Scalable Computing:
Practice and Experience

Scientific International Journal
for Parallel and Distributed Computing

ISSN: 1895-1767

Universitatea de Vest din Timişoara

Volume 26(5) September 2025

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

Scalable Computing: Practice and Experience

Volume 26, Number 5, September 2025

TABLE OF CONTENTS

Papers in the special issue on Deep Learning in Healthcare:

Papers in the special issue on Recent Advancements in Machine Intelligence and Smart Systems:

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 1943–1954, DOI 10.12694/scpe.v26i5.4717

ANALYSIS OF THE EFFECT OF THE APPLICATION OF MONTAGE ARTISTIC EXPRESSION OF VIDEO DATA INFORMATION TECHNOLOGY PROCESSING IN FILM AND TELEVISION CHOREOGRAPHY

XIAOXIAO LI *AND YANXIANG ZHANG †

Abstract. Film and television works blend the dynamic interplay of light and shadow within the camera, using a combination of visual composition, character portrayal, and artistic techniques to evoke emotions and convey deeper narratives to the audience. As artistic creation shifts towards perceptual and cognitive dimensions, traditional forms of expression in film choreography are being pushed to their limits. Among these, montage has remained a cornerstone of cinematic storytelling, continuously evolving throughout film history. It serves as both a sensory expression and a powerful artistic technique that harmonizes light, shadow, and narrative flow. This paper explores the artistic essence of montage, examines its application in film and television choreography, and integrates video information processing technology to enhance its expressive capabilities. Experimental results demonstrate that our approach significantly improves the effectiveness of montage in film and television editing and directing. Moreover, the incorporation of advanced video processing techniques further refines the precision and impact of montage, elevating the overall artistic and technical quality of film production.

Key words: Video processing; Application practice

1. Introduction. Montage is a French phonetic word, originally an architectural term meaning composition and assembly. With the development of multimedia technology, montage technique has been widely used in the field of film and television [1]. Montage includes three aspects: first, as a narrative way of film and television, a group of shots can form a montage sentence, and several montage sentences are combined to form a whole film; second, as a technique and means of film and television editing, montage combines various elements such as sound, acting and photography to form a complete visual image; third, as a unique way of thinking, montage thinking can be applied to The choreography of film and television works [2, 3]. Take film and television images as an example, montage technique can edit two or more shots in the film and television images together to form a new image and produce a stronger impact. The montage technique can make film and television works more vivid and transform them from mechanized sound and color art to a highly creative art. With the continuous development of film and television industry, the application of montage technique has become more and more widespread, but due to the lack of professional talents and insufficient attention to montage technique, the application of montage technique in China is not effective and its development is limited [4, 5].

In a narrow sense, montage is a creative means of expressing the plot of a film, consisting mainly of picture editing and picture composition. However, in the current film and television choreography, the value of montage is being redefined: when a story unfolds around a time line and a space line, it must be narrated around a specific structure, and the single line of film and television choreography becomes inevitable [6]. Montage provides creators and directors with more options: with the help of film editing technology, writers and directors can break the comprehensive restrictions of time and space and combine shots from different angles to form a more meaningful picture. Under montage, different shots echo and refute each other, enhancing the impact and structure of the story. For the creation of film and television choreography, montage is an artistic technique, but moreover a peculiar creative idea [7].

With the continuous development of the film and television industry, the application of montage technique is becoming more and more common, but in China. The application of montage technique is not effective

*School of Humanities and Design, Chengdu Technological University, Chengdu 611730, Sichuan, China (lanytsx1@163.com).
†Henan Polytechnic Public Art Teaching Department, Xinzheng 450046, China.

Table 2.1: Montage film and television choreography video choreography data encoding format classification.

Properties	Format
Video Encoding	H.264
Audio Encoding	PCM, AAC, MP2, MP3
Encapsulation Format	PS, TS
File Storage Format	AVI, MP4

and does not have its essence deeply [8]. On the one hand, domestic film and television choreographers do not use montage technique much. With the continuous progress of montage techniques and technologies, the montage technique is becoming more and more important in film and television choreography, but this technique originates from foreign countries and has a short history of development and insufficient experience in China, resulting in less use of the montage technique and no in-depth exploration of it [9]. At present, the montage technique used in domestic film and television choreography is mainly focused on photography and dubbing, which is more applied in the combination of shots in the early stage of film and television, and less applied in the post-production of film and television. On the other hand, many directors in China lack the sense of innovation and do not adjust the concept of film and television production in time, and still adopt the traditional shooting method when shooting works [10].

In the era of continuous development of information intelligence, various choreography systems are constantly being updated and improved, and the data generated is growing rapidly. The rapid development of the Internet has driven the production level of all industries, and at the same time, the era of intelligent networks has brought new challenges to film and television choreography. Artificial intelligence is widely used in choreography video processing systems, and by applying deep neural networks (DNN) can achieve high accuracy in choreography, but requires a lot of computing power and storage space to store video data. Based on this paper, we study the method of video information processing technology to apply montage representation technique to film and television choreography efficiently.

2. Methods.

2.1. Video data imputation and encoding conversion.

2.1.1. Automatic video data assembly. The data management scheme adopted by different choreographers' video choreography equipment varies greatly [11]. Considering that the choreographer's video choreography system has a network interface, a unified FTP service can be built through this interface to transfer the data at runtime; then the video data transfer equipment establishes a high-speed wireless network link with the server, and when it enters the access point (AP) coverage of the machine section, the video data in the transfer equipment is transferred to the server through the high-speed wireless network at one time to realize the automatic collection of different models of choreographer's video choreography data [12]. When it enters the wireless access point (AP) coverage area of the aircraft section, the video data in the transcoding equipment is transmitted to the server through the high-speed wireless network at one time, realizing the automatic collection of video programming data of different models.

2.1.2. Transcoding target format. Currently, the classification of video choreography data encoding formats used in domestic railway systems is presented in Table 2.1. These formats are selected based on factors such as compatibility with existing infrastructure, compression efficiency, real-time processing capabilities, and adaptability to different operational environments. Understanding these classifications is crucial for optimizing video storage, transmission, and playback in railway applications, ensuring both efficiency and reliability in real-world scenarios.

The target format of video transcoding needs to have good enough compatibility and universality, on the one hand, it should be able to be played by common player and support browser to decode and play directly to meet the flexible system linkage under B/S architecture; on the other hand, it needs to have a unified and standard file format and efficient target transcoding efficiency. On the other hand, it is necessary to have a uniform and standard file format and efficient transcoding efficiency.

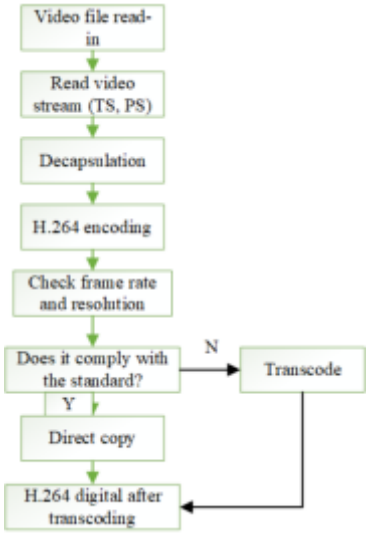


Fig. 2.1: Flow chart of video transcoding.

2.1.3. Encoding and encapsulation conversion. According to Table 2.1, there are various combinations of video, audio, encapsulation and file formats for video data. Among them, encapsulation and file formats do not involve re-encoding and decoding of video data, but only adding structured logos and container shells to the original data. In fact, the encapsulation of program stream (PS) or transport stream (TS) formats is used for all the choreographed video data and stored directly as non-standard MP4 format files. Therefore, for different video files, it is necessary to first decapsulate the video files using FFmpeg and then process the video and audio parts of the streams separately.

(1) Video transcoding. Each video coding system on the video stream part of the encoding are used a unified standard format, that is, the use of H.264 format for encoding, the main difference is reflected in the resolution and frame rate. The problem of different resolution can be stretched proportionally to ensure the same size of the video playback screen during the display playback, so it can be adjusted without the transcoding process. The frame rate difference can affect the synchronization of multi-screen fast forward, fast rewind and linked playback, which needs to be solved by interpolation compensation for low frame rate videos or frame extraction for high frame rate videos during transcoding. The flow of video transcoding is shown in Fig. 2.1, taking into account the smoothness of the image during motion and the storage space limit of the director's memory.

(2) Audio Transcoding. The transcoding of audio streams in the choreography video data is more complex and involves a variety of audio coding formats, mainly including AAC, PCM, MP2 and MP3. Especially in some early video coding products, some manufacturers use PCM_MULAW audio coding format, which makes it difficult for FFmpeg to correctly identify the sample accuracy of such audio coding format, resulting in the problem of sample bit mismatch when selecting audio decoder. Generally speaking, the current mainstream devices adopt 16bit sampling precision, but in the process of actual application, we found that some videos adopt the rare 8bit sampling precision. When transcoding, if the interface is used directly, the audio decoder will be selected automatically, and the PCM decoder with 16bit sampling precision will be used by default to decode, which will result in the problem of decreasing sampling rate, leading to sound quality degradation and more noise, and the length of the video will also be changed as a result. Therefore, when audio coding conversion is performed, if an abnormal sample rate is detected, an audio decoder needs to be specified in time to decode audio streams of different coding formats correctly. The transcoding flow of audio code streams is shown in Fig. 2.2.

After completing the encoding conversion of video and audio streams, the video stream is converted into a uniform H.264 format video encoding with a frame rate of 15 f/s; the audio stream is converted into an AAC

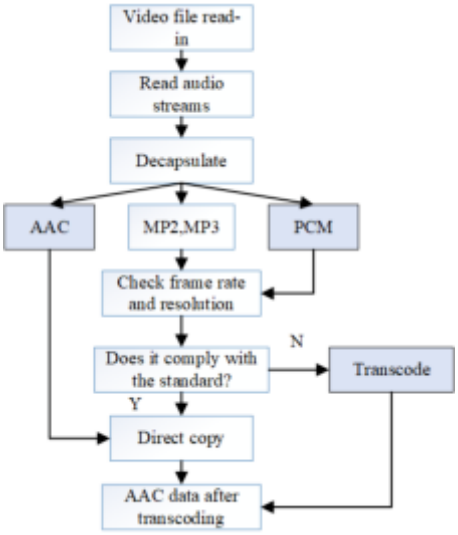


Fig. 2.2: Flow chart of transcoding audio stream.

audio encoding with a sampling rate of 8,000 Hz and a sampling precision of 16 bit, and both are encapsulated using the MP4 standard container.

2.2. Integrated Film and Video Programming Application System. The UEVideo(Usability Engineer Video) developed in this study is an integrated system based on edge computing, where video analysis can be performed in real time.In the process of real-time editing and directing of video capture, the process of data acquisition and transmission is inevitably affected by external factors. In order to reduce noise interference and improve data accuracy, Kalman algorithm is used to improve the filtering of video data. The Kalman algorithm is calculated as follows:

First, the state equation and the measurement equation are established at the key choreography nodes as:

$$\begin{cases} x_{k+1} = A_k x_k + w_k \\ y_k = H_k x_k + v_k \end{cases} \tag{2.1}$$

where K denotes a natural number that is not zero at a certain moment; x_k andy_k denote the state variables and measurement variables of the film and television choreography data signal at time k ; A_k and $H_k$denote the state transfer matrix and measurement coefficient matrix of the film and television choreography signal at time k; w_k and v_k denote the dynamic noise and measurement noise of the film and television choreography signal at time k , respectively.

Secondly, the error initialization equation is established according to Eq. 2.1 as:

$$\begin{cases} x(0 \mid 0) = E(x_0) \\ P_x(0 \mid 0) = E\left[(x_0 - x(0 \mid 0))(x_0 - x(0 \mid 0))^T\right] \end{cases} \tag{2.2}$$

where P is the covariance and E is the error. After Kalman's recursion, we get:

$$\begin{cases} x(k \mid k-1) = A_k x(k-1 \mid k-1) \\ P_x(k \mid k-1) = A_k P_x(k-1 \mid k-1) A_k^T \end{cases} \tag{2.3}$$

$$J_k = \frac{P_x(k \mid k-1)}{H_k P_x(k \mid k-1) H_k^T + R_v(k)} \tag{2.4}$$

$$Q_w = E\left[w_k w_k^T\right] R_v = E\left[v_k v_k^T\right] \tag{2.5}$$

In Eq. 2.3- Eq. 2.5, we denote the respective covariances of the dynamic noise and the measurement noise of the film and television choreography signal at time k, respectively; J denotes the filtered value of the final operation.

It can be seen from Eq. 2.1- Eq. 2.5 that the Kalman algorithm process is an iterative process, which realizes the noise reduction of the original video choreography data through repeated calculations.

2.3. Intelligent recognition of video processing. After the video data undergoes noise reduction using the Kalman algorithm, the computing unit of each edge node in the UEVideo system concurrently processes the data using a partitioned DNN model. This approach enables efficient real-time analysis, as the workload is distributed across multiple edge nodes, reducing latency and enhancing processing speed. The partitioned DNN model is specifically designed to handle various segments of the video data, allowing for parallel processing of different features such as motion detection, object recognition, and scene analysis. This technique not only improves the overall system performance but also ensures that the video data is processed in a timely and scalable manner, making it well-suited for dynamic and resource-constrained environments.

Firstly, let the three different structures of video choreography data set, N is the total number of data samples, and after correlation and fusion, we get the data sample set.

Secondly, a neural network model with three hidden layers is built to extract the feature information of the film and television choreography data, which mainly includes the encoding and decoding processes. In the encoding process, the Kth sample of film and television choreography data is encoded with the following equation:

$$a_k = f_\theta\left(d_k\right) = f\left(T_1 a_n + C_1\right) \tag{2.6}$$

$$b_k = f_\theta\left(d_k\right) = f\left(T_1 b_n + C_1\right) \tag{2.7}$$

$$c_k = f_\theta\left(d_k\right) = f\left(T_1 c_n + C_1\right) \tag{2.8}$$

where f_x denotes the activation function and θ denotes the parameter of autoencoder. Usually, the activation functions commonly used in encoding and decoding are the ReLU function and the sigmoid function [13], and in this paper, the ReLU function is used in the encoding process, while the two activation functions are mixed in the decoding process.

