinical and genetic data in the COVID-19's context.

Justification of Feature Engineering Decisions. Talk about the rationale behind the selection of odd-ratio statistics for detecting variation in genes and the significance of incorporating covariates such as gender and

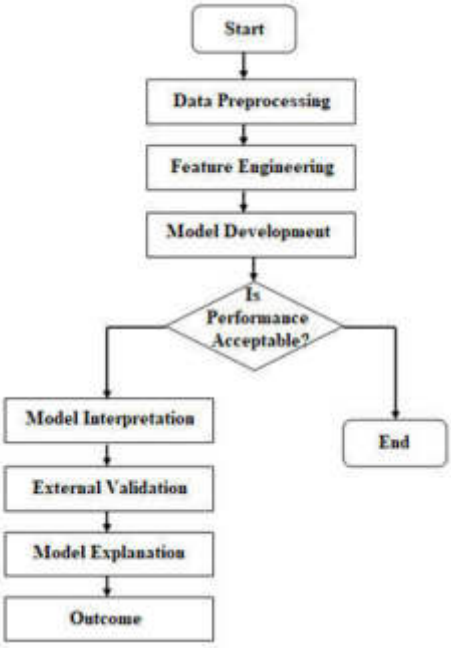


Fig. 1.1: Flow Graph

age. Describe the history of genetic variant evaluation and the relationship between specific variants and the severity or susceptibility of a disease.

1.2. Model's Basic Principles.

Ensemble Voting Fundamentals for Genomic Research. Describe ensemble voting classification techniques and the reasons that, in a multidimensional genetic context, combining multiple models improves predictive power. Talk about how the GCP model uses ensemble methods to take genetic data variation into account.

Genomic's Interpretability. In order to highlight significant clinical factors or genetic variants that contributes to the severity of COVID-19, describe how SHAP (SHapley Additive exPlanations) values are used for interpreting predictions from models and why feature importance is essential in a clinical-genomic context.

1.3. Appropriate Scenarios.

Medical Importance. Explain situations in which this model might be used directly, such as determining people at risk based on their genetic vulnerability or comprehending the impact of particular genetic variants on COVID-19.

Risk Management and Personalized Medicine. Stress the model's applicability to personalized healthcare, where patients with particular genetic profiles may benefit from tailored treatment regimens based on these predictions.

1.4. Contribution of the Research. By creating an innovative casting votes collective GenoCare Prognosticator (GCP) model, the present research addresses the crucial problem of comprehending the different levels of severity of infectious disease like COVID-19 among patients. This model combines clinical variables like gender and age with sophisticated machine learning methods, namely XGBoost and Random Forest classification techniques, trained on an extensive WES (Whole Exome Sequencing) dataset. In order to guarantee model reliability and stability, the study uses a 5-fold classified cross-validation method. The efficiency of the GCP model is verified with 18 essential characteristics, 16 of which are genetic variants found in earlier studies. To improve the predictability of the results, the study also uses ExplainerDashboard for ad hoc justifications. Further connections between variation in genes and ontologies, physiological pathways, and possible drug or

disease connection are made through the use of bioinformatic resources such as Enrichr and OpenTarget. The results show that although gender and age are important indicators of COVID-19 severity, complicated genetic relationships are also important for a subgroup of patients.

2. Related Work. Multiple investigations have determined the function of susceptibility and genetic factors to COVID-19 disease. For instance, the genetic makeup of the host could affect the susceptibility towards respiratory tract infections, together with additional risk variables, made this observation. It is thought that the mutation in the ACE2 (Angiotensin Converting Enzyme 2) genetic material is a marker of genetic susceptibility for infection with SARS-CoV-2 since the pathogen needs this particular gene to get into organisms. In addition to ACE2, other significant enzymes affecting illness severity include DPP4 (Dipeptidyl Peptidase-4) and TMPRSS2 (Transmembrane Protease Serine 2). The importance of host genetics for the SARSCoV-2 entry with replication and in mounting the immune response in the host demonstrated that several genes may be critically important in determining the severity dynamics of COVID-19 patients [28, 29, 37].

The three potential genetic entry points were suggested through their research: the genomes governing the complement system and toll-like receptor routes; the genome's Human Leukocyte Antigen locus, a key control of resistance toward an infection; and changes within the ACE2, which is the gene, which affect the patterns of the disease's spatial spread. The torrent of cytokines resulting from this then appears to be what causes the increased inflammatory processes that characterize the degree of COVID-19 severity in individuals. To forecast the extent of SARS-Cov-2 among individuals, they developed a model based on deep learning on computed tomography (CT) data. An AI-severity model that enhances prognostic efficiency was created by building an integrated AI-severity index that combines five biological and clinical parameters (platelet, sex, age, urea, and oxygenation) [30]. The results from the investigation indicated that artificial intelligence (AI) methods, like computed tomography neural network evaluation, may provide clinicians with specific prognostic data. The supervised machine learning (ML) techniques (Artificial Neural Networks, Decision Trees, Logistic Regression, Naive Bayes, and Support Vector Machine) with a dataset comprising positive and negative COVID-19 patients that have been epidemiologically labeled [24]. Their research demonstrated that the decision tree approach worked better when forecasting the course of the illness than any other approach. Simulation of COVID-19 sufferers' death using comprehensible machine learning (ML) methods, while the outcomes of their investigation revealed key mortality rates for the disease predictors [31].

Lymphocytes, hs-CRP (high-sensitivity C-reactive protein) and LDH (Lactic dehydrogenase) were chosen by the machine learning (ML) algorithms as the three types of markers which most accurately indicate an individual's mortality for ten days or more ahead. The majority of instances necessitating prompt healthcare intervention appear to be distinguished by excessive amounts of LDH alone. machine learning (ML) methods were developed to forecast the demand for critical care and ventilation machines using blood panel profile data and medical records. The results of their investigation proved that the three kinds of data are essential for medical facilities when preparing for emergencies for COVID-19 and allocating critical care and ventilatory therapy to individuals. A combined machine learning classifier and biological algorithm were coupled in the investigation to remove pertinent information and carry out classifications to determine COVID-19 using blood samples of patients [32]. The outcomes demonstrated that machine learning (ML) methods may enhance existing clinical procedures and equipment and advance cutting-edge approaches to combat the illness. Their study used an OGA-ELM using three criterions for selection for recognizing COVID-19 from X-ray imagery: roulette wheel, K-tournament, and random [33]. The study outcomes demonstrated that OGA-ELM may be employed to obtain 100 percent efficiency using quick time for computation. It showed that OGA-ELM (Optimized Genetic Algorithm-Extreme Learning Machine) is a successful technique over COVID-19 detection employing images from chest X-rays. It is essential to identify hospitalized Covid-19 sufferers at a higher risk of developing serious illnesses as soon as possible. Their research developed, confirmed, and externally tested a model using machine learning (ML) to predict initial hospital mortality or ventilatory needs based on laboratory and clinical information collected at arrival [5, 34]. The outcomes provided substantial proof that machine learning models might prove helpful in deciding which individuals should be admitted to hospitals and estimating the likelihood of developing a serious Covid-19 infection during emergencies in medicine, like disease outbreaks. Since the start of the global epidemic, investigators have used a variety of machine learning (ML) approaches to reduce the threat posed by the SARS-Cov-2 virus. Several of these approaches have proven

helpful in determining the presence of COVID-19 and in forecasting severity and death risk through easily accessible medical and laboratory data [7, 30].

Using the WES information about roughly 4000 SARS-CoV-2-positive individuals and also examined frequent and unusual variations [35]. Using this, a model based on machine learning that can predict the COVID-19 severity was defined. According to whether every gene had variations or not, those variations were transformed into distinct Boolean sets characteristics. To find among the most useful Boolean variables on the genetic basis of severity, they created a combination of LASSO logistic regression models.

Many investigations have focused on predicting the severity of COVID-19, and a variety of statistical and machine learning models have been established for recognizing key variables influencing outcome of patients. To predict the severity of a disease, conventional approaches frequently rely on medical information, including gender, age, and comorbidities. Although these models have yielded significant findings, they often fail to take into account the biological variables that can have a major effect on the course of a disease.

Forecasting models have been created using current approaches such as decision trees, logistic regression, and simple ensemble techniques. However, these approaches frequently have the following shortcomings:

1. Restricted Feature Consideration: A lot of current models mainly concentrate on medical characteristics without including extensive genetic data, which could lead to the loss of important genetic relationships that could affect the severity of the disease.
2. Model Overfitting and Stability: When working with small or unbalanced datasets, models that do not have strong validation procedures tend to become unstable and overfit.
3. Interpretability: Although certain machine learning models yield precise forecasts, they frequently function as "black boxes," providing minimal understanding of the decision-making procedure. This opaqueness may make it more difficult to comprehend the underlying biological mechanisms.
4. Generalizability and Scalability: The applicability of models developed on minimal or specialized datasets may be limited in distinct patient groups due to poor generalization to larger populations.
5. Data set description: Quantity of this dataset up to 2,000 COVID-19 patients are expected to be taking in this study.
   The quantity of clinical and genetic information gathered for every patient would determine the dataset's size. This could include medical history, demographic data, COVID-19 severity data, and whole exome sequencing data.
   The dataset's main components are:
   Clinical Data: Clinical data includes medical history, demographics (age, sex), and details about the severity of COVID-19 (e.g., oxygen requirements, hospitalization status).
   Genetic Data: Whole exome sequencing data, which offers details on the genome's protein-coding regions.
6. Investigation Design Essentials:
   A model Hyper-Parameter Contexts: List the hyper-parameters that are used for every model, including the number of Random Forest trees, the XGBoost learning rate, and other pertinent parameters. Declare if the hyper-parameters were set to their default values or adjusted using cross-validation.
   Uniformity Among Models: Underline that, in order to guarantee fair comparison, the identical hyper-parameter setting procedure was used consistently across models. Give specifics to explain any hypothetical differences if a particular hyper-parameter was changed.
   Dataset Size Consistency across Comparisons: To ensure a fair evaluation, make sure that every model in the comparable experiment has undergone testing and training on the identical dataset sizes. It should be clearly stated and supported by a justification if any changes were made to the dataset sizes for particular models.

3. Material and Methods. The qualitative collection employed in this research was organized the directed by the University of Siena and the GEN-COVID Multicenter Research Team and had a calculated enrollment of 2000 those participating as part of the population sample wherein the information on genetics obtained contained the healthcare information types that were employed. For training and creating the GCP framework, the original 2000 cohorts and variables (gender and age) dataset were used. The GCP framework

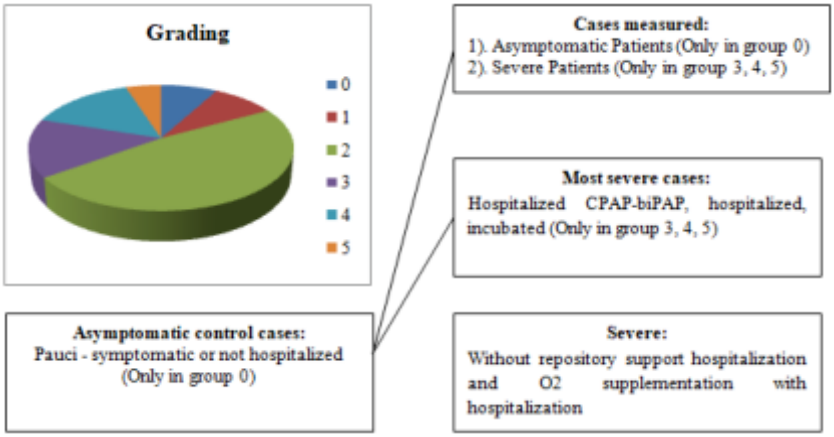


Fig. 3.1: Patients classifications

was created using 16 reported potential variations in genes and medical variables (gender and age) [20]. This research comprised out-of-sample data utilized for outside prognosis and contained data that the identical GEN-COVID team had used previously for validating the GCP framework. We employed the analogous data preparation techniques in this research that we operated in the training data analysis from 1920. We considered some classification scoring schemes first is unadjusted-by-age and second is adjust-by-age, to undertake subsequent analysis (such as post-hoc interpretations and external validation of the model) [3, 11, 35].

We selected the frequencies of alleles for each of the 16 discovered variants in genes in the Whole Exome Sequence dataset, along with the associated medical variables, to generate features vectors in outside validation of the model. The COVID-19 result severity ratings for the individuals had been binarized; ratings 3, 4, and 5 were assigned as "severe" along with a score of 1; however, scoring 0 was classed as "asymptomatic " and marked as 0. The study did not include individuals with severity grading of 1 or 2. The removal was done to reduce noise signals when variations in genes associated with illness severity or protection for patients were being filtered. We used a normal linear method to improve the scoring classification of sex-stratified individuals 36, using age as a parameter for input. We retained only individuals with scoring categories agreed upon alongside the class that took age into account. The age range of those taking part (Older Adult, Adult) starts at eighteen years older. Both genders have been taken into account for this research's gender. When using an additive approach for establishing severe and normal categories, individuals' genetic data includes alternative (Alt) or reference (Ref) alleles, with homogeneous genotypes (1/1) having double the protection (or risk) of a heterozygous allele (1/0 or 0/1) (see Figure 3.1).

Our earlier investigation regarding the 2000 population data has additional details and overviews of preprocessing of the Whole Exome Sequence (WES) dataset [20]. The out-of-sample dataset's 618 individuals who passed the selection criteria and filtering were utilized for the outside validation of the model. The parameter matrices have 18 features (2 covariates and 16 genetic variants) for CGP model third-party verification. Allelic frequency counts for every individual's genotyping data have been allocated to everyone from the 16 variations, which comprised the matrix of features (1/1 counted as 2, 1/0 or 0/1 counted as 1, and 0/0 counted as 0). The variables were subsequently combined into the feature matrix using the phenotypic individual's data. As indicated in Figure 3.1, the result of the parameter (classifying) is categorized as 1 and 0, where 1 denotes the disease's severity, and 0 is asymptomatic.

Employing the out-of-sample dataset. We used a four-scenario method for validating the GCP framework. The primary goal of the initial situation was to validate the hypothesis employing the matrix of features for the adjusted-by-age grade grouping system as shown in figure 3.2.

Hypothesis.

Generalizability of Model: Describe how the model is assumed to perform similarly through a range of datasets.

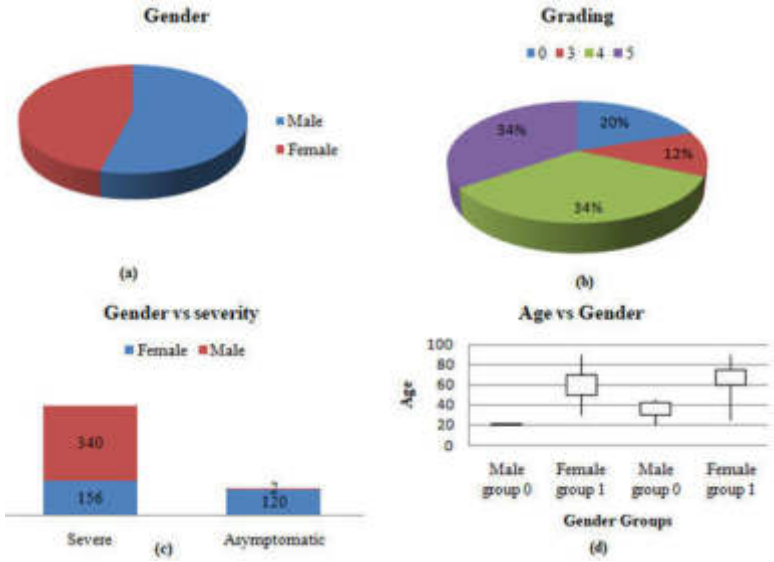


Fig. 3.2: Human phenotypic data from the following up WES dataset, graded according to age

It includes presumptions regarding population heterogeneity, like comparable clinical response patterns or genetic distributions.

Genetic Variant importance: Make it clear that those discovered genetic variations have a significant impact on the severity of COVID-19, even though you acknowledge that some associations may vary depending on the context or among individuals.

Requirements for External Validation: Make it clear that model reliability in an outside dataset is predicated on feature distributions and data quality that are similar as the distinctive training data, which may not be true for every application in real life.

Age-adjusted scoring dataset. That is a modification of the scoring classifier that uses a normal linear classifier that incorporates age as a parameter variable in sex-stratified individuals. It includes testing and training data. The testing and training data used for unprocessed scoring classifications are regarded as an "unadjusted with age scoring database." The initial train data set, or 20 percent of the testing data taken from every one of the five-fold CVs, is the collection of data employed for validation within a GCP framework. This group of samples, which uses an adjusted-by-age scoring system, contains 841 units of measurement.

Testing set. This is the name of the information set utilized to perform the GCP algorithm's validation from outside, and it falls within the "adjusted-by-age" rating system class. Each sampling section's Whole Exome Sequencing was eliminated in this group of genes to exclude 16 mutations using the identical standards as the initial data set used for training. It contains 618 units for sampling.

Observations. The GEN-COVID multimodal collaboration offered them the benchmark training dataset, which included the first sample with a previous investigation of 2000 individuals. They then conducted follow-up interviews with 3000 individuals, including 1000 individuals. Due to the commonality that emerged, multiple specimens were excluded from this investigation. Any instances identified in the subsequent data using the unadjusted-by-age classification method as asymptomatic or severe have been incorporated into the testing set for eliminated items. In this grouping, there are a total of 235 instances.

All samples identified as severe or asymptomatic using the initial training dataset's unadjusted-by-age grading system were omitted from the benchmark training dataset's adjusted-by-age scoring system. This grouping includes 357 units of sampling. Those who had been collectively removed were all rated employing the unadjusted-by-age rating system as asymptomatic or severe. It was the sum of all the removed initial testing and training instances. This group of samples includes 495 units for sampling.

Algorithm 7 GCP Approach

1 Get genetic information (X) using many features m and a number of samples n; additionally, gather labels (y) corresponding to every sample's severity.
2 Deal with X and Y values that are missing. Create representations of numbers for categorical variables in X (for example, using one-hot encoding). Standardize (zero mean, unit variance) the numerical properties of X.
3 Select XGBoost and Random Forest as your predictive machine learning algorithms for severity predictions.
4 For each of the ensemble's trees, i: Take a sample of the training data's bootstrap dataset (X_train_i, y_train_i) and trained a decision tree model using the training data for X and Y.
The Random Forest model's last forecast is: For all i in the ensemble

$$\texttt{y\_pred\_rf} = \frac{1}{\text{Number of Trees}} \times \Sigma(\texttt{Tree\_i.predict(X\_test)})$$

.
5 Training of Model (XGBoost): A sequential weak learner's ensemble, such as decision trees, is constructed via XGBoost.
Set a model's prediction initialized to zeros: y_pred_xgb = 0
Calculate the loss function's negative gradient (L) with regard towards the actual label (y_true) and the existing predictions (y_pred_xgb_t) for every boosting round (t): negative_gradient = - δ L / δ y_pred_xgb_t
To fix the mistakes in earlier rounds, match an entirely novel weak learner (such as a decision tree) to an existing negative gradient.
Revise the following prediction about the tth round of boosting: y_pred_xgb_t+1 = y_pred_xgb_t + learning_rate * new_weak_learner.predict(X_test)
The XGBoost model's last forecast is as follows: y_pred_xgb = y_pred_xgb_T # where T is the rounds of boosting.
6 Model Evaluation: To evaluate the performance of both models, compute evaluation measures (such as F1-score, precision, recall, accuracy, ROC-AUC).
Possible model combination: Make an ensemble prognosis by combining the predictions from the two models, including:
Ensemble prediction = (y_pred_rf + y_pred_xgb) / 2
7 Model Deployment, Interpretability, Updates, and Model Maintenance.

Modeling explanations by Post-Hoc. We used the subsequent methods to fulfill the objectives and goals of the research. Utilizing the explanation panel platform, we provided a post-hoc concept explanation utilizing the adjusted-by-age data.

Evaluate the significance of the features in the XGBoost and Random Forest models. All trained XGBoost and Random Forest algorithms should be kept for use in novel information prediction. Keep track of the model's performance periodically and, if required, change the simulations with fresh data. The stored GCP model created through training with all the variations and variables (gender and age) based on a straightforward stratified five-fold cross-validation splitting method used within the 2000 cohorts datasets research is loaded first. After that, we used the out-of-sample data to validate our hypothesis [19]. The separate classifiers were aggregated by the GCP using the "VotingClassifier" (an ensemble model) technique using the "sklearn.ensemble" py package according to the predicted probability (soft margin) regarding the final result. We discuss below some performance parameters:

Confusion matrix. A performance assessment statistic called the precision rating is employed to gauge the ability of a model to accurate prediction. It calculates the percentage of precise positive forecasts and all-around positive forecasts produced through the framework. The GCP model's performance based on the validation by an outside data was assessed using the precision score. It had been created to evaluate the degree to which the

system can accurately detect positive situations without overly frequently causing false positives to occur.

Accuracy score. An accuracy score represents a statistic that assesses a model's performance on a specific dataset within a machine learning classification problem. It calculates the percentage of the model's predictions that were generally accurate. Considering this instance, the preserved GCP model made predictions using separate data. The ratio of accurate predictions produced by the framework was subsequently determined using an accuracy score to assess how effectively the classifier worked with that data set.

$$\text{Accuracy Score} = \frac{TN + TP}{TN + TP + FN + FP} \tag{3.1}$$

Precision score. A performance assessment statistic called the precision rating is employed to gauge the ability of a model to make accurate predictions. It calculates the percentage of precise and all-around positive forecasts produced through the framework. The GCP model's performance based on the validation by outside data was assessed using the precision score. It had been created to evaluate the degree to which the system can accurately detect positive situations without frequently causing false positives.

$$\text{Precision Score} = \frac{TP}{TP + FP} \tag{3.2}$$

Where; TN = True Negative, TP = True Positive, FN = False Negative, FP = False Positive

Recall score. A recall score is a form of assessment measure that gauges an algorithm's sensitivities in accurately detecting positive cases. The recall rating measure was employed to evaluate the GCP simulation upon a new dataset. Its ratio of the model's accurate predictions that are positive to the overall number of real instances that are positive in the input data provides a gauge of the degree to which the framework recognizes true positives.

$$\text{Recall Score} = \frac{TP}{TP + FN} \tag{3.3}$$

F1-score. The model's accuracy in binary classification problems is measured using the f1-score, an assessment measure. The average harmonic of recall and precision constitutes the f1-score as the f1-score has an acceptable range of 0 to 1, with 0 denoting the least desirable efficiency and one representing flawless recall and precision. A common baseline employed to evaluate the effectiveness of various binary classification algorithms is known as the f1-score.

The Matthew Correlation Coefficient. The MCC (Matthew Correlation Coefficient) is a number which is varying from -1 to +1, whereby -1 denotes the most negative classification possible, 0 indicates the random type, and +1 represents the best class possible. Compared with other binary-based measures like F1 score or accuracy, MCC considers each of the four parts of the error matrix (TP, TN, FP, and FN). As a result, it is thought to have been an additional reliable measure.

$$MCC = \frac{(TN * TP - FN * FP)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

where TN, TP, FN, and FP represent the number of true negatives, true positives, false negatives, and false positives respectively.

The PR (Precision-Recall) Curve. The performance assessment measure, frequently employed to rate the effectiveness of a classification system with on binary form, corresponds to the PR curves—plotting the recall and Precision metrics for a binary classification algorithm at various probabilities threshold results in a precision-recall turn.

ROC Curve. A graphical depiction plot called the ROC (Receiver Operator Characteristic) curve demonstrates the diagnosing capability of classified jobs as binary. The receiver operating spectrum (ROC) curve was used to show how effectively the investigation's preserved GCP framework extrapolates from the outside data [37].

Table 4.1: Summarizes the results of outside validation of models for performance metrics

Study	MCC	Precision	F1-score	Recall	Accuracy
Aggregated excluded samples	65.792	90.513	88.143	85.892	84.441
Testing set	64.081	99.491	88.091	79.031	82.851
Excluded samples Training set	63.172	88.153	85.521	83.042	82.354
Excluded samples Testing set	62.121	95.574	88.820	82.971	83.832

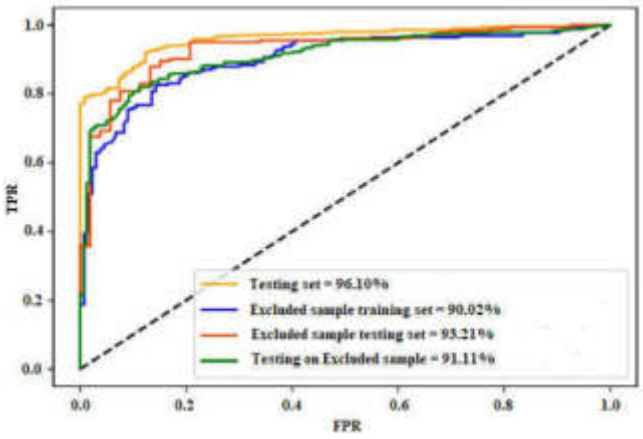


Fig. 4.1: Shows GCP model accuracy when out-of-sample validation of models is considered

Log-loss. We used this important classification measure to evaluate the performance of the ensemble classifier. The binary classifier's gloss measures the extent to which the probability of prediction matches the associated real or true possibility (0 or 1). The more significant the log-loss number, the further the expected chance deviates from the real number.

The ROC curve Interpretation. The Specificity (1 - FPR) and TPR (True Positive Rate) trade-offs of the GCP framework were displayed using the receiver operating characteristic (ROC) curves. A higher accuracy is indicated by the classifier used for voting-producing curves that are closest to its top-left quadrant. A random classifier will be expected to provide values (FPR = TPR) diagonally as benchmark reasons. The test may be inaccurate depending on how close the curve approaches the ROC space's 45-degree orthogonal.

4. Results.

Model Comparative Analysis. Clearly identified the metrics (e.g., precision, accuracy, recall, ROC-AUC, F1-score, log-loss) that are being used when comparing models. We make sure these metrics are computed consistently across all models and presented in visualizations or tables in an extensive manner.

We offered post-hoc modeling explanations regarding the GCP model predicted result, taking into account the system of rating that is not modified for age (see Fig. 3.1).

Validating hypothesis externally: Using the 16 discovered probable gene variations and the two related medical variables (gender and age), we consider the subsequent instances to validate the model.

To train decision trees (XGBoost and Random Forest classifier merged across a five-fold Cross Validation), we had previously developed the GCP model [19]. Advanced algorithms for machine learning, which demonstrate certain comprehension capabilities because of its recursion tree-based voting framework, were used to create the GCP simulation voting classifier. Given that the internal approach is challenging to grasp, we adopted this method instead of selecting an advanced framework. As part of the framework's explanations process, the explainer dashboard technique has been employed to examine and show critical regional measures, including overall SHAP permutations feature significance and dependency graphs. The cumulative application of every feature for every decision stage and the resulting permutations of each of them, sorted and shuffled according to

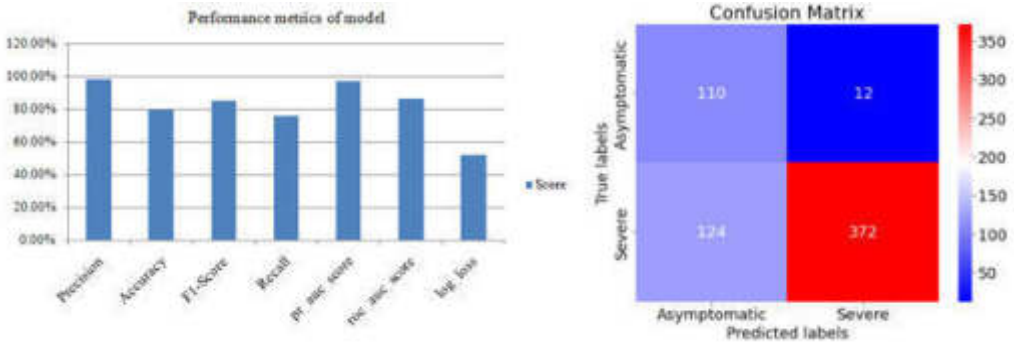


Fig. 4.2: Classification Statistics using ExplainerDashboard

their value in absolute terms, assign the SHAP features important value used in our ensemble system. We can better comprehend the weightings of each for all 18 variables by presenting them on a bar plot, which allows us to figure out which features provide the most accurate COVID-19 result severity prognosis in individuals. The efficacy of the ensemble approach upon the outside validation sample is summarized in the following table 1. The findings demonstrate this, independent of the cohort; the GCP simulation ensemble classifiers we built effectively replicate crucial data which forecasts a serious result in COVID-19 individuals. Additionally, it is important to note whether the outcome of the outside sample is comparable with the efficiency of both the validation and training sets found in the cohort used for the training dataset. It also indicates that the GCP framework ensemble classifier could collect the appropriate information that forecasts individuals' COVID-19 illness severity outcomes. Figure 4.1 shows the GCP model accuracy when out-of-sample validation of models.

Conclusions of Investigation of Post-hoc Interpretation and Justification of GCP Models: Subsequently, concerning GCP estimates, we conducted post-hoc model-agnostic explanations and interpretations at the individual level. To further comprehend the complicated relationships between clinical and genetic variables, we used the ExplainerDashboard explanation and interpretation technique, emphasizing both SHAP feature importance and dependence plots. Here, we aim to reveal hidden discoveries, including individuals where COVID-19 severity estimates are influenced through complicated interactions between genes among the 16 discovered continuous traits rather than variables (gender and age). The GCP Does Not Perform fundamental EDA techniques like descriptive statistical methods summaries and histograms or bar graphs. There are possibilities to examine various post-hoc interpreting outcomes across the Explainer Dashboard. The ExplainerDashboard dynamic interface's GCP framework outcomes are shown in Figure 4.2 to Figure 4.4.

Figure 4.2 displays an ExplainerDashboard's classifications stats findings for the GCP model for prediction, including plots for the classification, precision, and performance matrix. The proportion of positive is obtained for all the bins after the data points had organized into a set of approximately comparable expected probability. A model that has been adjusted precisely will display a straight line connecting the bottom right and left corners.

Using an outside following-up cohort's sample, the plots showed the ROC AUC and PR AUC (Precision-Recall Area under Curve) performances (see Figure 4.3). Its purpose was to assist us in determining the extent to which superior their created ensembles voting classifier was when compared to random.

You can choose an ID of any sample by selecting it from various options within the pop-up dialogue box, and you can press the select at random sample_ID option to choose one sample_ID randomly, which satisfies the requirements. It was done to aid in evaluating the overall percentage of false negatives and false positives for our prediction. For the chosen sample_ID for concern, a doughnut predictions graph (see Figure 4.4) displays the expected chance associated with every group descriptor. Figure 4.5 shows the results comparison of our proposed model with various existing model in terms of different performance parameters.

5. Discussion. Whenever considering the severity displayed by various SARS-CoV-2 individuals, it continues to be a great deal that has to be understood. For instance, why are some individuals, despite their young

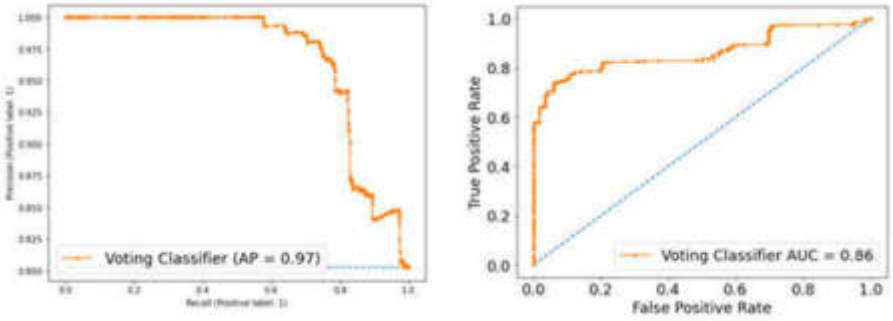


Fig. 4.3: Classification Statistics

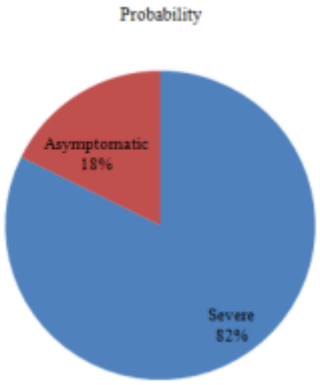


Fig. 4.4: Predictions plot

age and lack of comorbidities, more severely sensitive to the illness than others? Our findings are significant in three distinct respects. Firstly, utilizing information from an earlier investigation we conducted employing the 2000 sample WES and medical databases, we discovered 16 probable gene variations that probably drive COVID-19 severity results in individuals. The GCP model was then additionally enhanced by merging many conventional, comprehensible machine learning (ML) classifiers (XGBoost and Random Forest, that is, decision tree-based models over a straightforward stratification train a five-fold splits Cross Validations), and a second validation was performed employing a subsequent dataset. Secondly, we performed post-hoc explanations using an explanation dashboards freely available Py package. Through utilizing specific OpenTarget, the web interface and the domain expertise utilized in the phenome-wide correlation approach, we were able to link the 16 discovered variants in the genome to illness features that might produce a credible therapeutic trajectory of the COVID-19 disease in individuals.