In the decoding process, the film and television choreography data are reconstructed to obtain the output quantity , which is closest to the original film and television choreography data in the input layer, as shown in Eq. 2.9:

$$g_k = f_\theta(x) = f\left(T_2 x + C_2\right) \tag{2.9}$$

$$x = a_k, b_k, c_k \tag{2.10}$$

where x represents any one of the three film and television choreography datasets $T_1 = T_2^T, C_1 = C_2^T$.

Finally, the parameters are set. The parameters of the DNN model are mainly composed of the number of iterations, batch processing, and learning rate, and the loss function is called by the mean square loss function MSELoss in the PyTorch database to prevent overfitting of the data [14].

After feature extraction, the model decomposes the target recognition image into several sub-windows with different parts and scales, and the classifier determines whether the sub-windows match, classifies the judgment information, and finally merges all the judgment information to complete the video image recognition of film and television choreography [15].

The feature classification of film and television choreography video data is divided into strong classifier and weak classifier, and the source value of the recognized image is obtained after data processing before classification:

$$f(x, y) = \int_0^y \int_0^x g\left(i', j'\right) d_x d_y \tag{2.11}$$

where $f(x,y)$ is the source value of the recognition map, and $g\left(i',j'\right)$ is the source image. The calculation process is to use the sum of the pixel values located in (i,j) as the integral map value at the corresponding (i',j') . The computation process can be simplified as follows,

$$f(x,y)=g\left(x',y'\right) \tag{2.12}$$

The number of target samples is determined by using the integration diagram to quickly calculate the total sum of pixels in the target obtained region, and each feature of 20×20 pixel points is used as a weak classifier, while the number of positive and negative samples is guaranteed to be the same.

The minimum feature correlation parameters of the discrete samples are saved, and the sample feature data values and the corresponding data frequency values are obtained. After processing, the highest recognition rate of all features is obtained, and the parameters corresponding to the highest recognition rate are obtained, and the preliminary recognition of the image is completed, and the results of the weak classifier are obtained for the recognition of the sample choreography .

During the processing, the strong classifier initializes the sample weights as:

$$w_j=\frac{1}{M}, j=1,2,\cdots,M \tag{2.13}$$

Calculate the weighted error rate δ of the -th strong classifier, and get:

$$\delta_i=\sum_{j=1}^{M} w_{ij}\left(y_j-h_i(x_i)^2\right) \tag{2.14}$$

The meaning of $h_i\left(x_i\right)$ function is the classification result of the i-th strong classifier for the j -th target. If the classification is $h_i\left(x_i\right)=1$, we get a positive sample; if not, $h_i\left(x_i\right)=$ - 1.

According to the optimal strong and weak classification results, the target weights of each sample are adjusted to obtain:

$$w_{i+1,j}=w_{i,j}\exp\left(-y_j h\left(x_j\right)\right) \tag{2.15}$$

According to the calculation results, the classification target sample parameters are updated, the sample data parameters are differentiated, and the deep processing of video data for film and television choreography is completed.

3. Experiments.

3.1. Basic tests. In this section, we will conduct experiments to verify the applicability and effectiveness of the UEVideo system. The experimental hardware environment is Pentium (R) CPU, 8 cores and 16G memory, the hard disk capacity of the computer is 512G, the software operating system is Windows 10, JDK5. 0, and the simulation is performed by MATLAB software system.

3.1.1. Experimental setup. In this study, mobile camera smartphones (computing elements) are used as edge nodes, and Wi Fi routers are used as components of UEVideo system access points. Considering the radio signal strength on the mobile peripheral node, the research records the location of the wireless access point, and moves the border node to a certain range.

For video processing, this experiment configures two network bandwidths, and tests the performance of systems processing three types of video resolutions, 1280 *720, 640*480 and 320 *240, respectively. One of the relevant parameters involved in the computational framework is shown in Table 3.1.

3.1.2. Performance Evaluation. In this section, we analyze the performance of choreography and recognition in terms of execution time and power consumption for processing video data of film and video choreography using cloud computing model and UEVideo system, respectively.

Table 3.1: Relevant parameters in the computational framework.

Parameters	Code	Number
Number of iterations	Nmm_epochs	200
Number of batches	Batch_size	1280
Learning rate	lerning_rate	1e-3
Adam optimizer	Torch optim. Adam	1e-5
Mean square loss function	Criterion	MSEloss

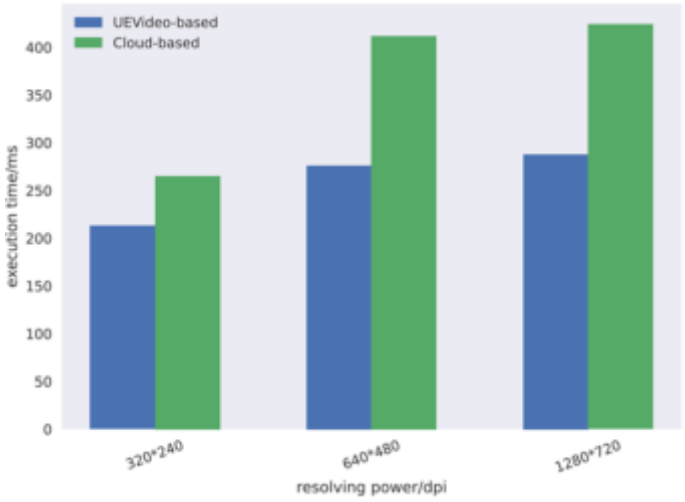


Fig. 3.1: Execution time at different resolutions (HBLD).

(1) Execution time. Fig. 3.1 and Fig. 3.2 show the implementation time of cloud computing model and UEVideo system in the same video format.

From Fig. 3.1 and Fig. 3.2, it can be seen that the execution time using the cloud computing model is always longer than the execution time of the UEVideo system in this study for deep processing of video data for film and video choreography. In the LBHD network, the implementation time of cloud based video conferencing is similar to that of UEVideo in the HBLD network, while the implementation time of UEVideo is relatively short, and almost remains unchanged with the improvement of resolution. On the contrary, the running time of cloud computing model increases gradually. In the HBLD based network, the UEVideo is implemented at a speed of about 309ms, with a resolution of 1280 * 720 and almost real-time. In contrast, the deployment time of cloud based frameworks is about 4500 milliseconds, almost 1.45 times that of UEVideo.

(2) Power consumption. For the same amount of film and TV choreography data, the power consumption of the cloud computing model and the UEVideo system for different frame sizes (1000 8000) of film and TV choreography data are compared in two different scenarios, and the results are shown in Fig. 3.3 and Fig. 3.4.

The power consumption of the UEVideo system in this study is always lower than that of the cloud computing model for deep processing of video data for film and video editing. In the LBHD network environment, for example, the difference between the cloud computing model and the UEVideo system is small at 1000 frames, but the difference increases exponentially as the number of frames increases, and at 8000 frames the power consumption of the cloud computing model is almost 2. 13 times that of the UEVideo system.

According to the experimental results, it can be concluded that the implementation time of UEVideo system in LBHD based network is almost the same. With the increase of network bandwidth, each frame will be transmitted to the server, so the running time will increase greatly. At the same time, the boundary computing system is used in the UEVideo system to make full use of computing resources in peripheral nodes

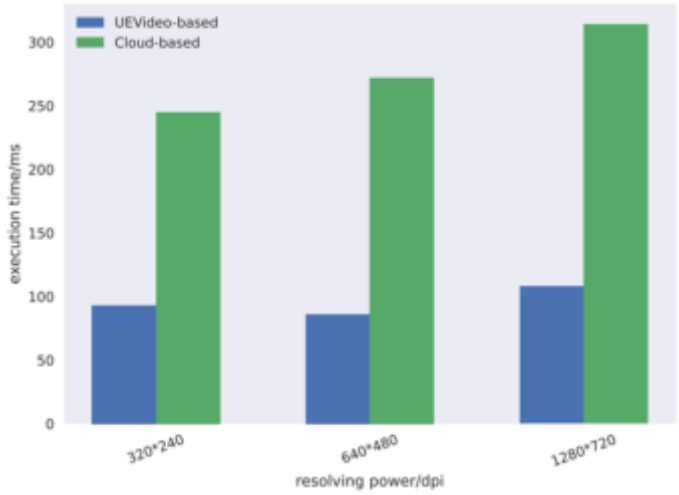


Fig. 3.2: Execution time at different resolutions (LBHD).

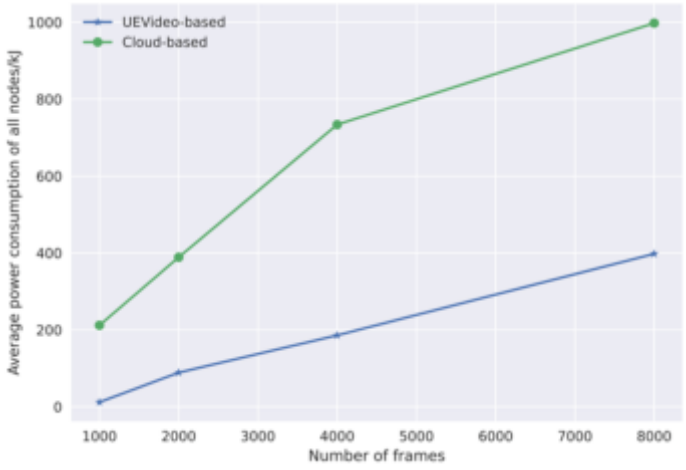


Fig. 3.3: Power consumption at different frame rates (HBLD).

to dance and dance.

In this study, detailed experiments were conducted to evaluate the enhancement effect of the proposed application of montage technique on artistic expression in movie and television choreography. The experimental data include the comparison of model performance under different experimental conditions, the performance evaluation of the model, and the statistical information of the experimental data. In the following, the experimental results will be analyzed from multiple perspectives, combining theory and practice to explore the specific application of montage artistic expression theory in film and television choreography and its validation.

3.2. Analysis of experimental data and results.

3.2.1. Comparing the results under different experimental conditions. We compared several different applications of montage techniques under different experimental conditions. For example, in the comparison of the results under traditional editing techniques and those processed in combination with video information processing techniques (e.g., deep learning networks, Gabor directional filters, etc.), it is found that

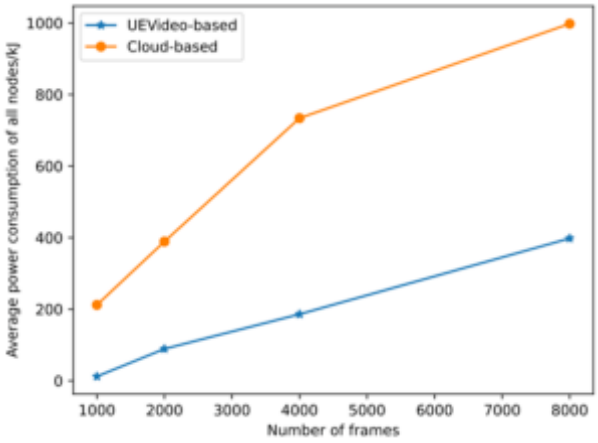


Fig. 3.4: Power consumption at different frame rates (LBHD).

Table 3.2: Comparison of experimental results.

Methodologies	Accuracy (%)	Recall rate (%)	F1 score (%)	Segmentation Effects (SSIM)	Processing time (ms)
Traditional methods	85.2	82.5	83.8	0.91	150
Using Deep Learning	90.7	89.1	89.9	0.96	120

the latter significantly improves the articulation and emotional expression between shots. Specifically, the articulation between scenes applying montage is smoother, the emotional transitions are more impactful, and in terms of the expression of details, the model using the weighted Gabor directional filter is able to better extract and highlight the directional and scale features in the picture.

3.2.2. Performance Evaluation. We conducted a comprehensive performance evaluation of the model through its evaluation metrics such as accuracy, recall, and F1 score. The results show that the montage editing after the introduction of video processing technology not only improves the precision of image segmentation, but also exhibits greater advantages in terms of the finesse of visual effects and scene adaptability. Specifically, compared with the traditional method, the montage editing based on the deep learning model shows a more significant improvement in both emotional expression and detail presentation(see Table 3.2).

3.3. Case Study. Vision is the primary means by which humans perceive and understand the world, serving as the most dominant sensory channel for receiving external information. In the realm of artistic film and television choreography, the integration of montage thinking revolutionizes visual storytelling by introducing a dynamic, layered approach to scene construction. This technique enhances the expressiveness of visual narratives, creating a richer, more immersive experience for audiences.

By seamlessly weaving together disparate images, perspectives, and temporal sequences, montage amplifies the emotional and psychological depth of a scene, allowing filmmakers to transcend traditional storytelling limitations. The strategic arrangement of shots not only heightens dramatic tension and visual rhythm but also stimulates the viewer's imagination, encouraging active engagement with the narrative.

Moreover, montage-infused film and television choreography possess a heightened aesthetic impact, leveraging light, shadow, and composition to evoke powerful emotions and reinforce thematic elements. This interplay between artistic vision and cinematic technique **broadens the scope of visual expression**, making the work more compelling, thought-provoking, and capable of capturing public attention in an era saturated with multimedia content.

Incorporating montage thinking into artistic film and television choreography ultimately results in a more

Fig. 3.5: A River Flows eastward

visually striking and emotionally resonant experience—one that not only enriches cinematic artistry but also redefines audience perception and engagement.

3.3.1. Contrast enhancement of images. Contrast refers to the difference between different visual elements in the picture. There are many contrasting factors in art film and television choreography, such as: light and dark, warm and cold, color, etc. The elements contrast with each other to set off and play a role in strengthening the contrast of the image, forming a strong visual tension and impact in the picture as a way to increase the visual activity and attention of the viewer.

The contrast thinking of montage is closely related to the collage technique. In order to make the final image present a strong contrast, the film choreographer chooses to juxtapose two or more images in the same image to create a clear ideographic image, which conveys the ideology of the creator and attracts the viewer to watch and resonate. Fig. 3.5 shows Zhang Zhongliang and her new sweetheart Lizhen, who are singing at night, are intoxicated with money and money, and their mother and his wife are suffering from hunger and cold in the war. Their sharp contrast strengthens the emotional tension of the film and lays the foundation for the outbreak of their conflicts.