We provide the following points to make sense of its application and improve the comprehension:

1. Goal and Justification: Describe the goal of 5-fold classified cross-validation, that is to guarantee that the variance of significant variables, including gender, age, and severity levels of COVID-19, is the same across all subsets (folds) used in the validation process. By using this method, evaluation biases in the model are mitigated and the model's effectiveness metrics are guaranteed to accurately reflect its capacity for generalization.
2. Specifics of Implementation: Explain the process used to apply the stratification. To be more precise, describe how the dataset was split into 5 folds of equal size, all of which kept the same proportion of important features (like demographic or severity levels factors) as before. A short overview of the method or instrument utilized to accomplish this, such as the StratifiedKFold works from machine

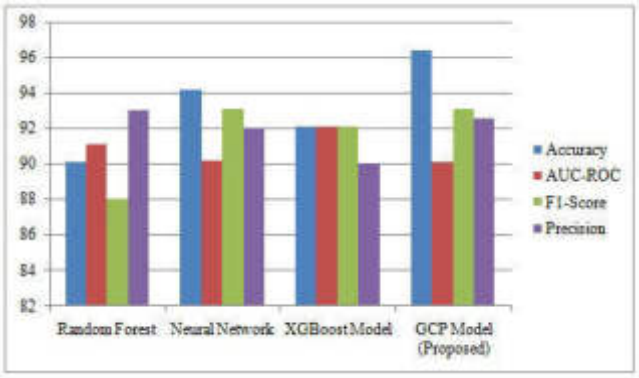


Fig. 4.5: Results comparison

learning library systems like scikit-learn, can be included in this.

3. Model Validation and Training: Clearly state that the model was validated on the fifth fold after being trained on four of the folds in every iteration of its 5-fold cross-validation. Five iterations of this process were carried out, using one fold per validation set. This provides a reliable indicator of the model's performance by guaranteeing that every statistic in the dataset is analyzed only one time for validation as well as training.
4. Data Leakage and Integrity Avoidance: Talk about any measures implemented to stop data leaks, making sure that no data from the training phase gets into the validation fold. This is especially crucial if each patient has several samples, or whether the dataset contains features that might unintentionally result in leakage.
5. Imbalanced Data Handling: If appropriate, describe any methods used to correct class imbalances within each fold. Make sure the model does not become biased in favour of more common classes by using class weighting, undersampling majority classes, or oversampling minority classes.

Last but not least, the team utilized the hierarchical clustering method to reveal information obscured from the explainer dashboard's SHAP feature significance rankings.

With interactive visualizations, ExplainerDashboard is a Python library that is open-sourced that makes machine learning algorithms easier to understand. In the present research, it was especially utilized to produce in-depth visual representations of the predictions made by the GCP model. Important elements consist of:

Feature Importance Visualization: The Feature Importance Visualization aids in identifying important variables influencing the severity of COVID-19 by displaying which features have the greatest impact on the model's predictions.

Interactive Exploration: By allowing users to interactively examine how various input values impact predictions, they can gain understanding of how the model behaves.

Decision-Making Transparency: By illustrating how every aspect influences the final predictions, it facilitates a better comprehension of the model's decision-making process.

Use of OpenTarget: The discovered genetic variations were mapped to established biological processes and disease associations using OpenTarget. This tool aids in determining the pathways in which these genes function and the potential correlation between them and the severity of COVID-19.

Use of Enrichr: To conduct enrichment analysis and connect genetic variations to particular biological pathways, operations, and conceptual frameworks, Enrichr was utilized. It sheds light on the connections between these genes and a range of biological processes and illnesses.

Integration of Findings: For interpreting the physiological importance of the genetic variants, the data from both tools were combined. In order to gain an improved comprehension of the biological processes underlying severity of COVID-19, for instance, the results were utilized to identify important pathways.

The GCP approach is additionally made available as a web-based tool to help experts evaluate their

hypothesis using the sixteen found variations in genetic and medical factors from COVID-19 WES datasets patients. In this case, a particular individual (refer to Figure 4.3) may be zoomed in to learn whether clinical and genetic variables combine to estimate the condition severity. Although the individual in need may be youth and in good health, exposure to the infection may put them in danger of contracting the illness. Illustrate that there has been a huge spike in the number of those infected throughout the outbreak's initial and subsequent crests, resulting in an enormous number of patients with serious illnesses who received medical care and a breakdown of medical facilities in numerous countries. By helping those making decisions select people who have actual COVID-19 severity. That GCP framework could thus serve as an asset for reducing the monetary cost of hospitalization. In addition, one may employ the GCP framework to clarify the result. The main objective of this research was to provide justifications and knowledge into the 16 genomic potential variants that have been discovered and utilized to build the GCP framework and verify it using an outside following-up dataset.

The GCP approach is also useful to justify the reasoning underlying each patient's severity prognosis results, including why they had been chosen. Creating an easy-to-use explanatory machine learning framework could assist doctors and medical decision-makers in creating more reliable and understandable algorithms that contribute to customized healthcare.

This process may come across as subjective. This might consist of:

1. Scientific Grounding: Describe each genetic variant's biological significance, including any known relationships to the severity of COVID-19 or associated pathways. Citing earlier research or databases that demonstrate the functions of these variants in the immune system, entry of viruses, or other pertinent mechanisms may be one way to do this.
2. Selection Criteria: Explain the decision-making criteria that were applied to these variants, including statistically significant effects in previous research, population frequency, or participation in important biological processes. One way to lessen the sense of subjectivity is to present a rationale that is both objective and clear.
3. Expert Consultation: Indicate whether clinicians or domain experts were consulted in order to guarantee that the variants selected have biological and clinical significance.
4. Comparative Analysis: Talk about whether or not other possible variations were taken into account and if not, why. This might entail contrasting the chosen variants' relevance or predictive ability with those of other variants.
5. Validation: Provide details on any validation actions performed to verify the significance of those variations, including connecting with databases such as OpenTarget and Enrichr or evaluating their impact on the model's accuracy through the use of machine learning methods.

Including these details would improve the research's general credibility and openness while strengthening the reasoning behind the selection of these particular gene variations.

6. Conclusion. The primary objective of this research was to provide explanations and perspectives concerning the 16 found genomic potential variations that we employed to create the GCP framework and evaluate its validity using a separate following-up dataset. The outcomes of this investigation add to our existing knowledge about the intricate biological relationships with which the 16 discovered variations in genes could have been interacting to determine the infectious disease like COVID-19 severity in the host's body. Gender, age variables, and additional modulating variables such as comorbidity have all added to the seriousness of the illnesses in most cases. Still, there's an isolated group of individuals whose severity is solely determined by genes. The community could be under exceptionally dangerous levels for this specific category of individuals who require protection from the infection caused by SARS-CoV-2. The subsequent studies will focus on analyzing the WES individual's dataset employing bioinformatics and statistical tools like regression-based SKAT-O investigation, polygenic variations rating, and deep neural networks to give experts in the field genetically prompted interpretation skills for reliable solutions based on data.

Limitations.

Missing values and data quality: Talk about the difficulty of imputation of missing data, which may result in bias if the values that are missing are not dispersed randomly.

Population-Specific Bias: Draw attention to possible biases that could affect the generalizability of the model,

including how connections between genes might differ between populations. Interpretation in a complex Genome-wide Models -Recognize that although SHAP values aid in interpretation, they might not adequately describe intricate causal chains or genomic interactions.

Innovation and Future Directions.

Uniqueness in Combining Clinical and Genomic Data: Discuss about how this novel approach, which is less prevalent in current models, combines clinical and WES data in an ensemble voting classification system designed to predict COVID-19 outcomes.

Potential Enhancements: To increase the power of prediction and model relevance across various categories of patients, suggest potential improvements like adding more medical covariates or additional data from omics (like proteomics). If data volume allows, propose sophisticated model processing methods for genomic data, such as deep learning.

REFERENCES

[1] Nawaz MS, Fournier-Viger P, Shojaee A, Fujita H., *Using artificial intelligence techniques for COVID-19 genome analysis*, in 1. Applied Intelligence. 2021 May;51:3086-103.

[2] J. Ahmed I, Jeon G, *Enabling artificial intelligence for genome sequence analysis of COVID-19 and alike viruses*,Interdisciplinary sciences: computational life sciences. 2022 Jun;14(2):504-19.

[3] Márquez S, Prado-Vivar B, Guadalupe JJ, Gutierrez B, Jibaja M, Tobar M, Mora F, Gaviria J, García M, Espinosa F, Ligña E, *Genome sequencing of the first SARS-CoV-2 reported from patients with COVID-19*, in Ecuador. MedRxiv. 2020 Jun 14 .

[4] Laamarti M, Alouane T, Kartti S, Chemao-Elfihri MW, Hakmi M, Essabbar A, Laamarti M, Hlali H, Bendani H, Boumajdi N, Benhrif O, *Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations*, 1. PloS one. 2020 Nov 10;15(11):e0240345.

[5] Mousavizadeh L, Ghasemi S, . *Genotype and phenotype of COVID-19: Their roles in pathogenesis. Journal of Microbiology, Immunology and Infection*, 2021 Apr 1;54(2):159-63.

[6] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding*, The lancet. 2020 Feb 22;395(10224):565-74 .

[7] Ray M, Sable MN, Sarkar S, Hallur V, *Essential interpretations of bioinformatics in COVID-19 pandemic*, Meta Gene. 2021 Feb 1;27:100844 .

[8] Quazi S, *Artificial intelligence and machine learning in precision and genomic medicine.*, 1. Medical Oncology. 2022 Jun 15;39(8):120.

[9] Ahmed I, Ahmad M, Jeon G, Piccialli F, *A framework for pandemic prediction using big data analytics.*,Big Data Research. 2021 Jul 15;25:100190 .

[10] Dubeya S, Kumar M, Verma DK, *Machine Learning Approaches in Deal with the COVID-19: Comprehensive Study*, ECS Transactions. 2022 Apr 24;107(1):17815 .

[11] Tripathi A, Chourasia U, Dubey S, Arjariya A, Dixit P, *A Survey: Optimization Algorithms In Deep Learning.*, InProceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020 Mar 31

[12] Soni S, Dubey S, Tiwari R, Dixit M., *Feature Based Sentiment Analysis of Product Reviews Using Deep Learning Methods*, 1. . International Journal of Advanced Technology & Engineering Research (IJATER). 2018. .

[13] Ullah K, Ahmed I, Ahmad M, Rahman AU, Nawaz M, Adnan, *A. Rotation invariant person tracker using top view*,Journal of Ambient Intelligence and Humanized Computing. 2019 Oct 4:1-7.

[14] Dubey, S., Verma, D. K., and Kumar, M., *Identification of Unique Genomic Signatures in Viral Immunogenic Syndrome (VIS) Using FIMAR and FCSM Methods for Development of Effective Diagnostic and Therapeutic Strategies*, Economic Computation & Economic Cybernetics Studies & Research. 2024 58(2).

[15] Ahmed I, Ahmad M, Ahmad A, Jeon G, *IoT-based crowd monitoring system: Using SSD with transfer learning*, Computers & Electrical Engineering. 2021 Jul 1;93:107226.

[16] Meraihi, Y., Gabis, A. B., Mirjalili, S., Ramdane-Cherif, A., Alsaadi, F. E., *Machine learning-based research for covid-19 detection, diagnosis, and prediction: A survey*, SN computer science. 2022; 3(4):286.

[17] Jia, P., Chen, L., Lyu, D., *Fine-Grained Population Mobility Data-Based Community-Level COVID-19 Prediction Model.*, Cybernetics and Systems. 2022; 1-19.

[18] Khan, S., Khan, H. U., Nazir, S., *Systematic analysis of healthcare big data analytics for efficient care and disease diagnosing.*, Scientific Reports. 2022; 12(1): 22377.

[19] Biswas, B., Chattopadhyay, S., Hazra, S., Hansda, A. K., Goswami, R., *COVID-19 pandemic: the delta variant, T-cell responses, and the efficacy of developing vaccines.*, Inflammation Research, 2022; 71(4): 377-396.

[20] Leal, V. N., Paulino, L. M., Cambui, R. A., Zupelli, T. G., Yamada, S. M., Oliveira, L. A., Pontillo, A., *A common variant close to the "tripwire" linker region of NLRP1 contributes to severe COVID-19.*, Inflammation Research, 2022; 1-8.

[21] Dubey, S., Verma, D. K., Kumar, M., *Severe acute respiratory syndrome Coronavirus-2 GenoAnalyzer and mutagenic anomaly detector using FCMFI and NSCE.*, International Journal of Biological Macromolecules, 2024; 258:129051.

[22] Dubey, S., Verma, D. K., and Kumar, M, *Real-time infectious disease endurance indicator system for scientific decisions using machine learning and rapid data processing.*, PeerJ Computer Science, 2024; 10, e2062.
[23] Li, J., Liu, H. H., Yin, X. D., Li, C. C., Wang, J., *COVID-19 illness and autoimmune diseases: recent insights.*, Inflammation Research, 2021; 70: 407-428.
[24] Zheng, Z., Wu, K., Yao, Z., Zheng, X., Zheng, J., Chen, J., *The prediction for development of COVID-19 in global major epidemic areas through empirical trends in China by utilizing state transition matrix model.*, BMC Infectious Diseases, 2020; 20: 1-12.
[25] Hassan, B., Izquierdo, E., Piatrik, T., *Soft biometrics: a survey: Benchmark analysis, open challenges and recommendations.*, Multimedia Tools and Applications, 2021; 1-44.
[26] Saleem, F., Al-Ghamdi, A. S. A. M., Alassafi, M. O., AlGhamdi, S., *Machine Learning, Deep Learning, and Mathematical Models to Analyze Forecasting and Epidemiology of COVID-19: A Systematic Literature Review.*, International journal of environmental research and public health, 2022; 19(9):5099.
[27] Alali, Y., Harrou, F., Sun, Y., *A proficient approach to forecast COVID-19 spread via optimized dynamic machine learning models.*, Scientific Reports, 2022; 12(1):1-20.
[28] Javed, I., Butt, M. A., Khalid, S., Shehryar, T., Amin, R., Syed, A. M., Sadiq, M., *Face mask detection and social distance monitoring system for covid-19 pandemic.*, Multimedia Tools and Applications, 2023; 82(9): 14135-14152.
[29] Harikrishnan, N. B., Pranay, S. Y., & Nagaraj, N., *Classification of SARS-CoV-2 viral genome sequences using Neurochaos Learning.*, Medical & Biological Engineering & Computing, 2022; 60(8): 2245-2255.
[30] Rohaim, M. A., Clayton, E., Sahin, I., Vilela, J., Khalifa, M. E., Al-Natour, M. Q., Munir, M., *Artificial intelligence-assisted loop mediated isothermal amplification (AI-LAMP) for rapid detection of SARS-CoV-2.*, Viruses, 2020; 12(9): 972.
[31] Rohaim, M. A., Clayton, E., Sahin, I., Vilela, J., Khalifa, M. E., Al-Natour, M. Q., Munir, M., *Artificial intelligence-assisted loop mediated isothermal amplification (AI-LAMP) for rapid detection of SARS-CoV-2.*, Viruses, 2020; 12(9): 972.
[32] Mohsan, S. A. H., Zahra, Q. U. A., Khan, M. A., Alsharif, M. H., Elhaty, I. A., Jahid, A., *Role of drone technology helping in alleviating the COVID-19 pandemic.*, Micromachines, 2022; 13(10):1593.
[33] Madhav, A. S., Tyagi, A. K., *The world with future technologies (Post-COVID-19): open issues, challenges, and the road ahead.*, Intelligent Interactive Multimedia Systems for e-Healthcare Applications, 2022; 411-452.
[34] Saleem, F., Al-Ghamdi, A. S. A. M., Alassafi, M. O., AlGhamdi, S. A., *Machine Learning, Deep Learning, and Mathematical Models to Analyze Forecasting and Epidemiology of COVID-19: A Systematic Literature Review.*, International journal of environmental research and public health, 2022; 19(9): 5099.
[35] Ratajczak, M. Z., Kucia, M., *Stem Cells as Potential Therapeutics and Targets for Infection by COVID19–Special Issue on COVID19 in Stem Cell Reviews and Reports.*, Stem Cell Reviews and Reports, 2021; 17:1-3.
[36] Kumar Saroha, S., Wang, Y.M. and Xuan Tung, N., *Investigation of future business opportunities for India and China after COVID-19.*, Environment, Development and Sustainability, 2024; 1-17.
[37] Dubey, S., Verma, D. K., and Kumar, M., *Severe acute respiratory syndrome Coronavirus-2 GenoAnalyzer and mutagenic anomaly detector using FCMFI and NSCE.*, International Journal of Biological Macromolecules, 2024; 258, 129051.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Apr 22, 2024
Accepted: Nov 19, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 940–949, DOI 10.12694/scpe.v26i2.3924

CLASSIFICATION OF ROYAL DELICIOUS APPLES USING HYBRID FEATURE SELECTION AND FEATURE WEIGHTING METHOD BASED ON SVM CLASSIFIER

SANDEEP KAUR*, MANOJ KUMAR SACHAN†, AND ASHWANI KUMAR AGGARWAL‡

Abstract. Fruit safety is a critical component of the global economy, particularly within the agricultural sector. There has been a recent surge in the incidence of diseases affecting fruits, leading to economic setbacks in agriculture worldwide. Conventional manual assessment methods are laborious, prompting the exploration of automated computerized techniques for evaluating fruit quality. This research presents a novel method for assessing the quality of golden delicious apples. A dataset comprising 1256 apple images was gathered under controlled conditions. Afterward Feature extraction focuses on texture features like LBP, GLCM, GLDM, DTF, and Gabor features, color features, and shape and size features. A total of 18654 features are extracted and normalized using z-score. A hybrid method for feature selection and weighting involves the mRMR algorithm to eliminate redundant features and the Sine Cosine Optimization Algorithm for feature weighting, enhancing classification performance. The SVM machine learning technique, augmented with optimized features, yielded a 10.53% improvement in accuracy compared to SVM alone. Validation against state-of-the-art methods using Friedman's mean rank test underscored the statistical significance of this approach across various metrics.

Key words: Fruit grading, normalization, feature selection, machine learning

1. Introduction. In the last decades, machine learning plays a vital role in the horticulture field. Various tasks like defect detection in products [33] , reliable yield prediction [19] , growth monitoring [36] , quality detection [27] , soil testing [8, 35] , etc. are performed with the help of machine learning techniques. The machine learning advances are mainly directed toward the fruit quality assessment [24, 3, 16]. To attain this goal two methods are used. In the first one,the grading is performed manually which is a too tedious, labor-intensive, inconsistent, and costly method [12] . The second method is grading performed with the help of machines. These machines will grade the apples based on size only, which is not an efficient method because other features such as color, texture, etc. also play a significant impact on the grading process.

To overcome these drawbacks machine learning is growing in the apple quality estimation field.The production of apples is the key player among all the fruits. India ranked fifth as a producer of apples all over the world [11]. There is a techniques in which a robotic system for fruit picking and grading, using computer vision for detection and a 4-DOF robotic arm for handling is developed. It enhances efficiency, reduces labor, and processes fruit in an average of 15 seconds for poor quality and 21 seconds for good quality, using Python and OpenCV [6]. The use of effective machine learning algorithms to combine image-based categorization approaches is at the leading edge of apple grading. Because of the uneven distribution of light on the apple's surface and the similar appearance of defects with stems and calyx, automatically grading the apples remains a difficult operation [25]. There is a study in which investigation is done that how color-balancing methods improve the classification of physiological disorders in apples under various lighting conditions using pre-trained CNN models. The ResNet50V2 model achieved 0.949 accuracy with green light sharpness data. Enhanced image quality was confirmed by improved PSNR, MSE, and SSIM metrics, demonstrating that color balancing significantly boosts classification performance.

Efficient and accurate classification is crucial for various applications, including quality assessment, sorting,

*Department of Computer Science and Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, 148106, Punjab, India
†Department of Computer Science and Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, 148106, Punjab, India
‡Department of Electrical and Instrumentation Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, 148106, Punjab, India

and marketing strategies Several scientists introduced various methods to grade the apple based on physical qualities such as defect, color, texture, shape, and size are considered important grading parameters [48, 45]. The classification of apples using feature selection and feature weighting techniques has been a subject of considerable research interest in agricultural and food industries. Feature selection is pivotal in enhancing classification accuracy by identifying the most relevant and discriminative features from input data. In apple classification, diverse methods such as filter, wrapper, and embedded approaches have been explored to accomplish this task. These methods focus on reducing the feature space dimensionality while preserving essential discriminatory information, thus boosting classification algorithm performance embedded [41, 43] . Conversely, feature weighting techniques allocate suitable weights to individual features based on their significance in classification tasks. This approach mitigates the influence of imbalanced or noisy features, thereby enhancing classifier performance. Notable methods include ReliefF [22] , and Chi-square [18] , which prioritize features according to their relevance and contribution to classification accuracy.

Lee et al. introduced color based technique for the determination of the apple maturity stage as well as the defective surface using Open CV software [24] .Yuan et al. also developed color based technique. For classification two methods were used one is SVM and the second one is the particle swarm optimization (PSO) algorithm [49] . Suresha M. et al. developed a color- based technique to classify the apples. The classification was performed by using an SVM classifier. A total of 90 images were used for experimentation [39] . Guo et al. introduced a back-propagation neural network method to classify multiple fruits and vegetables. The main focus of the classification is on color and texture. [15] . Bhatt and pant introduced Artificial Neural Network(ANN) based technique to classify the apples using size, shape, and color features. A total of 199 images were used to train the ANN [5] . Sofu et al. introduced A decision tree classification method based on color, weight,and shape features that were utilized for Granny Smith, Golden, and Starking Delicious apples. A total of 732 images are collected in which defect detection in granny smith and golden delicious was easy in comparison to starking delicious [38]. Moallem et al. introduced a technique for golden delicious apple grading based on surface features such as defect, texture, geometrical, and color features. The 120 images were collected for experimentation [30] . Ali and Thai introduced a technique for apple and mango grading by considering defects as the main features [2]. Bhargava and Bansal came up with fruit quality estimation techniques with the help of four classifiers k-nearest neighbor (k-NN), the sparse representative classifier (SRC), support vector machine (SVM), and artificial neural network (ANN)[4].

In recent literature, researchers have introduced machine learning methods for grading apples by combining multiple types of features. While this fusion can enhance the comprehensiveness of the analysis, it often complicates the grading process. The presence of redundant or irrelevant features can adversely affect the performance of classifiers. Moreover, the existence of outliers in the feature set can further diminish overall performance. These challenges serve as the primary motivation for our proposed approach, which aims to tackle these issues to boost classifier efficiency. Our method involves refining the feature set to eliminate redundancy and irrelevance while effectively managing outliers. By addressing these aspects, we aim to streamline the grading process, improve accuracy, and offer a more robust and reliable solution for apple grading using advanced machine learning techniques.

2. Proposed Methodology. In this work, a methodology is proposed to improve the classification of apples. The proposed approach is divided into six phases. In the first phase, the apple images are captured using the image acquisition mechanism. Then, the regions of the images are segmented using the image segmentation method in the second phase. In the next phase, the different features are extracted from the images based on the texture, color, shape and amp; size. A data normalization method is used to prevent larger numeric feature values from dominating smaller numeric feature values. The major objective of the normalization is to reduce the bias of those characteristics in pattern classes that contribute numerically more than others. After normalizing the features, the redundant and irrelevant features are removed with the help of a feature selection method. The reduced feature set is used for the classification of the apples in the last phase. The block diagram of the proposed approach has been illustrated in Figure 2.1.

2.1. Image acquisition system. The image acquisition system has been composed of 2 charged coupled devices (CCD) color cameras (SONY cyber shot DSC-W800). The system provides 2492x2492 color images. The lighting system has been composed of two T-shaped fluorescent tubes at the top of the square- shaped

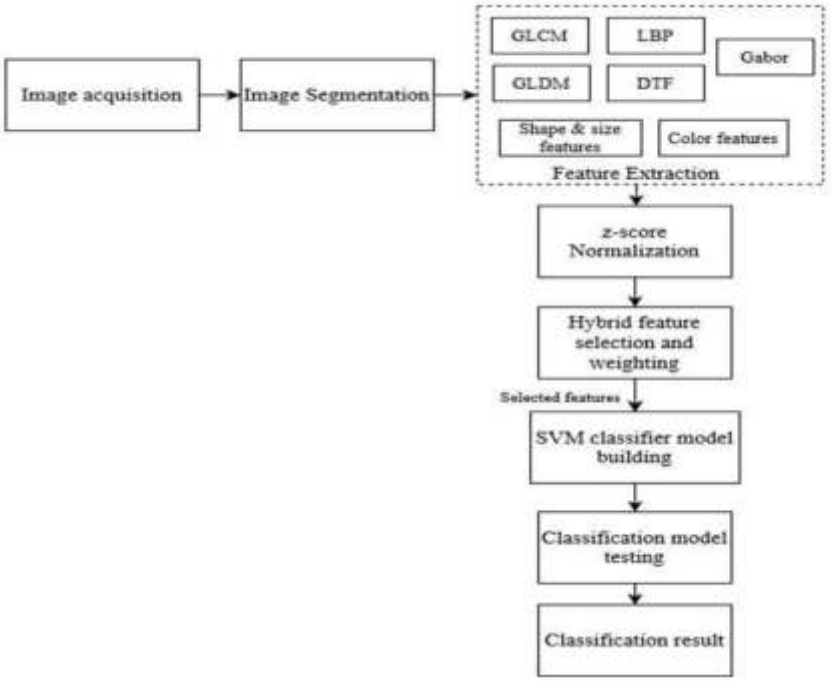


Fig. 2.1: Block diagram of the proposed approach

chamber. The chamber contains two holes for cameras, one at the top of the chamber and one at the left wall of the chamber. Reflecting surface protection has been placed between the fluorescent tube and apple to avoid direct light. Six images have been collected by rotating the apple thirty-degree horizontally using the top camera and two images from the side camera. The dataset has been created from 158 royal delicious apples collected from a Kashmir farm. A total of 1264 images have been captured which contain 256 images of grade A, 336 images of grade B, grade C, and grade D. The sample of images captured are shown in figure 2.2.

2.2. Image Segmentation. In the proposed approach, segmentation is used as a pre-processing method. The segmentation methods are used to divide the images into different isolated regions. In this approach, to divide the image into two regions (fruit area and background), the segmentation of 1256 images is performed by using the advanced photo editor Adobe Photoshop CC which is an image processing tool. The segmentation image result samples are shown in Figure 2.3.

2.3. Feature Extraction. In the proposed approach, different features are extracted from the apple images for the classification of apples into different classes. The features are extracted to learn the pattern between the different samples of apples related to the different classes. In this approach, features are extracted based on the texture, color and shape and amp; size of the apples. The details of the features are given below:

Texture Features. Image texture is a collection of two-dimensional arrays that have been generated to measure the visual texture of an image [32]. It gives information about how colors or intensities are arranged spatially in a region of an image. In this approach different texture features are extracted from collected images of apples. The texture which is used in this approach is linear binary patterns (LBP) [31] , Gray-Level Co-occurrence Matrices (GLCM) [17] , Gray- Level Difference Method (GLDM) [47] , difference theoretic features (DTF) [40] , Gabor [20] features are extracted for the classification. The description of texture feature extraction method is as follows:

Fig. 2.2: Different grades of apple: (a) Grade A, (b) Grade B, (c) Grade C, (d) Grade D

Fig. 2.3: Different grades of apple after segmentation: (a) Grade A, (b) Grade B, (c) Grade C, (d) Grade

Linear Binary Patterns (LBP). The texture features of the apple images are calculated using the LBP feature extraction method.The squared neighbourhood pixel values are computed in traditional LBP methods. Scale, illumination, and rotation invariance are computed in order to get all of the texture characteristics, then the magnitude vector in the difference vector is also taken into consideration [44] . The features can be calculated as follows:

Let pixel c_α be in the center of the neighbourhood image pixels.The P_x is considered as the index for the image pixel as $(g_0, \ldots, g_{P_x-1})$. Each binary pixel value is multiplied by a binomial factor to propduce the pixel's

LBP feature.

$$LBP_{PX} = \sum_{P_x=0}^{P_x=1} d(g_{p_x} - g_a), \tag{2.1}$$

where

$$d(n) = \begin{cases} 1, n \geq 0, \\ 0, n < 0. \end{cases} \tag{2.2}$$

Gray-Level Co-occurrence Matrices (GLCM). The GLCM features were based on energy, homogeneity, contrast, correlation, mean, entropy, and maximum probability etc. The GLCM features are calucated with the thirty-four statistical features for the different orientations $(0^0, 45^0, 90^0, 135^0)$.

Gray-Level Difference Method (GLDM). This feature extraction method is based on two pixels that were separated by a particular displacement δ and had an absolute difference in the grey levels. The motion vector and the probability density functionis defined as follows:

$$\delta = (\Delta x, \Delta y),\ S_\delta(x, y) = |S(x, y) - S(x + \Delta x, y + \Delta y)| \tag{2.3}$$

$$D(i|\delta) = Prob[S_\delta(x, y) = i] \tag{2.4}$$

where $\Delta x, \Delta y$ are the parameters of the methods, the input image is represented by $S_\delta(x, y), x, y$, are positions in image $S_{\delta(x,y)}$ with $1 \leq x \leq M$ and $1 \leq y \leq N$. Then, a feature vector is generated by concatenating the mean, contrast, entropy, and angular second moment that are computed from the probability density function.

Difference Theorectic Features (DTF). In this feature extraction method, scale, illumination- invariant, and rotation based features were created from the correlated distributions of local and global intensity differences at different grey levels in the image. DTF computes eleven fetures based on the correlation values along with the horizonal, diagonal, and vertical directions to obtained the feature set.

Gabor Features. Gabor feature are used to compute the texture feature based on the wavelet function. The following wavelet function is used in this work [21]:

$$h(x, y) = \exp\left(-\alpha^{2\omega}\frac{x^2 + y^2}{2}\right)\exp(\omega\pi\alpha^\omega(x\cos\theta + y\sin\theta) \tag{2.5}$$

where $\alpha = \frac{1}{sqrt2}, \omega = 0, 1, 2, ...$ and $\theta \in [0, 2\pi]$.

Based on the wavelet function, The gabor features are computed at different frequencies, locations and orientations by convolving the input image with filters:

$$m(x, y) = L_h(I(x, y)) = |I(x, y) \times h(x, y)| \tag{2.6}$$

- *Color Features:* A space can be used to specify, create, and visualize color into distinct data representation [1]. The Uniformity in color on the surface of the apple makes it good for the open market. In this study, HSV color space is used to extract color features [13]. The HSV color space is close to how humans perceive color. It contains three elements as Hue (values lie between 0 to 360), Saturation, and View. where hue contains the color values, saturation specifies a range of gray values and view is the brightness of color. The formulation for the HSV features of a RGB image (I) as follow:

$$mx_{(i,j)} = max(I_{R(i,j)}, I_{G(i,j)}, I_{B(i,j)}); mn_{(i,j)} = min(I_{R(i,j)}, I_{G(i,j)}, I_{B(i,j)}) \tag{2.7}$$

$$H(i, j) = \begin{bmatrix} \frac{60*(I_{G(i,j)} - I_{B(i,j)})}{mx-mn} I_{R(i,j)} > max(I_{G(i,j)}, I_{B(i,j)}) \\ \frac{180*(I_{B(i,j)} - I_{R(i,j)})}{mx-mn} I_{G(i,j)} > max(I_{R(i,j)}, I_{B(i,j)}) \\ \frac{300*(I_{R(i,j)} - I_{G(i,j)})}{mx-mn} I_{B(i,j)} > max(I_{R(i,j)}, I_{G(i,j)}) \end{bmatrix} \tag{2.8}$$

- *Shape and Size Features:* The size and shape of the apples are important features due to their considerable ability to distinguish between different types of apples. In general, farmers typically look for shape and size features for the apple grading.

In this proposed approach, 18654 features are extracted from the segmented images of the apples.

2.4. Normalization. Normalization, a crucial preprocessing step that avoids larger numeric attribute values from influencing smaller numeric feature values, converts features into a common range. The main goal of the normalization techniques is to lessen the bias of the features that are numerically more significant than others. The performance of the z-score is superior for the feature selection process, hence it is employed in the proposed approach [37].

2.5. Hybrid Feature Selection and Feature Weighting. High-dimensional datasets might contain duplicate and irrelevant features. The effectiveness of the machine learning algorithms is influenced by the feature set's quality. Thus, optimal set of features are essential for improving the performance of machine learning algorithms. In this approach, A hybrid method has been proposed to select optimal features based on feature selection and feature weighting approach. First the redundant and irrelevant features are eliminated using the minimal-redundancy-maximum-relevance (mRMR) technique. The choosen features using the mRMR method tare both maximally different from one another and have the highest relevance to the target class [34]. After the selection of k number for features based of mRMR method, a feature weighting method based on the metaheuristic method has been used to assign weigth to the optimal selected features. The feature weighting helps in the selection of features based on the feature relevance. In this method, Sine Cosine based optimization algorithm has been utilized for the feature weighting of the apple dataset features. The deatail description is as follows:

2.5.1. Minimal-redundancy-maximum-relevance based feature selection. The feature selection algorithm is preresented by Ding and Peng and the minimum-redundancy maximum- relevance (mRMR) algorithm ranks a set of features by maximising their relevance and minimising their redundancy within the subset of features [9]. The goal is to reduce the duplication of a subset that has been chosen using a relevance metric. The F-test is used as a relevance measure in the first phase of the mRMR algorithm, and the Pearson's correlation between features is calculated as a redundancy measure. The remaining set of features is repeatedly chosen based on the mRMR score aluclated using the following formula:

$$score = max_{ie\Omega S}\left\{F(i,s) - \left(\frac{1}{|S|}\sum_{j\in S}|c(i,j)|\right)\right\} \tag{2.9}$$

After the first feature, or the feature with the largest value of relevance (f-test ranked set), with the goal is chosen. F-test correlation difference (FCD) was the method employed in this investigation.After the selection of k number of feature usinh the mRMR feature selection method the selected feature then analysed using the feature weightning method for the selected of most relevant features.

2.5.2. Sine Cosine optimization algorithm based feature weightning. Sine cosine algorithm (SCA) is a population-based method for solving global optimisation issues [28]. SCA bases its search strategy on the properties of the sine and cosine trigonometric functions. SCA has demonstrated its effectiveness in both exploration and exploitation, and because of its prowess in exploration, it has been used to address a wide range of real-world problems. The approach uses the SCA algorithm as a weighting algorithm, giving features that are more important a larger weight than ones that are redundant or less relevant. In the SCA algorithm the position are updated based on the following formula after the random initialization of the position values:

$$P_i^{j+1} = P_i^j + rand_1 \times \sin(rand_2) \times |rand_3 \times D_i^j - P_i^j| \tag{2.10}$$

$$P_i^{j+1} = P_i^j + rand_1 \times \cos(rand_2) \times |rand_3 \times D_i^j - P_i^j| \tag{2.11}$$

where the current solution's position vector is represented by P_i^j at j^{th} iteration and i^{th} dimension. The $rand_1$, $rand_2$, and $rand_3$ denotes the random variables. The D_i^j shows the values of the position vector at the destination. The above equations are combined in the algorithm based on following formula:

$$P_i^{j+1} = \begin{cases} P_i^j + rand_1 \times \cos(rand_2) \times |rand_3 \times D_i^j - P_i^j|, rand_4 < 0.5, \\ P_i^j + rand_1 \times \cos(rand_2) \times |rand_3 \times D_i^j - P_i^j|, rand_4 \geq 0.5. \end{cases} \tag{2.12}$$

where $rand_4$ represents a random variable and its range in [0,1].

In the SCA algorithm, the initial features values are multiplied with weight values using a linear weight assignment technique [7] to increase accuracy. The following is the formulation of the linearly feature weighted data:

$$feat'_l = Weight_l \times feat_l \tag{2.13}$$

where $feat_l$ is the l^{th} feature and $Weight_l$ denotes the corresponding weight of l^{th} feature calcukated using SCA algorithm. By increasing the space of highly weighted features and decreasing the space of nominally weighted features, it modifies the feature space of the classification problem. Along with the selection of the relevant features using weighting method, the SCA method also selects the optimal parameteric values for the SVM classifier for the classification task.

2.6. Classification. SVM is a supervised learning model that analyses data used for regression and classification through associated learning computations. The main focus of the SVM classifier is on the hyper-plane from N-dimensional space that splits the data points associated with different classes [29]. In the proposed approach, the SVM classifier is utilized for the classification of apples. As the SVM looks for the most effective hyper-plane to use feature values to split the categories.

3. Results and Discussions. In this work, an approach for classification improvement of the apples is proposed. The experiments are performed on a Windows 10 Pro operating system running on the computer having the Intel® Xenon® CPU E5-2650 v3 (2.30 GHz) and 8 GB of RAM using the MATLAB 2019a platform. Ten-fold cross-validation is used for the classification performance evaluation.

In figure 3.1, classification accuracies have been achieved using the proposed approach for the different percentages of features selected using the mRMR approach illustrated.

It has been observed from figure 3.1 that the proposed approach achieves the highest accuracy is 92.27% at the 10% of features selected from the original set of apple features.

In figure 3.2, the accuracy comparision of the proposed approach with ReliefF [26], Mut-Inf [23], FSASL[10], zero-norm[46], Fisher[14], and Lasso [42] feature selection methods is illustrated.

Figure 3.2 indicates If we use all the features the accuracy was 87.94% but, the proposed approach achieves the highest 92.27% by selecting 10% of the features. This proves that the proposed approach is better in comparison to the other state-of-the-art methods as discussed in introduction section.

Friedman's mean rank test is used to verify the statistical significance of the proposed approach. In Table 3.1, Friedman's rank of the proposed approach in comparison to the well-regarded methods has been shown.

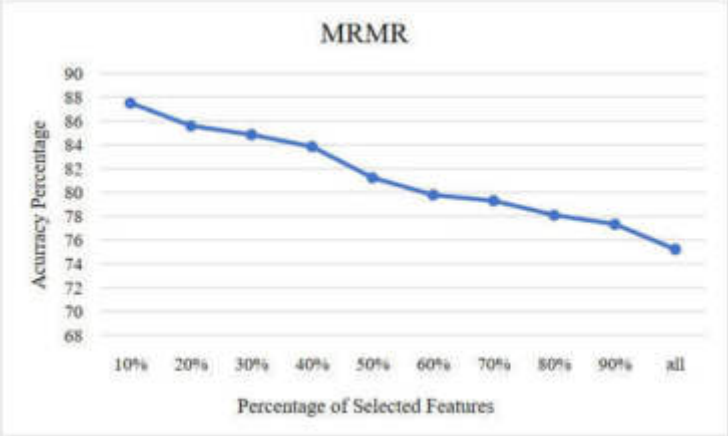


Fig. 3.1: Accuracy achieved using the proposed approach

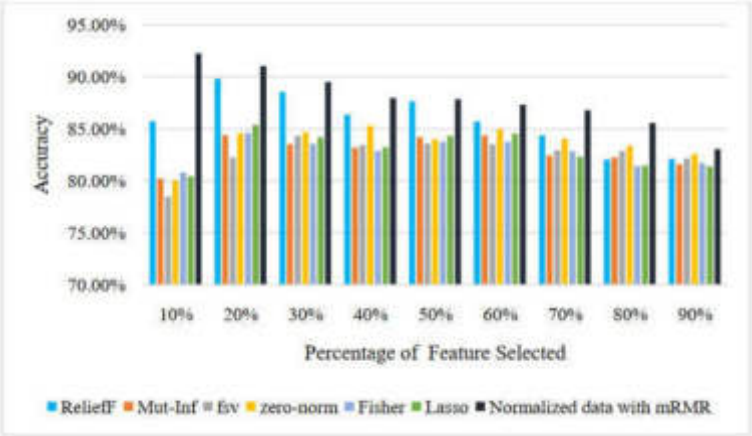


Fig. 3.2: Accuracy comparison of the proposed approach with state-of-the-art methods

Table 3.1: Friedman's mean rank of methods

Feature selection method	**Friedman Mean rank**
Normalized data with mRMR	1.11
Lasso	5.89
Fisher	6.33
Zero-Norm	3.67
Fsv	3.11
Mut-Inf	5.78
ReliefF	2.11

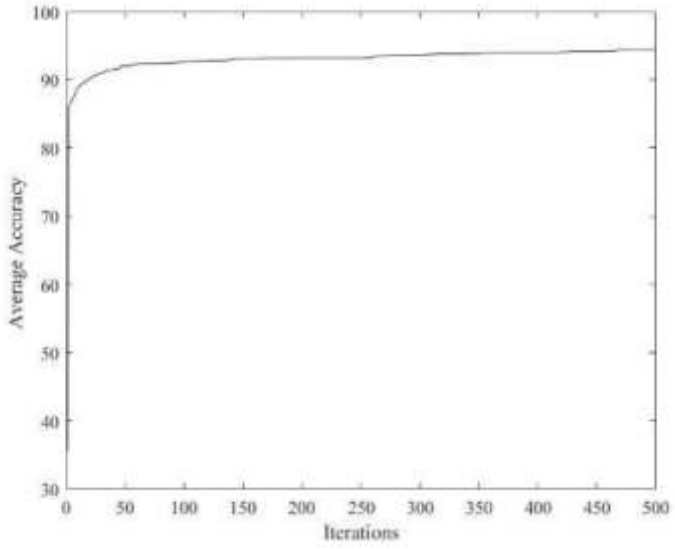


Fig. 3.3: Convergence curve of SCA based feature weighting

From Table 3.1, it has been observed that the proposed approach attains the lowest Friedman mean rank value with a p-value of 1.0825E-08 at a 5% of a confidence interval for the error values which shows the statistical significance of the normalized features selected using the mRMR approach. The seleceted 10% feature then fed into the SCA based feature weighting method and the convergence curve of attain using the algorithm is shown in the figure 3.3.

From the figure 3.3, it has been observed that the proposed SCA based feature weightning converge fater and attain accuracy of 94.3708% with the standard deviation value of 0.82 with optimized SVM classifier.

4. Conclusions. In this work, a methodology was proposed to enhance the classification of apples, structured into six phases. Initially, apple images were acquired, followed by segmentation to isolate regions. Features including texture, color, and shape & size were extracted and normalized to mitigate numerical biases. Redundant features were subsequently removed through a feature selection process. The final phase utilized a reduced feature set for classification. The method was validated using images of 158 royal delicious apples, showcasing different grades (A, B, C, D). SVM classifier is used for the classification of the apples into different classes. For the validation proposed approach is compared with the different existing state-of-the-art algorithms. The feature selection method attains the lowest Friedman's mean rank with the highest classification accuracy of 92.27% which proves the significance of using mRMR approach.The SCA based feature weighning approach attain accuracy of 94.3708% on the optimal set features obtained using the mRMR method. In the future, the proposed approach can be applied to the classification of the different fruits.

Data Availability. The data is made available on demand.

REFERENCES

[1] H. Afrisal, M. Faris, G. Utomo P., L. Grezelda, I. Soesanti, and M. Andri F. Portable smart sorting and grading machine for fruits using computer vision. In *Proceeding - 2013 Int. Conf. Comput. Control. Informatics Its Appl. "Recent Challenges Comput. Control Informatics", IC3INA 2013*, pages 71–75, 2013.

[2] M. A. H. Ali and K. W. Thai. Automated fruit grading system. In *2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation (ROMA)*, pages 1–6, 2017.

[3] S. K. Behera, A. K. Rath, and P. K. Sethy. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. *Inf. Process. Agric.*, 8(2):244–250, Jun. 2021.

[4] A. Bhargava and A. Bansal. Automatic detection and grading of multiple fruits by machine learning. *Food Anal. Methods*, 13(3):751–761, Mar. 2020.

[5] A. K. Bhatt, D. K. Ghosh, S. J. Darokar, J. S. Sikarwar, and P. R. Kadu. Analysis of volatile compounds in ripening of apple (malus domestica) fruits using gas chromatography–mass spectrometry (gc–ms). *Curr. Sci.*, 115(5):891–895, 2018.

[6] M.H. Dairath, M.W. Akram, M.A. Mehmood, H.U. Sarwar, M.Z. Akram, M.M. Omar, and M. Faheem. Computer vision-based prototype robotic picking cum grading system for fruits. *Smart Agric. Technol.*, 4:100210, 2023.

[7] S. Dalwinder, S. Birmohan, and K. Manpreet. Simultaneous feature weighting and parameter determination of neural networks using ant lion optimization for the classification of breast cancer. *Biocybern. Biomed. Eng.*, 40(1):337–351, Jan. 2020.

[8] F. A. Diaz-Gonzalez, J. Vuelvas, C. A. Correa, V. E. Vallejo, and D. Patino. Machine learning and remote sensing techniques applied to estimate soil indicators – review. *Ecol. Indic.*, 135:108517, Feb. 2022.

[9] C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. In *Computational Systems Bioinformatics. CSB2003. Procs of 2003 IEEE Bioinformatics Conference*, 523–528. IEEE Comp.Soc,2003.

[10] L. Du and Y. D. Shen. Unsupervised feature selection with adaptive structure learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 209–218, New York, NY, USA, Aug. 2015. ACM.

[11] S. D. Golombek and M. M. Blanke. Orchard management strategies to reduce bruises on apples in india: a review. *Vegetos*, 35(1):1–8, Mar. 2022.

[12] J. F. S. Gomes and F. R. Leta. Applications of computer vision techniques in the agriculture and food industry: a review. *Eur. Food Res. Technol.*, 235(6):989–1000, 2012.

[13] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing Using MATLAB: Pearson Prentice Hall.* Pearson Prentice Hall, Up. Saddle River, New Jersey, 2004.

[14] Q. Gu, Z. Li, and J. Han. Generalized fisher score for feature selection. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, UAI 2011*, pages 266–273. AUAI Press, Feb. 2011.

[15] H. Guo, Y. Tan, and W. Li. Surface texture detection of double-feature apple based on computer vision. In *Computational Intelligence in Robotics and Automation*, pages 117–127. 2014.

[16] M. Halstead, C. Mccool, S. Denman, T. Perez, and C. Fookes. Fruit quantity and quality estimation using a robotic vision system. 2018.

[17] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.*, (6):610–621, 1973.

[18] Jin Huang and C. X. Ling. Using auc and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.*, 17(3):299–310, Mar. 2005.

[19] K. Jain and N. Choudhary. Comparative analysis of machine learning techniques for predicting production capability of crop yield. *Int. J. Syst. Assur. Eng. Manag.*, 13(S1):583–593, Mar. 2022.
[20] P. Jolly and S. Raman. Analyzing surface defects in apples using gabor features. In *Proceedings - 12th International Conference on Signal Image Technology and Internet-Based Systems, SITIS 2016*, pages 178–185. Institute of Electrical and Electronics Engineers Inc., Apr. 2017.
[21] Z. Juan and C. Xiao-Ping. Field pest identification by an improved gabor texture segmentation scheme. *New Zeal. J. Agric. Res.*, 50(5):719–723, 2007.
[22] K. Kira and L. A. Rendell. A practical approach to feature selection. In *Machine Learning Proceedings 1992*, pages 249–256. Elsevier, 1992.
[23] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, 69(6):066138, Jun. 2004.
[24] D. J. Lee, J. K. Archibald, and G. Xiong. Rapid color grading for fruit quality evaluation using direct color mapping. *IEEE Trans. Autom. Sci. Eng.*, 8(2):292–302, Apr. 2011.
[25] Q. Li, M. Wang, and W. Gu. Computer vision based system for apple surface defect detection. *Comput. Electron. Agric.*, 36(2–3):215–223, Nov. 2002.
[26] H. Liu and H. Motoda. *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2007.
[27] T. Y. Melesse, M. Bollo, V. Di Pasquale, F. Centro, and S. Riemma. Machine learning-based digital twin for monitoring fruit quality evolution. *Procedia Comput. Sci.*, 200:13–20, 2022.
[28] S. Mirjalili. Sca: A sine cosine algorithm for solving optimization problems. *Knowledge-Based Syst.*, 96:120–133, Mar. 2016.
[29] A. Mizushima and R. Lu. An image segmentation method for apple sorting and grading using support vector machine and otsu's method. *Comput. Electron. Agric.*, 94:29–37, Jun. 2013.
[30] P. Moallem, A. Serajoddin, and H. Pourghassem. Computer vision-based apple grading for golden delicious apples based on surface features. *Inf. Process. Agric.*, 4(1):33–40, 2017.
[31] G. Muhammad.Date fruits classification using texture descriptors&shape-size features. *Eng.Appl.Artif.Intell.*,37:361–36,2015.
[32] S. Naik and B. Patel. Machine vision based fruit classification and grading - a review. *Int.J.Comput.Appl.*,170(9):22–34,2017.
[33] J. F. I. Nturambirwe and U. L. Opara. Machine learning applications to non-destructive defect detection in horticultural products. *Biosyst. Eng.*, 189:60–83, Jan. 2020.
[34] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1226–1238, Aug. 2005.
[35] B. T. Pham, L. H. Son, T. A. Hoang, D. M. Nguyen, and D. Tien Bui. Prediction of shear strength of soft soil using machine learning methods. *CATENA*, 166:181–191, Jul. 2018.
[36] A. Sengupta, A. Mukherjee, A. Das, and D. De. Agristick: An iot-enabled agricultural appliance to measure growth of jackfruit using 2-axis joystick. *IEEE Instrum. Meas. Mag.*, 25(3):58–62, May 2022.
[37] D. Singh and B. Singh. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.*, 97:105524, Dec. 2020.
[38] M. M. Sofu, O. Er, M. C. Kayacan, and B. Cetişli. Design of an automatic apple sorting system using machine vision. *Comput. Electron. Agric.*, 127:395–405, 2016.
[39] M. Suresha, N. A. Shilpa, and B. Soumya. Apples grading based on svm classifier. *Int. J. Comput. Appl.*, 975:8878, 2012.
[40] S. Susan and M. Hanmandlu. Difference theoretic feature set for scale-, illumination- and rotation-invariant texture classification. *IET Image Process.*, 7(8):725–732, Nov. 2013.
[41] H. A. Le Thi, V. V. Nguyen, and S. Ouchani. Gene selection for cancer classification using dca. In *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, volume 5139 LNAI, pages 62–72, 2008.
[42] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, 58(1):267–288, Jan. 1996.
[43] M. Tripathi. Analysis of convolutional neural network based image classification techniques. *J. Innov. Image Process.*, 3:100–117, 2021.
[44] S. Veerashetty and N. B. Patil. Novel lbp based texture descriptor for rotation, illumination and scale invariance for image texture analysis and classification using multi-kernel svm. *Multimed. Tools Appl.*, 79(15–16):9935–9955, Apr. 2020.
[45] T. Vijayakumar. Posed inverse problem rectification using novel deep convolutional neural network. *J. Innov. Image Process.*, 2:121–127, 2020.
[46] J. Weston, A. Elisseeff, B. Schölkopf, and M. Tipping. Use of the zero-norm with linear models and kernel methods. *J. Mach. Learn. Res.*, 3:1439–1461, 2003.
[47] J. S. Weszka, C. R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Trans. Syst. Man Cybern.*, SMC-6(4):269–285, 1976.
[48] M. Yang, P. Kumar, J. Bhola, and M. Shabaz. Development of image recognition software based on artificial intelligence algorithm for the efficient sorting of apple fruit. *Int. J. Syst. Assur. Eng. Manag.*, 13(S1):322–330, Mar. 2022.
[49] Jinli Yuan, Zhitao Guo, and Dawei Yue. The apple color grading based on pso and svm. In *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, pages 5198–5201. IEEE, Aug. 2011.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: May 13, 2024
Accepted: Jun 11, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 950–963, DOI 10.12694/scpe.v26i2.4059

DDOS ATTACK DETECTION AND PERFORMANCE ANALYSIS IN IOT NETWORK USING MACHINE LEARNING APPROACHES

DEPRIYA PANDA*, NEELAMADHAB PADHY† AND KAVITA SHARMA‡

Abstract. The Internet is the most common connecting tool for devices, such as computers, mobile phones, smart watches, etc. These devices communicate with designated servers to provide information. Here we refer to the system that connects numerous autonomous devices known as the Internet of Things (IoT). As the devices are of diverse categories and sometimes very small, it becomes challenging to provide comprehensive security to those in need. However, the sensors on the IoT collect huge amounts of data and the huge network becomes an attractive target for assaulters. One of the several assaults on IoT is Distributed Denial of Service (DDoS). Machine learning can play a crucial role in identifying these attacks in the IoT because of its ability to analyse large amounts of data. Machine learning models can learn the pattern of legitimate traffic and later identify malicious packets that deviate from the learned pattern. Classification techniques can distinguish malicious packets from genuine ones based on several attributes associated with them. This work uses classification techniques such as Random Forest, Gradient Boosting, and XgBoost to determine the malicious packets in traffic. The analysis shows that balancing techniques such as SMOTE and ADASYN are vital in improving the performance of techniques.

Key words: ADASYN, DDoS, Gradient Boost, IoT, Random Forest, XgBoost

1. Introduction. The Internet of Things provides ubiquitous computing power across a network of devices. The devices in an IoT network can be any object, referred to as a 'thing.' It can be a sensor, an embedded device, a mobile phone, etc. The basic architecture of an IoT network can be visualized as a combination of four layers [26]. A network is built to communicate among these layers and the server to provide information and help us reach a universal objective [6]. IoT is spreading across various domains, but the issue of security is a significant concern. A small part of any extensive IoT network can be targeted to launch a DDoS attack. By overloading the network's bandwidth, a group of vulnerable IoT nodes can cripple a full-scale network with powerful servers.

In IoT, the nodes have minimal resources, whether computing power, memory, or any other. Hence, the nodes in IoT networks are very vulnerable to attacks. Attackers find the devices attached to the IoT network attractive. Various types of attacks are possible in an IoT network. One of these is 'Distributed Denial of Service' (DDoS). It can be performed using botnets such as Mirai, launched in 2016 [13]. In the same year, Mirai affected the heating systems in some buildings in Finland and crashed those systems [3]. The greater the number of vulnerable devices, the greater the probability of creating botnets for the attackers. The magnitude of the DDoS attack is proportionate to the botnet's size. When a huge number of nodes participate in a botnet, the fierceness of the attack becomes increasingly dreadful. The bots generate false requests for the server, making it busy and unable to serve genuine requests. As a result, genuine users keep waiting for the service for an uncertain period.

In a UDP-based DDoS assault, the attacker transmits many UDP packets to the victim device. After receiving, the system finds the appropriate application to service the packets. But the system getsa huge number of packets, and in attempting to service these, the server becomes unavailable for other clients, launching a DDoS attack.The second case describes how the LDAP protocol can launch a DDoS attack. In this case, the target's IP address is spoofed. An attacker pretends to be the intended victim and sends a packet requesting service. When the server tries to respond, it transmits the responses to the target machine, as spoofing forces the server to think the attacker is legitimate. Portmap attacks are launched by exploiting the vulnerability

*GIET University, Odisha, India (devpriya.panda@giet.edu).
†GIET University, Odisha, India (dr.neelamadhab@giet.edu).
‡Galgotias College of Engineering, Greater Noida, India (kavitasharma_06@yahoo.co.in).

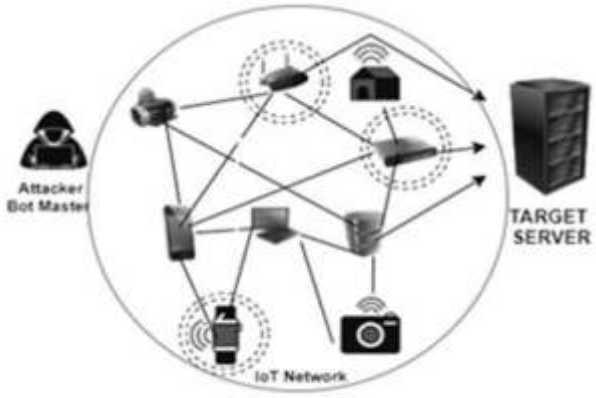


Fig. 1.1: DDoS Architecture.

of the port mapper. In port mapping, the request is expanded to include a significant response. This feature makes it suitable for DDoS attacks as it can increase the fierceness of the attack. Machine learning classification techniques are extensively used by various researchers to segregate one category of data from other categories [11]. The same concept can be applied to identify malicious packets in a chunk of packets transferred over an IoT network. Based on the DDoS attacks, several datasets have been prepared. CICDDoS2019 is one of the most recent datasets, describing several aspects [9].

In this work, this dataset is considered to make the most of it. This work analyses three types of models to detect DDoS attacks. Random forest has been used in some earlier works; the 'XgBoost' model has been considered in this work in addition to gradient Boosting and the formerly mentioned model. Also in the pre-processing, in addition to the 'Random Over Sampling' method, SMOTE and ADASYN sampling models were considered.

After going through related literature in Section 2, the objective of this investigation is defined in Section 3. Then, the workflow is explained in Section 4. After the workflow, the results are being analysed in Section 5. The conclusion and the future scopes are mentioned at the end in Section 6.

2. Literature Survey. In an IoT environment, the security mechanisms available at the nodes are negligible. Hence, the vulnerability of these nodes is exploitable. The nodes in IoT networks are deprived of resources compared to standard computing devices because of the constraint of power and as those are designed for specific purposes. DDoS is one of the most frequent attacks performed on different IoT networks. Previous works in the context under discussion are studied and are described below.

Jain et al. in [9] suggested an ensemble of different machine learning techniques to get better performance in detecting DDoS attacks. A voting mechanism at the run time has been proposed to choose the best method for identifying the malicious data packets in the traffic. Naïve Base, Random Forest, K Nearest Neighbour, and Support Vector Machine methods have been used for classifying affected packets. But instead of considering all those at a time, the authors' proposed voting system chooses the best result.

In this work [12], the concept of Cyber Security is discussed by Khari et. al. with an overview alongside the history of Cyber Security that has evolved from information security to encompass individual confidential data. It defines the human being as not only a victim and perpetrator of cyber-criminals but also as a defender, categorizes cybercrimes and explores their effects and measures. It also includes cyber security policy and management, risk and compliance, laws and regulations, accreditation, and courses related to the profession.

Wang et al. proposed a multilayer perceptron-based framework where a concept of handling the errors generated in the classification of DDoS-affected packets has been introduced [28]. The authors have considered parameters such as the IP of the source, destination port, and some other attributes of TCP packets for classification.

Jia et.al in their work have proposed an all-inclusive framework in [10]. A combination of recurrent LSTM (Long Short–Term Memory) neural network and CNN (Convolutional Neural Network) model is prepared and applied to the CIC-DDoS2019 dataset.

Latifet. al. have used an EVFDT ('Enhanced Very Fast Decision Tree') to identify DDoS assault in the WBAN ('Wireless Body Area Network') scenario [15]. The work generates highly accurate results, and the number of false alarms is also very low.

Saini et. al. [21] used three different techniques to detect DDoS attacks in an IoT scenario. The first technique used is J48, a decision tree model. 'Random Forest' and 'Naïve Base' classifiers were being used to segregate the malicious packets. It was found that the results obtained were very accurate.

Suresh et al. [27] used two different methods to identify the important characteristics of a packet. One is chi-square, and the other is information gain. The next step involved using models like Naïve Bayes, C4.5, and KNN to detect the DDoS attack. The last model outperformed all other models for the dataset under consideration.

Shieh et.al. tried to address the OSR('Open Set Recognition') issue [24]. The attackers use various techniques to launch a DDoS attack. When a detection system is designed based on some known methods and the attacks are devised differently, then the system may fail. At least the performance would not be the same as that of detecting the known method of attack. The authors in their work have used GMM ('Gaussian Mixture Model') and BI-LSTM ('Bidirectional Long-Short Term Memory') to design the detector models. It is found that GMM is very effective in identifying new categories of attacks.

Al-Hadhrami et. al. [1] have studied several scenarios leading to DDoS attacks in the IoT. Various mitigation techniques were considered, and the effectiveness of those was analysed. Some points regarding the open problems in these scenarios were also discussed.

Marvi et. al. [16] have suggested a three-step procedure to select the features before applying the data to the detection modules. The proposed framework was being designed using a decision tree-based LGBM ('Light Gradient Boosting Machine') algorithm. Source IP, Mean ACK, Header length, etc. are the parameters the model considers. Unseen DDoS attacks can also be identified using the model.

Shrivastava et. al. investigated the role of different Android devices in connection with IoT networks [25]. The authors mainly focused on the applications running on any Android device, interfacing with IoT. They used sensitivity analysis techniques to assess the effectiveness of Android intents as a distinguishing characteristic to identify malicious apps. Additionally, a substantial number of samples gathered from Android app markets are used in the proposed study. Several criteria are assessed and contrasted using the methods currently in use.

Authors in [18] surveyed several manuscripts on attacks such as DDoS and MitM in networks such as IoT, IoMT and Vanet. The works of various authors were being studied and methods suggested for identification and/or mitigation of these attacks were listed for better illustration.

Sharma et.al. [23] in their work investigated the risks faced by Android devices which are most of the time connected to IoT networks. It has been suggested that these devices are prone to various attacks due to the lack of legitimacy audits. They achieved excellent accuracy, using machine-learning models on the M0Droid dataset.

The above discussion is summarized in Table 2.1.

3. Objective. In a DDoS attack, the attacker tries to exhaust the recipient device by sending a huge quantity of packets. Different types of packets can be used for the above purpose. Hence, discovering the attack involves identifying the packets used in the attack. Our objective is to concentrate on three categories of packets. The abnormality in the received packets is going to be identified if any of these types of packets are being used in the transmitted data.

4. Workflow. Various related works have been analysed. It is decided that more than one type of classifier will be applied to the data set under consideration. The CICDDoS2019 dataset is used in this work, which is very close to the packets transmitted in the real network. The attributes of the packets under transmission are being considered. The 'Information Gain Ranking' [29] method arranges the attributes in order. 33 attributes are selected, leaving behind other insignificant attributes. After preprocessing the data, three different classification methodologies are applied to the said dataset. The results obtained using all three methods are compared at the end. The steps described above are represented in Figure 4.1.

Table 2.1: Literature survey.

Author Name	Description	Method	Parameter	Advantages	Future Scope
Jain et. al. (2021)	The authors considered the capabilities of different machine learning classifier techniques and then used avoting mechanism for the best.	Naïve Bayes Random Forest KNN SVM	Source IP Destination IP Packet Size Source Port Flow Bytes Header length etc.	Ensemble performs better than individual classifiers	Instead of the existing dataset, a simulated dataset with a higher size is to be considered.
Wang et.al. (2020)	The authors tried to optimize the detector using feedback.	Multi-Layer Perceptron	Source IP TCP attributes Destination port.	Handling of detection errors in detecting DDoS is novel.	To extend the work for SDN
Jia et.al. (2020)	A comprehensive framework for security against DDoS attacks in IoT was proposed	LSTM and CNN	Source IP Destination IP Packet Size Source port Destination Port	Comprehensive Approach	Implementing the proposed work in parallel systems.
Latif et. al. (2015)	The authors used a version of the decision tree in WBAN, a primitive step towards IoT.	Very Fast Decision Tree	Packet Loss Delay Jitter	High accuracy Low false alarm	Simulation -based approach to be implemented in real-world scenario.
Saini et.al. (2020)	Authors have analysed the performance of the different machine-learning techniques in identifying malicious packets.	J48 Random Forest Naïve base	SRC Address DES Address PKT ID PKT AVG SIZE	High accuracy	To work on more types of attacks.
Shieh et.al. (2021)	Authors in this work tried to address the Open Set Recognition issue in DDoS attack identification.	Gaussian Mixture Model BI-LSTM	MI_dir_L5_weight MI_dir_L5_mean MI_dir_L3_variance etc.	GMM is effective in both trained and novel attacks.	The proposed model can be validated on more datasets.
Marvi et al. (2020)	Authors suggested a three-step feature selection before applying the model. LGBM procedure is used for identifying DDoS attacks.	LGBM algorithm	Source IP Mean ACK Header length etc.	The proposed framework even works for unseen DDoS attacks of some specific types.	The work is to be applied to other types of DDoS attacks.

The whole process can be broadly divided into three phases. 'Preparation is the primary phase, then 'Pre-processing' is applied, followed by 'Classification' and 'Analysis' phases.

4.1. Preparation. In the preparation phase, the dataset is selected to work on. The CICDDOS2019 dataset is a well-defined dataset that represents the context and is very close to the actual scenario. Hence, that is considered for this work. Various categories of DDoS attacks are possible on a network. Each of these datasets contains numerous attributes. So, there is a need to select the vital ones. One of the ranking algorithms

Fig. 3.1: Objective.