3.3.2. Continuous construction of time and space. The continuous application of montage thinking in artistic film and television choreography plays a crucial role in establishing a seamless spatio-temporal flow within a visual narrative. By strategically structuring scenes, montage not only enhances the continuity of time and space but also expands the dimensional depth of storytelling, allowing for a more immersive and dynamic cinematic experience.

From the perspective of spatio-temporal construction, montage thinking transcends conventional narrative constraints, enabling filmmakers to manipulate time, space, and perspective fluidly. This approach introduces a heightened sense of rhythm and visual harmony, where the sequencing of frames and transitions between shots create a compelling aesthetic cadence. The interplay of motion, composition, and editing rhythm in montage-infused choreography evokes a poetic visual resonance, reinforcing the thematic and emotional undertones of the work. By integrating montage as a storytelling device, artistic film and television choreography achieve a unique fusion of continuity and fragmentation, enhancing narrative expressiveness while captivating audiences with a dynamic and cohesive visual language.

In the creation of artistic film and television choreography, the film and television choreographer will use the continuous thinking of montage throughout it, using the camera to shoot the action continuously, so that the action or trajectory of a single still photo is finally presented in the form of a composite image. This kind of image using multiple exposure technique has details that the human eye cannot capture quickly, and the continuous layout of the image will bring the viewer a powerful visual impact. Although the continuous thinking of montage emphasizes the continuity of reality, its application to film and television choreography means more than simply recording reality; Mitri believes that the reality expressed by images ”is never metaphysical.

Fig. 3.6 shows the changes of the street corner scene, and observes the changes of the street corner from multiple angles through editing, recording a momentary movement while at the same time expressing the continuity of visual space, thus preserving the temporal combination of the picture while showing a space.

Fig. 3.6: Change of street corner scene

Fig. 3.7: The beautiful legend of Sicily

3.3.3. Expression of diversified techniques. The creation of artistic film and television choreography cannot be separated from the expression of film and television choreography techniques. When fusing montage thinking with artistic film and television choreography, the diversity of creative techniques plays a crucial role. The film and television choreographer often uses collage or superimposition techniques to deconstruct and reorganize a large number of visual elements to create artistic film and television choreography works that are full of temporal and spatial contradictions and rich in allegory and emotion. With the purpose of reflecting montage thinking, the use of diverse film and television choreography techniques to uniquely and novelly process the images of works can enhance the visual impact, attract the viewer's attention, fully display the artist's unique ideological point of view, and leave people with infinite reverie.

Fig. 3.7 shows Ma Lianna's lens is inserted into the picture of teenagers burning ants with a magnifying glass, and the metaphor of ants echoes the end of the film. It not only implies her tragic experience, but also conveys the idea of the creator. Beautiful things will be ruthlessly destroyed in the face of a strong secular torrent.

4. Conclusion. The evolution of the information age has expanded the possibilities for applying montage techniques in diverse directions. Montage plays a crucial role in the audiovisual processing of film and television productions, aiding writers and directors in structuring their works more cohesively. Furthermore, it enhances thematic expression, highlights character development, and elevates the artistic quality of visual storytelling. Given its significant impact, film and television choreographers should actively study montage techniques, integrate them effectively into their creative processes, continuously explore new editing methods, and apply montage principles to refine visual composition and enhance the emotional appeal of their works. Despite its undeniable advantages, the application of montage in film and television still faces challenges such as technical immaturity and a lack of conceptual focus. To fully realize the artistic and creative potential of montage, filmmakers must reassess and reinterpret its essence, innovate its application, and seamlessly integrate it with both cinematic narrative and emotional depth. By leveraging technology and camera techniques in tandem with montage principles, creators can transform montage into a powerful and meaningful artistic tool that enhances storytelling and enriches the viewer's experience.

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

REFERENCES

[1] KAIMAL, Girija, *Examining associations between montage painting imagery and symptoms of depression and posttraumatic stress among active-duty military service members.* Psychology of Aesthetics, Creativity, and the Arts, 2022, 16.1: 16.
[2] LUERS, Will.*Making and breaking space: Rethinking montage in digital writing.* Rhizomes: Cultural Studies in Emerging Knowledge, 2020, 36.
[3] WIEDER, Christina.*Montages of exile. Photographic techniques and spatial dimensions in the artwork of Grete Stern.* Jewish Culture and History, 2020, 21.1: 42-65.
[4] WHITE, Michael.*Mustering Memory: George Grosz's Late Montages.* Art History, 2020, 43.4: 674-709.
[5] VON CHAMIER-WAITE, Clea. *Somatic Montage for Immersive Cinema.* Cultural Science Journal, 2021, 13.1: 113-128.
[6] ZHANG, Hongbo, *Tile selection method based on error minimization for photomosaic image creation.* Frontiers of Computer Science, 2021, 15.3: 1-8.
[7] Guo, L., & Sun, Y. *Economic Forecasting Analysis of High-Dimensional Multifractal Action Based on Financial Time Series.* International Journal for Housing Science and Its Applications, (2024). 45(1), 11-19.
[8] Altay, A., & Mirici, İ. H. *Efl Instructors' Implementations of 21st Century Skills in Their Classes.* International Journal for Housing Science and Its Applications, (2024).45(2), 37-46.
[9] Wu, Y. *Exploration of the Integration and Application of the Modern New Chinese Style Interior Design.* International Journal for Housing Science and Its Applications, (2024). 45(2), 28-36.
[10] Chen, P. *Research on Business English Approaches from the Perspective of Cross-Cultural Communication Competence.* International Journal for Housing Science and Its Applications, (2024). 45(2), 13-22.
[11] Wang, W. *Esg Performance on the Financing Cost of A-Share Listed Companies and an Empirical Study.* International Journal for Housing Science and Its Applications, (2024). 45(2), 1-7.
[12] Z. Guo, K. Yu, N. Kumar, W. Wei, S. Mumtaz and M. Guizani, "*Deep-Distributed-Learning-Based POI Recommendation Under Mobile-Edge Networks,"* in IEEE Internet of Things Journal, vol. 10, no. 1, pp. 303-317, 1 Jan.1, 2023.
[13] H. Liao et al.,*"Cloud-Edge-Device Collaborative Reliable and Communication-Efficient Digital Twin for Low-Carbon Electrical Equipment Management," in IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1715-1724, Feb. 2023
[14] GAO, Zhenping. *Application of 3D Virtual Reality Technology in Film and Television Production Under Internet Mode.* In: The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2023. p. 341-349.
[15] AJIWE, Uchechukwu Chimezie; CHUKWU-OKORONKWO, Samuel O.*Navigating Nollywood Filmmaking in Covid-19 Pandemic through New Technology.* UJAH: Unizik Journal of Arts and Humanities, 2023, 24.1: 23-44.

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/ Emerging Wireless Networks
Received: Aug 26, 2024
Accepted: Mar 1, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 1955–1963, DOI 10.12694/scpe.v26i5.4752

AFM AUTOINT INTELLIGENT RECOMMENDATION SYSTEM BASED ON ATTENTION MECHANISM AND AUTOMATIC INTERACTION MODELING

QIANG HAN *AND LIANG DONG †

Abstract. With the rapid growth of Internet data, intelligent recommendation systems are crucial for enhancing user experience and platform efficiency. Traditional algorithms struggle with high-dimensional sparse data and complex feature interactions. To address this, we propose the AFM-AutoInt model, integrating deep learning, attention mechanisms, and automatic feature interaction modeling. It utilizes embedding layers for dimensionality reduction, attention mechanisms for adaptive learning, and multi-layer self-attention for capturing high-order interactions. Experimental results show that AFM-AutoInt outperforms traditional methods in accuracy and robustness, making it a promising solution for next-generation recommendation systems.

Key words: Intelligent recommendation system; Deep learning; Attention mechanism; AFM AutoInt model; Data sparsity

1. Introduction. In recent years, the rapid development of the Internet and the surge in user behavior data have made intelligent recommendation systems a crucial technology for online platforms [1, 2]. These systems not only enable users to quickly discover relevant content but also enhance user satisfaction and platform revenue [3, 4]. However, as data volume grows and user preferences diversify, traditional recommendation algorithms face increasing difficulties in processing complex data structures and capturing high-order feature interactions. To overcome challenges such as data sparsity, high-dimensional feature interactions, and adaptive learning of feature weights, researchers have been actively exploring more efficient recommendation algorithms.

The existing recommendation algorithms can be mainly divided into three categories: content-based recommendation, collaborative filtering, and hybrid recommendation [5, 6]. Content based recommendation algorithms focus on analyzing the characteristics of items or users, while collaborative filtering methods rely on the similarity of user behavior [7, 8]. However, these two methods often struggle to achieve ideal results when faced with data sparsity and feature interaction complexity. Especially when dealing with high-dimensional sparse data, the interaction between features cannot be fully captured, resulting in poor recommendation performance.

With the rapid advancement of deep learning technology, new possibilities have emerged for enhancing recommendation systems. Deep learning models can effectively extract complex feature interaction patterns from vast amounts of data, thereby improving recommendation accuracy [9, 10]. However, these models also face challenges such as overfitting, limited interpretability, and prolonged training times. In particular, when handling high-dimensional sparse data, designing models that can effectively capture and leverage feature interactions has become a critical research focus in both academia and industry. To address these challenges, this paper proposes an intelligent recommendation system that integrates deep learning, attention networks, and clustering injection algorithms. The primary contribution lies in the fusion of attention mechanisms with automatic interaction modeling. The suggested method successfully gets over the drawbacks of conventional recommendation systems in learning high-order feature interactions by implementing an adaptive learning mechanism. The two main components of the AFM AutoInt model are AutoInt (Automatic Feature Interaction) and AFM (Attention Factorization Machine). By using an attention mechanism to provide weights to second-order feature interactions, the AFM module improves recommendation accuracy by allowing the model to automatically learn the importance of various feature combinations. Meanwhile, the AutoInt module further refines high-order feature interactions through a multi-layer self-attention mechanism, significantly improving the model's representation capability and overall performance [11].

*Qiongtai Normal University, Haikou, Hainan 570100, China.
†Qiongtai Normal University, Haikou, Hainan 570100, China (151439438390163.com).

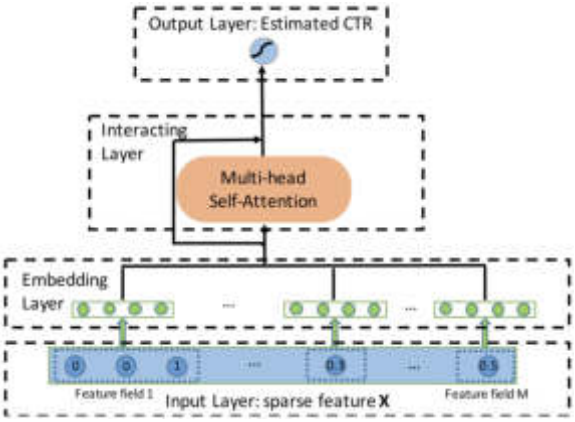


Fig. 2.1: AFM AutoInt model structure diagram.

Compared to existing recommendation algorithms, the AFM AutoInt model has significant advantages in handling data sparsity, high-dimensional feature interaction, and feature weight adaptive learning [12]. Firstly, the model proposed in this article reduces data dimensionality and improves the convergence speed and prediction accuracy of the model by embedding sparse data.Secondly, the AFM module utilizes attention mechanism to weight the second-order feature interactions, enabling the model to effectively distinguish the importance of different feature combinations for recommendation results and reduce the noise caused by invalid feature interactions [13]. Finally, the AutoInt module enhances the predictive ability of the model by introducing a multi-layer self attention mechanism to further capture high-order feature interaction information.

Although the AFM AutoInt model proposed in this article has demonstrated superior performance in experiments, there are still some challenges. For example, the complexity of the model is high, the training time is long, and further optimization may be required when dealing with large-scale data. In addition, the interpretability of the model is weak, and future research can consider combining interpretability techniques to make the recommendation results more transparent and understandable [14].

In summary, this article proposes a new recommendation algorithm by introducing deep learning and attention mechanisms, and verifies its superiority on different datasets through experiments. This study not only provides new ideas for solving the problems of data sparsity and feature interaction in recommendation systems, but also lays the foundation for further improving the performance of recommendation systems. Future research can further optimize the efficiency and interpretability of the model based on existing work, providing more effective solutions for practical applications.

2. AFM AutoInt recommendation algorithm. The AFM AutoInt model proposed in this article is shown in Fig.2.1. The overall structure of the AFM AutoInt model consists of AFM module and AutoInt module, with sparse feature data input layer, embedding layer, model layer, and output layer from bottom to top.

1. AFM (Attention-based Factorization Machine). AFM is mainly used to learn the interactions between features, while introducing an attention mechanism to give different importance to different feature interactions. Its core structure is as follows: Input layer: the input is a feature vector $x \in R^d$, where ddd is the feature dimension, and each feature is mapped to a low-dimensional space through the embedding layer. Second-order feature interaction layer: the AFM uses a factorization machine (FM) to model second-order feature interactions, calculated as follows:

$$FM(x) = \sum_{i=1}^{d} \sum_{j=i+1}^{d} (v_i \cdot v_j) x_i x_j \tag{2.1}$$

where v_i, v_j are the embedding vectors of features i and j, and, x_i, x_j are the corresponding feature values.

Attention mechanism: weights are assigned to different feature interactions, and weighted summation is computed by attention weights:

$$h = \sum_{i=1}^{d} \sum_{j=i+1}^{d} a_{ij} \left(v_i \cdot v_j\right) x_i x_j \tag{2.2}$$

where the attention weights a_{ij} are computed by an MLP network and normalized by softmax.

Output Layer: outputs the final prediction after fully connected layer and sigmoid activation function.

2. AutoInt (Automatic Feature Interaction). AutoInt mainly learns higher-order feature interactions automatically through the Self-Attention mechanism, avoiding the tedious process of manually designing feature interactions. Its network structure is as follows:

Input layer: as in the AFM, the input is the feature embedding vector matrix $X \in R^{d \times k}$, where d is the number of features and k is the embedding dimension. Multi-Head Self-Attention (MHSA): the multi-head self-attention mechanism in the Transformer structure is used to compute the importance of each feature with respect to other features:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = soft\max\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{2.3}$$

where are obtained from the input feature embedding mapping and d_k is the scaling factor.

Feature Interaction Layer: multi-layer self-attention captures higher-order interaction information and ultimately optimizes the information flow through Residual Connection and Layer Normalization.

Output layer: a fully connected layer is performed on the learned feature interaction results to compute the final predicted values.

3. Model implementation and optimization. *Loss function*: Binary Cross-Entropy (BCE) is usually used for optimization:

$$L = -\sum_{i=1}^{N} \left(y_i \log\left(\hat{y}_i\right) + \left(1 - y_i\right) \log\left(1 - \hat{y}_i\right)\right) \tag{2.4}$$

Optimization algorithms: Use the Adam optimizer with a learning rate decay strategy (e.g. Warmup + Cosine Decay) to improve convergence speed.

Regularization: Dropout and L2 regularization are used to prevent overfitting.

2.1. Sparse Feature Data Input Layer. Data sparsity is a challenge faced by recommendation systems. In this paper, the user attribute information and movie attribute information in the dataset moveielens-1M are merged based on user IDs and movie IDs to obtain N samples S, where $S = \{s_1, s_2, \cdots s_i, \cdots s_N\}$, where N is a positive integer and s_i represents the i-th user viewing record. In each user viewing record, there is only one active user, for example, s_i is the viewing record of Dirty Dancing movie by user ID 2896. s_i has m features, including both user and movie features, and one hot encoding is used to process each feature [15].

2.2. Embedding Layer. After one-hot encoding, each user and movie is represented as a unique high-dimensional sparse vector. However, such representations pose challenges, including slow convergence, excessive computational complexity, and an overwhelming number of model parameters, which makes direct input into the model impractical. To address these issues, an embedding layer is introduced to map high-dimensional sparse vectors into lower-dimensional dense representations [16]. This transformation not only reduces the model's complexity and computational burden but also preserves meaningful semantic relationships between features, thereby enhancing the model's learning ability and improving recommendation performance.

Before inputting the data into the model for training, these encoded features are divided into m different feature domains (m is a positive integer), and then matrix mapping is used to transform the features into dense and appropriately long vectors on each feature domain, thereby alleviating the problem of data sparsity. The embedding layer, in simple terms, is an initialized matrix, and its mapping process is actually a matrix multiplication. The output of the embedding layer of the jth feature domain is shown in Eq. 2.5:

$$e_j = x_j v_j \tag{2.5}$$

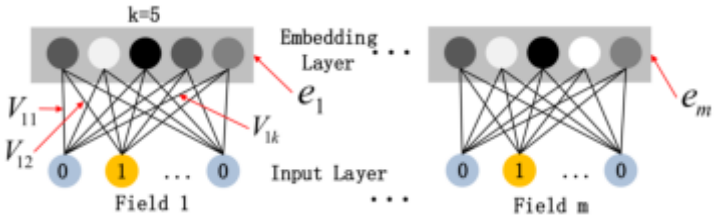


Fig. 2.2: Structure of Embedded Layer.

Among them, x_j is the vector in the jth feature field, and $j \in \{1, 2, \cdots, m\}, v_j$ is the embedding matrix corresponding to feature field j. Simply put, as shown in Eq. 2.6 and Eq. 2.7, the left part of the equal sign is the one hot encoded sparse vector multiplied by the initialized embedding layer matrix. The initialized matrix is updated after each training, resulting in continuously updated results.

$$[01] \times \begin{bmatrix} 3174 \\ 2843 \end{bmatrix} = [2843] \tag{2.6}$$

$$[0100000] \times \begin{bmatrix} 7529 \\ 1748 \\ 3913 \\ 1576 \\ 2395 \\ 6473 \\ 8912 \end{bmatrix} = [1748] \tag{2.7}$$

As illustrated in Fig. 2.2, the sub-network structure from the sparse feature input layer to the embedding layer is clearly depicted. At this stage, it can be observed that while the vector representations of different input feature domains vary, the neurons in their respective embedding layers maintain a consistent dimensionality of k . This consistency arises from the fact that the embedding matrix's dimensions are determined by the product of the feature domain's dimensionality and the embedding size k . Notably, for the same feature domain, a shared embedding matrix is utilized, ensuring parameter efficiency and facilitating effective learning of feature representations. By transforming sparse high-dimensional data into dense lower-dimensional vectors, the embedding layer not only reduces computational complexity but also preserves semantic relationships, thereby enhancing the model's ability to capture intricate feature interactions.

However, features may have multiple values, so they need to be calculated separately. Multiple values are encoded separately using one hot encoding, and then input into the embedding layer. The average value of the corresponding feature embedding vector element is taken. Taking movie viewing prediction as an example, the type of a movie may have multiple values. The movie Titanic is of drama, romance, and adventure genre, and three values need to be encoded separately using one hot encoding, and then embedded. The corresponding elements of the three values are added and divided by 3, as shown in Eq. 2.8:

$$e_i = \frac{1}{q}\left(x_i^{(1)} v_i + x_i^{(2)} v_i + \cdots x_i^{(q)} v_i\right) \tag{2.8}$$

Among them, q is the number of median values in the multi valued feature domain, $q \in \{1, 2, \cdots\}$.

2.3. Model Layer.

2.3.1. AFM Model. The FM model improves the performance of recommendation models by combining individual features and introducing cross term features through pairwise combinations of features. This article

improves its formula as shown in Eq. 2.9:

$$\hat{y}_{FM} = w_0 + \sum_{i=1}^{m} w_i^T e_i + p^T \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} e_i \odot e_j \tag{2.9}$$

Among them, w_0 is a global offset, w_i^T is the weight vector of different feature domains i, m represents the number of feature domains, $\odot$ is the Hadamard product, representing the multiplication of vector corresponding positions, p is the parameter vector, used to represent second-order combined features as a scalar.

The left part of Figure 2.1 is the AFM model, which is an improvement based on Eq. 2.9. Compared to other methods, FM can more effectively capture second-order feature interactions on sparse datasets. However, when the FM model outputs, the weights of second-order combined features are all 1, making it difficult to adaptively learn weights. Some second-order combined features are meaningless, which can bring unnecessary noise to the model and affect user behavior judgment.For example, girls prefer to watch romantic movies, while boys prefer sports movies. Therefore, combination features such as "girls" and "romance", "boys" and "sports" have a positive impact on movie recommendation, while combination features such as "girls" and "sports", "boys" and "romance" do not have a significant positive impact on movie recommendation and may also bring noise to the data. Therefore, this article introduces an attention mechanism at the output of FM second-order combination features to form a new model AFM, which enables the model to adaptively learn weights and improve its performance.

The attention mechanism model has two inputs, one is the target movie vector, and the other is the second-order combination feature output by the FM model. The purpose of using attention mechanism here is to discover which of all second-order combination features is more helpful for predicting the target movie rating.This article uses a scaled inner product attention mechanism model to solve the problem. Firstly, the correlation between the two is calculated through inner product operation. To prevent excessively long input vectors from causing the inner product to be too large, it is scaled and then normalized to obtain the weight of the second-order combined feature, which is a_{ij} in Eq. 2.10.

$$a_{ij} = softmax\left(\frac{e_i \odot e_j \cdot e_M}{\sqrt{d_k}}\right) \tag{2.10}$$

By using a soft attention mechanism to calculate the weighted distribution of second-order combined features, where each second-order combined feature is weighted according to its own weight,C_{ij} is obtained as shown in Eq. 2.11.

$$C_{ij} = a_{ij} \cdot e_i \odot e_j \tag{2.11}$$

The AFM model is shown in Eq. 2.12:

$$\hat{y}_{AFM} = w_0 + \sum_{i=1}^{m} w_i^T e_i + p^T \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} C_{ij} \tag{2.12}$$

2.4. Output Layer. The **AFM-AutoInt** model builds upon enhancements to both the **traditional AFM** and **AutoInt** models, making it highly adaptable for prediction tasks across diverse application scenarios. In this study, we evaluate its performance using the **Movielens-1M** and **Douban movie** datasets, focusing on predicting **users' movie ratings**, a classic **regression task** in recommendation systems. To ensure accurate predictions of continuous rating values, we employ a **linear activation function** in the output layer. The model's effectiveness is assessed through a **comparative analysis** against benchmark models, measuring key performance metrics such as **Root Mean Square Error (RMSE)** and **Mean Absolute Error (MAE)**. Additionally, we explore the impact of different hyperparameter settings, training strategies, and feature interaction mechanisms to **optimize predictive accuracy and model robustness**. Future work will extend the model's applicability to **multi-modal data** (e.g., incorporating text and image features), explore **real-time recommendation scenarios**, and enhance model interpretability for better user trust and system transparency.

Table 3.1: Experimental Data Field Information.

Field Name	Field information
User ID	User ID, numerical code 1-6060
Gender	Gender, M/F (M is male, F is female)
Age	User age, 7 stages (1, 18, 25, 35, 45, 50, 55)
Occupation	20 professions, numbered 0-20
Movie ID	Movie ID, numerical code 0-201-3590
Genres	Movie genres, 18 types (Action, Adventure,...)
Rating	User ratings for movies, ranging from 1-5

The output of the AFM AutoInt model is shown in Eq. 2.13.

$$\hat{y} = \mu_1 \hat{y}_{AFM} + \mu_2 \hat{y}_{AutoInt} \tag{2.13}$$

Among them, μ_1 and μ_2 are the weight coefficients of AFM and AutoInt, respectively. The values of $\mu_1 + \mu_2 = 1$ and μ_2 can be used to determine the importance of low order combined features and high-order combined features.

3. Experimental results.

3.1. Experimental data. The datasets used in this experiment are Movielens-1M dataset and Douban movie dataset. The Movielens-1M dataset not only contains user attribute information and movie data information, but also over one million rating information from 6060 users for 3888 movies. The Douban movie dataset contains 29030 valid movie data, including movie data information and rating information.

3.2. Data Preprocessing. This article takes the Movielens-1M dataset as an example, merges the user attribute information and movie attribute information of the Movielens-1M dataset based on user ID and item ID. The merged data fields are shown in Table 3.1. Then, using the five fold cross validation method, the dataset is randomly divided into two categories: training set and testing set. The training set selects 80% of the data, and the testing set selects the remaining 20% of the data. Then, the features of the training set and the testing set are encoded one hot separately.

3.3. Evaluation indicators. The experiment in this article selects four evaluation indicators, namely mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and R^2 (R-Square) coefficient of determination, to evaluate the effectiveness of predictive scoring. MSE is used to measure the degree of dispersion of a set of data itself, while RMSE is the arithmetic square root of MSE. Compared with MSE, RMSE is more sensitive to dimensionality. MAE is used to measure the degree of deviation between predicted scores and actual scores. The smaller the values of MSE, RMSE, and MAE, the smaller the model prediction error and the better the prediction result. The larger the R^2 (R-Square) coefficient of determination, the better the model prediction result.

3.4. Experimental Comparison Method. In order to demonstrate the value of this study and make the experimental results more informative and persuasive, the following five comparative algorithms were adopted in the experiment:

(1) NFM: Adding a deep neural network to the second-order feature interaction layer and introducing a feature cross pooling layer to "add up" the embedding vectors of different feature domains. High order feature interactions can be captured through the nonlinearity of the neural network, combining the modeling ability of FM for low-level feature interactions with the learning ability of DNN for high-order feature interactions.

(2) AFM: An extension of FM that uses attention mechanism to assign different weight indices to each interaction feature vector to distinguish the different importance of second-order combined features.

(3) DeepFM: combines traditional second-order factorization machines with feedforward neural networks, sharing the same sparse data input and embedding layers to extract low - and high-order features.

Table 3.2: Results of MSE and MAE under Different λ Values.

λ	MSE	MAE
0.00001	0.8020	0.7290
0.0001	0.7706	0.6966
0.001	0.7661	0.6880
0.01	0.7660	0.6863
0.1	0.7711	0.6931

Table 3.3: Comparison of evaluation indicators for each model on the test set.

Model	MSE	RMSE	MAE	R^2
NFM	0.7920	0.8902	0.7007	0.3677
AFM	0.8168	0.9035	0.7138	0.3482
Deep FM	0.7655	0.8746	0.6911	0.3888
xDeep FM	0.7672	0.8757	0.6838	0.3878
Auto Int	0.7811	0.8838	0.6977	0.3760
AFM-AutoInt	0.7603	0.8719	0.6746	0.3933

(4) xDeepFM: A new method for explicitly crossing high-order features based on vector wise pattern is proposed, which can construct finite order crossing features.
(5) AutoInt: Maps the original sparse high-dimensional feature vectors to a low dimensional space while modeling high-order feature interactions.

3.5. Results. In order to verify the effectiveness of the algorithm proposed in this article, several main parameters in the algorithm model, including regularization parameter λ, embedding layer dimension k, weight coefficientsμ_1 and $\mu_2$of AFM and AutoInt, were carefully analyzed. Multiple comparative experiments were conducted to comprehensively analyze and compare the performance of each algorithm based on four evaluation indicators: MSE, RMSE, MAE, and R^2 (R-Square) determination coefficient.

3.5.1. Regularization parameter λ. Different regularization parameters λ can also have a certain impact on the overall performance of our model. Table 3.2 lists the MSE and MAE results obtained by the AFM AutoInt model at different λ values, with an embedding layer dimension of 4.

From Table 3.2, it can be seen that as the value of λ continues to increase, the results of MSE and MAE show a continuous downward trend between λ 0.00001-0.01. At λ=0.1, the results of MSE and MAE increase again, and both MSE and MAE reach their lowest values at λ=0.01. Therefore, the model performs best when λ=0.01 is chosen.

If the regularization parameter λ value is too small, it will cause overfitting during the training process of the model. When the λ value gradually increases beyond a threshold, underfitting will occur, resulting in the loss of too many features. Therefore, when λ=0.01, it can prevent overfitting and make the model more robust during training.