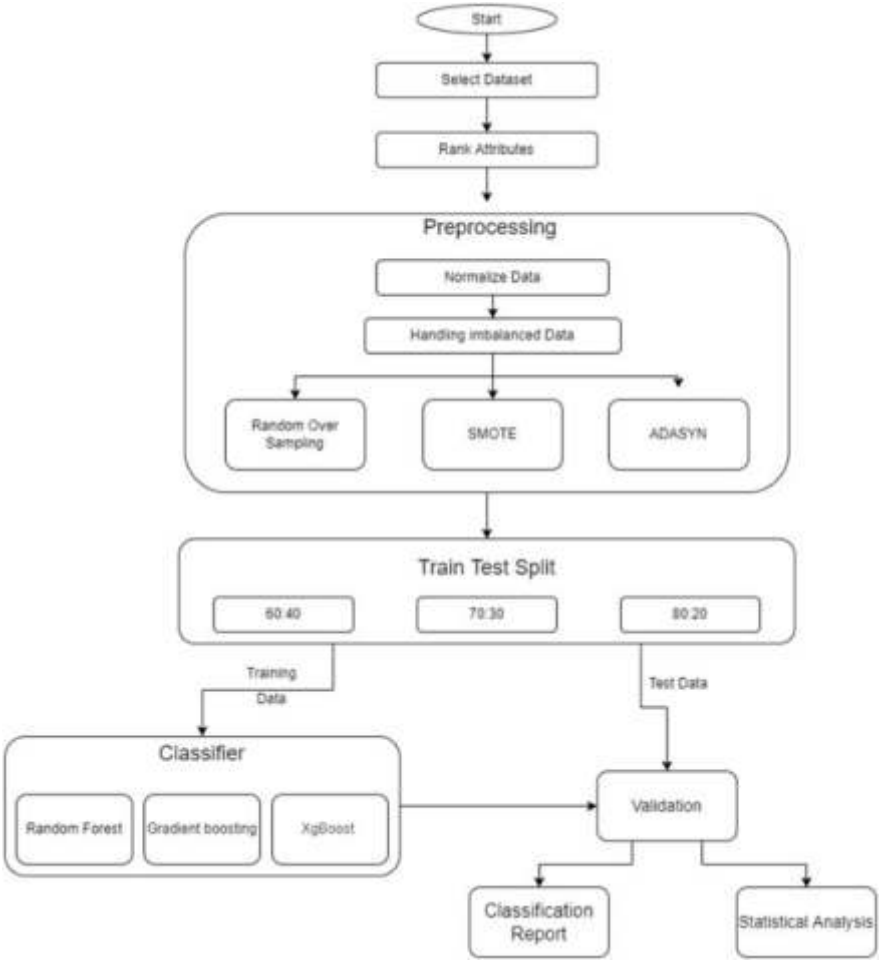


Fig. 4.1: Workflow.

called 'Information Gain Ranking' is being applied [29]. After finalizing the set of attributes, the final dataset is prepared and passed to the second stage.

4.2. Preprocessing . In this phase, the data are first normalized to make it appropriate for the machine learning model. The next step is to handle the imbalanced data set. As the process flow diagram demonstrates, three sampling techniques are introduced to the dataset. Random Oversampling [2] is a simple method to balance a data set. In this case, the minority class data are randomly replicated to bridge the gap.

SMOTE is an oversampling method [4]. Sampling by replacement is not being used in this technique. Rather, several examples are synthesized and used for oversampling. According to the amount of oversampling,

the necessary records are synthesized based on the k nearest neighbours.

'ADASYN' is an advanced approach for synthetic sampling [7]. The minority class data are considered to be either 'easier to learn' or 'difficult to learn'. The dataset will include more synthesized data if it falls under the first category. It helps to reduce the imbalance.

All the classification models are considered for each version of the re-sampled dataset.

4.3. Classification. After finalising the dataset, three different classification techniques are applied to it. Random Forest, Gradient Boosting, and Extreme Gradient Boosting classification models are applied to this work's dataset. The proposed model, i.e., using the XgBoost algorithm with ADASYN for classifying affected packets, outperformed the other models under investigation. All three classifiers used for classification in this work are described here.

4.3.1. Random Forest. Random Forest [8] is popularly used as a classification model. A random forest can be constructed by combining several decision trees. Here we use a voting concept to decide the class of the data under consideration. The result is based on the aggregation of the predictions.

The steps can be expressed as described in Algorithm 1.

Algorithm 1 Random Forest

Initialize $i = 1$
while $i \leq M$ **do**
  Make D_i = A sub dataset decided randomly
  Make Nd_i = A node with sub dataset D_i
  Invoke **Construct(**Nd_i**)**
end while
Construct(Nd_i**)**
if p and q belong to C and N respectively and $p = q$ **then**
  return
else
  Verify every possible splitting criteria
  Choose the feature F that has the highest information gain
  Create n sub-nodes of N {Here n represents the different possible features of F}
  Assign $c = 1$
  while $c \leq n$ **do**
    Assign $Nd + i = D_i$
    Invoke **Construct(**Nd_i**)**
    $c = c + 1$
  end while
end if

4.3.2. Gradient Boosting. 'Gradient Boosting' is an ensemble technique [17]. It is a combination of several weak learners (in this case, decision trees) to create a powerful classification model. This is a boosting ensemble technique in which several homogeneous weak learners work sequentially, improvising the model to get better results. It works by optimizing a loss function to minimize prediction errors. The basic steps are listed in Algorithm 2.

4.3.3. Extreme Gradient Boosting. Extreme Gradient Boosting (XgBoost) [5] comprises steps to optimize 'Gradient Boosting' training. 'Boosting' is different from the 'Bagging' concept. Bagging uses a voting mechanism to finalize the output, whereas boosting is an ensemble of phases. In the case of boosting, each phase learns from the previous phase. The 'XgBoost' uses the same concept, and the steps can be briefly explained as follows in Algoritm 3.

4.4. Analysis. For validating the investigation, different ratios of the train-to-learn data set are considered. The same model is applied to three different combinations of train-to-learn ratios, i.e., 60:40, 70:30, and 80:20.

After getting the result by applying the three different classifiers mentioned earlier, the result is represented using a confusion matrix. Then the different parameters, including accuracy, precision, recall, and f1-value,

Algorithm 2 Gradient Boosting

Input:

- Training set $\{(x_i, y_i)\}$
- Loss function L
- Number of iterations: M
- Base learner model: $h(x)$

A. Initialize the model:

$$f_0(x) = \arg\min_{\gamma} \sum L(y_i, \gamma)$$

B. Perform iterations to learn:

for $m = 1$ to M **do**

a. Calculate pseudo-residuals:

$$r_{i,m} = \left(\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right)_{f=f_{m-1}} \quad \text{for } i = 1 \text{ to } n$$

b. Train a weak learner:

Fit the base learner to $h_m(x)$

c. Calculate multiplier γ_m using the given optimization problem:

$$\gamma_m = \arg\min_{\gamma} \sum L(y_i, f_{m-1}(x_i) + \gamma h_m(x_i))$$

d. Update the model:

$$f_m(x) = f_{m-1}(x) + \gamma_m h_m(x)$$

end for

C. Output $f_m(x)$

are calculated. These measures are then compared with some related work. Also, the results obtained are statistically analysed, and comparing different models' performances is established.

5. Experimentation & Result Analysis. The CICDDoS2019 dataset is being considered in this work for classifying malicious and benign packets. The three methodologies explained in the earlier section are used. The dataset has 88 attributes for each of the entries. Using the ranking process explained earlier, 33 attributes have been considered in this work. In that process, the attributes selected are based on the entropy value. The decrease in entropy of each attribute is considered to evaluate the gain with respect to the target, which in turn helps in ranking the attributes. The method associates each of the attributes with a weight. For this work, we fixed a standard weight value and considered the 33 attributes with a weight value greater than the threshold. The attributes considered are: *Flow ID, Destination IP, Source IP, Avg Fwd Segment Size, Fwd Packet Length Mean, Average Packet Size, Sub flow Fwd Bytes, Total Length of Fwd Packets, Packet Length Mean, Destination Port, Flow Bytes/s, Bwd Packets/s, Bwd Header Length, Sub flow Bwd Packets, Packet Length Variance, Bwd IAT Mean, Flow IAT Mean, Flow Packets/s, Flow Duration, Init Win bytes forward, Total Length of Bwd Packets, Sub flow Bwd Bytes, Fwd Packets/s, Avg Bwd Segment Size, Bwd Packet Length Mean, Protocol* etc.

Before the implementation of any of the classification techniques, the dataset is first normalised. The dataset is normalised to make the classifiers work better than non-normalized data. As in the non-normalized data, the attributes tend to have values on different scales. Hence, the statistical technique of normalization is used to convert the values to a particular range.

Then the dataset is made to undergo another statistical method called sampling. In this work, the over-sampling method is chosen as the minor class has a relatively lower count than the major class data. The classification techniques are applied after the sampling. 'Random oversampling', 'SMOTE' and 'ADASYN' are the three sampling methods applied to the dataset.

Algorithm 3 Extreme Gradient Boosting Algorithm

Input:

- Training set $\{(x_i, y_i)\}$
- Loss function L
- Number of learners with count M
- Learning rate denoted as α

A. Initializing the model:

$$f_0 = \arg\min_{\theta} \sum L(y_i, \theta)$$

B. Loop to update the model:

for $q = 1$ to M **do**

a. Evaluate gradients:

$$g_q(x_i) = \left(\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right)$$

b. Evaluate Hessians:

$$h_q(x_i) = \left(\frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)^2} \right)$$

c. Use the following training set to fit a base learner:

$$\left\{ x_{i,q} - \left(\frac{g_q(x_i)}{h_q(x_i)} \right) \right\} \text{ by resolving the following problem}$$

$$\phi_q = \arg\min_{\phi} \sum \frac{1}{2} h_q(x_i) \left(\left(-\frac{g_q(x_i)}{h_q(x_i)} \right) - \phi(x_i) \right)^2$$

$f_q(x) = \alpha \phi_q(x)$

d. Update the model:

$$f_q(x) = f_{q-1}(x) + f_q(x)$$

end for

C. Output:

$$f(x) = \sum f_q(x)$$

After the dataset is sampled, three different classification techniques are applied. Those are Random Forest, Gradient Boosting, and XgBoosting. Each version of sampling is combined with each classification technique under consideration in this work. Hence, nine different models are considered; for example, SMOTE is combined with Random Forest, Gradient Boosting, and XgBoosting to form three different classification models. After applying each of the methods, the quality of classification is judged by the measures widely used, and those are explained further.

First, the numbers of 'True Malignant', 'False Malignant', 'True Benign', and 'False Benign' instances are recorded for each type of packet.

Based on these values, 'precision', 'accuracy', 'recall' and 'F1 Value ' are calculated to verify results obtained by applying the methods. The concepts behind these measures are discussed next.

Accuracy represents the percentage of the total number of packets identified as compared to the total number of samples under consideration. It can be expressed as equation quation 5.1.

$$.\text{Accuracy} = \frac{TM + TB}{TM + FM + FB + TB} \tag{5.1}$$

Another measure, 'precision' is used to represent the total number of truly identified positive cases against the total number of predicted positive cases. It may be described as equation 5.2.

$$\text{Precision} = \frac{TM}{TM + FM} \tag{5.2}$$

Table 5.1: Result Analysis of Random Forest with different cases of Sampling

MODEL	SPILT RATIO	TM	FM	FB	TB	ACCU RACY	PRECI SION	RECALL	F1	TOTAL
RF- NO SAMPLING	60:40	6490	1107	1756	2405	0.76	0.85	0.79	0.82	11758
	70:30	4657	840	1336	1985	0.75	0.85	0.78	0.81	8818
	80:20	6283	382	916	1298	0.85	0.94	0.87	0.91	8879
RF-ROS	60:40	6108	1718	1451	5994	0.79	0.78	0.81	0.79	15271
	70:30	4848	802	1145	4657	0.83	0.86	0.81	0.83	11452
	80:20	6245	725	864	2901	0.85	0.90	0.88	0.89	10735
RF-SMOTE	60:40	6199	1727	1374	6070	0.80	0.78	0.82	0.80	15370
	70:30	4734	1069	1069	4314	0.81	0.82	0.82	0.82	11186
	80:20	6130	687	785	3130	0.86	0.90	0.89	0.89	10732
RF-ADASYN	60:40	6994	1435	1794	6257	0.80	0.83	0.80	0.81	16480
	70:30	6619	968	802	4696	0.86	0.87	0.89	0.88	13085
	80:20	6254	687	764	2939	0.86	0.90	0.89	0.90	10644

Table 5.2: Result Analysis of Gradient Boosting with different Cases of Sampling

MODEL	SPILT RATIO	TM	FM	FB	TB	ACCU RACY	PRECI SION	RECALL	F1	TOTAL
GB- NO SAMPLING	60:40	6970	727	947	3214	0.86	0.91	0.88	0.89	11858
	70:30	5848	649	551	2771	0.88	0.90	0.91	0.91	9819
	80:20	6582	649	802	2642	0.86	0.91	0.89	0.90	10675
GB-ROS	60:40	6650	1176	107	7337	0.92	0.85	0.98	0.91	15270
	70:30	5004	532	360	5557	0.92	0.90	0.93	0.92	11453
	80:20	8863	764	507	2901	0.90	0.92	0.95	0.93	13035
GB-SMOTE	60:40	6994	832	336	7108	0.92	0.89	0.95	0.92	15270
	70:30	5199	337	465	5452	0.93	0.94	0.92	0.93	11453
	80:20	8901	534	725	3474	0.91	0.94	0.92	0.93	13634
GB-ADASYN	60:40	7032	756	260	6841	0.93	0.90	0.96	0.93	14889
	70:30	5710	283	360	4832	0.94	0.95	0.94	0.95	11185
	80:20	8939	678	514	3901	0.92	0.93	0.95	0.94	14032

The third measure used for the analysis of classification is known as 'recall'. It is the total number of positive cases identified as true against the total positive samples. It is expressed as equation 5.3.

$$\text{Recall} = \frac{TM}{TM + FB} \tag{5.3}$$

Another measure, the 'F1 Score', was used to check the balance between the previous two measures. It may be measured as equation 5.4.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.4}$$

The readings and the measures obtained in this work are mentioned in Tables 5.1, 5.2 and 5.3.

The classification techniques are applied to the non-sampled data along with the sampled data. The findings are represented in the tables above. The random forest classifier performs most accurately when combined with ADASYN and with the split ratios 70:30 and 80:20. But the precision is highest when no sampling is applied with a split ratio of 80:20. The recall is highest when RF-SMOTE is used with 80:20 ratios, as well as for the RF-ADASYN combination with 70:30 and 80:20 split ratios. In the case of Random Forest, the maximum accuracy gained is 86%.

Table 5.3: Result Analysis of Extreme Gradient Boosting with Different Cases of Sampling

MODEL	SPILT RATIO	TM	FM	FB	TB	ACCU RACY	PRECI SION	RECALL	F1	TOTAL
XG- NO SAMPLING	60:40	8490	707	565	2596	0.90	0.92	0.94	0.93	12358
	70:30	8102	487	898	2736	0.89	0.94	0.90	0.92	12223
	80:20	8092	687	764	2336	0.88	0.92	0.91	0.92	11879
XG-ROS	60:40	8108	680	683	6299	0.91	0.92	0.92	0.92	15770
	70:30	8443	764	525	3901	0.91	0.92	0.94	0.93	13633
	80:20	8345	764	525	3901	0.90	0.92	0.94	0.93	13535
XG-SMOTE	60:40	8681	345	565	5879	0.94	0.96	0.94	0.95	15470
	70:30	8489	640	573	3825	0.91	0.93	0.94	0.93	13527
	80:20	8398	640	573	3825	0.91	0.93	0.94	0.93	13436
XG-ADASYN	60:40	8261	527	422	5879	0.94	0.94	0.95	0.95	15089
	70:30	9245	424	278	3749	0.95	0.96	0.97	0.96	13696
	80:20	9245	524	278	3749	0.94	0.95	0.97	0.96	13796

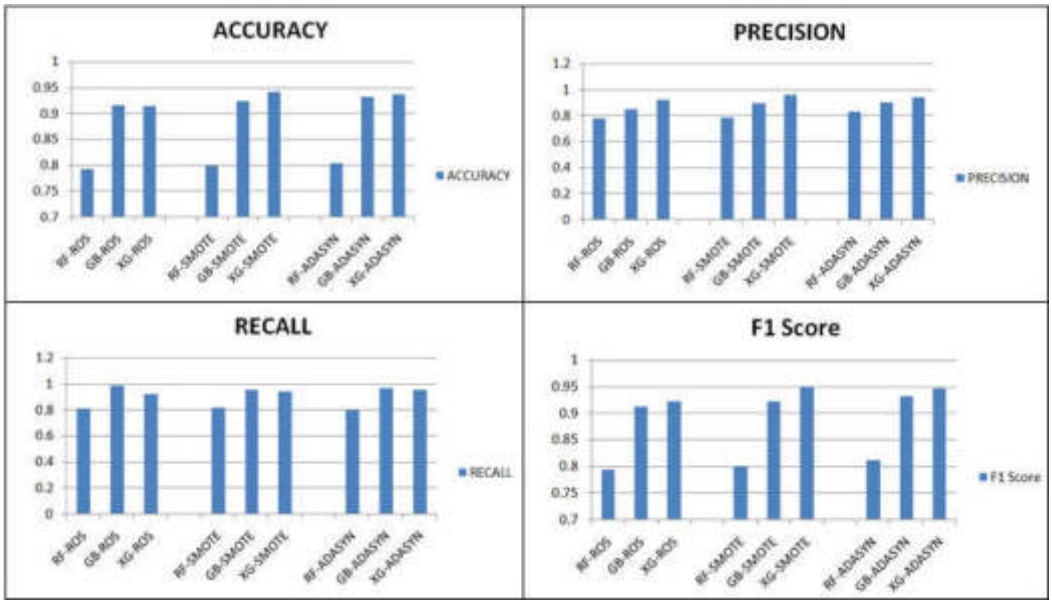


Fig. 5.1: Overall analysis with 60:40 split

The next classifier under consideration is the gradient-boosting method. It provides the highest accuracy of 94% when combined with the ADASYN sampler, and the train test split is 70:30. It also gives the most precise result for the same combination. Though recall is highest for the Gradient Boosting-Random Oversampling combination with a 60:40 ratio, this combination doesn't have better accuracy or precision. From the observation, it may be concluded that the Gradient Boosting-ADASYN model with a 70:30 split of data is the best combination with a sufficient value of recall and f1 score.

The last technique used for classification is 'Extreme Gradient Boosting'. When this is combined with the ADASYN sampler with 70:30 split ratios, the best accuracy of 95% is obtained. The same combination also has the best precision of 96% with better recall and f1 score, i.e., 97% and 96%.

The analysis of three different cases (train-test data ratio-wise, i.e., 60:40, 70:30, and 80:20) is further represented using Figures 5.1, 5.2, and 5.3.

The analysis shows that the XgBoost–ADASYN model results in the highest accuracy and F1 score. Also,

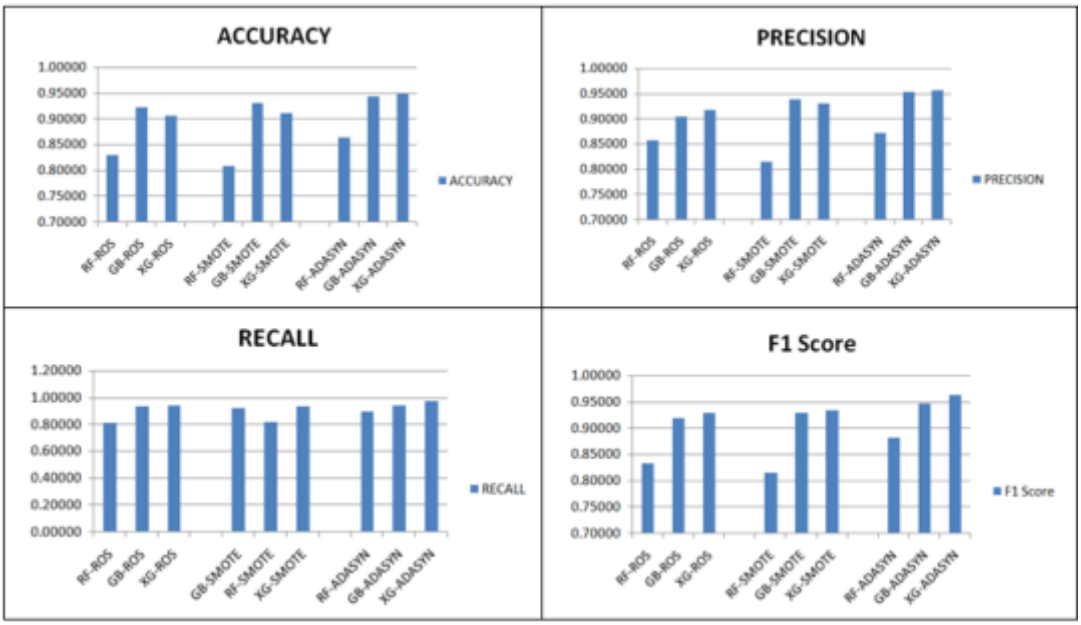


Fig. 5.2: Overall analysis with 70:30 split

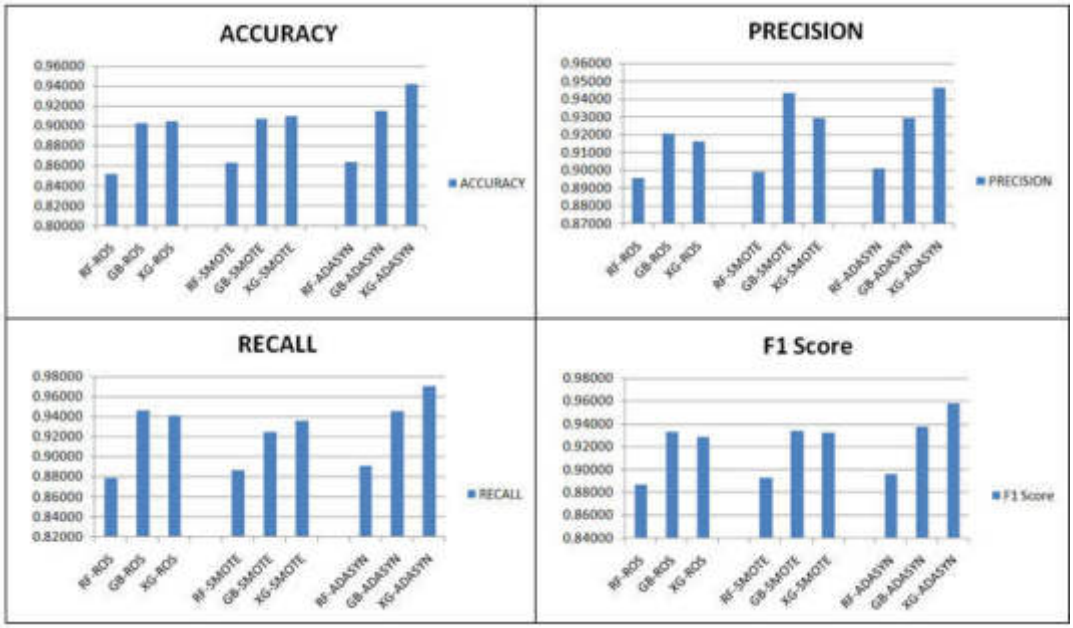


Fig. 5.3: Overall analysis with 80:20 split

this model yields the second-best precision and recall metrics. The same classifier, i.e., XgBoost gives the best precision when used with the SMOTE technique.

Statistical Analysis. It is evident from the plot in Figure 5.4 that the median improvement in GB-SMOTE accuracy is close to 0.915, and the median rise in XG-SMOTE accuracy is near 0.93. These results

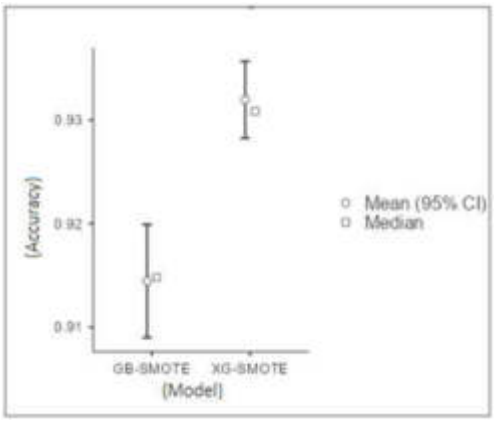


Fig. 5.4: Model vs accuracy plot

Table 5.4: Wilcoxon's Paired Samples T-Test

Model 1	Model 2	Wilcoxon W	P value	Mean difference
GB-SMOTE-60-40	GB-ADASYN-60-40	8.00	0.049	-0.0105
XG-ROS-60-40	XG-SMOTE-60-40	0.00	0.002	-0.0253
GB-SMOTE-60-40	XG-SMOTE-60-40	0.00	0.002	-0.0172
GB-ADASYN-60-40	XG-ADASYN-60-40	8.00	0.049	-0.00392
GB-ROS-60-40	XG-ROS-60-40	55.0	0.002	0.00935
GB-SMOTE-70-30	GB-ADASYN-70-30	1.00	0.004	-0.0125
XG-ROS-70-30	XG-SMOTE-70-30	27.00	1.00	4.23e-4
GB-SMOTE-70-30	XG-SMOTE-70-30	54.0	0.004	0.0224
GB-ADASYN-70-30	XG-ADASYN-70-30	17.0	0.322	-0.00401
GB-ROS-70-30	XG-ROS-70-30	55.0	0.006	0.0202
GB-SMOTE-80-20	GB-ADASYN-80-20	42.0	0.160	0.00596
XG-ROS-80-20	XG-SMOTE-80-20	24.0	0.770	0.00263
GB-SMOTE-80-20	XG-SMOTE-80-20	40.0	0.232	0.00812
GB-ADASYN-80-20	XG-ADASYN-80-20	1.00	0.004	-0.0283
GB-ROS-80-20	XG-ROS-80-20	9.00	0.064	-0.00689

suggest that XG-SMOTE outperforms GB-SMOTE in terms of effectiveness. We still need to determine whether our discovery is statistically significant. For a paired sample t-test, the Wilcoxon rank test [20] [19] is performed with the null hypothesis H_0 set to "there is no difference in accuracy among the two models". The total of the signed ranks, as given by equation 5.5, is the test statistic W in the Wilcoxon signed-rank test.

$$W = \sum_{i=1}^{N} [\text{sgn}(x_{2,i} - x_{1,i}) \cdot R_i] \tag{5.5}$$

The i^{th} value of N measurement pairs is indicated by $x_i = (x_{1,i}, x_{2,i})$, while the pair's rank is shown by R_i.

The test statistic W and P values for the paired sample T-test are shown in Table 5.4. At the 5% significance level, the null hypothesis can be rejected if the P value is less than 0.05. It is therefore inclined to believe the alternative hypothesis, according to which model 2 performs better than model 1 or vice versa.

From the Table 5.4, it is observed that the GB-ADASYN model outperforms the GB-SMOTE and the XG-ADASYN model is better than the first one. So it may be concluded that the XG-ADASYN model works better than other models under consideration.

Table 5.5: Comparison with existing work

Author	Percentage with Method & Parameter	
Kumar et. al.	88.8% (RF)(Accuracy)	77.78% (Naïve Bayes) (Accuracy)
Sharafaldin et. al.	77%(RF) (Precision)	78% (ID3) (Precision)
Shieh et. al.	89.80%(BI-LSTM) (Accuracy)	86.8%(BI-LSTM-GMM) (Accuracy)
Our Work	94.8% (Accuracy)	96.1% (Precision)

Comparison. From the experiment, it was found that XgBoost, when used with the ADASYN sampling method, outperforms other models under consideration with respect to accuracy and F1 score, where as the XgBoost-SMOTE method has better precision over other models. Further, the results obtained are compared with some of the existing work, such as [24], [14] and [22]. The comparison is demonstrated in Table 5.5.

6. Conclusion and Future Work. The behaviour of the DDoS dataset is extensively studied, and then a few state-of-the-art classification techniques are investigated. It may be suggested that not only the classification technique but also the sampling method play a great role in obtaining better results. From the overall observation, the conclusion obtained is that the XgBoost-ADASYN combination helps in obtaining the best results in this scenario.

Future research will consider other datasets related to various protocol packets. Some other techniques for identifying malicious packets may also be considered for identifying proper attributes, and then normalizing and sampling, as well as new techniques for classification, may be considered to improve efficiency.

REFERENCES

[1] Y. Al-Hadhrami and F. K. Hussain, *Ddos attacks in iot networks: a comprehensive systematic literature review*, World Wide Web, 24 (2021), pp. 971–1001.
[2] G. E. Batista, R. C. Prati, and M. C. Monard, *A study of the behavior of several methods for balancing machine learning training data*, ACM SIGKDD explorations newsletter, 6 (2004), pp. 20–29.
[3] E. Bursztein, *Inside the infamous mirai iot botnet: A retrospective analysis*, Cloudflare Blog, Aug, (2020).
[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, *Smote: synthetic minority over-sampling technique*, Journal of artificial intelligence research, 16 (2002), pp. 321–357.
[5] T. Chen and C. Guestrin, *Xgboost: A scalable tree boosting system*, in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
[6] D. Giusto, *A. lera, g*, Morabito, l. Atzori (Eds.) The Internet of Things, (2010), p. 11.
[7] H. He, Y. Bai, E. A. Garcia, and S. Li, *Adasyn: Adaptive synthetic sampling approach for imbalanced learning*, in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), Ieee, 2008, pp. 1322–1328.
[8] T. K. Ho, *Random decision forests*, in Proceedings of 3rd international conference on document analysis and recognition, vol. 1, IEEE, 1995, pp. 278–282.
[9] A. K. Jain et al., *Ddos detection using machine learning ensemble*, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12 (2021), pp. 1647–1655.
[10] Y. Jia, F. Zhong, A. Alrawais, B. Gong, and X. Cheng, *Flowguard: An intelligent edge defense mechanism against iot ddos attacks*, IEEE Internet of Things Journal, 7 (2020), pp. 9552–9562.
[11] G. Kesavaraj and S. Sukumaran, *A study on classification techniques in data mining*, in 2013 fourth international conference on computing, communications and networking technologies (ICCCNT), IEEE, 2013, pp. 1–7.
[12] M. Khari, G. Shrivastava, S. Gupta, and R. Gupta, *Role of cyber security in today's scenario*, in Detecting and mitigating robotic cyber security risks, IGI Global, 2017, pp. 177–191.
[13] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, *Ddos in the iot: Mirai and other botnets*, Computer, 50 (2017), pp. 80–84.
[14] A. Kumar and T. J. Lim, *Edima: Early detection of iot malware network activity using machine learning techniques*, in 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), IEEE, 2019, pp. 289–294.
[15] R. Latif, H. Abbas, S. Latif, A. Masood, et al., *Evfdt: an enhanced very fast decision tree algorithm for detecting distributed denial of service attack in cloud-assisted wireless body area network*, Mobile Information Systems, 2015 (2015).
[16] M. Marvi, A. Arfeen, and R. Uddin, *A generalized machine learning-based model for the detection of ddos attacks*, International Journal of Network Management, 31 (2021), p. e2152.
[17] A. Natekin and A. Knoll, *Gradient boosting machines, a tutorial*, Frontiers in neurorobotics, 7 (2013), p. 21.

[18] D. Panda, B. K. Mishra, and K. Sharma, *A taxonomy on man-in-the-middle attack in iot network*, in 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), IEEE, 2022, pp. 1907–1912.
[19] R. R Core Team et al., *R: A language and environment for statistical computing*, 2013.
[20] M. Şahin and E. Aybek, *Jamovi: an easy to use statistical software for the social scientists*, International Journal of Assessment Tools in Education, 6 (2019), pp. 670–692.
[21] P. S. Saini, S. Behal, and S. Bhatia, *Detection of ddos attacks using machine learning algorithms*, in 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2020, pp. 16–21.
[22] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, *Developing realistic distributed denial of service (ddos) attack dataset and taxonomy*, in 2019 international carnahan conference on security technology (ICCST), IEEE, 2019, pp. 1–8.
[23] K. Sharma and B. B. Gupta, *Towards privacy risk analysis in android applications using machine learning approaches*, International Journal of E-Services and Mobile Applications (IJESMA), 11 (2019), pp. 1–21.
[24] C.-S. Shieh, W.-W. Lin, T.-T. Nguyen, C.-H. Chen, M.-F. Horng, and D. Miu, *Detection of unknown ddos attacks with deep learning and gaussian mixture model*, Applied Sciences, 11 (2021), p. 5213.
[25] G. Shrivastava and P. Kumar, *Sensdroid: analysis for malicious activity risk of android application*, Multimedia Tools and Applications, 78 (2019), pp. 35713–35731.
[26] S. G. H. Soumyalatha, *Study of iot: understanding iot architecture, applications, issues and challenges*, in 1st International Conference on Innovations in Computing & Net-working (ICICN16), CSE, RRCE. International Journal of Advanced Networking & Applications, vol. 478, 2016.
[27] M. Suresh and R. Anitha, *Evaluating machine learning algorithms for detecting ddos attacks*, in Advances in Network Security and Applications: 4th International Conference, CNSA 2011, Chennai, India, July 15-17, 2011 4, Springer, 2011, pp. 441–452.
[28] M. Wang, Y. Lu, and J. Qin, *A dynamic mlp-based ddos attack detection method using feature selection and feedback*, Computers & Security, 88 (2020), p. 101645.
[29] E. Zdravevski, P. Lameski, A. Kulakov, B. Jakimovski, S. Filiposka, and D. Trajanov, *Feature ranking based on information gain for large classification problems with mapreduce*, in 2015 IEEE Trustcom/BigDataSE/ISPA, vol. 2, IEEE, 2015, pp. 186–191.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: May 29, 2024
Accepted: Aug 4, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 964–972, DOI 10.12694/scpe.v26i2.3905

EXPLORATION ON GRASSROOTS PARTY BUILDING INNOVATION DRIVEN BY BIG DATA

LINGLING MIAO*

Abstract. Big data has become a disruptive factor in many fields in the digital age, including grassroots party development and political organizing. This study examines the creative use of big data to support grassroots party building initiatives with the goal of defining its effects and possibilities in promoting more effective, flexible, and responsive political systems. This study looks at how big data analytics can be incorporated into grassroots community mobilization, policy formation, and political involvement processes using a thorough analytical methodology. We investigate several case studies where the application of big data tools and methodologies has improved engagement strategies, streamlined decision-making processes, and improved communication between party leaders and voters. The RNN-LSTM algorithm's principle is discussed in this work from an Internet+ standpoint, and the objective function and regularization term are used in the Taylor expansion to maximize the algorithm's objective function. The RNN-LSTM model is then trained to identify its ideal splitting nodes, and the ten-fold cross-validation technique is used to assess the model's performance. This multifaceted research explains how big data encourages creativity in grassroots party mobilization, coalition building, voter engagement, and issue advocacy through case studies, quantitative measures, and qualitative observations. incorporating curriculum thinking's ingrained ideological science into the creation of grassroots party formation to activate curriculum thinking's nurturing function. The study ends with tactical suggestions for incorporating big data into grassroots party formation successfully, opening the door for additional data-driven approaches in governance and political processes. This study adds to the growing body of knowledge about the relationship between digital technology and political innovation, namely at the grassroots level.