3.5.2. Iteration times. After selecting the regularization parameter λ and embedding layer dimension, Fig.3.1 shows the changes in MSE and MAE of the algorithm test set with increasing iteration times.

From Fig.3.1, it can be seen that the MSE and MAE of the algorithm converge to their minimum values after 15 iterations. At this point, the model predicts that the user's rating error for the movie is minimized.

3.5.3. Comparative experimental results. Table 3.3 shows the comparative results of various algorithms.

According to Table 3.3, the AFM-AutoInt model outperforms both the AFM and AutoInt models individually, reducing prediction errors by 5.65% and 2.12%, respectively. This demonstrates that a hybrid approach combining both low- and high-order feature interactions is more effective than models that rely solely on either.

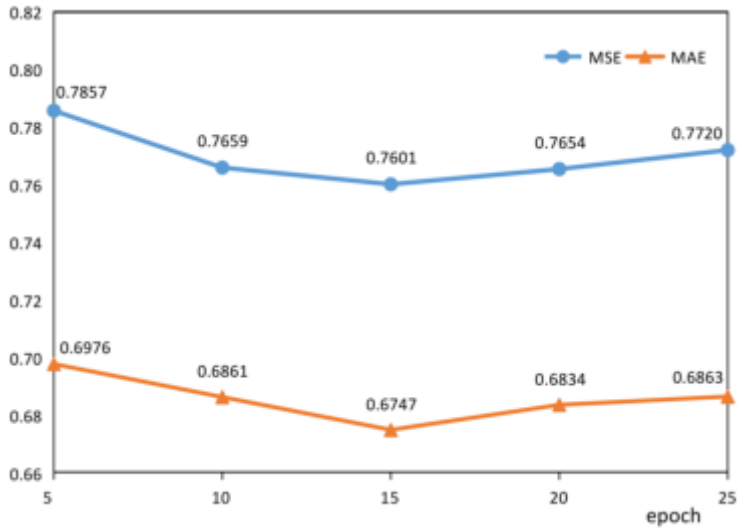


Fig. 3.1: Changes in MSE and MAE with increasing iteration times.

Table 3.4: Comparison of evaluation indicators of various models on the Douban movie dataset test set.

Model	MSE	RMSE	MAE	R^2
NFM	0.7947	0.8918	0.6945	0.3656
AFM	0.8323	0.9120	0.7086	0.3353
Deep FM	0.7908	0.8890	0.6967	0.3689
xDeep FM	0.8647	0.9300	0.7241	0.3096
Auto Int	0.7867	0.8873	0.6877	0.3715
AFM-AutoInt	0.7832	0.8847	0.6875	0.3746

However, AFM-AutoInt requires a longer training time due to its comprehensive consideration of both feature types. Extensive parameter tuning plays a crucial role in minimizing loss, which is a key factor in the model's superior performance. Additionally, AFM-AutoInt surpasses the DeepFM model by 0.52%, benefiting from the integration of attention mechanisms that refine the weighting of second-order feature combinations. Experimental results confirm that the proposed model effectively captures movie data characteristics and enhances the accuracy of user rating predictions.To fully verify the advantages of the proposed model in various evaluation indicators, this paper processed and trained the Douban movie dataset, and finally obtained a comparison of the evaluation indicators of each model on the Douban movie dataset test set, as shown in Table 3.4.

According to Table 3.4, compared to other models, the AFM AutoInt model also achieved the best prediction results on the Douban movie dataset.

4. Conclusion. This article presents a novel intelligent recommendation system model, **AFM-AutoInt**, which integrates deep learning, attention mechanisms, and automatic feature interaction modeling to address key challenges in recommendation systems, such as **data sparsity** and **complex feature interactions**. The **AutoInt module** enhances the model by leveraging a multi-layer self-attention mechanism to mine **high-order feature interactions**, significantly improving predictive accuracy. Experimental results on real-world datasets, including **Movielens-1M** and **Douban movie**, validate the superiority of AFM-AutoInt over traditional recommendation algorithms, demonstrating **lower prediction errors** and achieving state-of-the-art performance across multiple evaluation metrics. This study highlights the crucial role of incorporating both **low- and high-order feature interactions** in recommendation systems, offering a promising advancement for personalized content recommendations. Future work can explore several directions: (1) **Optimizing

model complexity** to improve computational efficiency without compromising accuracy, (2) **Enhancing generalization** by adapting AFM-AutoInt to diverse datasets and real-world applications, such as e-commerce and personalized healthcare, (3) **Integrating additional contextual information**, such as temporal dynamics and user behavior patterns, to refine recommendations, and (4) **Developing explainable recommendation strategies** to enhance user trust and system transparency. By addressing these areas, AFM-AutoInt can further advance the development of next-generation intelligent recommendation systems.

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

Funding Statement.

1. 2024 Hainan Provincial Higher Education Teaching Reform Research Project, "Construction and Practical Research on the Teaching Model of Pair Programming Based on Shared Regulation" (Hnjg2024-154)
2. 2023 Hainan Provincial Higher Education Teaching Reform Research Project, "Practical Research on the Teaching of the Public Course 'Modern Educational Technology' Based on the DELC Model in a Blended Learning Environment" (Hnjg2023-139)
3. 2024 Hainan Provincial Social Science Planning Project, "Research on Strategies for Enhancing Digital Literacy of Primary and Secondary School Teachers in Rural Hainan under the Background of Educational Digital Transformation" (HNSK(ZC)24-173)

REFERENCES

[1] Deng, L. P., Guo, B., & Zheng, W. *A Service Recommendation Algorithm Based on Self-Attention Mechanism and DeepFM.* International Journal of Web Services Research (IJWSR), (2023),20(1), 1-18.

[2] Yin, P., Ji, D., Yan, H., Gan, H., & Zhang, J. *Multimodal deep collaborative filtering recommendation based on dual attention.* Neural Computing and Applications, (2023), 35(12), 8693-8706.

[3] Gao, C., Zheng, Y., Wang, W., Feng, F., He, X., & Li, Y. Causal inference in recommender systems: A survey and future directions. ACM Transactions on Information Systems, (2024), 42(4), 1-32.

[4] Dong, H., & Wang, X. Hoint: Learning explicit and implicit high-order feature interactions for click-through rate prediction. Neural Processing Letters, (2023), 55(1), 401-421.

[5] Bi, Z., Sun, S., Zhang, W., & Shan, M. *Click-through rate prediction model based on graph networks and feature squeeze-and-excitation mechanism.* International Journal of Web Information Systems, (2024), 20(4), 341-357.

[6] Gan, M., Li, D., & Zhang, X. *A disaggregated interest-extraction network for click-through rate prediction.* Multimedia Tools and Applications, (2023), 82(18), 27771-27793

[7] Jiang, Z., Li, L., & Wang, D. *MCGM: A multi-channel CTR model with hierarchical gated mechanism for precision marketing.* World Wide Web, (2023), 26(4), 2115-2141.

[8] Zheng, J., Chen, S., Cao, F., Peng, F., & Huang, M. *Explainable recommendation based on fusion representation of multi-type feature embedding.* The Journal of Supercomputing, (2024), 80(8), 10370-10393.

[9] Zhang, L., Liu, F. A., Wu, H., Zhuang, X., & Yan, Y. *CFF: combining interactive features and user interest features for click-through rate prediction.* The Journal of Supercomputing, (2024), 80(3), 3282-3309.

[10] Liliang, Z. H. O. U., Shili, Y. U. A. N., Zijian, F. E. N. G., Guilan, D. A. I., & Guofu, Z. H. O. U. *A Lambda Layer-Based Convolutional Sequence Embedding Model for Click-Through Rate Prediction.* Wuhan University Journal of Natural Sciences, (2024), 29(3), 198-208.

[11] Zhou, H., Zhang, S., Qiu, L., Wang, Z., & Hu, K. *A factorisation-based recommendation model for customised products configuration design.* International Journal of Production Research, (2023), 61(19), 6381-6402.

[12] Zheng, Q., Peng, Z., Dang, Z., Zhu, L., Liu, Z., Zhang, Z., & Zhou, J.*Deep tabular data modeling with dual-route structure-adaptive graph networks.* IEEE Transactions on Knowledge and Data Engineering, (2023), 35(9), 9715-9727.

[13] Yang, L., Zheng, W., & Xiao, Y. *Exploring different interaction among features for CTR prediction.* Soft Computing, (2022), 26(13), 6233-6243.

[14] Wu, Y. *Exploration of the Integration and Application of the Modern New Chinese Style Interior Design.* International Journal for Housing Science and Its Applications, (2024), 45(2), 28-36.

[15] Chen, P. *Research on Business English Approaches from the Perspective of Cross-Cultural Communication Competence.* International Journal for Housing Science and Its Applications, (2024), 45(2), 13-22.

[16] Wang, W.*Esg Performance on the Financing Cost of A-Share Listed Companies and an Empirical Study.* International Journal for Housing Science and Its Applications, (2024), 45(2), 1-7

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/ Emerging Wireless Networks
Received: Aug 28, 2024
Accepted: May 27, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 1964–1973, DOI 10.12694/scpe.v26i5.4790

MULTI PATH GAIT CONTROL METHOD FOR BIPEDAL ROBOTS BASED ON DEEP REINFORCEMENT LEARNING

LEHUI LIN *AND PINGLI LV †

Abstract. We propose a multipath gait control strategy based on deep reinforcement learning (DRL) for bipedal robot motion planning on diverse and challenging terrains. Traditional control methods, such as PID controllers and model-based motion planning, often struggle in complex environments. These approaches typically underperform because they rely on precise mathematical models or predefined rules, making them ill-suited for nonlinear, uncertain, and dynamic settings. Conventional techniques also have difficulty adapting their control strategies in unpredictable and fluctuating terrains, where robots may encounter unforeseen disturbances, leading to instability or failure.

Deep reinforcement learning is able to independently acquire optimal control methods from environmental feedback without requiring a precise model since it combines deep learning and reinforcement learning. In this work, we leverage deep reinforcement learning algorithms (DDPG, TRPO, PPO, A3C, SAC, etc.) based on actor-critic (AC) architectures to enable reliable gait control of bipedal robots in a continuous motion environment. The issue that traditional approaches have in challenging to converge complicated environments is solved by DRL, which, when compared to traditional methods, can effectively cope with the high nonlinearity of complex terrain and adaptively alter the strategy through continuous contact with the environment.

Using goal-conditional techniques, we created a motion planning model and tested it on the actual hardware platform Cassie. According to the experimental results, the approach successfully transfers the simulation strategy to the actual environment, and the robot can accurately complete the goal task without global location feedback. It can also perform a variety of complex tasks, like jumping on discontinuous and flat terrain. Furthermore, the method exhibits significant robustness and adaptability through multithreaded asynchronous training and randomized strategy selection, which solves the shortcomings of conventional motion planning methods in hyperparameter tuning and strategy convergence.

Key words: Reinforcement learning; Bipedal robot; Multi path gait control; Actor critic; Robot robustness

1. Introduction. The potential use of bipedal robots in a variety of complicated contexts has steadily gained significant attention as a result of the rapid advancement of robotics technology [1]. Nevertheless, bipedal robot motion control remains extremely difficult, particularly in dynamic, uneven, or uncharted territory [2]. Robust gait control techniques are essential for bipedal robots to be able to move independently and adapt to a variety of situations [3].

Traditional robot control approaches, such as pre-defined motion trajectories, reveal difficulties when coping with complicated environmental changes. This is because the robot might not be able to finish the task if the environment changes or if there are inaccuracies in the sensor data, as this approach usually depends on exact sensor data and correct environment modeling [4]. Consequently, in an effort to attain more autonomous gait control, an increasing number of researchers have focused on deep reinforcement learning (DRL) and other reinforcement learning-based control techniques in recent years.

Among various DRL algorithms, the actor critic framework based algorithm has received widespread attention due to its superior performance in continuous action space. Deep Deterministic Policy Gradient (DDPG), Trust Region Policy Optimization (TRPO), Near End Policy Optimization (PPO), Asynchronous Advantageous Actor Critic (A3C), and Soft Actor Critic (SAC) are a few examples of this kind of technique [5]. These algorithms each have their own characteristics and are suitable for different task scenarios. For example, DDPG has been successfully applied in continuous action planning for mobile robots by using two neural networks to estimate the policy function and value function separately [6]. However, DDPG is prone to poor convergence in practical applications [7]. TRPO improves the convergence of the algorithm by optimizing the step size of policy updates, but its high computational complexity limits its application scope [8]. PPO simplifies the calcu-

*Dongbei University of Finance and Economics, Dalian 116025, China (linlehui@dufe.edu.cn).
†Xuzhou Industrial Vocational and Technical College, Xuzhou 221140, China.

lation process of TRPO, although it improves computational efficiency, there are still shortcomings in terms of exploration [9]. A3C improves training efficiency through multi-threaded parallel training, but requires a large amount of training data, which increases the cost in practical applications [10]. SAC improves the sensitivity of the policy gradient algorithm to hyperparameters and enhances the robustness of the algorithm by introducing the maximum entropy strategy [11].

While these DRL algorithms work well for controlling robot motion, there are still significant obstacles when trying to directly apply them to controlling the gait of bipedal robots. First of all, it is challenging to directly use known algorithms in real-world applications due to the high dimensionality and complexity of bipedal robot motion control. Second, there is an urgent need to find a solution to the issue of how to guarantee the algorithm's robustness and universality while transferring between various jobs and contexts. Furthermore, present algorithms still perform inadequately when it comes to handling the impacts of uncertainty and randomness in the motion of bipedal robots in dynamic situations.

This research suggests a deep reinforcement learning-based multi-path gait control technique for bipedal robots as a solution to these problems. This method aims to enhance the adaptability and stability of bipedal robots in different environments by introducing a strategy that combines multi-path planning and DRL. The main contributions of this article include the following aspects:

The multi-path gait control method proposed in this article not only considers the stability of the robot on a single path, but also enhances the adaptability of the robot in complex environments through multi-path planning. Multi path planning allows robots to quickly select the optimal path when facing environmental changes, thereby improving the robustness of their motion.

By combining DRL and multi-path planning, this paper constructs an adaptive control framework that enables robots to flexibly switch between different tasks and environments. This framework not only considers various possible environmental changes during the training process, but also has good generalization ability in practical applications.

This research carried out numerous experiments on both real hardware platforms and simulated environments to confirm the efficacy of the suggested approach. The experimental findings demonstrate that the suggested multi-path gait control approach outperforms existing DRL algorithms and conventional techniques in complicated situations. Simultaneously, the approach showed its promise in real-world applications by performing better in various tasks and situations.

2. Actor critic-based DRL motion planning. In the task of robot motion planning in continuous action space, the output of the robot is no longer a finite set of discrete actions, but the execution probability of each action in the continuous action domain [12]. The speed and acceleration changes of wheeled robots acting on the active wheels, and the output of force and torque of legged robots acting on each joint. For this type of task, the actor critic based DRL motion planning method has stronger applicability, and commonly used algorithms include DDPG, TRPO, PPO, A3C, SAC, etc., which will be introduced one by one below.

In practice, multipath control methods have shown significant advantages in several areas, especially in robot motion control in complex environments. The following are some typical application cases to demonstrate how the method can solve real-world motion control problems and its potential application areas:

Rescue robots: After natural disasters or accidents, rescue robots are often required to perform tasks in complex terrains, such as rubble piles, mountain slopes, or forest environments. Traditional motion control methods often perform poorly in such uncertain and changing environments. In contrast, deep reinforcement learning-based multipath control methods are able to adapt to various complex terrains through continuous learning and adjustment. Through training and strategy optimization in different terrains, the rescue robot can achieve more stable and flexible motion control, easily cross obstacles, quickly reach the disaster site, and improve rescue efficiency.

Industrial robots: In modern production lines, industrial robots often need to perform high-precision motion planning in limited space, such as handling heavy objects and assembling parts. Multipath control methods in such scenarios allow for more flexible and efficient operations based on real-time changes in the environment, such as avoiding obstacles or quickly adjusting paths. In addition, when the robot is performing multiple tasks, the algorithm is able to effectively respond to the switching of complex tasks through multipath planning, ensuring the efficiency and continuity of industrial production.

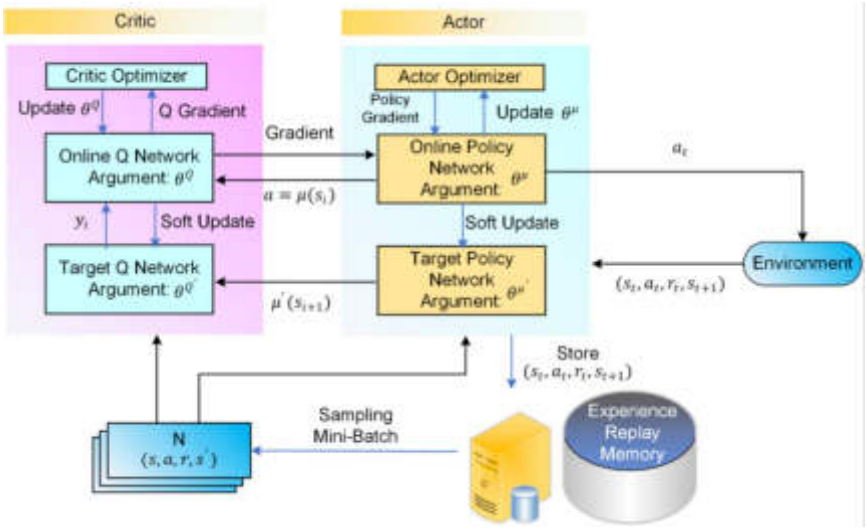


Fig. 2.1: Motion Planning Method Based on DDPG Algorithm.

Medical robots: During surgery, medical robots require extreme precision and flexibility, especially when operating complex surgeries. Motion planning methods based on multipath control can simulate and learn from the anatomical structures and surgical requirements of different patients to provide optimal operation paths for surgical robots while avoiding errors. The application of this technology can significantly improve the success rate of surgery and reduce the operation time.

Driverless vehicles: Driverless technology also faces the challenge of complex environments, especially path planning in dynamic traffic scenarios. Multipath control methods can help driverless vehicles adaptively adjust driving paths in complex urban roads, safely avoid pedestrians and obstacles, and optimize driving routes through learning to improve driving efficiency and safety. These typical application cases show that the multipath control method has obvious practicality and advantages in solving robot motion control problems in complex environments, and its potential application areas cover many key fields such as rescue, industry, medical care, and autonomous driving.

2.1. DDPG. A model free deep deterministic policy gradient algorithm (DDPG) based on the actor critic (AC) algorithm framework is proposed for motion planning of mobile robots in continuous action spaces, which has the ability of continuous action planning for mobile robots [13].The DDPG algorithm constructs four neural networks for mobile robots, while estimating the policy function and value function, including the target network of actor and critic and the current network, respectively. The actor's current network is responsible for iteratively updating the robot strategy network parameter θ, and the parameter θ' in the actor's target network is copied from the current network based on a fixed step size; The current network of critic is responsible for iteratively updating the value network parameter w and calculating the current Q value. The target network parameter w' of critic is periodically copied from w. The motion planning method for mobile robots is shown in Fig.2.1.

DDPG uses memory replay units for random selection, breaking the correlation of data and improving algorithm efficiency. It adopts a deterministic strategy to directly output the action corresponding to the maximum value function of the mobile robot, and defines the optimization objective function of DDPG algorithm as:

$$J(\theta^\mu) = E_{\theta^\mu}\left[r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots\right] \tag{2.1}$$

where μ is the parameter of the strategy network that generates deterministic actions.

With the use of a dual network structure and priority experience replay mechanism, DDPG has successfully

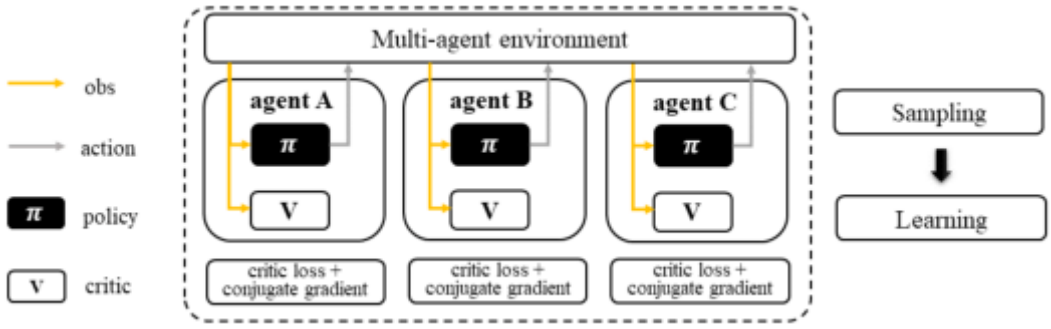


Fig. 2.2: Motion Planning Method Based on TRPO Algorithm.

tackled the problems of actor critic convergence and continuous control of mobile robots [14].

2.2. TRPO. In the optimization process of DDPG strategy gradient, the update step size will directly determine whether the mobile robot can quickly and accurately reach the target point [15]. Some researchers have suggested a reinforcement learning algorithm (TRPO) based on trust domain strategy optimization to determine the proper step size [16]. TRPO is based on calculating the KL dispersion range between the new strategy and the old strategy, with the goal of maximizing the difference between the action value function and the state value function, that is, taking the advantage function as the objective. The objective function is defined as:

$$J_{TRPO}^{\theta'}(\theta) = E_{\pi_\theta}\left[\pi_\theta / \pi_{\theta'} A^{\theta'}(s_t, a_t)\right] \tag{2.2}$$

Among them, θ and θ' are the network parameters of the new and old strategies, respectively.

A motion planning technique for a mobile robot based on the TRPO algorithm is shown in Fig.2.2. TRPO outputs an action probability distribution, and the mobile robot can optimize its motion parameters based on this probability distribution.

TRPO (Trust Region Policy Optimization) addresses the challenge of determining the appropriate update step size in policy gradient methods. It does so by using a trust region to ensure stable updates, which prevents large, destabilizing changes to the policy. However, in practice, the implementation of TRPO involves certain approximations that can complicate the process. These approximations often lead to labor-intensive calculations and, in some cases, substantial inaccuracies, particularly when the trust region is not well-calibrated. Despite its theoretical advantages, the computational complexity and potential for approximation errors make TRPO challenging to apply in real-time systems or environments with high-dimensional state spaces. These limitations highlight the need for further optimization of the algorithm to improve both accuracy and computational efficiency.

2.3. PPO. In response to the problem of excessive complexity in the calculation process of TRPO algorithm, which leads to deviation between the motion planning path and the optimal path of mobile robots, some scholars have proposed a reinforcement learning algorithm (PPO) for proximal strategy optimization, which enables mobile robots to have better exploratory ability in action selection [17]. The PPO algorithm directly uses the $KL(\theta, \theta')$ divergence of the old and new strategies as the penalty term, and its objective function is updated to:

$$J_{PPO}^{\theta}(\theta) = E_{\pi_j}\left[\frac{\pi_\theta}{\pi_\theta} A^{\theta}(s_t, a_t)\right] - \beta KL(\theta, \theta') \tag{2.3}$$

where β is the parameter for updating dispersion. Compared with the TRPO reinforcement learning algorithm, the PPO algorithm significantly simplifies the calculation steps of TRPO while retaining the exploration method of random strategies. In the case where the sampled samples satisfy the maximum likelihood probability, the robot's motion planning method will have better exploration and robustness [18].

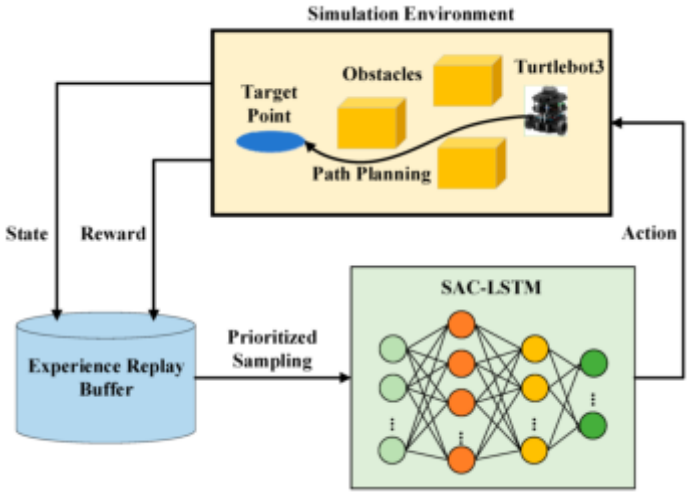


Fig. 2.3: SAC Algorithm-Based Robot Motion Planning Principle.

2.4. SAC. SAC is an offline learning reinforcement learning method based on maximum entropy, and mobile robots also use random strategies to select actions, which can learn from past experiences as well as other task experiences. As shown in Fig.2.3, the reinforcement learning motion planning method based on SAC maximizes the objective function through the maximum entropy method, where the objective function based on maximum entropy is:

$$J(\theta) = E_{\pi}\left[\sum Q(s_t, a_t) - \hat{\alpha} log \pi(a_t|s_t)\right] \tag{2.4}$$

Among them, $\hat{\alpha}$ are entropy regularization coefficients, which contain dynamic parameters from the real robot system.

The SAC motion planning algorithm uses a strong learning framework to reduce the need for hyperparameter change under the entropy constraint. In order to maximize expected return on the mobile robot's objective function, it updates the action selection strategy network of the robots using regularization coefficients rather than hyperparameters. This increases the policy gradient method's sensitivity and weak convergence to hyperparameters, improving the stability and applicability of the SAC motion planning algorithm in real-world settings. For legged and other mobile robots, it can quickly learn and plan walking skills.

3. Experiments. We have successfully deployed simulated target condition strategies and flat ground strategies for different target tasks on Cassie hardware, covering tasks such as jumping to different positions and turning directions, as well as discrete terrain strategies for variable positions and heights.

This section aims to verify two hypotheses: 1) whether the simulated training strategy can perform the same tasks in the real world; 2) Can the target condition strategy stably utilize the learned tasks to control the robot after being transferred to reality. The robot did not receive global position feedback in any of the studies, so once it moves, it will not be able to determine how far it is from the target landing location or the earth.

1. Validation of Hypothesis 1: Whether a simulation-trained strategy can perform the same task in reality.

Experimental Goal: To verify whether the strategies trained in the simulation environment can be directly applied to real robot systems to ensure that the robots perform well when performing the same tasks in the real environment.

Experimental Design:

Experimental conditions: In the simulated environment, the robot will be trained on multiple tasks, including operations such as walking, jumping and steering on different terrains. After the training is completed, the same tasks will be executed directly on the real hardware platform using the

same strategy, and the simulated training environment matches the dynamic characteristics of the real hardware environment.

Experimental Tasks:

Task 1: Jumping to a fixed target point. The robot needs to jump from its starting position to a fixed target point 1 meter away, the goal of the task is to test its accuracy in completing the task in both simulated and real environments.

Task 2: Backward jump. The robot needs to jump backward 0.3 meters from the starting position and accurately land on the ground marking point to verify whether the strategy can be stably executed in different directions of motion.

Task 3: In-situ steering and jumping. The robot first turns -60 degrees in place and then jumps back to the starting position to test its adaptability in complex motions.

Experimental Steps:

Simulation training phase: complete training of the robot on all tasks in a simulated environment, using deep reinforcement learning strategies for optimization until the robot demonstrates efficient action planning and execution in the simulated environment.

Realistic task execution phase: the trained strategies are applied to the real robot platform, and the above three tasks are repeated under the same physical conditions (same start position and goal position). Record the robot's movement trajectory, jumping distance and landing point, and compare the accuracy, time and stability of task completion in the simulated and real environments.

Validation metrics:

Task success rate: whether the robot successfully completed the tasks in the real environment.

Jumping accuracy: measure the deviation of the robot from the target point in the real environment.

Motion Stability: Observe whether the robot is unstable or deviates from the original strategy during the actual task execution.