Key words: Grassroots Party Building, Innovation Driven, Big Data, deep learning

1. Introduction. A greater awareness of the evolution of curriculum thinking and popular grassroots party building has resulted from the Internet era[4]. Because of the people's desire for a better life, the construction of socialism with Chinese characteristics, which is situated at a new historical point, has raised the bar for talent standards and called for the selection of superior builders and successors who possess both excellent business acumen and a strong moral code [8]. Curriculum Civics addresses the issue of "two skins" between knowledge and values education in traditional education and adapts to the practical demands of social development [16]. We actively consider the problem of "siloing" political education and curriculum thinking, realizing that ideological and political education must be covered in its whole for the purpose of creating grassroots parties[2].

The integration of curricular thinking, political education, and grassroots party formation enables these three domains to jointly take on the responsibilities of teaching, consensus building, and convergence of educational synergy [11]. The literature [18] argues that to achieve the goal of comprehensive and high-quality student growth, ideological and political education must permeate every course, including higher mathematics. This makes the study of curriculum politics and ideology conducive to the all-around development of ideas. A model for assessing the efficacy of integrated physical education ideology and politics was put forth in the literature [15]. It specified evaluation indicators like teaching content, activity planning and organization, and sports teams, and it also included data collection and evaluation analysis based on each evaluation indicator.

A Zhejiang Province City an investigation into the adoption of digital telemedicine technology in Guizhou Province's community-based medical facilities[13]. A study on the information disclose- sure method of grassroots governance in digital Transformation from the perspective of rural revitalization. Impact of Digital Grassroots Social Governance Transformation on Digital Vulnerable Groups and Reactions. Using digitalization to hide murky issues: Technology as a grassroots governance mechanism's enabler. Resolving cooperative

*Department of teaching and research of C.P.C history and construction, Party School of CangZhou Municipal Committee of C.P.C, Cangzhou, 061000, China (linglingmiandata@outlook.com)

issues in local government: the route of digital platform integration[19]. Examination and evaluation of the Guizhou Province's grassroots hospitals' digital construction. Organizing and developing a digital platform to address cooperative issues in local government. The existing situation and the digital transformation of grassroots social governance's optimization approach[1].

Big data has become a powerful force in today's digital landscape, drastically altering processes and boosting the effectiveness of decision-making systems across a range of industries. Its incorporation into political organizing and grassroots party growth represents a critical turn toward more responsive, intelligent, and dynamic political processes. Acknowledging the significant influence of big data, this research aims to thoroughly investigate its capacity to redefine grassroots political tactics and improve democratic participation.

Finding and communicating the revolutionary potential of big data in political systems—especially at the local level—is the main driving force behind this research. The study aims to outline the wide range of advantages that such technology can provide to political organizations looking to capitalize on and adapt to the digital revolution by looking at the creative uses of big data analytics in community mobilization, policy development, and political participation.

The main contribution of proposed method is given below:

1. The fundamental innovation of this work is how it optimizes decision-making processes within grassroots political organizations by applying the RNN-LSTM algorithm, which was modified from the Internet+ perspective.
2. Using ten-fold cross-validation and a Taylor expansion of the objective function with a regularization term, the study not only demonstrates significant improvements in the strategic planning and execution of political initiatives but also validates the usefulness of big data tools in improving engagement strategies and communication between party leaders and voters.
3. This study adds to the growing body of knowledge about how digital technology can support political innovation, especially at the local level, and provides new directions for the study of data-driven political engagement and organizing.

The rest of our research article is written as follows: Section 2 discusses the related work on various Grassroots Party Building, Innovation Driven, Big Data and deep learning methods. Section 3 shows the algorithm process and general working methodology of proposed work. Section 4 evaluates the implementation and results of the proposed method. Section 5 concludes the work and discusses the result evaluation.

2. Related Works. The literature [16] extends and transcends the literature on radicalism geography and party organization to provide a developmental vision of grassroots party growth. It argues that geography supports grassroots party building through three interconnected processes. The literature [3] makes the case that fostering grassroots party organization advances party organization, speeds up the development of public infrastructure and services, and enhances grassroots party organization to advance basic modernization and party construction. It also helps to foster relationships between the party and the people as well as between cadres and the people. The literature[5] demonstrates how the CCP has attempted to emphasize the delivery of goods and services at the grassroots level to broaden its sphere of influence and restore political relevance.

College and university party building, as a crucial component of party building in the new era, needs to accurately understand the laws governing party building and enhance the caliber of party building and the degree of scientization with higher standards and more proactive practicality[6]. Party construction in the new era has presented new and higher requirements for college and university party building work in the face of the ever-changing internal and external situation, higher requirements for party-building work, and the issues that persist in certain college and university party building work[7]. It is particularly crucial to actively develop a new model of systematic party-building work and encourage innovative practices.

The researcher examined how school systems homogenize via the lens of the new constitutional theory, demonstrating how schools respond to difficulties by means of coercion, imitation, and normative isomorphism. Combining competition, market failure risk, coercive accountability measures, and schooling paths regarding the gold standard, this isomorphism process offers significant insights into educational innovation[10]. The author examines how innovation and technology are affecting education, and she suggests that educational institutions and policymakers take action to address technological illiteracy and infrastructure issues to improve the environment for teaching and learning [9].

The literature [14] contends that student party branches in colleges and universities are prompted to base themselves on the current construction situation; continuously strengthen the construction of party branch committees; enhance the ideological and political leadership of students; and strictly cultivate and educate student party members to ensure the work of party branches is completed. This is because of the overall layout of the party construction work in colleges and universities as well as the realistic needs of the reform and development of colleges and universities [12, 17]. This article explores three parts to investigate how to establish the nurturing task of integrating curriculum thinking into grassroots party formation from the perspective of Internet+. The Internet+ and its accompanying algorithm are the main topics of the first section.

Big data is becoming more and more integrated into many industries, but there is still much to learn about how these technologies may be used for party development and grassroots political organization, especially when it comes to in-depth, empirically supported research. The majority of the literature that has already been written about big data in politics concentrates on its general uses, including voter behavior research and election forecasts, and pays less attention to how these technologies might especially support grassroots movements and party development. This study aims to close a number of significant gaps in the present scholarly discourse.

Comprehensive studies that demonstrate the methodical integration of big data analytics into the several aspects of grassroots party development—such as direct political involvement, policy drafting, and community mobilization—are hard to come by. The goal of this research is to give a detailed understanding of these procedures together with a well-defined approach for using big data tools in these situations.

Although there is a wealth of literature on general data analysis methods, there is a dearth of studies on the use of particular algorithms, such as RNN-LSTM, in political data analysis. This paper explores how these algorithms might be tailored to meet the particular needs and difficulties of political datasets, thereby improving their usefulness and relevance.

Current research frequently makes extensive use of either qualitative or quantitative methods. Studies that combine the two are necessary to give a comprehensive picture of the effects of big data. In order to close this gap, this study evaluates the usefulness of big data tools in grassroots politics using both qualitative and quantitative methods, including case studies and anecdotal evidence, statistical analysis, and model performance indicators.

3. Proposed Methodology. The proposed methodology for Grassroots Party Building Innovation Driven by Big Data using RNN-LSTM is used for the evaluation. Initially the data is collected from the evolution of political theory and politics incorporated into grassroots party building, as well as an examination of the nurturing style and purpose of higher education. Next the data is trained by using RNN-LSTM. In figure 3.1 shows the architecture of proposed method.

3.1. Data Collection. Obtain primary sources and historical documents that trace the development of political beliefs and their use in grass-roots movements. Possible sources include old periodicals, historical groups, and libraries. Get information on how political theories have been incorporated into higher education courses across time by consulting university archives, course catalogs, and syllabi. Perform qualitative data collection by means of interviews with political scientists, educators, and party organizers regarding the perceived influence and efficacy of political education within grassroots movements. Give examples of how political ideologies that were introduced in classrooms have impacted grassroots movements. These can offer comprehensive insights on how educational theories are really use in the real world. Analyze survey data using statistical techniques, searching for patterns, correlations, and trends that will help you measure the influence of political education on grassroots initiatives.

3.2. RNN. We use recurrent neural networks (RNNs) as the foundation for attention-based deep neural networks. An expansion of the traditional feed-forward neural network is a recurrent neural network. Models of long short-term memory (LSTM) are also built using RNN design. They enhance the learning capacity for long-time sequencing information and address the gradient-related weaknesses of RNNs. The distinction is that there are four layers communicating in a particular way rather than just one neural network layer. An input gate, an output gate, a forget gate, and a cell make up an LSTM unit. The newly added memory cell can maintain its state for extended periods of time, and the three gates control the information that enters and

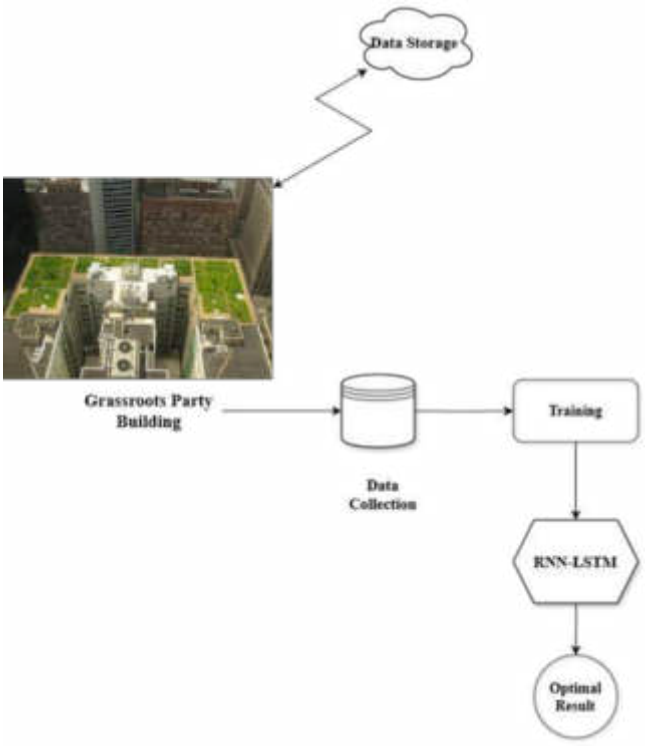


Fig. 3.1: Architecture of Proposed Method

leaves the cell:

$$X = [h_{t-1}, x_t] \tag{3.1}$$

$$f_t = \sigma(W_f.X + b_f) \tag{3.2}$$

$$i_t = \sigma(W_i.X + b_i) \tag{3.3}$$

$$o_t = \sigma(W_o.X + b_o) \tag{3.4}$$

$$c_t = f_t \Theta c_{t-1} + i_t \Theta \tanh(W_c.X + b_c) \tag{3.5}$$

$$h_t = o_t \Theta anh(c_t) \tag{3.6}$$

A variant of LSTMs known as gated recurrent units (GRUs) became available. They produce a simpler model than LSTMs by adding a gating mechanism, combining the ”forget” and ”input” gates into a signal update gate, and making a few minor modifications.

$$z_t = \sigma(W_z.X + b_z) \tag{3.7}$$

$$r_t = \sigma(W_r.X + b_r) \tag{3.8}$$

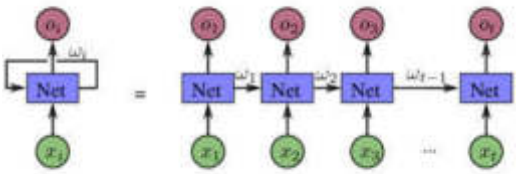


Fig. 3.2: Structure of RNN

$$h_t = \tanh(W_h.(r_t \Theta X)) \tag{3.9}$$

$$h_t = (1 - z_t)\,\Theta h_{t-1} + z_t \Theta h_t \tag{3.10}$$

A deep recurrent neural network differs from a recurrent neural network primarily in that it consists of several layers of separate recurrent networks stacked on top of one another. The issue that, although RNNs are capable of depth, this concept isn't going to entail a hierarchical analysis of the information is what gave rise to the concept for the current implementation. Each element and sentence can be seen in the same space by using the same algorithm applied recursively to determine the contributions of children to their parents and the identical computation to provide an output response. Compared to traditionally arranging deep learners, this technique differs in that each hidden layer theoretically lies in a distinct representation time and may be more the summary depiction of the input data than the layer before it. This hierarchy among hidden representations is a significant advantage of depth.

Additionally, we used the Adam optimizer, which is renowned for its solid performance on a variety of tasks, as part of our optimization approach to modify the settings of our model during training. Adam precisely calculates and modifies learning rates based on feature sets. Unlike other approaches, it uses speed in addition to storing a declining mean of previous gradients to calculate current gradients. The Adam algorithm has a healthy rate of learning that does not disappear, quick integration, and minimal variation. The experimental investigation provides a detailed presentation of the training procedure's variables. In figure 3.2 shows the structure of the Recurrent Neural Network.

3.3. Long Short-Term Memory (LSTM). An LSTM is a type of artificial neural network designed to process sequential data, such as text, audio, or time series. It is highly beneficial when information undergoes processing with persistent dependencies, i.e., when the result at a particular time step depends on data from previous time stages. For LSTM networks to be able to remember this information for a long time, they use memory cells, forget gates, input gates, and output gates. The gates let a network store and retrieve data based on requests by controlling the data that enters and exits the storage cells. LSTMs are widely used for tasks including speech recognition, language translation, and stock price predictions. The constituents of an LSTM network comprise.

Data entry into the memory cell is managed by the input gate.

$$input = \sigma(Wi^* \,[ht-1, xt] + bi) \tag{3.11}$$

The forget gate regulates the data flow that leaves the memory cell.

$$forget = \sigma(Wf^* \,[ht-1, xt] + bf) \tag{3.12}$$

The output gate regulates how the memory cell's output is sent to the remaining components of the network.

$$output = \sigma(Wo^* \,[ht-1, xt] + bo) \tag{3.13}$$

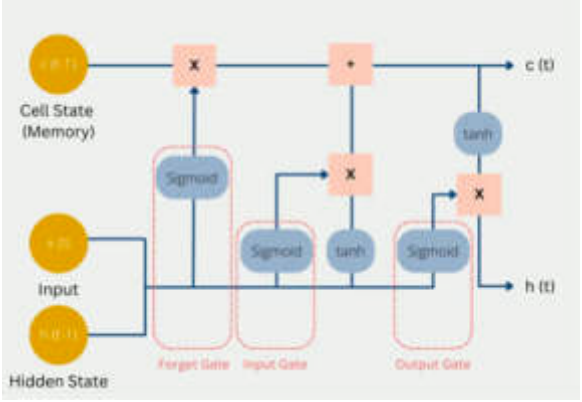


Fig. 3.3: Structure of LSTM network

The memory cell is where the information is stored.

$$memory = ft^{*}ct - 1 + it^{*}\tanh(Wc^{*}\,[ht - 1, xt] + bc) \tag{3.14}$$

The LSTM unit's output is utilized to generate predictions or transfer data to the following LSTM unit.

$$hidden = ot^{*}\tanh(ct) \tag{3.15}$$

In figure 3.3 shows the structure of LSTM network used for the prediction of industrial data.

Long-term memory retention is a key feature of LSTM units, which is advantageous for applications where prediction accuracy depends on prior knowledge. Standard RNNs are hampered by the vanishing gradient problem, whereas LSTMs are able to capture long-term dependencies in the data thanks to this memory feature. Because LSTMs can store historical data points and dynamically alter their internal state based on historical inputs, they are naturally well-suited for time series prediction. Because of this, they are perfect for applications like weather forecasting, stock market prediction, and any other situation where past trends have a substantial impact on future results.

4. Result Analysis. The generalization error of the learner is typically tested experimentally to assess a model; however, in a real-world experimental situation, other factors such as detection accuracy, time overhead, and performance overhead must also be considered. To help ensure stability and fidelity of test findings and assess the learner's capacity to discriminate between fresh samples, the training and test sets should be split and the test set should be mutually exclusive with the training set to the greatest extent feasible.

It looks like a bar chart in the image you sent shows performance measures for each of the ten trials, Exp1 through Exp10. Accuracy (Bars in Yellow): calculates the percentage of real results—both true positives and true negatives—among all cases that were looked at. Precision (Blue Bars): Shows the percentage of actual positive outcomes among all positive cases that the algorithm was able to identify. Recall (Red Bars): Indicates the percentage of real positive outcomes that are true positives, demonstrating the model's capacity to identify all pertinent cases. F1-score (Purple Bars): A single metric that balances accuracy and recall is the harmonic mean of precision and recall.

The graph indicates that the metrics are largely consistent between the experiments, indicating that the model or system under test is operating steadily. Visually speaking, recall (Red) and F1-score (Purple) tend to be lower than accuracy (Yellow) and precision (Blue) in the majority of tests. This pattern may indicate that, despite the model's high precision in properly predicting positive cases, its recall in catching a large percentage of all potential positive cases may be lower, which could have an impact on the F1-score. It is possible that some of the observed variability is the result of variations in the data sets, settings, or environmental circumstances of each experiment. In domains like machine learning, where these metrics are essential for comprehending and

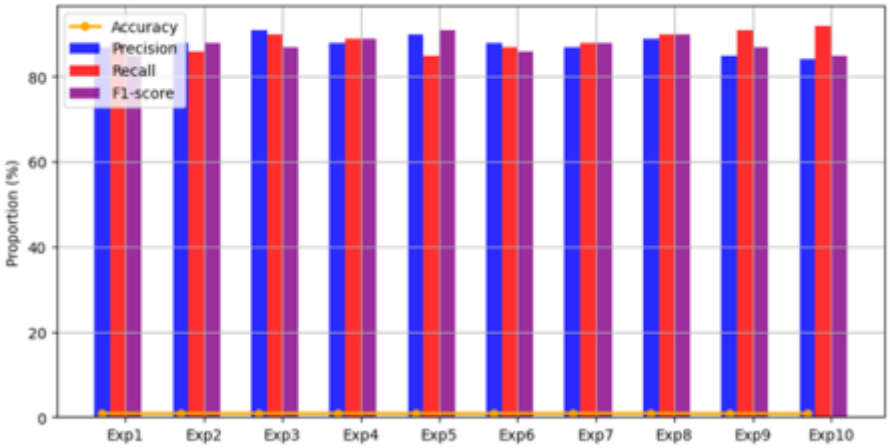


Fig. 4.1: Evaluation of various performance metrics

enhancing model behavior, the visualization helps assess how well a model or system is performing. Finding experiments that perform worse on some measures may point to areas that might be optimized. For example, if the F1-score is not optimal, strategies to balance recall and precision could be used. In figure 4.1 shows the result of various performance metrics.

The bar and line chart in the image you gave shows survey results on distinct motivational and ideological characteristics based on varying degrees of agreement. Responses are divided into four levels by it: "Very agree," "More agree," "General agree," as well as "Disagree." Determine the correct three views (Red Bars): This probably means to match up with the three main viewpoints or beliefs that the surveying body deems to be accurate. Hold fast to your beliefs and ideals (Green Bars): evaluates the respondents' commitment to their political or personal convictions. Boost political literacy (Blue Bars): Shows if respondents agree with statements regarding improving their knowledge of politics. Enhance Learning Motivation (Purple Bars): Indicates the degree to which participants concur that they can or have enhanced their learning motivation.

High Agreement on Core viewpoints: "Be firm in your ideals and convictions" and "Establish the correct three views" are two statements that typically obtain greater levels of agreement, particularly in the "Very agree" and "More agree" categories. Decline in Agreement for Increasing Political Literacy and Learning Motivation: From "More agree" to "General agree" and "Disagree," there is a discernible decline in agreement on enhancing political literacy and learning motivation. Overall Decline in Mean Values: As the degree of agreement declines, the mean value clearly indicates a downward trend, indicating that, overall, respondents are less in agreement with or supportive of the survey statements as they become more neutral or negative. In figure 4.2 shows the result of Analysis of learning initiatives and the function of education.

The picture you gave shows data on various political and educational traits broken down into five groups using a grouped bar chart. Three educational initiatives are used to measure each category: "Educate all staff," "Whole process education," and "All-round education." Red Bars (Educate all staff): Most likely denotes a fundamental strategy meant to give all staff members a basic education in politics. The concept of "Blue Bars" (whole process education) advocates for a more thorough and ongoing educational program that covers all facets and phases of political participation. A holistic educational strategy that incorporates different facets of political education and personal development is implied by Light Blue Bars (all-round education).

The ratios change dramatically amongst the various categories and educational programs. For example, the "Educate all staff" initiative's relatively high shares for "Knowledge of Politics" and "Political Participation" indicate that basic political knowledge is effectively disseminated, and involvement is encouraged. Particularly under "All-round education," "Sense of Political Efficacy" and "Sense of Political Responsibility" have lower proportions, suggesting possible areas where educational efforts might be less successful or require improvement. When comparing the three projects, "Educate all staff" typically has larger proportions, which could mean that

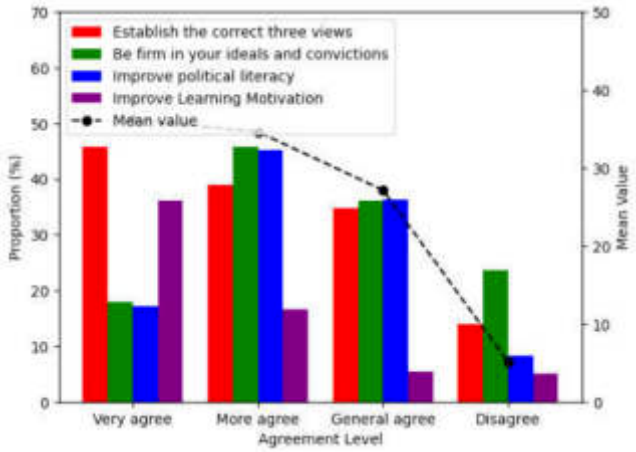


Fig. 4.2: Analysis of learning initiatives and the function of education

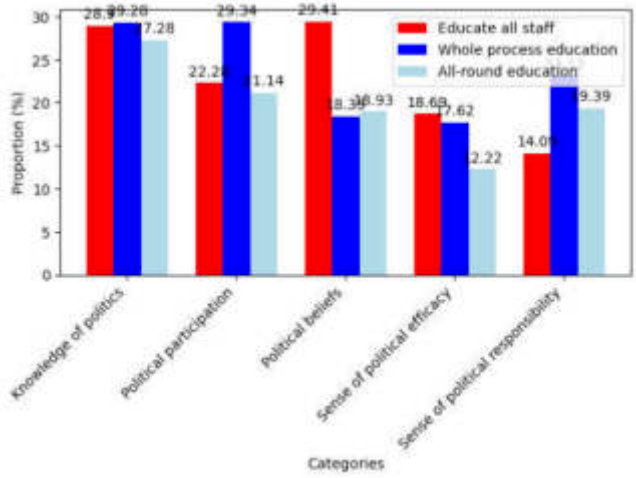


Fig. 4.3: The three-pronged educational paradigm, which combines politics, philosophy, and grassroots party formation

a simple, all-inclusive educational strategy works better in some situations, such as increasing involvement and knowledge. In figure 4.3 shows the result of The three-pronged educational paradigm, which combines politics, philosophy, and grassroots party formation.

5. Conclusion. In terms of comprehending and applying data-driven tactics inside political grassroots movements, the research on "Exploration on Grassroots Party Building Innovation Driven by Big Data and RNN-LSTM" is a major advancement. This study has carefully described how big data can significantly improve the efficacy and responsiveness of grassroots party building initiatives when combined with cutting-edge machine learning techniques like Recurrent Neural Networks and Long Short-Term Memory (RNN-LSTM) algorithms. The discovery of significant patterns and trends in voter behavior and emotion was made possible by the deployment of RNN-LSTM models to large data sets. This has made it easier to implement more focused and significant engagement tactics, which has raised activism and participant involvement in grassroots movements. Big data and RNN-LSTM integration into grassroots party construction not only changes the way political

organizations interact with their constituents but also changes the field of political strategy and communication. The information gleaned from this study will be extremely important as political groups continue to navigate a world that is becoming more and more data driven. The study's suggestions for resolving issues with data privacy, technological complexity, and guaranteeing ongoing advancements in data handling capacities highlight the necessity of a flexible and comprehensive strategy for innovation in political grassroots movements.

Future studies could investigate the incorporation of sophisticated machine learning models, like Transformer models or GPT-3 for natural language processing, which could provide enhanced precision and comprehension of context in political communication analysis. Creating models that can analyze and respond to data in real-time may be essential for campaign plans that are dynamic and enable quick adjustments depending on engagement metrics and real-time voter input.

REFERENCES

[1] X.-k. Chen and J. Yu, *Evaluation model of physical education integrated ideology and politics based on principal component analysis*, Mobile Networks and Applications, 27 (2022), pp. 1240–1251.
[2] Y. DUAN, *The practical approach for rural grassroots party building to lead to poverty alleviation in the new era*, Journal of Jiangxi University of Finance and Economics, (2021), p. 8.
[3] S. Halvorsen, *The role of territory in grassroots party-building: insights from argentina*, Territory, Politics, Governance, 11 (2023), pp. 1303–1323.
[4] J. Huang, *Research on the integration of curriculum thinking and politics into the construction of grassroots party building and nurturing work under the perspective of internet+*, Applied Mathematics and Nonlinear Sciences, (2023).
[5] D. Jing, *The party building leads the innovative exploration of grassroots governance in rural pastoral areas under the concept of co-governance*, Scientific and Social Research, 3 (2021), pp. 125–131.
[6] K. Kan and H. B. Ku, *Serving the people, building the party: Social organizations and party work in china's urban villages*, The China Journal, 85 (2021), pp. 75–95.
[7] W. Liu, L. Lei, Z. Zhang, et al., *Research on countermeasures for the construction of student party branches in higher education institutions*, Frontiers in Educational Research, 4 (2021).
[8] L. Lv and D. Shi, *Innovative development and practice of digital rural governance model based on green ecology*, Sustainability, 15 (2023), p. 2955.
[9] P. K. R. Maddikunta, Q.-V. Pham, D. C. Nguyen, T. Huynh-The, O. Aouedi, G. Yenduri, S. Bhattacharya, and T. R. Gadekallu, *Incentive techniques for the internet of things: a survey*, Journal of Network and Computer Applications, 206 (2022), p. 103464.
[10] Z. Ren, L. Shen, and Z. Zhou, *Research on the reform of digital economy professional training mode from the perspective of" internet plus curriculum thinking"*, Frontiers in Business, Economics and Management, 2 (2021), pp. 32–35.
[11] C. J. Sells, *Building parties from city hall: Party membership and municipal government in brazil*, The Journal of Politics, 82 (2020), pp. 1576–1589.
[12] T. Tao, Y. Liu, Y. Qiao, L. Gao, J. Lu, C. Zhang, and Y. Wang, *Wind turbine blade icing diagnosis using hybrid features and stacked-xgboost algorithm*, Renewable Energy, 180 (2021), pp. 1004–1013.
[13] G. Tian and W.-H. Tsai, *"beautiful countryside construction," policy inspection teams, and grassroots political participation in china*, Journal of Contemporary China, 32 (2023), pp. 951–962.
[14] K. K. Yun, S. W. Yoon, and D. Won, *Prediction of stock price direction using a hybrid ga-xgboost algorithm with a three-stage feature engineering process*, Expert Systems with Applications, 186 (2021), p. 115716.
[15] H. Zhang, *Who serves the party on the ground? grassroots party workers for china's non-public sector of the economy*, Journal of Contemporary China, 27 (2018), pp. 244–260.
[16] P. Zheng, X. Wang, and J. Li, *Exploration and practice of curriculum ideological and political construction reform——take" information security" course as an example*, ASP Transactions on Computers, 1 (2021), pp. 1–5.
[17] Q. Zhou and Y. Sun, *High order one-step methods for backward stochastic differential equations via itô-taylor expansion.*, Discrete & Continuous Dynamical Systems-Series B, 27 (2022).
[18] Y. Zhou, *The application of curriculum ideology and politics in the training of judicial vocational education talents*, Journal of Higher Education Research, 3 (2022), pp. 155–159.
[19] H. Zhu, Z. Zhang, Y. Hua, et al., *Teaching exploration and practice of advanced mathematics based on curriculum ideology and politics*, Curriculum and Teaching Methodology, 4 (2021), pp. 74–78.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: May 11, 2024
Accepted: Jun 9, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 973–982, DOI 10.12694/scpe.v26i2.3911

EXPLORATION ON RESOURCE SCHEDULING OPTIMIZATION STRATEGIES IN CLOUD COMPUTING ENVIRONMENT

XIAOLAN LI* AND WENFU YANG†

Abstract. The rapid expansion of cloud computing has brought opportunities and difficulties for resource schedule optimization to enhance productivity and decrease expenses. This study offers a thorough analysis of resources planning optimisation techniques in cloud computing settings, emphasizing the most recent developments and approaches. First, we go over the basic ideas of resource scheduling, such as resource sharing, load balancing, and task distribution. The task scheduler's execution-time mappings among the future demand and cloud resources makes application planning one of the major difficulties in cloud computing. By decreasing makespan and improving resource utilization, an effective scheduling system is required to schedule the diversified workload and enhance performance metrics. Numerous scheduling methods that were previously in use only considered makespan and utilization of resources metrics, ignoring other important factors that have an immediate effect on cloud service performance, such as energy consumption and migration time. To address the problems, the authors have developed the Hybridize Whale Optimization Algorithm (H-WOA), a nature-inspired multi-objective task scheduling algorithm that can make scheduling decisions at runtime depending on the availability of cloud resources and impending workload requirements. Furthermore, the suggested method distributes the resources according to task priorities and end users' budgets. The workload for the proposed H-WOA technique, which is based on the Cloudsim toolkit, is created by creating datasets (da01, da02, da03, da04) with various task densities and workload records from NASA (da05, da06) and HPC2N (da01, da02) parallel workload repositories. An extended experiment's findings demonstrate that the suggested H-WOA technique enhanced the important variables and performed better than alternative baseline policies.

Key words: Resource Scheduling, Optimization, Cloud Computing Environment, Hybridize Whale Optimization Algorithm

1. Introduction. The ability to offer resources and services on demand over fast computer networks is one of the main advantages of cloud computing. As network technologies and infrastructure continue to evolve and advance, cloud computing has proven to perform better at handling various large-scale and complicated end-users' (clients') workloads. Modern physical and virtualized technology, coupled with advanced development platforms and applications, satisfy the needs of heterogeneous clients. The cloud and the internet of things (IoT) have emerged as key ideas that establish new standards for industry advances at the current stage of technological breakthroughs [4].

The term "cloud computing" describes the provision of several computer services via the internet, including networking, processing, and storage [8, 13]. Individuals and organizations can now access and use computing resources without having to manage or own the hardware that underlies them because to this approach to change in architecture. Rather, these resources are supplied and overseen by other organizations, sometimes known as cloud service providers. The services that are provided fall into many categories, such as Platform as a Service (PaaS), Software as a Service (SaaS), and Infrastructure as a Service (IaaS), all of which cater to different operating requirements. In today's modern world, cloud computing is extremely important. Cloud environments' agility, adaptability, and scalability are becoming essential as digital change gains traction for enterprises [9].

Swarm intelligence is a well-known class of metaheuristics that draw inspiration from natural occurrences. Techniques to swarm intelligence are continuous, unpredictable, and population based. Swarm intelligence has proven effective in solving a wide range of real-world issues, including the location issue in wireless sensor networks (WSNs) [10, 17], drone placement [7], robot path planning [20, 12], the network planning issue in radio frequency identification (RFID) networks [6], machine learning optimization, image processing, computer aided diagnostic systems, portfolio optimization, and more. Swarm intelligence algorithms have many uses, but

*College of Internet of Things, Jiangxi Teachers College, JTC, Yingtan, 335000, China (xiaolanlicloud@hotmail.com)
†College of Internet of Things, Jiangxi Teachers College, JTC, Yingtan, 335000, China

they have also undergone continuous modification, hybridization, and parallelization to produce the optimal outcomes.

A multiobjective task scheduling system is presented in [6] to reduce the amount of time and energy required for task scheduling. Whale Optimization Algorithm is a technique used to efficiently schedule jobs based on voltage frequency while considering variables like task execution duration and operation order. It is compared to the current PSO algorithm, is simulated on MATLAB, generates workload at random, and significantly reduces execution time and energy usage. The authors of [16] presented a task-scheduling system to manage energy usage and makespan. An algorithm known as Ant Colony Optimization is applied in a hybrid fashion to handle the parameters.

Cloud computing's explosive growth has brought opportunities and difficulties for resource scheduling optimization that maximizes efficiency and minimizes costs. With an emphasis on the most recent advancements and methodologies, this paper provides a thorough review of resource scheduling optimization techniques in cloud computing systems. First, we review the basic ideas of resource scheduling, such as work allocation, load balancing, and resource sharing. Application planning is essential because one of the main issues in cloud computing is the execution-time mappings of the task scheduler between future demand and cloud resources. To manage the diverse workload, reduce makespan, and optimize resource usage, an efficient scheduling system is required, which will improve performance metrics.

Our solution to these problems is the Hybrid Whale Optimization Algorithm (H-WOA), a multi-objective task scheduling algorithm inspired by nature that can schedule tasks in real time depending on cloud resource availability and workload demands in the future. Additionally, the suggested approach distributes resources based on end users' budgets and work priorities. The workload for the H-WOA approach is built by using datasets with different task densities and workload records from NASA and HPC2N parallel workload repositories, using the Cloudsim toolbox. Comprehensive test outcomes show that the suggested H-WOA method surpasses alternative baseline policies and enhances important factors.

The main contribution of the proposed algorithm is given below:

1. This study presents a new multi-objective work scheduling system that draws inspiration from nature. The H-WOA is intended to dynamically optimize scheduling choices in response to the workload needs that are approaching and the real-time availability of cloud resources.
2. The novel resource distribution method included in the suggested H-WOA algorithm ensures that resource allocation is in line with end-user budgets and task priorities.
3. This feature makes the scheduling system more useful and applicable to real-world cloud computing scenarios where operational and budgetary constraints are important factors to consider.
4. The study provides comprehensive experimental results that substantiate the H-WOA algorithm's efficacy. The analysis shows a considerable improvement in key performance measures when compared to baseline policy.

The rest of our research article is written as follows: Section 2 discusses the related work on Resource Scheduling Optimization, Cloud Environment. Section 3 shows the algorithm process and general working methodology of proposed work. Section 4 evaluates the implementation and results of the proposed method. Section 5 concludes the work and discusses the result evaluation.

2. Related Works. An energy-efficient task scheduling method was developed in [11] that considers varied workloads and divides tasks into two stages. A genetic algorithm is employed as a scheduling process implementation technology. It was executed in two stages. The first phase was created without taking deadline constraints into account, with tasks mapped onto virtual resources. Using task reassignment strategy based on task priority tasks mapped to virtual machines (VMs) in the second phase [14]. It was incorporated into the Amazon Cloud platform. It was contrasted with baseline methods such as SJF, FCFS, and GA variants. Eventually, based on the data, it was determined that the suggested strategy uses less energy and accurately meets task deadlines.

An energy-saving scheduling technique was put forth by the authors in [9] to address energy usage in a heterogeneous cloud environment. This scheduling system was developed using a methodology based on vacation queuing theory. To reduce the amount of time tasks spend waiting in queues, they have scheduled tasks so that when virtual machines (VMs) are busy, they must go into a queue where they can sleep [18].

When resources become available, they can then resume their work in the active queue, which reduces both the amount of time tasks spend waiting in the queue and the amount of energy they use. The simulation was done using MATLAB. It was tested against Min-Min and TSAST, and the results indicated that it had an advantage over the comparative methods for the specified metrics [3].

The technique of mapping and scheduling jobs or tasks from a scientific workflow onto distributed computing resources to accomplish goals, including limiting execution time or cost, is known as scientific workflow scheduling, or SWFS [5, 1]. Scientific processes are arranged, systematic sets of calculations intended to accomplish specific scientific objectives. These activities could include simulations, data processing, or intricate analysis involving several computational processes [2]. Because distributed computing resources are heterogeneous, with different processor kinds, memory capacities, and network bandwidths, SWFS seeks to efficiently divide these workloads among the resources at hand. Managing task dependencies, balancing numerous objectives, overcoming resource failures or uncertainties, and considering the dynamic nature of tasks and resources are some of the difficulties associated with SWFS [15, 19].

Resource scheduling methods have come a long way, yet there are still some holes in the literature. Many traditional scheduling techniques frequently ignore other important variables like energy usage and migration time in favour of maximizing makespan and resource utilization. The importance of these aspects is growing as cloud computing systems aim for increased sustainability and efficiency. Moreover, the majority of current algorithms are not adaptable enough to make decisions about scheduling that take workload demands and resource availability into account in real-time. Studies on the distribution of resources according to task priorities and user budgets are similarly few, despite the fact that these factors are critical for maximizing cloud service performance from both a technical and financial standpoint.

3. Proposed Methodology. This study introduces and evaluates the Hybrid Whale Optimization Algorithm (H-WOA), which aims to enhance resource allocation in cloud computing environments. Several key components of our proposed method are designed to assess the performance and effectiveness of H-WOA relative to traditional scheduling methods. Inspired by nature, the H-WOA is going to be developed as a multi-objective scheduling system. This strategy will maximize a number of scheduling parameters, including makespan, resource utilization, energy usage, and migration time. The flexibility to decide on runtime scheduling in response to workload requirements and real-time resource availability will be a key feature of H-WOA. To improve resource scheduling in cloud computing environments, the Hybridize Whale Optimization Algorithm (H-WOA) is being introduced and assessed in this research. A number of crucial elements in our suggested technique are intended to evaluate the efficacy and efficiency of H-WOA in comparison to conventional scheduling algorithms. The H-WOA will be created as a multi-objective scheduling system with inspiration from nature. Several aspects of scheduling, such as makespan, resource use, usage of energy, and migrate time, will be optimized by this approach. One of the main components of H-WOA will be the capacity to decide on runtime scheduling in response to workload needs and real-time resource availability. In figure 3.1 shows the architecture of proposed method.

Budgetary concerns and resource allocation based on priorities are also included in this method to guarantee a well-rounded and economical scheduling solution. The suggested method's architecture is depicted in Figure 1, whereby workload prediction models, dynamic resource monitoring, and adaptive task prioritization mechanisms are seamlessly integrated. To ensure H-WOA is reliable and flexible, its performance will be put to the test in a variety of workload conditions. Furthermore, the algorithm will be evaluated by benchmarking it against industry-standard scheduling strategies in order to show its superiority in a number of important performance parameters. To support the efficacy of the suggested approach, a thorough comparative analysis of the simulation results will be provided.

3.1. Problem Formulation. This section begins with a clearly stated problem and moves on to a discussion of the suggested system design. Assuming for the moment that there are k tasks, which are represented as $ta_k = \{ta_1, ta_2, ta_3, \ldots.., ta_k\}$, n number of virtual machines (VMs) $v_n = \{v_1, v_2, v_3, \ldots.., v_n\}$, i amount of physical hosts $h_i = \{h_1, h_2, h_3, \ldots.., h_i\}$, and lastly we have considered j datacenters, i.e. $d_j = \{d_1, d_2, d_3, \ldots.., d_j\}$. These k jobs must be meticulously scheduled or mapped onto n virtual machines (VMs), each of which is housed in i physical hosts and j datacenters. The assignment or scheduling of these jobs must minimize makespan, migration time, and consumption of energy while considering the priorities of the tasks and VMs based on the

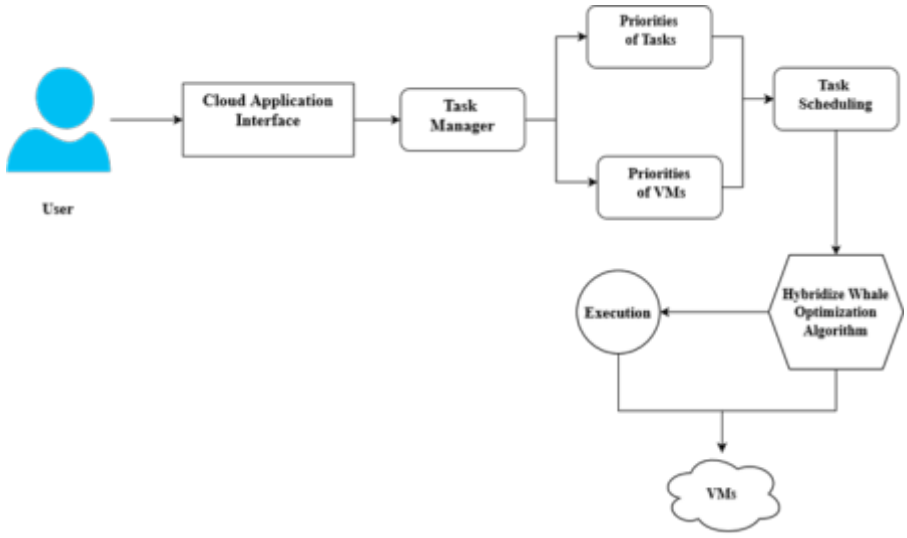


Fig. 3.1: Architecture of Proposed Method

cost of electricity per unit.

Initially, users upload tasks to the cloud console, and a broker submits them on their behalf. Brokers send these jobs to task manager, which checks if they're granted permission to use virtual resources before assigning them to scheduler. Should they own virtual resources, it is imperative to verify the service level agreed upon by the cloud provider and the user regarding the specific resources to which they have accessibility. Both the cloud service supplier and the cloud user should gain from a Service Level Agreement (SLA) that forces the service owner to implement effective multi-objective scheduling algorithms, like makespan, to reduce processing cost, and so on. In this case, the customers benefit by finishing their jobs more quickly, while the cloud owner makes more money by offering a variety of cloud services to a larger number of clients using these multi-objective scheduling tactics. The scheduler in the cloud computing standard architecture will typically assign the matching user request if the SLA between the user and the cloud provider is valid. Here, we are introducing a method in our architecture to determine tasks and virtual machine priorities depending on the cost of electricity units for datacenters. job priorities are determined because not every job has a comparable size, processing speed, or duration capability.

Task priorities are determined by the task's length and processing capacity. As a result, we must first compute the total load operating on the virtual machines to determine the priorities for the jobs. whole load on virtual machines computed using the following Eq.(3.1)

$$I_{vm} = \sum l_n \tag{3.1}$$

where l_n is the current load across all n virtual machines (VMs).

Every VM's overall load has been computed. since all virtual machines (VMs) are housed on physical hosts, for the purpose of calculating the total load on those machines. Next, we computed the total workload on the physical hosts utilizing the formula below, Equation(3.2).

$$l_h = l_{vm} / \sum h_i \tag{3.2}$$

While h_i is the total number of physical hosts considered in our model, l_h denotes the load on all physical hosts, and l_{vm} denotes the overall load on all VMs.

When many users request virtual resources from various sources, load balancers are required in the cloud computing paradigm. If a VM is unable to handle the volume of requests, a load balancer will transfer user

requests or tasks to the next virtual machine (VM) in the same pool or to a different pool. Setting a threshold value is necessary to balance the load. The performance of the cloud will be negatively impacted by static threshold values, which make putting one up in cloud computing a difficult operation. Now of the SLA, the threshold value is established based on several factors, including the volume of work, future requests, resource capacity, and so on.

Therefore, to shift load to the following virtual machine or pool, we have taken into consideration the changing threshold value. On the physical hosts, the dynamic threshold (thr_n) value is regarded as an average load.

$$thr_n = \frac{\sum_{i=1}^{n} lh_i}{n} \tag{3.3}$$

where n is the number of physical hosts and $lh_i$is the load at each physical host. The dynamic demands on the actual machine cause periodic changes in the dynamic threshold value. We determine physical hosting underutilization, overutilization, or balancing situations using this threshold.

3.2. Task Scheduling using Hybridize Whale Optimization Algorithm (H-WOA). We attempted to enhance the initial WOA execution by tackling the exploitation–exploration trade-off adjustments, keeping in mind that the two most crucial mechanisms of any swarm algorithm are exploitation (intensification) and exploration (diversification), and that the efficiency of a swarm algorithm relies strongly on the adjusted equilibrium between both steps in terms of both convergence speed and solution quality.

Modest and/or substantial enhancement tactics can be applied to any swarm algorithm to make it better. Changes to a few search equation components and adjustments to the algorithm's control variable behavior (several studies add dynamic parameter behavior) are examples of small enhancements. The term "major improvements" typically refers to combining one heuristic or metaheuristic with another. By substituting the shortcomings of one strategy with the merits (strengths) of another, hybrid algorithms integrate the best elements of two or more techniques. It is evident from looking through the literature that hybrid algorithms can be quite effective in solving various kinds of issues.

We created and put into effect a hybridized WOA method that addresses the shortcomings of the original version, building on our previous work with hybrid algorithms for swarm intelligence.

3.2.1. Encircling Prey. The purpose of modeling such actions is to replicate the algorithm's exploration and exploitation stages, in which potential solutions converged to the most optimal solution thus far. This behavior's mathematical model is:

$$\overrightarrow{D} = |\overrightarrow{C}.\overrightarrow{X} * (t) - \overrightarrow{X}(t)| \tag{3.4}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X} * (t) - \overrightarrow{A}.\overrightarrow{D} \tag{3.5}$$

The best solution so far is represented by $\overrightarrow{X}*(t)$, a solution is represented by $\overrightarrow{X}(t)$, coefficient vectors A and C are used to adjust the search's intensity and unpredictability, and D is the distance between the best solution and the current solution.

3.2.2. Spiral Bubble-Net Attacking Mechanisms. This is a simulation of the whales' spiral approach to their prey, as modeled by:

$$\overrightarrow{D'} = |\overrightarrow{X} * (t) - \overrightarrow{X}(t)| \tag{3.6}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{D'}.e^{b.l}.\cos(2\pi l) \; + \overrightarrow{X} * (t) \tag{3.7}$$

where l is a random number in [−1,1], b is a constant determining the spiral's shape, and $e^{b.l}$and $\cos(2\pi l)$ determine the logarithmic spiral route.

Large-scale simulations were run with a range of benchmark datasets that had varying task densities and workload characteristics in order to verify the efficacy of H-WOA. When compared to traditional scheduling algorithms, the findings showed a considerable improvement in makespan reduction, resource utilization, energy efficiency, and migration time. Based on the comparison analysis, H-WOA proved to be a reliable solution for real-time task scheduling in cloud computing systems, continuously outperforming baseline techniques.

Table 4.1: Simulation Configuration and Environment

Parameter	**Values**
Number of tasks	100-1000
Length of tasks	700,000
Physical host RAM	16,384 MB
Physical host storage	1,000,000
Bandwidth	10,000
Number of VMs	20
RAM of VM	512 MB
OS	Linux
Number of datacenters	2

4. Result Analysis. A full discussion of simulation and findings is covered in this section. We contrasted the ACO, GA, PSO, and CS algorithms with our H-WOA. The Cloudsim [4] simulator was used to conduct the experiment. We have considered the workload of NASA [4] and HPC2N [4] computing work logs, and we have created datasets into four categories, which we go over in the section below.

An extremely well-liked online study simulator called cloudsim [4], created at the University of Melbourne, was used to simulate this study. The simulator, which is extremely helpful for a wide range of scheduling techniques, was created entirely in Java. This simulator is installed on a laptop that has an i7 processor, 1 TB hard drive, and 16 GB of RAM. Our suggested method generates workload in two different ways. At first, we created fake datasets by distributing jobs according to uniform, normal, left-, right-, and uniform distributions.

Following that, we took into consideration parallel workload traces from HPC2N [4] and NASA [4] to correctly compute parameters and determine the effectiveness of our approach. As previously noted, datasets are constructed into four categories, which are represented by the symbols da01, da02, da03, and da04, accordingly, denoting uniform, normal, left, and right distributions that are skewed. In table 4.1 shows the simulation environment.

We have assessed makespan using workload from da01, da02, da03, da04, da05, and da06, using the configuration parameters listed in Table 4.1 above.

As the number of tasks increases, all algorithms exhibit an increase in makespan, which makes sense given that more tasks often take longer to finish. Up to about 800 jobs, the suggested H-WOA appears to outperform (lower makespan) the other algorithms; after that, it dramatically improves. Out of all the algorithms examined in this situation, the ACO algorithm has the largest makespan throughout, suggesting that it may be the least efficient. For most tasks, the performance of the GA, PSO, and CS algorithms is comparable; however, for higher task counts, CS significantly outperforms GA and PSO. In figure 4.1 shows the result of Uniform Distribution of Tasks.

It seems sense that when there are more jobs to finish, more time will be required, hence as the number of tasks increases, so does the makespan for all algorithms. The ACO algorithm increases at the steepest pace and begins as the least efficient algorithm with the longest makespan. When there are fewer tasks involved, the H-WOA has the lowest makespan; however, as the number of tasks increases, its efficiency decreases, and it becomes the second least efficient when there are 1000 tasks involved. In between the ACO and H-WOA are the GA, PSO, and CS algorithms, with CS surpassing PSO and GA at about 800 tasks. In figure 4.2 shows the result of Normal Distribution.

When comparing this "Normal Distribution" plot to the last graph you shown with a "Uniform Distribution," it indicates a different situation for task distribution. A probability distribution with more tasks concentrated around a central point and less duties as you travel away from the center may be implied by a normal distribution of tasks. awareness how the algorithms behave in real-world situations requires an awareness of how their curves differ from one another. This graph and the previous one show that the algorithms function differently under various task distributions.

Each line shows how well an algorithm performed, as determined by "Makespan (ms)," or the total amount of time needed to do a specific number of tasks. The number of tasks, which is represented by the x-axis and

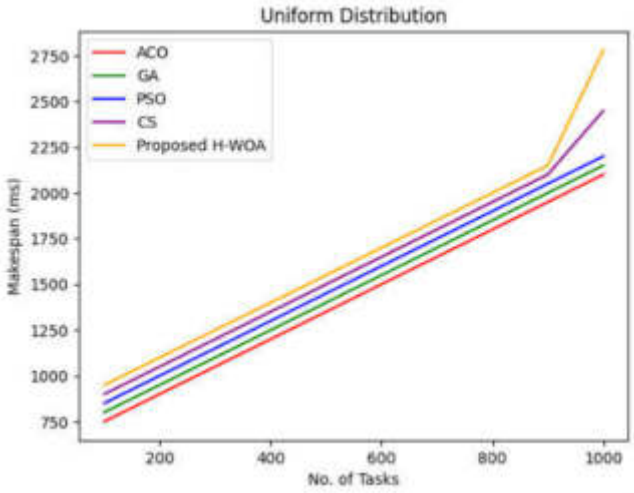


Fig. 4.1: Uniform Distribution of Tasks

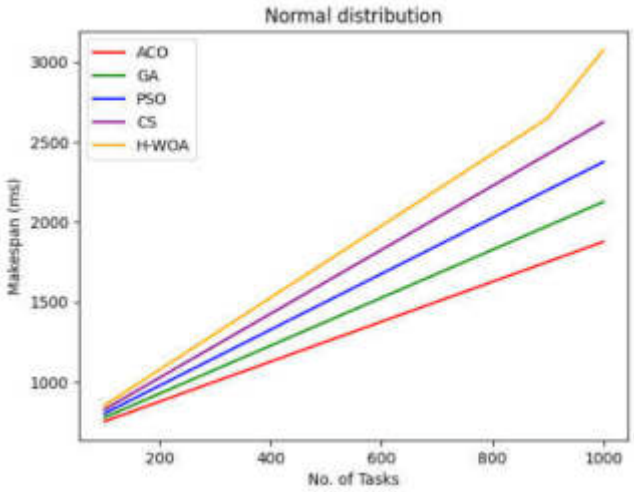


Fig. 4.2: Normal distribution of tasks

varies from 0 to 1000, is shown against the milliseconds-long makespan on the y-axis. The graph is probably used to compare how effective these algorithms are, with better performance shown by a lower makespan for a given number of tasks. As the number of tasks increases, it appears that all algorithms increase their makespan, which makes sense given that more work often requires more time. The orange-colored H-WOA method appears to outperform the other algorithms (lower makespan) at approximately 1000 tasks; nevertheless, its relative performance fluctuates along the task range. In figure 4.3 shows the result of HPC2N workload.

The "NASA Workload" graph is another line chart that bears a resemblance to the earlier ones you have presented. The makespan, or total time needed to finish a set of activities, is plotted against the number of tasks in the chart. The makespan for all algorithms grows as the number of tasks increases, indicating that larger or more complicated tasks take longer to finish. Because its line is always the lowest across the range of jobs, showing it has the shortest makespan, the H-WOA algorithm, represented by the orange line, appears to perform consistently better than the other algorithms. The GA and PSO lines exhibit a noticeable crossover

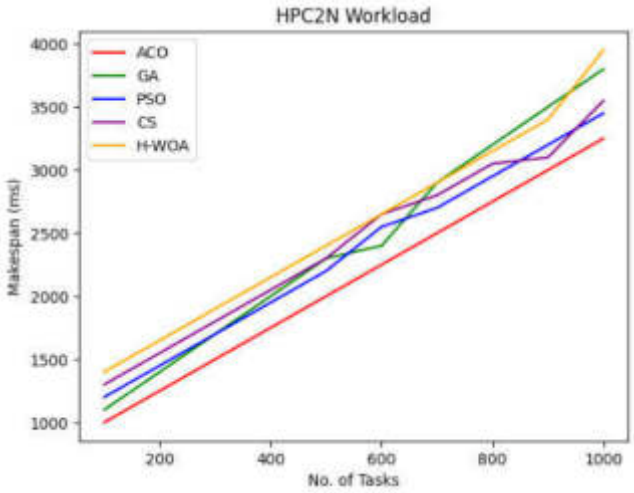


Fig. 4.3: HPC2N Workload

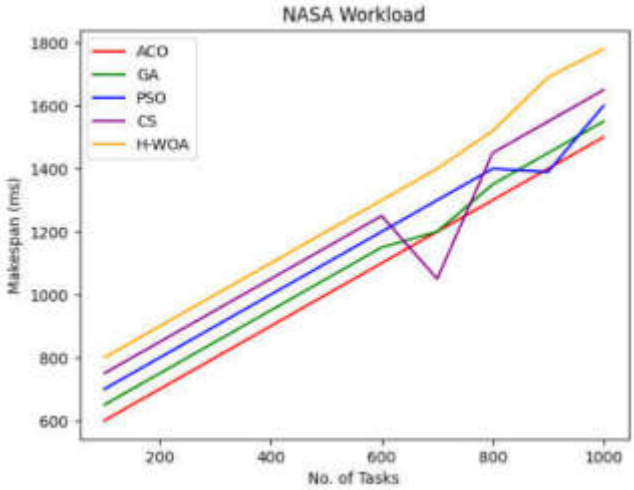


Fig. 4.4: NASA workload

point, suggesting that GA performs better than PSO for a lower number of tasks. However, PSO becomes more effective when the number of tasks goes above a specific threshold. In figure 4.4 shows the result of NASA workload.

5. Conclusion. This research has carefully examined the intricate dynamics of resource scheduling optimization tactics in cloud computing settings, showcasing important developments and approaches that tackle both established and novel problems. We found that load balancing, resource sharing, and job distribution are critically intertwined in our exploration of the fundamentals of resource scheduling. A significant development in the subject is the invention of the Hybridize Whale Optimization Algorithm (H-WOA), which provides a reliable answer to the problems associated with application scheduling. Through the integration of multi-objective optimization features, the H-WOA considers crucial elements like energy consumption and migration time, which are essential for improving cloud service performance, in addition to making makespan and resource utilization a priority. Empirical findings from tests with datasets from multiple workload repositories, includ-

ing NASA and HPC2N, and the Cloudsim toolkit confirm the superiority of the H-WOA over conventional scheduling techniques. The H-WOA guarantees optimal performance and resource efficiency by dynamically adjusting to the shifting needs of cloud resources and managing varied workloads effectively. Future research must continue to create and improve resource scheduling algorithms in order to keep up with the rapid growth of technology and the increasing complexity of cloud settings as cloud computing continues to change. As a result, the study's conclusions not only add to the body of knowledge on cloud computing optimization, but they also offer useful information that cloud service providers can use to improve service quality and operational effectiveness. In the future, scheduling techniques could potentially undergo a further revolution in resource management inside cloud computing environments with the incorporation of newer technologies like artificial intelligence and real-time data analytics.

To further enhance H-WOA's performance, future studies may investigate the integration of other meta-heuristic algorithms. To improve outcomes, combining H-WOA with Genetic Algorithms (GA) or Particle Swarm Optimization (PSO) could take advantage of each algorithm's advantages at different stages of the optimization process. By adding machine learning techniques to the H-WOA architecture, its predictive and adaptable qualities can be improved. The algorithm can make better scheduling decisions if machine learning models, for instance, are employed to forecast future workload patterns and resource availability. One area of research could be prediction models that dynamically modify H-WOA parameters depending on feedback from real-time users and historical data.

REFERENCES

[1] M. Abdullahi, M. A. Ngadi, S. I. Dishing, and S. M. Abdulhamid, *An adaptive symbiotic organisms search for constrained task scheduling in cloud computing*, Journal of ambient intelligence and humanized computing, 14 (2023), pp. 8839–8850.

[2] K. Ajmera and T. K. Tewari, *Energy-efficient virtual machine scheduling in iaas cloud environment using energy-aware green-particle swarm optimization*, International Journal of Information Technology, 15 (2023), pp. 1927–1935.

[3] D. A. Amer, G. Attiya, and I. Ziedan, *An efficient multi-objective scheduling algorithm based on spider monkey and ant colony optimization in cloud computing*, Cluster Computing, 27 (2024), pp. 1799–1819.

[4] I. Behera and S. Sobhanayak, *Task scheduling optimization in heterogeneous cloud computing environments: A hybrid ga-gwo approach*, Journal of Parallel and Distributed Computing, 183 (2024), p. 104766.

[5] A. Y. Hamed, M. K. Elnahary, F. S. Alsubaei, and H. H. El-Sayed, *Optimization task scheduling using cooperation search algorithm for heterogeneous cloud computing systems.*, Computers, Materials & Continua, 74 (2023).

[6] S. M. Kak, P. Agarwal, M. A. Alam, and F. Siddiqui, *A hybridized approach for minimizing energy in cloud computing*, Cluster Computing, 27 (2024), pp. 53–70.

[7] M. I. Khaleel, *Efficient job scheduling paradigm based on hybrid sparrow search algorithm and differential evolution optimization for heterogeneous cloud computing platforms*, Internet of Things, 22 (2023), p. 100697.

[8] M. Khenwar, A. Sisodia, S. Vishnoi, and R. Kumar, *Exploration: Cloud computing scheduling techniques*, Scandinavian Journal of Information Systems, 35 (2023), pp. 673–679.

[9] S. Mangalampalli, G. R. Karri, and M. Kumar, *Multi objective task scheduling algorithm in cloud computing using grey wolf optimization*, Cluster Computing, 26 (2023), pp. 3803–3822.

[10] R. F. Mansour, H. Alhumyani, S. A. Khalek, R. A. Saeed, and D. Gupta, *Design of cultural emperor penguin optimizer for energy-efficient resource scheduling in green cloud computing environment*, Cluster Computing, 26 (2023), pp. 575–586.

[11] H. Mikram, S. El Kafhali, and Y. Saadi, *Hepga: A new effective hybrid algorithm for scientific workflow scheduling in cloud computing environment*, Simulation Modelling Practice and Theory, 130 (2024), p. 102864.

[12] S. Pal, N. Jhanjhi, A. S. Abdulbaqi, D. Akila, F. S. Alsubaei, and A. A. Almazroi, *An intelligent task scheduling model for hybrid internet of things and cloud environment for big data applications*, Sustainability, 15 (2023), p. 5104.

[13] P. Pirozmand, H. Jalalinejad, A. A. R. Hosseinabadi, S. Mirkamali, and Y. Li, *An improved particle swarm optimization algorithm for task scheduling in cloud computing*, Journal of Ambient Intelligence and Humanized Computing, 14 (2023), pp. 4313–4327.

[14] F. S. Prity, M. H. Gazi, and K. A. Uddin, *A review of task scheduling in cloud computing based on nature-inspired optimization algorithm*, Cluster computing, 26 (2023), pp. 3037–3067.

[15] M. Raeisi-Varzaneh, O. Dakkak, A. Habbal, and B.-S. Kim, *Resource scheduling in edge computing: Architecture, taxonomy, open issues and future research directions*, IEEE Access, 11 (2023), pp. 25329–25350.

[16] Y. Wang, S. Dong, and W. Fan, *Task scheduling mechanism based on reinforcement learning in cloud computing*, Mathematics, 11 (2023), p. 3364.

[17] M. Yadav and A. Mishra, *An enhanced ordinal optimization with lower scheduling overhead based novel approach for task scheduling in cloud computing environment*, Journal of Cloud Computing, 12 (2023), p. 8.

[18] M. Zeedan, G. Attiya, and N. El-Fishawy, *Enhanced hybrid multi-objective workflow scheduling approach based artificial bee colony in cloud computing*, Computing, 105 (2023), pp. 217–247.

[19] G. Zhou, W. Tian, R. Buyya, and K. Wu, *Growable genetic algorithm with heuristic-based local search for multi-dimensional resources scheduling of cloud computing*, Applied Soft Computing, 136 (2023), p. 110027.
[20] G. Zhou, W. Tian, R. Buyya, R. Xue, and L. Song, *Deep reinforcement learning-based methods for resource scheduling in cloud computing: A review and future directions*, Artificial Intelligence Review, 57 (2024), p. 124.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: May 12, 2024
Accepted: Jun 9, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 983–992, DOI 10.12694/scpe.v26i2.3936

RESEARCH ON MODELLING ARCHITECTURAL HERITAGE OF THIRD-LINE CONSTRUCTION BASED ON HIERARCHICAL ANALYSIS AND DATA FUSION USING RAT SWARM TUNED ARTIFICIAL NEURAL NETWORK

YANLONG LIU[*], XUAN LIANG[†], RONG YU[‡], AND JIE LI[§]

Abstract. This study explores the novel use of Artificial Neural Networks (ANNs) with Rat Swarm Optimization (ANN-RSO) to model the architectural legacy of third-line construction projects. A multi-layered, combined approach to thoroughly assess and protect valuable historical buildings is constructed by the study through the application of hierarchical analysis and data fusion tools. The approach makes use of ANN-RSO power to maximize the analysis and understanding of a variety of data sets, from historical and cultural relevance to structural details and material compositions. Through the systematic division of complicated data into digestible layers, hierarchical analysis improves the neural network's power by concentrating on distinct aspects at various levels. Data fusion combines different data kinds at the same time, such as verbal descriptions, architectural plans, and photographic proof, to create a rich, consistent database that feeds into the neural network. By fine-tuning the ANN parameters, the RSO approach greatly increases the model's efficiency and accuracy in predicting and modeling architectural features. This study provides a standard for the use of modern computational techniques in the protection of cultural heritage in addition to showcasing the potential of ANN-RSO in architectural heritage modeling. The results show that these technologically advanced models can be essential resources for architectural historians and preservationists working on reconstruction projects, giving them a better understanding of the past and more precise reconstructions of historical buildings.

Key words: Architectural Heritage, Third-Line Construction, Hierarchical Analysis, Data Fusion, Rat Swarm Optimization, Artificial Neural Networks, Model Accuracy, Cultural Preservation

1. Introduction.

1.1. Importance of Protecting Cultural Heritage and Digital Preserving Technologies. Due to their historical, cultural, and educational significance, third-line constructions have to have their cultural heritage protected. These structures often denote important historical periods and capture the architectural styles, materials and techniques of their eras[6]. By keeping these buildings in active, future generations will be able to connect with their past in a tangible way and will be better able to understand and value their cultural identity. The conservation of architectural history now requires the use of digital preservation techniques. It is possible to produce accurate digital models of these historic structures using modern technology[17, 16]. These models fulfill two functions: they are essential for recording minute details that are subjected to erosion or destruction, and they act as a digital backup in case the actual structures are lost to time. These digital models are also very helpful for historical construction research and restoration[19]. Accurate restorations are made easier by their ability to undertake comprehensive assessments by architects and conservationists without compromising the quality of the original structures. These models can also simulate the results of various restoration treatments and environmental changes, which can assist decision-makers in more efficiently designing preservation plans[12, 8]. Digital models improve learning in educational settings by allowing experts and students to realistically study historical constructs. Learning becomes more accessible and engaging as a result, as these models may be shared worldwide across all physical and geographic barriers.