2. Validate Hypothesis 2: Whether the goal-conditional strategy can control the robot stably after transferring to reality.

Experimental Goal: To verify whether the robot trained by the goal-conditional strategy can stably execute the learned task in the real environment, especially when facing complex terrain.

Experimental Design:

Experimental conditions: The goal-conditioned strategy will be trained on a variety of complex tasks in a simulated environment that contains irregular terrain and different task requirements, such as jumping to target points at different heights or crossing obstacles at different heights.

Experimental Tasks:

Task 1: Jumping to a 1-meter distant target point. The robot is required to jump 1 meter from the starting point to a target point on flat ground to verify its stability under target conditions.

Task 2: Jump to 1.4 meter away target point. Increase the jumping distance to verify whether the target condition strategy can maintain accuracy and stability under increasing distance.

Task 3: Jump to target points at different heights. The robot needs to jump to a 0.44-meter-high platform to test its adaptability under different height conditions.

Experiment Steps:

Training in the simulated environment: the robot is trained using goal-conditional strategies to ensure that the robot learns the corresponding strategies under different tasks and executes them accurately at different target locations and heights.

Realistic task execution phase: transfer the strategies in the simulated environment to the actual hardware platform, repeat the same jumping task, record the robot's trajectory and completion, and test its stability and accuracy in different tasks.

Validation metrics:

Task completion rate: whether the robot can complete the task in the real environment.

Target accuracy: measure the accuracy of the robot in different jumping distance and height tasks.

Stability and Adaptability: assess the robot's ability to adapt in the real world by observing its performance in different complex environments.

Analysis of experimental results : By comparing the results of the robot's task execution in the simulated and real-world environments, the following points can be verified: Whether the simulation training strategy is able to achieve seamless transfer in reality and the stability of task execution.
Whether the target conditioning strategy is able to maintain high accuracy and stability under complex terrain conditions, especially when the jumping distance and height change. These experiments are designed to verify the performance and stability of the simulated strategies in real-world environments, thus better explaining the two hypotheses in the article.

3.1. Execution of Realistic Tasks. Initially, we conducted three challenges to evaluate the robot's flat ground strategy: Jumping forward to a target point one meter ahead, jumping 0.3 meters backward, and jumping in place after turning negative sixty degrees.

These tests highlight two major benefits of the suggested approach: Initially, it can adapt to different system dynamics and transition from simulation to real environments; Secondly, it can flexibly deviate from the reference motion trajectory and complete predetermined tasks through multiple touchpoints.

Next, we validated the performance of the discrete terrain strategy in three tasks: jumping 1 meter forward, jumping 1.4 meters forward, and jumping to a target 0.88 meters ahead with a height of 0.44 meters. This tactic illustrates the capacity to accurately manipulate the robot to land on a specific target,. The robot took off and landed later during the 1.4-meter jump compared to the 1-meter jump, suggesting that the robot needs longer flight phases and higher takeoff speeds to land at further places.Furthermore, to ensure accuracy in height, the robot made a more vertical jump and raised its leg position when jumping to a platform 0.44 meters high . This method can still stabilize the robot's performance at various landing places even with an early landing time.

These experimental results indicate that the proposed strategy can flexibly adapt to the dynamic characteristics of robot hardware and achieve precise target landing by adjusting takeoff attitude and acceleration. This ability is particularly critical when dealing with ballistic movements during flight, as small errors during takeoff can lead to significant deviations in landing position.

3.2. Semi sealed gap. To further reveal the difficulty of the successful jumping experiment of the robot in Fig. 3.1, we conducted a thorough analysis of the gap between simulation and the real world. Fig.3.1 shows the joint position changes when the robot performs a turning task . Comparing the recorded data, it was found that there is a significant deviation between the joint positions of the robot in the simulated environment (blue curve) and the actual positions in the real world (red curve).

For example, the maximum deviation of the tarsal joint connection position q_6 exceeds 0.35 radians. Considering that this joint is not driven by a leaf spring with a nominal stiffness of 1250 Nm/rad, this deviation has a significant impact on the dynamic performance of the robot. Similar deviations also occur in other key joints, such as the rotation joint q_2, thigh joint q_3, and knee joint q_4, which play crucial roles in jumping and turning processes. In addition, similar differences were observed in other experiments using discrete terrain strategies. These differences emphasize the huge gap between simulation and the real world, but also demonstrate that despite such significant differences.

As shown in Fig.3.1, when the robot executes the command to jump and rotate -60 °, its joint positions show significant differences between simulated and real environments. For example, the tarsal joint driven by passive leaf springs exhibits significant deviations during the transition from simulation to actual operation. In addition, the flight phase in the real environment is significantly delayed compared to the simulated world.

3.3. Diversified and Strong Strategies under Target Condition Policies. To further overcome the limitations of the proposed control strategy, we conducted more complex dynamic jump experiments, as shown in Fig. 3.3. In these multi axis jumps, robots exhibit more complex motion behaviors. For example, in Fig. 3.3, the robot tilts laterally while jumping forward, and rotates -45 ° after landing, accurately landing at the target position 0.5 meters in front of the starting point and 0.2 meters to the left. In some challenging tasks, as shown in Fig. 3.3, the robot adjusts its body posture through small jumps after unstable landing to maintain balance.

In addition, to verify the robustness of the strategy, we applied a backward perturbation force at the pelvic vertex of the robot and tested its performance in dynamic environments, as shown in Fig.3.3. After

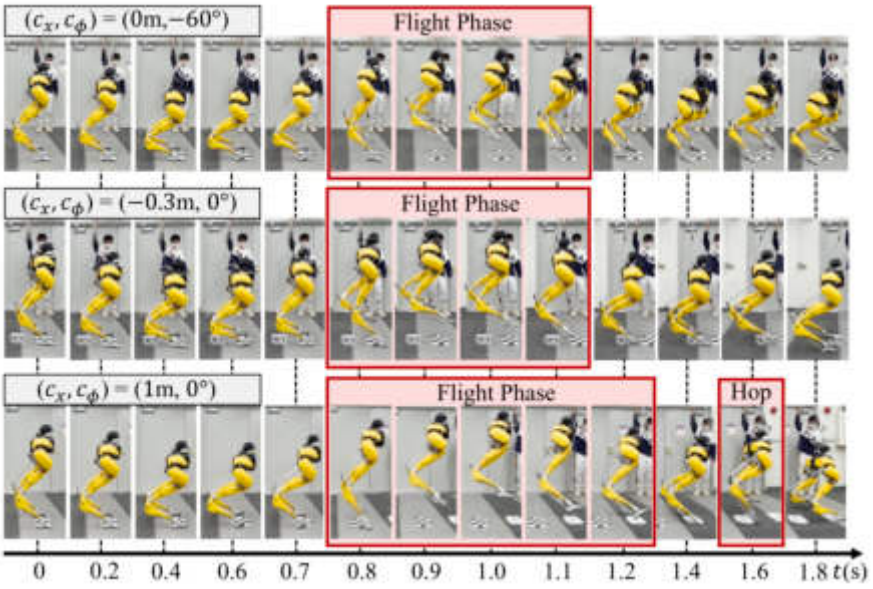


Fig. 3.1: Simulation differences in robot jumping and rotation -60 °.

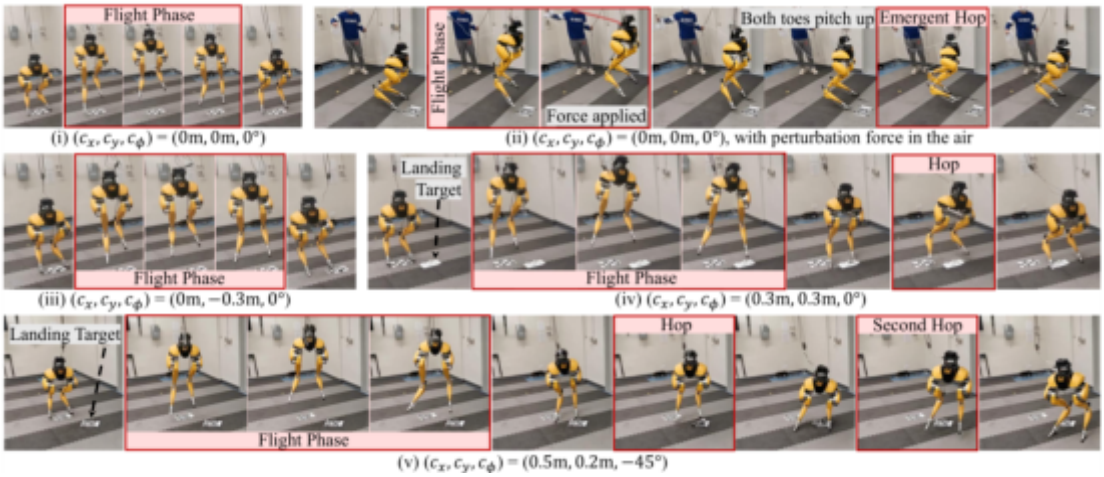


Fig. 3.2: Different jumps using flat ground strategy.

being disturbed, the robot tilted backwards during descent and caused its toes to tilt upwards during landing, resulting in insufficient driving force at the contact point. However, robots quickly adjust themselves by jumping backwards, a response learned in multi-objective training. During this jumping process, the robot is able to readjust its posture during the flight phase, achieving stability upon landing. The original goal of this test was to jump in place, but interestingly, the robot deviated from its initial task to avoid falling.

4. Conclusion. In this study, we propose a deep reinforcement learning (DRL) based multipath gait control technique for bipedal robots to overcome the difficulties of motion planning over complex terrain. The irregularity and variety of the terrain present a challenge to established control approaches, and the methods' limitations in terms of resilience and adaptation in continuous action space are just two of the many challenges

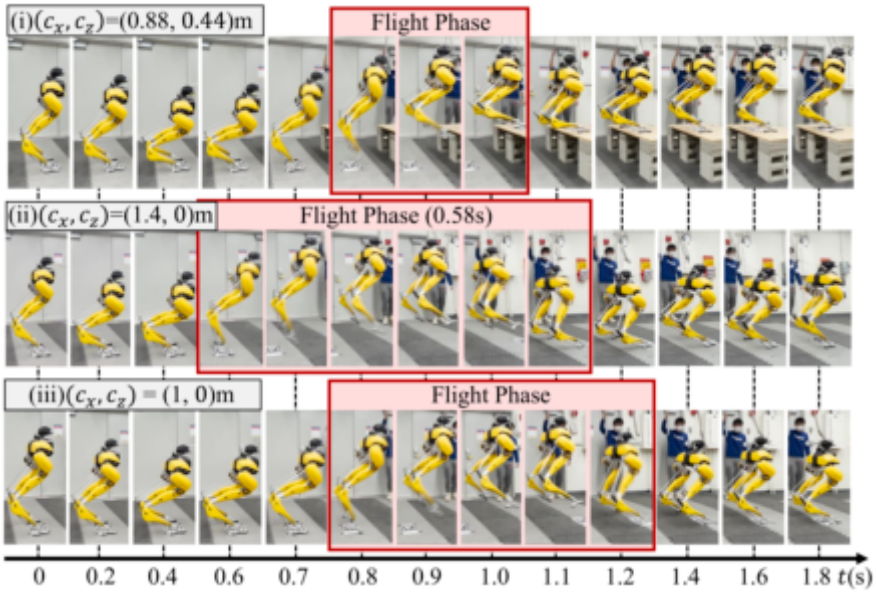


Fig. 3.3: Different jumps using discrete terrain strategy.

that come with motion planning in complex terrain. These factors hinder the robot's ability to walk steadily and accurately. We achieve accurate gait control of robots in continuous motion space by methodically examining various deep reinforcement learning algorithms based on actor-critic structure (e.g., DDPG, TRPO, PPO, A3C, SAC, etc.). The method's superior adaptability and robustness in a variety of complex tasks and environments, along with its ability to migrate strategies trained in a simulated environment to the real hardware platform Cassie, are demonstrated by the experimental results. This improves the performance of traditional methods in complex terrains to a significant degree.

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

Funding Statement. Xuzhou Industrial Vocational and Technical College High level Talent Research Launch Project Project name: Research on Deep Reinforcement Learning Method Based on Value Function Exploration Strategy Project Number: XGY2021EG02

REFERENCES

[1] Roka, S., Diwakar, M., Singh, P., & Singh, P. *Anomaly behavior detection analysis in video surveillance: a critical review.* Journal of Electronic Imaging, (2023), 32(4), 042106-042106.

[2] Wang, Y., Ren, Y., & Yang, J. *Multi-Subject 3D Human Mesh Construction Using Commodity WiFi.* Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, (2024), 8(1), 1-25.

[3] Shen, L. H., Feng, K. T., & Hanzo, L. *Five facets of 6G: Research challenges and opportunities.* ACM Computing Surveys, (2023), 55(11), 1-39.

[4] C. Zhang, M. Li and D. Wu. *"Federated Multidomain Learning With Graph Ensemble Autoencoder GMM for Emotion Recognition".* IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 7, pp. 7631-7641, July 2023, doi: 10.1109/TITS.2022.3203800.

[5] Rinchi, O., Ghazzai, H., Alsharoa, A., & Massoud, Y. *Lidar technology for human activity recognition: Outlooks and challenges.* IEEE Internet of Things Magazine, (2023), 6(2), 143-150.

[6] Abuhoureyah, F., Wong, Y. C., Isira, A. S. B. M., & Al-Andoli, M. N. *CSI-Based Location Independent Human Activity Recognition Using Deep Learning.* Human-Centric Intelligent Systems, (2023), 3(4), 537-557.

[7] Jamil, H., & Kim, D. H. *Optimal fusion-based localization method for tracking of smartphone user in tall complex buildings.* CAAI Transactions on Intelligence Technology, (2023), 8(4), 1104-1123.

[8] Lian, J., Ren, W., Li, L., Zhou, Y., & Zhou, B. *Ptp-stgcn: pedestrian trajectory prediction based on a spatio-temporal graph convolutional neural network*. Applied Intelligence, (2023), 53(3), 2862-2878.