1.2. Previous Techniques under Preserving and its Limitations. The preservation of architectural heritage has always depended on physical restoration and traditional recording procedures, by considering the importance of heritage, these traditional techniques have significant drawbacks [3, 5]. Extensive physical

[*]School of Art, Chongqing Technology and Business University, Chongqing 400067, China (yanlongliufusion1@hotmail.com)
[†]School of Art, Chongqing Technology and Business University, Chongqing 400067, China
[‡]School of Education, Chengdu College of Arts and Sciences, 610401, China
[§]School of General Education, University for Science & Technology Sichuan, 610051, China

involvement is a common component of traditional restoration, which can be demanding and occasionally result in the loss of original materials and designs. This can be a dangerous technique because structures with substantial historical value may be completely altered. While necessary, traditional documentation techniques like physical measurements and photos only provide a limited amount of detail and do not fully express the complex nature of architectural details. Additionally, these approaches produce stable records that are not suitable to active analysis and may be challenging to update[8, 10]. They can also disappear as a result of accidents, and they are capable of decay over time. Because of these restrictions, the dynamic and detailed elements of architectural history are not always completely preserved, which makes it difficult to carry out precise restorations or carry out in-depth historical study. These approaches are also inadequate in terms of accessibility, since conservation requirements and geographical limitations may restrict physical access to sites, which in turn restricts educational opportunities and wider public involvement with cultural heritage.

1.3. Machine Learning Involvement and its advantages. The field of architectural heritage preservation can benefit greatly from machine learning, mainly because of its capacity to accurately and efficiently handle vast amounts of difficult data. The improvement of condition monitoring and predictive maintenance is one major benefit. Through the analysis of data trends over time, machine learning algorithms are able to identify possible degradation and structural difficulties [18, 15, 2]. This allows for preventive interventions to prevent damage to historical structures. Furthermore, machine learning makes it easier to create large, advanced 3D models and simulations that accurately depict architectural elements. This feature is essential for the digital reconstruction of pieces that have been damaged because it allows researchers and to see and modify precise representations of the original structures without running the danger of additional physical interference. In addition, architectural styles and elements can be automatically classified by machine learning algorithms, which simplifies the documentation process and improves information management[11]. This technology provides extensive study of historical sites from distant locations, as well as improved accessibility and interactive learning opportunities. Overall, machine learning improves the global understanding and enjoyment of architectural history in addition to increasing the efficacy and efficiency of preservation efforts.

1.4. Proposed RSO-ANN and its advantages. By considering the modern technology development and its advantages in various applications, particularly AI and machine learning, this study, presents a new novel approach called ANN-RSO, Artificial Neural Network enabled Rat Swarm Optimization algorithm. In the field of cultural heritage preservation, the RSO-ANN is an innovative strategy that combines the advantages of neural networks and optimization approaches[7]. RSO is a well-known optimization algorithm that is effective in locating the best answers to challenging issues. It was inspired by the group behavior of rats. RSO efficiently modifies the weights and biases of a network to improve an ANN's accuracy and performance. When working with the difficult data of architectural heritage, where accuracy is critical, this improvement is especially important. Because RSO-ANN combines the optimization skills of RSO with ANN's capacity to model nonlinear interactions, it is especially well-suited for producing digital models of historical constructions that are incredibly precise and accurate. Architectural features require exact details and patterns to be captured by these models in order to be analysed and restored. There are several benefits to using RSO-ANN for cultural heritage preservation. It can handle huge datasets with a variety of data kinds, increases prediction accuracy, and streamlines the modeling process overall. This eventually contributes to the ongoing appreciation and conservation of cultural heritage by improving preservation tactics, improving restoration choices, and deepening our understanding of historical architecture.

The difficulties presented by the many and intricate datasets related to architectural heritage—which range from intricate architectural drawings to transient historical texts and images—were the driving force behind this project. Conventional analytical techniques frequently fail to combine this diverse data into a logical, useful framework. Through the use of a combination of data fusion and hierarchical analysis methods, this study attempts to take advantage of the strong capabilities of ANN-RSO in order to efficiently analyze and comprehend these layered datasets. This approach intends to establish a new benchmark in the computational preservation of cultural assets, while simultaneously improving the depth and accuracy of architectural study.

2. Related Work. The combination of advanced computational tools and optimization strategies to improve the preservation and sustainable reuse of cultural material is a common theme across the previous works

Table 2.1: Previous studies Contributions in Cultural Heritage Preservation

Study	Source	Techniques Used	Improvements
[9]	Decision making for cultural heritage protection using Matlab and Python	Combined documentation protocols, advanced software programming, mathematical models for decision making	Easy to implement, expandable, compatible across different OS
[14]	Sustainable reuse of heritage buildings with Diva-grasshopper and Octopus plugin	Thermal and daylight simulation, multi-objective genetic optimization	Improved energy and daylight performance, tested and validated techniques
[1]	Priority setting in the conservation of Florence's old city center	Monitoring of degradation, optimization of construction costs, strategic planning within management plans	Harmonized with local plans, promotes cost-effective joint construction sites
[4]	Securing and repairing historical objects using FLAC2D	Structural, conservation, and architectural criteria (S-C-A method), Finite Difference Method for stability analysis	Applicable to various structures, aids in stabilizing historical sites

which is mentioned below. These studies highlight the value of data-driven decision-making, the applicability of these techniques to many types of heritage conservation projects, and the application of simulation and optimization for energy efficiency and structural stability. Additionally, they stress the significance of creative approaches and the necessity of compliance with current frameworks and systems for efficient management and preservation of cultural resources.

The current approaches for preserving architectural history frequently have trouble integrating and evaluating disparate data sources efficiently. Different processing strategies are needed for textual records, architectural plans, and photographic evidence. Many current models are unable to combine these disparate data sources into a cohesive analytical strategy. This may result in fragmented insights and the omission of important relationships.

The multi-layered complexity of data that is frequently encountered in architectural heritage projects is too much for many conventional and computational methods to handle. Methods without hierarchical analysis run the risk of overlooking important information that require more in-depth, structured data investigation to identify.

The current computational techniques may not be sufficiently scalable or computationally demanding for large-scale heritage projects with substantial datasets. Their practical application is limited by this constraint, particularly in resource-constrained conditions common to many preservation efforts. The models used to forecast and reconstruct architectural elements frequently differ significantly in terms of accuracy and precision. Errors in restoration due to modeling inaccuracies may cause irrevocable damage to priceless historical buildings.

3. Methodology.

3.1. Methodology General Outline. Several crucial elements are included in the proposed ANN-RSO technique for architectural heritage preservation in order to guarantee the accuracy and efficiency of the model. The process starts with data gathering, when several kinds of information about the architectural history are collected. Examples of this information include structural dimensions, material properties, historical context, and visual documentation. After that, this large dataset is pre-processed to standardize and normalize the data so that neural network processing may use it. After the data is prepared, the neural network's architecture is created, usually consisting of multiple layers that are arranged in a way that best captures the intricacies of the architectural data. Through the RSO process, which simulates rats' hunting behavior, this tuning is continuously improved in order to find the ideal set of parameters that reduce error and improve the model's predicted accuracy. The prepared dataset is then used to train the optimal neural network, allowing it to identify and recreate the complex patterns and features of the architectural history. Before used in real-world

Table 3.1: Dataset Features

Features	**Details**
Total Number of images	10,235
Training set size	8,188 images (80%)
Validation set size	2,047 images (20%)
Independent Test set size	1,404 images
Image Resolution	Original varied sizes, 128x128 pixels, 32x32 pixels, 64x64 pixels
Rescaling Method	Smaller dimension adjusted to 128 pixels, central region trimmed to 128x128 pixels
Dataset Availability	Publicly available on DataHub: DataHub Link
Source of Dataset Categories	Getty Art & Architecture Thesaurus (AAT)
Comparative Tests	Compared with another dataset of 5,000 images classified into 25 architectural styles from Wikimedia Commons

Fig. 3.1: Collected Image Samples

heritage preservation applications, the model is validated and tested after training to guarantee its accuracy and dependability. This approach makes use of ANN advantages in processing difficult, multi-dimensional data, and RSO guarantees that the network is running to its full potential. The end product is a reliable instrument for cultural heritage analysis and digital preservation.

3.2. Data Collection and Preprocessing. Several kinds of cultural heritage images are collected and are fed to pre-processing to make sure about its ability. The collected are normalized using preprocessing techniques to remove noise. Data collection process cover several kinds of images form the AHE dataset which is described below.

The collection of images from Architectural Heritage Elements (AHE)-Dataset is adapted from the study[13], the features of the adapted dataset are presented in Table 3.1, Figure 3.1 graphically depicts the collected images from AHE Dataset.

Careful data collection and preprocessing are required in the first phases of the ANN-RSO process since they are essential for the analysis that follows. Through the collection of several types of data, such as structural measurements, material attributes, historical context, and visual documentation, the model guarantees an extensive dataset that accurately represents the complex nature of architectural history. Standardization and normalization are two preparatory techniques that convert this heterogeneous input into a format that is ideal for neural network computation. This guarantees consistency between various forms of information and improves the quality of the data, both of which are critical for precise analysis.

3.3. Artificial Neuarl Network (ANN). Artificial Neural Networks (ANN) are parallel connectionist architectures developed to simulate the neural network of the human brain in order to solve difficult issues. ANN are composed of three layers: input, hidden, and output which is visually depicted in Figure 3.2. Neurons, which are complex mathematical processing units, make up each layer. Weights and biases connect the neurons, allowing information to move more easily throughout the network. Particularly, the research has shown that one hidden layer in the design is usually enough to handle a wide range of difficult functions.

We use a structured procedure with our ANN-RSO model for the protection of architectural heritage. The input layer receives data and transfers it to the hidden layer, where processing takes place. The hidden layer is essential because it uses an activation function and a combination of weights to process inputs.

$$H_{oj} = F_j \left(\sum_{i=1}^{r} iw_{j,i}\ x_i + hb_j \right) \tag{3.1}$$

is the computational formula for the hidden layer. The weight between input neuron i and hidden neuron j is represented by $iw_{j,i}$, the bias at the hidden layer is represented by hb_j, r is the total number of input neurons, the input data is represented by x_i, and the transfer function is represented by F. The output is computed at the output layer after the hidden layer, where the output of each neuron is given by

$$Y_k = F_k \left(\sum_{j=1}^{N} Hw_{k,j} H_{oj} + ob_k \right) \tag{3.2}$$

where $ob_k$is the bias in the output layer and $Hw_{k,j}$ represents the weight between hidden neuron j and output neuron k.

After the architecture of the ANN is established, training is done using input and output data sets that are known to provide the proper weights and biases for the network. Finding the ideal values for the weights and biases of the network is referred to as "network training." The right weights and biases for the ANN are usually found using a variety of methods. By varying the values of neural weights and biases, the unstructured optimization issue of ANN training is to reduce the global error. In order to come closer to the goal, a learning algorithm iteratively modifies the values of the network parameters for the input–output vector training data that is supplied. Usually, to carry out this update process, the error signal is back-propagated layer by layer, and the parameters are adjusted according to the error signal's magnitude. The most popular learning algorithm that has been developed is back-propagation, which has been successfully applied to represent a wide range of processes.

3.4. Rat Swarm Optimization Algorithm (RSO). A unique metaheuristic algorithm called RS) was developed after observing and attacking rats. Male and female rats live in swarms and are considered regional creatures. The rats can be very violent in many situations, which can lead to multiple animal deaths. Rats' aggressive and subsequent behaviors are mathematically modeled in this method to carry out optimization. The rat swarm optimizer begins with a set of random solutions that reflect the rat's position in the search space, just like the other population-based optimization methods. This random population is continuously calculated by an objective function and refined according to the aggressive and follower behaviors of rats. The initial positions of eligible solutions, or the positions of the rats, are chosen at random in the search space in the original RSO technique as follows

$$xi = xi\ min + rand \times (xi\ max - xi\ min),\ i = 1, 2, \ldots,\ N \tag{3.3}$$

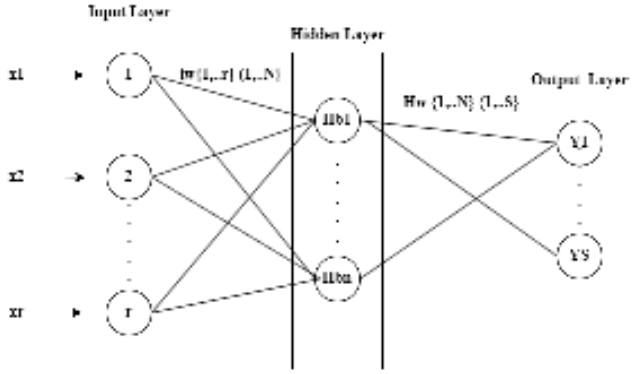


Fig. 3.2: ANN Architecture

where the lower and higher bounds of the ith variable are denoted, respectively, by $xi\ min$ and $xi\ max$. Rats typically follow the bait in a group by engaging in painful social behaviors. It is mathematically assumed that the best search agent is aware of where to place prey in order to explain the behavior of rats. As a result, the other search agents can share their positions based on the best search agent found thus far. It has been proposed that the rat's updated next position and the procedure by which it attacks with food are represented by

$$\overrightarrow{pi}\,(t+1) = |\overrightarrow{pr}\,(t) - \overrightarrow{p}\,| \tag{3.4}$$

where t is the number of iterations, $\overrightarrow{pi}\ (t)$ is the most optimal solution identified thus far, and $\overrightarrow{pi}\,(t+1)$ specifies the updated positions of ith rats.

$$\overrightarrow{p} = A \times \overrightarrow{pi}\,(t) + c \times (\overrightarrow{pr}\,(t) - \overrightarrow{pi}\,(t)) \tag{3.5}$$

where $\overrightarrow{pi}\,(t)$ defines the position of ith rats and the calculation of parameter A and C is expressed as follows

$$A = r - t \times \left(\frac{r}{tmax}\right), t = 1, 2, 3 \ldots, tmax \tag{3.6}$$

$$c = 2 \times rand \tag{3.7}$$

A random number between [1,5] and [0,2] makes up the parameter r. The parameter c is also random. The optimization process is currently in iteration t, and the maximum number of iterations is $tmax$. Equation (3.4) saves the optimal solution and updates the search agents' locations. While RSO outperforms other evolutionary algorithms such as Moth-fame Optimization (MFO), Grey Wolf Optimizer (GWO), and Gravitational Search Algorithm (GSA) in finding global optima, it may have issues finding optimal solutions when examining difficult functions. This study combines opposition-based learning (OBL) into an algorithm of RSO in order to improve the efficiency of RSO. RSO starts with a set of initial responses and works its way toward the best solution iteratively. Equation (3.3) generates random initializations without the need for prior solution knowledge

$$xi = rand\,[xi\ min,\ xi\ max] \tag{3.8}$$

Notably, when these initial solutions approach the optimal solution, better performance and faster convergence are usually attained. By using OBL, the MRSO algorithm produces an opposing number for every initial position, which could move the initial solutions closer to the global optimum.

$$X = (x1, x2, \ldots, xn) \tag{3.9}$$

is the definition of an N-dimensional vector X, where each xi lies inside a given range $[xi\ min, xi\ max]$,

$$\overrightarrow{xi} = (xi\ max + xi\ min) - xi\ ,\ i = 1, 2, \ldots,\ N \tag{3.10}$$

which states that $\overrightarrow{xi} = (xi\ max + xi\ min) - xi$, is used to determine xi, which is the opposite of xi. In RSO, the objective function $f(\cdot)$ evaluates both a random solution and its inverse (xi). $xi$is changed to $\overrightarrow{xi}$ if $F(\overrightarrow{xi}) < F(xi)$; this decision-making procedure is carried out on the first iteration to start the algorithm with more promising solutions. Like many algorithms, RSO may still converge to local conditions despite its efficiency. In order to overcome this, RSO uses another approach in which the least advantageous solution as defined by the goal function is expressed as

$$xworst = \begin{cases} rand1 \times \overrightarrow{pr}\ (t)\, if\ rand\ 3\ \leq 0.5 \\ (xi\ max + xi\ min) - xi\ if\ rand\ 3 > 0.5 \end{cases} \tag{3.11}$$

This technique promotes diversity and examines new areas inside the issue space by replacing the position of the least desirable rat with either a random new position based on the current best solution or its reverse.

Algorithm 11 Proposed RSO Algorithm

```
Define algorithm parameters: population size N, maximum iterations t_max
for i = 1 to N do
    Initialize the positions of rat agents x_i using a culturally informed distribution method
    Evaluate the opposite positions of the rats x̄_i based on cultural metrics using Equation (3.10)
    if x̄_i < f(x_i) then
        Replace x_i with x̄_i
    end if
end for
Initialize learning factors and coefficients A, c, and r
Identify and record the best search agent P_r ← best for heritage data
while t < t_max do
    for i = 1 to N do
        Update A and c adaptively for cultural context using Equations (3.6) and (3.7)
        Update the positions of rat agents x_i for cultural data using Equation (3.4)
        Evaluate the fitness of each rat agent with respect to heritage preservation
        if a rat agent reaches beyond predefined cultural constraints then
            Adjust its position
        end if
    end for
    Identify and replace the worst-performing rat agent with a new culturally feasible solution using Equation (3.11)
    Update the best search agent p⃗_r if a better candidate is found
    t = t + 1
end while
```

When RSO's optimization skills are combined with ANN's proficiency in handling intricate, multidimensional data, a potent instrument for cultural heritage preservation is produced. This collaboration makes it possible to take a dynamic, flexible approach that not only addresses the unique requirements of architectural data analysis but also pushes the envelope in terms of what is possible for digital preservation and restoration. To sum up, the ANN-RSO strategy offers a strong, dependable, and effective way to examine and conserve architectural history. For architectural historians, preservationists, and anybody else concerned in the conservation of cultural heritage, it is an essential tool due to its capacity to manage intricate datasets and continuously optimize its performance.

4. Results and Experiments.

4.1. Evaluation Criteria. The accuracy and loss presented in Figure 4.1 and 4.2 for training and validation show the effectiveness of the ANN-RSO model. The training accuracy increases steadily over the course of

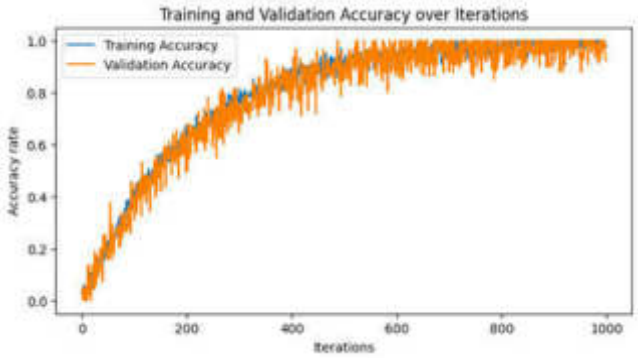


Fig. 4.1: Training and Validation Accuracy of tested samples

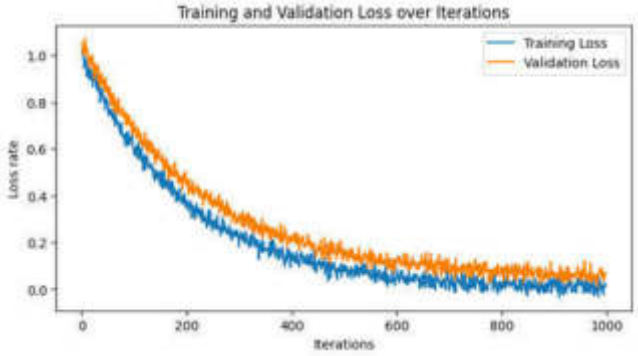


Fig. 4.2: Training and validation loss over samples

the iterations, gradually getting closer to a high degree of accuracy, showing the model's capacity to efficiently learn from the dataset. This is a positive sign that the RSO tuning mechanism and the model's design are capable of identifying the basic trends in the architectural heritage data. Concurrently, the trajectory of the validation accuracy shows a similar rising trend, while being significantly lower, indicating that the model has high adaptation skills when exposed to unknown data. When training and validation accuracy converge, the model is well-fitted and has minimal to no overfitting. The training and validation loss measurements show decreases with iterations in terms of loss, supporting the trends in accuracy. The model's effectiveness in reducing errors during the learning phase is demonstrated by a sharp drop in training loss. In the meantime, there is a noticeable decline in the validation loss in addition to a minor offset that suggests a greater error rate on the validation set. The relatively small difference between the validation and training losses supports the model's stability and its ability to function reliably in practical applications for the preservation of cultural assets. Together, these patterns show the way the ANN-RSO model learns intricate representations while preserving the reliability needed for accurate and reliable predictions in the field of architectural heritage.

An ANN-RSO model's learning trajectory through iterative training phases is shown in Figure 4.3 of Feature Matching Score. The model's feature matching performance initially begins at a lower baseline, showing that it is still in the early stages of learning and is only now starting to identify the distinctive features of the architectural heritage dataset. The capacity of the model to match features from the real heritage site with its digital reconstruction grows gradually with the number of iterations, showing a successful learning curve. The score shows some fluctuations, but overall, it is consistent with the expected training process of a deep learning model. Variability in the score can be caused by the algorithm's exploration of the solution space, the introduction of more difficult patterns in the dataset and variations in the quality of the data. In order to

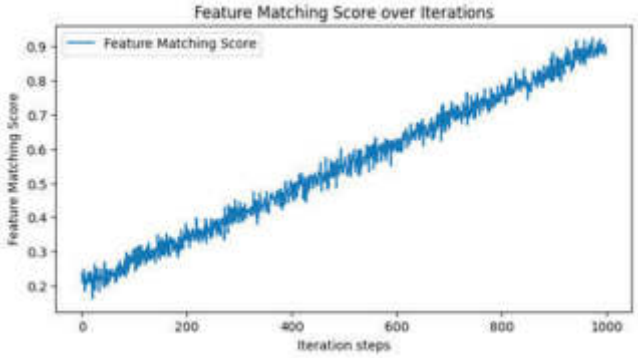


Fig. 4.3: Feature Matching Score

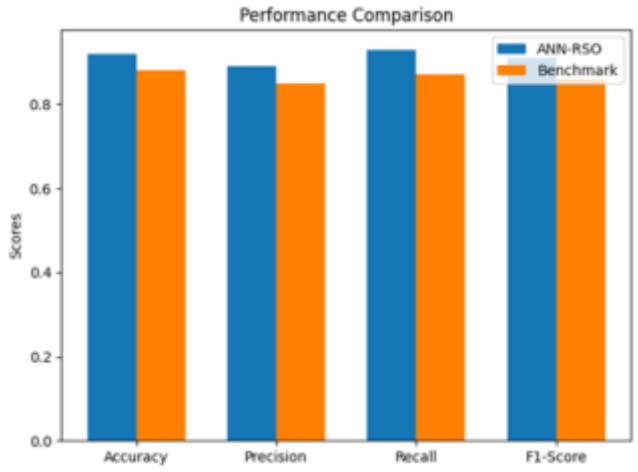


Fig. 4.4: Comparison Analysis

improve its feature matching skills, the model must balance discovering new patterns with taking advantage of established ones, as seen by the fluctuations. At the end the score shows the higher values marks the ANN-RSO effectiveness in adapting the architectural data.

Figure 4.4 shows the comparative effectiveness of the proposed ANN-RSO model, which highlights a significant improvement in performance over a number of important parameters. With a higher accuracy, the ANN-RSO model is able to identify a larger proportion of architectural historical characteristics properly, both in terms of their existence and absence. This is particularly important in the field of heritage preservation, since correct documentation and analysis depend on the precise identification of architectural features. With a precision metric that beats the benchmark, the ANN-RSO model is able to identify features with a greater percentage of true positives, demonstrating its capacity to reduce false positives and accurately recreate heritage sites. Additionally, the recall and F1-Scores also shows the highest scores when compared with the benchmarks.

5. Conclusion. With a focus on third-line construction projects, this study illustrates the creative use of ANN-RSO in the field of architectural heritage preservation. Through the incorporation of hierarchical analysis and data fusion methodologies, the study presents a strong methodological foundation for assessing and preserving historical structures. The application of ANN-RSO improves the examination of various datasets, covering topics ranging from structural compositions to cultural value, thereby promoting a more comprehensive comprehension of architectural history. Whereas data fusion combines multiple data forms textual

descriptions, architectural drawings, and photographic evidence into a unified dataset, hierarchical analysis efficiently breaks down complex data into manageable layers, allowing the neural network to concentrate on particular features at successive levels. This all-encompassing method not only uses the RSO technique to improve the neural network's parameters, increasing model accuracy and efficiency, but it also establishes a precedent for using cutting-edge computational techniques in the conservation of cultural assets. The results of this study highlight the important advanced modeling tools like ANN-RSO are for supporting preservationists and architectural historians. This represents a major advancement in the preservation and comprehension of our architectural past, as these techniques provide improved insights into historical contexts and facilitate more realistic reconstructions of ancient buildings.

REFERENCES

[1] G. Acampa, F. Battisti, and M. Grasso, *An evaluation system to optimize the management of interventions in the historic center of florence world heritage site: from building preservation to block refurbishment*, Land, 12 (2023), p. 726.

[2] K. Aravinda, K. Sandeepa, V. Sedara, A. Chamodya, T. Dharmasena, and P. K. Abeygunawardhana, *Digital preservation and noise reduction using machine learning*, in 2021 3rd International Conference on Advancements in Computing (ICAC), IEEE, 2021, pp. 181–186.

[3] E. Balliana, G. Ricci, C. Pesce, E. Zendri, et al., *Assessing the value of green conservation for cultural heritage: Positive and critical aspects of already available methodologies*, International journal of conservation science, 7 (2016), pp. 185–202.

[4] P. Chudoba, J. Przewłócki, P. Samól, and L. Zabuski, *Optimization of stabilizing systems in protection of cultural heritage: The case of the historical retaining wall in the wisłoujście fortress*, Sustainability, 12 (2020), p. 8570.

[5] S. Georgiou, D. Anglos, and C. Fotakis, *Photons in the service of our past: lasers in the preservation of cultural heritage*, Contemporary Physics, 49 (2008), pp. 1–27.

[6] W. M. W. Isa, N. A. M. Zin, F. Rosdi, and H. M. Sarim, *Digital preservation of intangible cultural heritage*, Indonesian Journal of Electrical Engineering and Computer Science, 12 (2018), pp. 1373–1379.

[7] M. Khajehzadeh, S. Keawsawasvong, and M. L. Nehdi, *Effective hybrid soft computing approach for optimum design of shallow foundations*, Sustainability, 14 (2022), p. 1847.

[8] A. Khalid, *Conservation challenges and emerging trends of digital preservation for unesco architectural heritage, pakistan*, Conservation, 2 (2021), pp. 26–37.

[9] A. Kioussi, M. Karoglou, E. Protopapadakis, A. Doulamis, E. Ksinopoulou, A. Bakolas, and A. Moropoulou, *A computationally assisted cultural heritage conservation method*, Journal of Cultural Heritage, 48 (2021), pp. 119–128.

[10] V. Kutut, N. Lepkova, and S. Źróbek, *Immovable cultural heritage usage modes: theoretical approach*, European research studies journal., 24 (2021), pp. 1136–1151.

[11] N. Lisin and S. Zapechnikov, *Methods and approaches for privacy-preserving machine learning*, in Advanced Technologies in Robotics and Intelligent Systems: Proceedings of ITR 2019, Springer, 2020, pp. 141–148.

[12] J. Liu et al., *Digitally protecting and disseminating the intangible cultural heritage in information technology era*, Mobile Information Systems, 2022 (2022).

[13] J. Llamas, P. M. Lerones, R. Medina, E. Zalama, and J. Gómez-García-Bermejo, *Classification of architectural heritage images using deep learning techniques*, Applied Sciences, 7 (2017), p. 992.

[14] M. Marzouk, M. ElSharkawy, and A. Eissa, *Optimizing thermal and visual efficiency using parametric configuration of skylights in heritage buildings*, Journal of Building Engineering, 31 (2020), p. 101385.

[15] R. U. Rahman, D. S. Tomar, G. P. Kacharea, M. M. Gawde, T. Tsundue, P. Kumar, and H. A. E. W. Khalifa, *Advanced techniques for digital evidence preservation: The power of blockchain and machine learning*, in Sustainable Security Practices Using Blockchain, Quantum and Post-Quantum Technologies for Real Time Applications, Springer, 2024, pp. 99–124.

[16] S. R. Shimraya and C. K. Ramaiah, *Issues in preservation of digital cultural heritage*, (2017).

[17] T. Trencheva and E. Zdravkova, *Intellectual property management in digitization and digital preservation of cultural heritage*, in EDULEARN19 Proceedings, IATED, 2019, pp. 6082–6087.

[18] A. Yasser, K. Clawson, C. Bowerman, M. Lévêque, et al., *Saving cultural heritage with digital make-believe: machine learning and digital techniques to the rescue*, in HCI'17: Proceedings of the 31st British Computer Society Human Computer Interaction Conference, no. 97, ACM, 2017, pp. 1–5.

[19] Y. Zhou, J. Sun, and Y. Huang, *The digital preservation of intangible cultural heritage in china: a survey*, Preservation, Digital Technology & Culture, 48 (2019), pp. 95–103.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: May 14, 2024
Accepted: Jun 9, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 993–1003, DOI 10.12694/scpe.v26i2.3938

RESEARCH ON PHYSICAL EDUCATION TEACHING IMPROVEMENT STRATEGIES AND ALGORITHMS BASED ON BIG DATA ANALYSIS

YU TIAN*

Abstract. The study "Physical Education Teaching Improvement based on Deep Learning (PETDEL) which Combines Decision Tree with Fuzzy Level Algorithm" presents a novel method for improving teaching methods in physical education (PE) by involving advanced machine learning techniques. The novel framework PETDEL, which this study creates, combines fuzzy logic and decision trees to produce a strong model that can handle the difficult structures present in PE data. By defining exact pathways based on measurable data from physical education contexts, such as student attendance, performance indicators and exam outcomes, the decision tree algorithm helps structured decision-making. Similarly, the fuzzy logic feature provides variation and flexibility to the model by taking into consideration highly individualized and subjective variables, which may be difficult to accurately quantify, such effort and student involvement levels. With this combination, PETDEL is able to process and analyze large amounts of data collected in educational environments in an efficient manner, which results in more exact forecasts of student outcomes and more customized teaching approaches. Because of its ability to combine and analyze both clear and fuzzy data, the system is very good at giving useful information that can guide changes to programs and instructional strategies. The findings of the experiments show that PETDEL delivers notable increases in the variation of teaching approaches and greatly improves the prediction accuracy of student performance, both of which have a favorable impact on the overall quality of physical education. This research proves that physical education may be more responsive to the varied requirements of students by paving the path for more advanced instructional tools and improving teaching methods through data-driven insights.

Key words: Physical Education, Deep Learning, Decision Tree, Fuzzy Logic, Big Data, Student Performance, Teaching Strategies, Machine Learning

1. Introduction.

1.1. Physical Education Teaching Improvement Strategies. In recent years, there has been a substantial evolution in the teaching and improvement techniques of PE, with a growing focus on specific methods that address the various requirements of learners [6, 20, 15]. Using technology to monitor and evaluate student performance and health parameters in real-time such as through interactive software and wearable fitness devices is one of the key methods. Personalized workout plans that take into account each student's unique talents and goals may be created with the help of this data-driven approach, which improves engagement and results. In addition, modern PE programs are going beyond traditional sports to include things like yoga, dance and adventure sports, which support student's lifetime fitness habits and support a variety of interests[13, 17]. The professional development of PE teachers through workshops and ongoing education, providing them with the newest teaching methods and technology resources, is another essential improvement option. By addressing children with special needs and diverse learning preferences, this training assists educators in implementing education approaches in an efficient manner. Through team sports and cooperative games, focus is also placed on the development of social skills and emotional resilience, which are essential elements of a well-rounded education [10]. Additionally, it is becoming more increasingly observed that student and parent feedback channels are important for improving PE planning and instructional strategies, this fosters a more dynamic and responsive physical education environment by assisting in the ongoing adaptation and improvement of the educational experience to meet changing student expectations and educational standards.

1.2. Traditional Techniques Limitations. Traditional PE systems usually has a number of problems that can reduce student involvement and impair the efficacy of instruction. PE programs have traditionally placed a high value on standardized fitness exams and competitive sports, which may not be appropriate for all students' interests and can cause dissatisfaction among students who perform poorly in traditional athletics[8].

*Sports Department, Tarim University, Aral, Xinjiang, 843300, China (yutianimprovemenr@outlook.com)

This one-size-fits-all method ignores the unique requirements and possibilities of every student, including those who are more suited for varying physical capabilities. Furthermore, the quantitative metrics that are often the focus of traditional PE assessments such as the variety of seated poses recorded or the time recorded in races may not adequately represent a student's overall physical development[2, 16]. Furthermore, these systems commonly lack the adaptability needed to involve new technologies that may improve engagement and learning through adapted and interactive experiences. Furthermore, children' complete growth suffers by the traditional programs limiting focus on physical activities that address psychological, social and emotional development components[1]. Consequently, these conventional physical education programs may not be able to inspire students to follow a lifetime of physical activity and to lead healthy lifestyles, which will ultimately affect the overall educational goals.

1.3. Machine Learning Techniques and its Advantages Now a Days. In the field of PE, machine learning techniques provide benefits that allow for a more individualized, data-driven approach to teaching and learning[11]. Teachers can identify patterns and trends that are not immediately obvious by using these strategies to examine large volumes of data from student activities and performance. Machine learning, for example, might assist in predicting which students might find it difficult to engage in a given physical activity, allowing for protective intervention to provide extra support[18]. In order to increase engagement and efficacy, it can also modify exercise programs to meet the needs of specific students based on their performance metrics, health information, and personal preferences. Moreover, machine learning algorithms have the ability to automate the process of tracking and evaluating student progress, giving teachers accurate, useful feedback that is typically necessary[22]. This feature not only saves time but also improves evaluation accuracy, resulting in assessments that are more impartial and objective. Furthermore, based on continuous data from student activities, machine learning can allow dynamic real-time modifications to training programs and teaching methods[9]. By being responsive, physical education programs can effectively support students' total physical welfare and establish a lifelong habit of physical activity and health awareness.

1.4. Proposed PETDEL and its advantages. Considering advantages and impacts of the previous techniques and technologies, the proposed study introduced the novel approach called PETDEL. Decision trees and fuzzy logic are combined in PETDEL to produce a strong model that can handle the complexity of PE data [5, 4]. By setting out distinct pathways based on quantifiable data, such as student attendance and performance markers, decision trees facilitate systematic decision-making. Fuzzy logic, on the other hand, provides flexibility by through taking along random variables like student effort and involvement levels, which can be difficult to measure accurately. Because of this special fusion ability, PETDEL can handle big datasets quickly, leading to more precise predictions of student outcomes and individualized teaching strategies. Furthermore, PETDEL is an expert in analyzing data that is both clear and fuzzy, providing detailed evaluation that can inform program modifications and instructional methods. According to experimental findings, PETDEL considerably broadens the range of instructional strategies and improves student performance prediction accuracy, which eventually raises the standard of physical education as a whole. In order to address the varied requirements of students in physical education settings, this research emphasizes the possibility for advanced instructional tools and data-driven insights.

Numerous activities in physical education provide complicated data, including qualitative elements like student effort and involvement as well as quantitative ones like exam results and attendance. This complexity is frequently too much for traditional analytical tools to handle, especially when it comes to integrating and making sense of both organized and unstructured data sources. Better Decision-Making Is Necessary: In physical education, choices about curriculum modifications and instructional strategies are crucial, but they're frequently made without the backing of solid data analysis. Structured decision-making that can make use of tangible data to optimize and inform instructional techniques is desperately needed.

2. Related Work. This paper[21] presents the iSAES-DL system, which uses deep learning to monitor students in physical education. Convolutional neural networks (CNN) are used to categorize harmful behaviors, providing insights into student learning and recommendations for improvements. The study focus how learning analytics systems in physical education can be improved by using deep learning algorithms and Internet of Things devices (IoT). This article[7], which focuses on college physical education, suggests a deep learning-

Table 2.1: Techniques and Results of Previous Studies

Source	Techniques Used	Results	Need for Improvement
[21]	Deep Learning, CNNs, IoT	Improved student monitoring, Risky action classification	Further validation in real-world settings
[7]	Deep Learning, Neural Networks	Analysis of practical training, Innovation paths	Evaluation of scalability and applicability
[19]	Deep Learning, Yolo Algorithm, U-Net	Foul recognition, Real-time coaching assistance	Evaluation in diverse race-walking scenarios
[12]	Deep Learning, Human Center of Gravity Calculation, ALSTM-LSTM Model	Improved sports teaching analysis	Validation in larger-scale studies
[14]	Adaptive Genetic Algorithm, AGA-BP Model	Improved physical education evaluation	Exploration of long-term impact and scalability

based educational system reform and an analysis of practical training features. It finds and analyzes training characteristics and innovation paths in education systems using deep learning algorithms and neural networks. This work [19] presents a deep learning based foul recognition technique to address the technical issues in race walking. To help coaches make real-time training modifications, it employs a U-Net network with attention mechanisms to identify fouls and the Yolo algorithm for preprocessing. This research[12] explores the use of deep learning techniques to improve physical education instruction in colleges and universities. Using deep learning for position estimate and human center of gravity computation, it outperforms conventional methods in sports instructional movement analysis. This study [14] uses the AGA-BP model to assess physical education quality by using an adaptive evolutionary algorithm. Through the use of an adaptive evolutionary algorithm to improve a BP neural network and entropy measurement, it reveals notable gains in students' levels of physical activity following reform.

Table 2.1 Presents the techniques of the above studies and its improvement and drawbacks

Effective integration of different types of data is a common shortcoming of traditional educational approaches. For example, disparate processing of quantitative data, like exam results, and qualitative assessments, like student involvement levels, results in fragmented insights. This fragmented approach may make it difficult for teachers to have a comprehensive grasp of the requirements of their students.

The strict, rule-based decision-making procedures that are commonly used in educational institutions today typically fail to take into consideration the subtleties and variations that exist in student data. This inflexibility may result in less-than-ideal teaching methods that don't address the unique demands of each pupil.

3. Methodology.

3.1. Method Outline. To improve the efficacy of instruction in physical education, fuzzy logic and decision trees are combined into the proposed PETDEL architecture. The decision tree component of this creative framework is the first step in creating distinct pathways based on quantifiable data, such as exam results, performance indicators and student attendance. It does this by using structured decision-making. These data segmentation trees efficiently enable focused interventions and individual instructional approaches. In addition to the decision tree, the fuzzy logic component adds flexibility and adaptability of aspects like student effort and engagement levels that are harder to measure and less quantifiable. With the use of this dual approach, PETDEL can effectively handle both confusing and cut down data, producing more accurate predictions of student outcomes. Through the processing of vast amounts of data gathered from educational settings, PETDEL is able to modify teaching strategies in real time, increasing the variety of instructional approaches and raising the accuracy of student performance predictions. Overall, the design promotes an optimal learning environment in physical education by supporting an engaging environment in which decisions are based on data and sensitive to the specific needs of each student. The proposed methodology outline is visually presented in Figure 3.1.

Careful data collection and preprocessing are required in the first phases of the ANN-RSO process since they

Fig. 3.1: Proposed PETDEL Outline

are essential for the analysis that follows. Through the collection of several types of data, such as structural measurements, material attributes, historical context, and visual documentation, the model guarantees an extensive dataset that accurately represents the complex nature of architectural history. Standardization and normalization are two preparatory techniques that convert this heterogeneous input into a format that is ideal for neural network computation. This guarantees consistency between various forms of information and improves the quality of the data, both of which are critical for precise analysis.

3.2. Decision Tree Analysis. The decision tree technique is widely used in data analysis and mining applications due to its ability to handle high-dimensional data and the fact that its creation does not need domain knowledge, making it ideal for exploratory knowledge mining. The biggest benefit of decision trees, among many other data mining and statistical analysis algorithms, is that they generate a set of rules from root to branch (or leaf), which analysts and business staff can readily comprehend. Furthermore, these general guidelines which include ready-to-use business optimization approaches and strategies even need to be somewhat sorted out. Moreover, decision tree technology is highly adaptive to data distribution and even a lack of data and is not easily hitted by extreme values.

The biggest information gain rate attribute is chosen as the splitting attribute to build branches using the fundamentals of the decision tree method. The algorithm is then called repeatedly for each branch until it is unable to divide any more branches. This can be expressed as

$$(GR\ \ (A) = \frac{G(A)}{Split\ I(A)} \tag{3.1}$$

The information gain of attribute A is represented by the formula above as $G(A)$, which may be computed using the decision tree algorithm's information gain. Split information is denoted by $Sp\ _tI(A)$. And the calculation of $Sp\ _tI(A)$ is expressed as follows

$$Sp\ _tI(A) = (D) = -\sum_{j=1}^{v} \frac{|E_r|}{D} \times log_2\left(\frac{|E_r|}{D}\right) \tag{3.2}$$

A training dataset Ci $(i\ =\ 1, 2, ...m)$ has m classes in D.If D is divided into v subsets, $\{E_1, E_2, ..E_v\}$, and A is a split attribute with v values $v_1, v_2, ..v_v$ if the number of tuples in subset E_r that belong to class Ci is $|E_r|$, then the probability pi that belongs to class Ci is $|E_r|/|D|$.

For training sample, the error between the actual and expected output is expressed as

$$err = \frac{1}{2}\sum_{k=1}^{c}\left(t_k - o_k\right)^2 \tag{3.3}$$

Here c is the number of output where t_k is the expected output and o_k is the actual output.

Let S be a set of s data samples, the class attribute contains m distinct classes Ci, and si is the number of samples in the Ci class. The following expression provides the information entropy needed to classify a given sample

$$I\left(s_1, s_2, \ldots s_m\right) = -\sum_{i=1}^{m} pi\ log_2(pi) \tag{3.4}$$

The dataset is divided into subsets according to attribute values during the process, and each attribute's information entropy is calculated using

$$I\left(s_1, s_2, \ldots s_m\right) = \sum_{i=1}^{m} -pi\ log_2(pi) \tag{3.5}$$

S is divided into v subsets $\{s_1, s_2, \ldots s_v\}$, where s_j comprises the data samples whose attribute A in set S takes a_j L value. This is the case if A is chosen as the test attribute and there are v values $\{a_1, a_2, ..a_v\}$. To partition the current sample set by attribute A, one uses the information entropy necessary, assuming that Sij is the number of samples in subset Sj that correspond to category Ci.

$$E\left(A\right) = \sum_{j=1}^{v}\frac{s_{1j} + s_{2j} + \cdots + s_{mj}}{s} \times I(s_{1j}, s_{2j}, ..s_{mj}) = -\sum_{j=i}^{v}\sum_{i=1}^{m}\frac{s_{1j} + s_{2j} + \cdots + s_{mj}}{s} \tag{3.6}$$

It calculates the information gain obtained by portioning the dataset based on the current attribute which is used to compute the gain.

$$G\left(A\right) = I\left(s_1, s_2, \ldots s_m\right) - E(A) \tag{3.7}$$

Limit theory understanding shows that when $0 \leq x \leq 1$, n goes to ∞. The previous formula's output after the third term will also get smaller as the power degree increases. The latter terms are roughly 0 when compared to the first two terms, and the final function $F(X)$ is reduced to

$$In\left(1 + X\right) = X - \frac{X^2}{2} \tag{3.8}$$

The calculation of information entropy is expressed as

$$I\left(s_1, s_2, \ldots s_m\right) = -\sum_{i=1}^{n} -pi\ log_2\ pi = -\sum_{i=1}^{n}\frac{s_i}{s}\ \log_2\ \frac{s_i}{s} \tag{3.9}$$

The above formulae is simplified into

$$I\left(s_1, s_2, \ldots s_m\right) = \sum_{i=1}^{n}\frac{s_i}{s}\ \log_2\ \left(1 + \left(\frac{s - s_i}{s}\right)\right) \tag{3.10}$$

While updating statistical data, the decision tree algorithm inserts training samples from the root node to the relevant leaf node. In the event that every sample that arrives is of the same class, the leaf node splits, and the number of decision attributes defines how many new leaf nodes are added. Through this method, decisions

are made within the PETDEL framework in an efficient and effective manner, improving physical education teaching strategies through data analysis.

Structured decision-making based on measurable data, such as exam results, performance indicators, and student attendance, is made possible by the PETDEL framework's integration of decision trees. This methodical approach facilitates the development of focused interventions and defined routes, allowing teachers to modify their lesson plans to better suit the individual requirements of their pupils. Fuzzy logic adds the flexibility needed to manage more arbitrary components, such as student effort and engagement, which are important but more difficult to gauge. This two-pronged strategy guarantees that choices are informed by data and flexible enough to take into account the unique circumstances of each student.

3.3. Fuzzy Level Algorithm.

Step 1. Start by using the appropriate vector table activities py to calculate the cognitive function $s(py)$. Based on their participation in a variety of activities, this function represents the cognitive ability of learners.

$$s\left(py\right) = \frac{vpy\ \left(1\right)}{vpy\ \ \left(1\right) + vpy\ \left(-1\right)} \tag{3.11}$$

Step 2. The usage vector S, which has values between 0 and 1 and represents the extent to which learners use various activities, can be derived using the computed cognitive function.

$$S = (s_1, s_2, .., s_n) \in (0, 1) \tag{3.12}$$

Step 3. Create a correct rate vector to represent the assessment accuracy. The values in this vector, which range from 0 to 1, show how accurate the evaluation results were, which is expressed as

$$S = \left\{\frac{2/3}{2, 1/2}, 1, 0, 1\right\} \tag{3.13}$$

Step 4. To indicate the relevant range to apply, divide area M into fuzzy subsets that correspond to student performance ratings, such as "great," "good," and "bad." The "good" fuzzy subset's membership feature Mw(m) should be computed.

$$m_w\left(m\right) = \begin{cases} \frac{1}{1+m/0.14},\ m > 0, \\ 0, else \end{cases} \tag{3.14}$$

Step 5. Using the suitable utilization vector, verify the sub-cognitive abilities' participation rating. In this stage, the relative value of various activities for cognitive growth is weighted.

$$w = \{w_1, w_2, \ldots, w_n\} \in (0, 1) \tag{3.15}$$

Step 6. Create an evaluation matrix (HM) that includes a number of assessment vectors for different types of activities. Every row in the matrix represents an evaluation vector for a set of questions, illustrating the many functions that various activities perform, the HM matrix are expressed as

$$HM = \begin{bmatrix} w11 & w12 & \cdots & w1n \\ w21 & w22 & \cdots & w2n \\ \vdots & \vdots & \vdots & \vdots \\ wp1 & wp2 & \cdots & wpn \end{bmatrix} \tag{3.16}$$

Step 7. Present a weight matrix B, in which each entry bi denotes the weight of a certain task. Make sure that the weight matrix's total element count is always one.

$$B = \{b_1, b_2, \ldots b_p\}\, and\ \sum_{y=0}^{p} b_y = 1 \tag{3.17}$$

Step 8. Make use of the weighted mean approach to provide a thorough evaluation of a vector that represents learning ability. The weights given to various activities and their contributions to overall learning outcomes are taken into account in this stage.

$$k = B \times HM = \{k_1, k_2, \ldots k_n\} \tag{3.18}$$

Step 9. Map test results, age, the effectiveness of learning, psychological conditions, and other input variables to output values that indicate learners' mastery of concepts, abilities, and application. A multimodal network that nonlinearly converts input data into output values achieved by this mapping.

$$o_t = \{grade\ k_1, k_2, \ldots k_n, w_1, w_2, \ldots w_n\} \tag{3.19}$$

$$c = \{c1, c2, c3\} \tag{3.20}$$

The process is started in Step 1 by computing a cognitive function that is dependent on engagement in different activities. This approach offers a mathematical foundation for dynamically measuring cognitive ability while taking learners' involvement variation into account. More individualized educational interventions are made possible by ensuring that cognitive evaluations are based on real, quantifiable participation metrics. The second step is to derive a usage vector that illustrates how much each activity is used by students. Teachers can better determine which activities are most engaging students and how those activities affect learning outcomes thanks to this level of detail. It makes it easier to identify highly productive activities, which allows for more focused enhancements to instructional strategies and material.

In order to provide a more accurate measure of student performance and guarantee that assessments accurately reflect the capacities of learners, this method recognizes and accounts for the variability in assessment accuracy. By segmenting the evaluation area into fuzzy subgroups that correlate to varying performance scores. The small differences in student performance are accommodated by this fuzzy classification, which permits nuanced interpretations of performance levels beyond conventional binary or rigid categorizations. Weighing the relative importance of different activities for cognitive growth is the main emphasis. This stage helps improve resource allocation and instructional emphasis by selecting activities based on their educational benefit, ensuring that the most beneficial activities are prioritized within the learning environment.

4. Results and Experiments.

4.1. Simulation Setup. The dataset used to evaluate the proposed PETDEL is adapted from the source[3], we extract only the valid data features for proceed the evaluation which is clearly presented in Table 4.1.

4.2. Evaluation Criteria. By comparing physical test results between the experimental group, which received advanced teaching techniques with a smart sports platform, and the control group, which received conventional teaching techniques, Figure 4.1 illustrates the effectiveness of the suggested PETDEL method. Figure shows that the experimental group beats the control group in every physical performance show that was tested, including strength, flexibility, and endurance. In particular, the experimental group outperforms the control group in endurance, scoring 75.74 as opposed to 65.63. Strength is another area where this pattern is evident, with the experimental group scoring 80.25 points, much higher than the control group's 70.89. This pattern is also shown in flexibility, with the experimental group scoring 78.94 points compared to the control group's 68.47. These findings highlight the benefits of the PETDEL strategy and imply that incorporating technology-enhanced and data-driven physical education techniques can greatly improve student performance. The increases in all of these variables point to a rise in physical prowess as well as possible gains in student motivation and engagement, two important aspects that make physical education programs successful in the long run.

The suggested PETDEL approach's efficacy is clearly shown by Figure 4.2 that compares the experimental and control groups' final test results. Across the three academic assessments, the experimental group consistently outperformed the control group due to the combination of innovative teaching methods and technology-based improvements into the physical education syllabus. With marks of 88.24, 91.86, and 89.21, the experimental group's scores show a consistent high performance, showing not just a gain in physical capabilities but

Table 4.1: Dataset Features

Feature	**Description**
Experimental Group	Group receiving experimental teaching, combining traditional methods with a smart sports platform
Control Group	Group receiving traditional teaching methods as per college syllabus
Experimental Teaching Hours	Total hours of experimental teaching provided to the experimental group (36 hours)
Number of Experimental Subjects	Total number of subjects in the experimental group (160 subjects)
Comparison of Physical Test Data	Evaluation of physical performance measures between experimental and control groups
Comparison of Final Test Results	Evaluation of academic outcomes between experimental and control groups
Comparison of Exercise Attitudes	Evaluation of attitudes towards physical exercise between experimental and control groups
Evaluation Criteria	Criteria divided into five dimensions: very reasonable, relatively reasonable, general, unreasonable, very unreasonable
Expert Evaluation of Questionnaire	Evaluation of teacher questionnaire by 10 experts: 6 experts found it generally reasonable, 4 experts found it more reasonable
Reliability Testing	Test-retest method used for reliability testing of the questionnaire, involving 15 physical education teachers
Correlation Coefficient	Correlation coefficient of the questionnaire, showing high reliability 0.88 and statistical significance ($P < 0.05$)

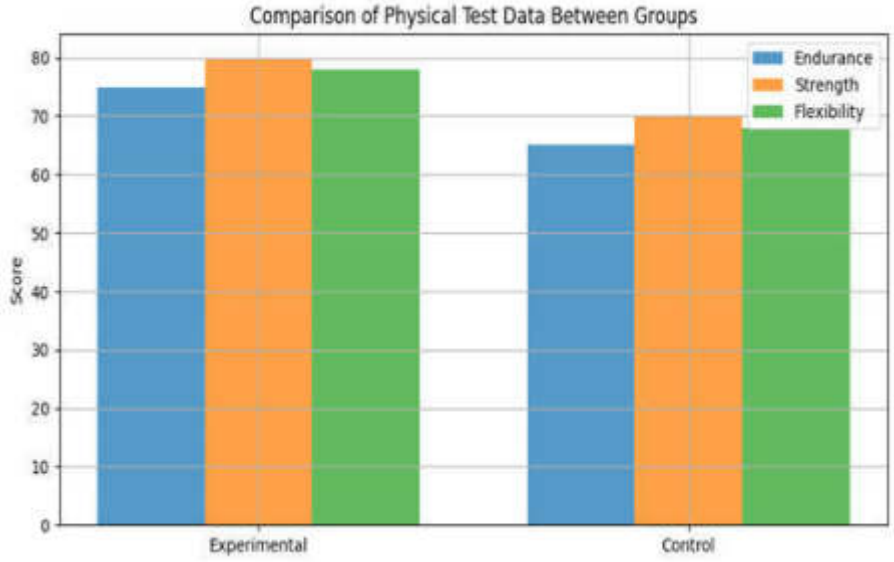


Fig. 4.1: Physical Test data Results

also increased understanding and memory of physical education ideas. The control group's ratings, which were 71.99, 75.76, and 70.48, demonstrated a more conventional and ineffective teaching approach, even though they did gradually improve. This pattern highlights the effectiveness of the PETDEL approach is in producing improved academic results, which are probably the result of more individualized and engaging teaching strategies that inspire students to perform better. The figure visual trend shows the technology and data-driven training

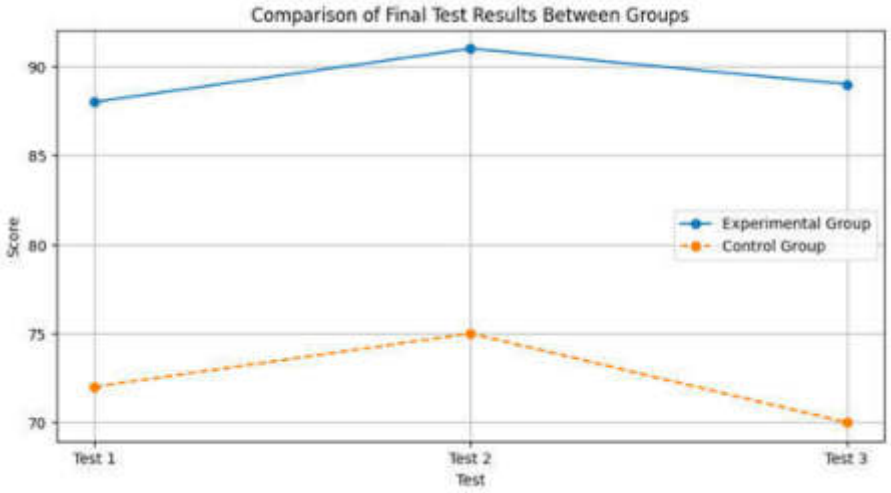


Fig. 4.2: Final test results comparison

can result in better learning experiences and outcomes, confirming the PETDEL's ability to significantly and sustainably alter physical education teaching methods.

Figure 4.3 shows the result on exercise attitudes clearly displays the efficacy of the proposed PETDEL strategy in influencing students' attitudes towards physical exercise. Regarding the figure, the experimental group (85.21%) has a significantly higher percentage of positive attitudes than the control group (65.23%), which was not showing to the innovative teaching methods that included technology and personalized strategies. This suggests that the PETDEL technique has a major impact on establishing a more favorable view towards physical education among students. Next, the experimental group had a lower percentage of neutral (10%) and negative attitudes (5%) than the control group, highlighting the novel teaching approaches' beneficial psychological and motivational impacts. The control group, on the other hand, used conventional teaching methods and had a higher incidence of neutral (25%) and negative attitudes (10%), which suggests poorer motivating impact and less engagement. This shows the PETDEL framework's integration of data-driven insights advanced teaching tools not only improves students' physical performance but also dramatically improves their attitudes toward exercise, which are critical for fostering long-term healthy behaviors and a supportive learning environment.

5. Conclusion. The proposed study is a novel approach which investigates about the physical education teaching improvement strategies based on big data. By combining the advanced machine learning algorithms, we introduce the novel approach called PETDEL a unique technique which combines decision tree and fuzzy logic to obtain the better result in PE improvement. Here the decision tree algorithm helps for structured decision making, at the same time fuzzy logic is used to provides variations and flexibility to the model. By combining these two strengths the proposed model achieves the ability of expected improvement in PE teaching and training. Regarding simulations the datasets are split in two groups experimental group and control group, by using the proposed model the experimental group achieves the expected improvements than the control group which uses the traditional methods. Overall, the proposed model sets a new way in the domain of physical education domain by using the advanced technologies to achieves the expected development not only the present but in future. Subsequent studies can concentrate on enhancing PETDEL's algorithms to increase processing speed and computational efficiency. In order to handle larger datasets more efficiently, fuzzy logic techniques and decision tree structures must be improved. Improving the decision tree forecasts' and fuzzy logic interpretations' accuracy will help to increase the system's dependability.

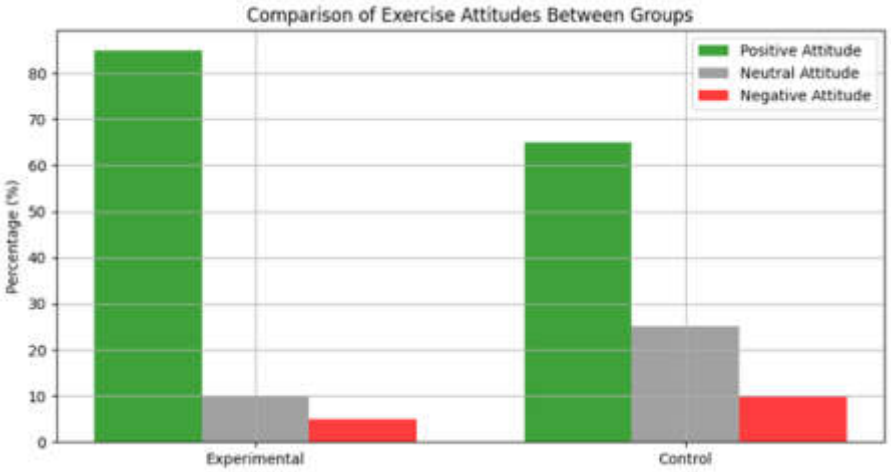


Fig. 4.3: Exercise Attitudes

REFERENCES

[1] J. Alpen, F. Dlis, L. Apriani, E. Kurniawan, D. Sofyan, et al., *Trends in scientific publication of traditional game learning models in physical education and sports in indonesia: A bibliometric analysis*, Journal Sport Area, 7 (2022), pp. 214–226.

[2] A. Azlan, N. Ismail, N. F. M. Fauzi, and R. A. Talib, *Playing traditional games vs. free-play during physical education lesson to improve physical activity: a comparison study*, Pedagogy of Physical Culture and Sports, 25 (2021), pp. 178–187.

[3] F. Cao, M. Lei, S. Lin, M. Xiang, et al., *Application of artificial intelligence-based big data ai technology in physical education reform*, Mobile Information Systems, 2022 (2022).

[4] F. Cao, M. Xiang, K. Chen, M. Lei, et al., *Intelligent physical education teaching tracking system based on multimedia data analysis and artificial intelligence*, Mobile Information Systems, 2022 (2022).

[5] Y. Che, K. Che, Q. Li, et al., *Application of decision tree in pe teaching analysis and management under the background of big data*, Computational intelligence and neuroscience, 2022 (2022).

[6] B. Dyson, *Quality physical education: A commentary on effective physical education teaching*, Research Quarterly for Exercise and Sport, 85 (2014), pp. 144–152.

[7] P. Fang et al., *Analysis of practical training characteristics and teaching system reform path of college physical education curriculum based on deep learning*, Security and Communication Networks, 2022 (2022).

[8] G. Griban, K. Prontenko, T. Yavorska, S. Bezpaliy, T. Bublei, M. Marushchak, L. Pustoliakova, V. Andreychuk, P. Tkachenko, Y. Zhukovskyi, et al., *Non-traditional means of physical training in middle school physical education classes*, International Journal of Applied Exercise Physiology, (2019), pp. 224–232.

[9] C. Hu, *Evaluation of physical education classes in colleges and universities using machine learning*, Soft Computing, 26 (2022), pp. 10765–10773.

[10] N. Indahwati, A. R. S. Tuasikal, and M. A. Al Ardha, *Developing project based learning (pbl) as a teaching strategy in physical education for preservice physical education teacher*, in 1st International Conference on Education Social Sciences and Humanities (ICESSHum 2019), Atlantis Press, 2019, pp. 490–497.

[11] L. Jinfeng and Y. Bo, *Design of evaluation system of physical education based on machine learning algorithm and svm*, Journal of Intelligent & Fuzzy Systems, 40 (2021), pp. 7423–7434.

[12] R. Liang, *Research on the path of enhancing physical education teaching in colleges and universities based on the background of deep learning*, Applied Mathematics and Nonlinear Sciences, (2023).

[13] L. J. Lieberman, C. Houston-Wilson, and M. Grenier, *Strategies for inclusion: Physical education for everyone*, Human Kinetics, 2024.

[14] L. Miao, *Current situation and reform trends of physical education teaching evaluation in the context of deep learning*, Applied Mathematics and Nonlinear Sciences.

[15] P. Morgan and V. Hansen, *Recommendations to improve primary school physical education: Classroom teachers' perspective*, The journal of educational research, 101 (2007), pp. 99–108.

[16] Y.-H. Pan, C.-H. Huang, I.-S. Lee, and W.-T. Hsu, *Comparison of learning effects of merging tpsr respectively with sport education and traditional teaching model in high school physical education classes*, Sustainability, 11 (2019), p. 2057.

[17] R. Reina, M. C. Íñiguez-Santiago, R. Ferriz-Morell, C. Martínez-Galindo, M. Cebrián-Sánchez, and A. Roldan, *The effects of modifying contact, duration, and teaching strategies in awareness interventions on attitudes towards inclusion in physical education*, European Journal of Special Needs Education, 37 (2022), pp. 57–73.

[18] C. Wang and C. Du, *Optimization of physical education and training system based on machine learning and internet of things*, Neural Computing and Applications, (2022), pp. 1–16.
[19] W. Wang, J. Jiang, et al., *A novel deep learning-enabled physical education mechanism*, Mobile Information Systems, 2022 (2022).
[20] P. Ward, I. Kim, B. Ko, and W. Li, *Effects of improving teachers' content knowledge on teaching and student learning in physical education*, Research quarterly for exercise and sport, 86 (2015), pp. 130–139.
[21] L. Zhang, S. Sengan, and P. Manivannan, *The capture and evaluation system of student actions in physical education classroom based on deep learning*, Journal of Interconnection Networks, 22 (2022), p. 2143025.
[22] Z. Zhang and Y. Zhang, *Research on effective strategies of college physical education interactive teaching based on machine learning*, Applied bionics and biomechanics, 2022 (2022).

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: May 14, 2024
Accepted: Jun 9, 2024

AIMS AND SCOPE

The area of scalable computing has matured and reached a point where new issues and trends require a professional forum. SCPE will provide this avenue by publishing original refereed papers that address the present as well as the future of parallel and distributed computing. The journal will focus on algorithm development, implementation and execution on real-world parallel architectures, and application of parallel and distributed computing to the solution of real-life problems. Of particular interest are:

Expressiveness:

- high level languages,
- object oriented techniques,
- compiler technology for parallel computing,
- implementation techniques and their efficiency.

System engineering:

- programming environments,
- debugging tools,
- software libraries.

Performance:

- performance measurement: metrics, evaluation, visualization,
- performance improvement: resource allocation and scheduling, I/O, network throughput.

Applications:

- database,
- control systems,
- embedded systems,
- fault tolerance,
- industrial and business,
- real-time,
- scientific computing,
- visualization.

Future:

- limitations of current approaches,
- engineering trends and their consequences,
- novel parallel architectures.

Taking into account the extremely rapid pace of changes in the field SCPE is committed to fast turnaround of papers and a short publication time of accepted papers.

INSTRUCTIONS FOR CONTRIBUTORS

Proposals of Special Issues should be submitted to the editor-in-chief.

The language of the journal is English. SCPE publishes three categories of papers: overview papers, research papers and short communications. Electronic submissions are preferred. Overview papers and short communications should be submitted to the editor-in-chief. Research papers should be submitted to the editor whose research interests match the subject of the paper most closely. The list of editors' research interests can be found at the journal WWW site (`http://www.scpe.org`). Each paper appropriate to the journal will be refereed by a minimum of two referees.

There is no a priori limit on the length of overview papers. Research papers should be limited to approximately 20 pages, while short communications should not exceed 5 pages. A 50–100 word abstract should be included.

Upon acceptance the authors will be asked to transfer copyright of the article to the publisher. The authors will be required to prepare the text in LaTeX 2$_\varepsilon$ using the journal document class file (based on the SIAM's `siamltex.clo` document class, available at the journal WWW site). Figures must be prepared in encapsulated PostScript and appropriately incorporated into the text. The bibliography should be formatted using the SIAM convention. Detailed instructions for the Authors are available on the SCPE WWW site at `http://www.scpe.org`.

Contributions are accepted for review on the understanding that the same work has not been published and that it is not being considered for publication elsewhere. Technical reports can be submitted. Substantially revised versions of papers published in not easily accessible conference proceedings can also be submitted. The editor-in-chief should be notified at the time of submission and the author is responsible for obtaining the necessary copyright releases for all copyrighted material.