[9] Bian, S., Liu, M., Zhou, B., Lukowicz, P., & Magno, M. *Body-Area Capacitive or Electric Field Sensing for Human Activity Recognition and Human-Computer Interaction: A Comprehensive Survey*. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, (2023), 8(1), 1-49.

[10] Zoghlami, C., Kacimi, R., & Dhaou, R. *5G-enabled V2X communications for vulnerable road users safety applications: a review*. Wireless Networks, (2023), 29(3), 1237-1267.

[11] Rizk, H., Yamaguchi, H., Youssef, M., & Higashino, T. *Laser range scanners for enabling zero-overhead wifi-based indoor localization system*. ACM Transactions on Spatial Algorithms and Systems, (2023), 9(1), 1-25.

[12] Prateek, K., Ojha, N. K., Altaf, F., & Maity, S.*Quantum secured 6G technology-based applications in Internet of Everything*. Telecommunication Systems, (2023), 82(2), 315-344.

[13] Liu, Y., Yang, D., Wang, Y., Liu, J., Liu, J., Boukerche, A., ... & Song, L. *Generalized video anomaly event detection: Systematic taxonomy and comparison of deep models*. ACM Computing Surveys, (2024), 56(7), 1-38.

[14] Yang, J., Chen, X., Zou, H., Wang, D., & Xie, L. *Autofi: Toward automatic wi-fi human sensing via geometric self-supervised learning*. IEEE Internet of Things Journal, (2022), 10(8), 7416-7425.

[15] Shastri, A., Valecha, N., Bashirov, E., Tataria, H., Lentmaier, M., Tufvesson, F., ... & Casari, P.. *A review of millimeter wave device-based localization and device-free sensing technologies and applications*. IEEE Communications Surveys & Tutorials, (2022), 24(3), 1708-1749.

[16] Wu, Y.*Exploration of the Integration and Application of the Modern New Chinese Style Interior Design*. International Journal for Housing Science and Its Applications, (2024), 45(2), 28-36.

[17] Wang, W.*Esg Performance on the Financing Cost of A-Share Listed Companies and an Empirical Study*. International Journal for Housing Science and Its Applications, (2024), 45(2), 1-7.

[18] Z. Guo, K. Yu, N. Kumar, W. Wei, S. Mumtaz and M. Guizani.*"Deep-Distributed-Learning-Based POI Recommendation Under Mobile-Edge Networks"*. IEEE Internet of Things Journal, vol. 10, no. 1, pp. 303-317, 1 Jan.1, 2023.

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/ Emerging Wireless Networks
Received: Aug 30, 2024
Accepted: Apr 21, 2025


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 1974–1984, DOI 10.12694/scpe.v26i5.4847

DEEP ANALYSIS ON THE COLOR LANGUAGE IN FILM AND TELEVISION ANIMATION WORKS VIA SEMANTIC SEGMENTATION TECHNIQUE

YANXIANG ZHANG*

Abstract. Color functions as a distinct type of ideographic symbol in animation for film and television, playing a crucial role in enhancing visual narratives and conveying emotions. In different types of animation, such as fantasy, horror, or children's genres, color language influences audience perception and can convey meanings beyond the capabilities of image language alone. For instance, bright colors may symbolize innocence in children's animation, while darker shades may evoke tension or fear in horror. However, current approaches to representing color in animation often fail to capture its full semantic richness and ideographic potential. Existing methods primarily focus on image-based analysis, overlooking the deeper layers of meaning encoded in color language. In this paper, we address these gaps by utilizing semantic segmentation techniques to combine the three modalities of color, content, and text to establish a consistent representation of color language in animation. We propose a method for semantic segmentation of color-depth (RGB-D) images using two-stream weighted Gabor convolutional network fusion. A weighted Gabor orientation filter builds a deep convolutional network (DCN) capable of extracting feature information adaptive to changes in orientation and scale, allowing for orientation- and scale-invariant features. Dual-stream picture features—color and depth—are extracted using a broad residual-weighted Gabor convolutional network and then combined into a lightweight feature extraction network. To evaluate the ideographic functions of color language quantitatively, we conducted extensive experiments using open databases. Our proposed method outperforms existing RGB-D picture semantic segmentation algorithms, demonstrating its effectiveness in representing color language in animation.

Key words: Semantic Segmentation, Video Animation, Color Language

1. Introduction. Text symbols and colour have a definite relationship in animation for film and television, and they can also naturally blend together with visuals. It is a kind of code in film and television works and has the function of expressing meaning. Ideology is the main feature of color [1]. The ability to combine the language of television and films, as well as its meaning, are qualities bestowed by life itself. People and colour have already had a significant impact on people's daily lives [2]. There is a certain experience in the expression of different colors, and different intentions are formed between different colors [3]. Not only can colour expressions convey meaning in the language of cinema and television animation, but colour also functions as a type of code with symbolic meaning [4]. It has distinct qualities. Colour can communicate diverse information when working in tandem with voice and vision. For example, as shown in Fig.1.1, due to the different shades of colors, the emotions of the characters are also expressed. Through the distinction of color and brightness, the contrast of the emotions of the protagonists in the film can be formed.

In animation production, color is not only an aesthetic element, but also a symbolic language that can convey emotions and suggest the direction of the plot. For example, in the same scene, by adjusting the brightness or hue of a color, the audience's perception of the emotion will change significantly. However, traditional image processing methods often fail to capture the potential meaning of color language in depth, especially when there is a difference between the semantic expression of color and image content. Simple color analysis methods cannot adequately express the semantic information in complex animations, which makes it challenging to study the color language.

Semantic segmentation techniques provide a way to solve this problem. By segmenting an image into regions with independent semantics, semantic segmentation enables a better understanding of objects and backgrounds in an image and associates colors with their symbolic meaning in a particular scene. Especially in animation, semantic segmentation can not only extract the boundaries and shapes of objects, but also identify the colors of different regions and their corresponding semantic information, so as to dig deeper into the expressive function

*HeNan Polytechnic Public Art Teaching Department, Zhengzhou 450046, China.(727842564@qq.com).

Fig. 1.1: Snow White Fragment.

of color language.

Therefore, this paper proposes to study the color language through semantic segmentation techniques and use the color-depth (RGB-D) image segmentation method to better understand the symbolic meaning of color in animation. This approach can effectively deal with the complex relationship between color language and image semantics, and provide a new perspective for understanding and analyzing the application of color in animation works.

If the color segments of different regions can be segmented, similar regions will present the same semantics, and information transmission can even be accomplished without subtitles. Consequently, the ideographic function of colour language in animation for film and television can be efficiently interpreted by semantic segmentation[5]. Traditional semantic segmentation methods generally use classifiers to perform pixel-level classification of artificial features and use conditional random fields (CRF) for refinement. However, the design of the classifier is generally aimed at a single category, and the classifier has a large training difficulty and high computational complexity when it is used for multi-category segmentation tasks [6]. In addition, the manually designed features have great limitations, resulting in poor model generalization ability and low segmentation accuracy. Deep learning-based semantic segmentation has demonstrated significant benefits thus far. It is capable of end-to-end training in addition to multi-category segmentation. Its varieties can be broadly categorised as color-depth (RGB-D), codec-based, and mixed with CRF. Three types of image fusion segmentation frameworks[7]. We propose a semantic segmentation method based on color-depth (RGB-D) images for better representation of color language in animations. When dealing with this kind of task, it is a major challenge to capture both color and depth information in an image and establish the semantic relationship between them. Especially in scenarios where color language and image language express different meanings, it is difficult for traditional methods to handle these complex semantic information effectively.

To solve this problem, we design a two-stream weighted Gabor convolutional network fusion method. First, a deep convolutional network is constructed by a Gabor directional filter to adapt to the orientation and scale variations of the image, so as to extract orientation- and scale-invariant features. Then, we extract image features from the color and depth channels separately and combine them into a lightweight feature extraction network.

The contributions of this paper are three points: (1) We present semantic segmentation as a means of understanding the ideographic role of colour language in animation for film and television. (2) Our suggestion is to use a two-stream weighted Gabor convolutional network for color-depth (RGB-D) picture fusion. Method of semantic segmentation (3) our approach is better than the RGBD picture semantic segmentation methods currently in use.

2. Related works. Colour matching in animated films and television shows must be determined by the story's narrative. The meanings that are created when various colours are combined also differ. It doesn't matter if it's colour matching within an image or colour matching across multiple photos. able to communicate the story's meaning.

2.1. In a single-frame composition, colour matching. When diverse colours are mixed in animation for film and television, the resultant colours frequently lose the original colour of the object. Together, these hues can convey a variety of connotations that together convey the main theme of the movie, particularly Certain spatial properties of colour elements can give a single image a dynamic beauty through various colour compositions, combining the storyline of the film and more effectively expressing its message. These spatial features are found in both cinema and television animation works [8]. When matching the colors of a single composition, the matching of colors can play a role in foiling, and several colors with relatively large contrast can be matched together, to form opposing colors, and people may feel opposed to one another when they see distinct colours. powerful visual impression that can support the plot's development [9]. When producing animation for film and television, the audience can be made to feel the emotions in the works by manipulating the colours. This can help to communicate feelings and emotions while also giving the viewers a powerful visual impact while they watch the animation. There will be enjoyment while working [10].For instance, the animation "Prince of Egypt" used a stark contrast between cool and warm tones to help the viewer distinguish between the good and the evil. When Jesus was crucified in Egypt, the film appeared black; during their fight, Moses appeared orange; and Pharaoh Ramses II is blue [11]. Characters and visual aesthetics in animated films and television shows can be shaped by colour choices. A single-frame composition can convey the stress of animation by combining contrasting colours. It usually has no visual effect in pure colours. When colours are combined, it can give the spectator a sense of movement [12]. For example, mixing cool and warm tones can work well and create great tension.

2.2. Colour coordination between images or lenses. Combining various colours in a single composition can produce a more logical impact and intensify the tension in the image. Especially in cinema and television animation works, continuity is created between the images by matching different colours to make the colour differences between the images obvious and enable more natural connections between the images, and compared with a single color, it can be more effective. It reflects the transmission effect of complex events [13]. After combining different colors between the pictures, it can show the sense of space of the picture, and allow the viewer to experience the change in time between various images. In animation, colour serves as a "time chain" in both film and television. It has the ability to produce many settings [14]. Colour is a type of expressive symbol that may be used to represent a movie's theme directly, set the mood, change and match colours, and then work in tandem with character language and actions to create the colour and theme of the movie. By coming to a consensus, it might make the movie's theme easier to explain [15]. For instance, the sun emerges from behind clouds in the opening scene of the motion picture and television cartoon "Prince of Egypt". The colours are used extremely well in this picture. The shot uses a mixture of white and blue colors. This kind of Cool tones give a sense of unease and make the viewer feel like something is going to happen right away.

As an important part of the model in this paper, the main role of the "pyramid pooling module" is to capture image information at different scales through multi-scale feature pooling. Simply put, pyramid pooling pools the input image features at different scales, enabling the model to focus on both global and local information in the image. This is important for complex scenes, especially in the presence of multi-scale objects, and the module can effectively alleviate problems caused by differences in object size and location.

Specifically for the combination of RGB and depth image features, the pyramid pooling module is able to process these two types of features separately to extract rich semantic information at different scales. For RGB images, it can help capture color and texture features; for depth images, it can better extract spatial contours and edge information. By pooling RGB and depth features at different scales and fusing them in subsequent steps, the model is able to more accurately deal with the problem of scale differences of objects in the scene, which in turn improves the accuracy and robustness of semantic segmentation.

This multi-scale processing makes the pyramid pooling module have obvious advantages in complex scenes, especially in dealing with objects of different sizes and shapes, and it can provide more accurate feature representation capability.

2.3. Semantic segmentation. The encoding and decoding-based segmentation method is divided into two parts: encoding and decoding. The encoding part is used to extract features, while the decoding part is used to gradually recover the lost spatial information, including U-Net, fully convolutional neural network (FCN),

Signet and DeepLabv3+ et al. The paper [16] regards the semantic segmentation problem as an instance segmentation problem and proposes a deep deconvolution network Deconned. The method firstly performs pixel-by-pixel category label recognition, predicts the segmentation mask, and then sends it into the network to obtain a combination of segmentation results through training, so this method can handle objects of different scales and enhance the processing of image details.

However, the extraction of target candidate boxes requires a lot of time and storage space, and the process is complicated, making it difficult to achieve fast and accurate segmentation. The paper [17] added the global context information to the fully convolutional network, and proposed a Parse Net network, which uses the average feature of any layer in the network to represent the feature of each position, It is employed to enhance the model's segmentation performance by capturing the image's global semantic information. The paper [18]proposed an improved symmetric encoding-decoding network SIPRNet, It builds an end-to-end semantic segmentation network by fusing semantic information and image data using the pooling index and convolution. Performance segmentation. In order to solve the problem that the up-sampling operation in the fully convolutional network cannot compensate for the loss of information, aurous convolution is used to achieve multi-level context aggregation, and the resolution of the feature image is not reduced. This method improves the performance based on FCN. Suggested an FCN model for the semantic segmentation of high-resolution remote sensing pictures that takes class imbalance into account. The end-to-end accuracy of small class prediction was improved by using the adaptive threshold method and the weighted cross-entropy loss function. Researchers have suggested techniques like aurous convolution and deconvolution to try and recover as much of the feature information that was lost during the downsampling process. However, aurous convolution requires a lot of processing power and storage space, while deconvolution is unable to restore low-level features. Both are not helpful for quick and precise semantic segmentation. suggested a network of many extraction stages. Refine Net to combine feature maps with varying resolutions and combine data that was lost during downsampling by using a high number of residual connections. Furthermore, residual connection and identity mapping are used to achieve end-to-end training. This can improve the segmentation impact by giving each layer's features a distinct pertinence. suggested an approach for semantic segmentation of polarization synthetic aperture radar images that combines the depth characteristics of each. Designed with a dense up sampling convolution (DUC) and hybrid dilated convolution (HDC) structure, the former can be utilized to expand the network and provide predictions at the pixel level, while the latter can be used to capture and decode missing information in bilinear up sampling. In order to prevent grid issues brought on by aurous convolution procedures, the receptive field is employed to aggregate global information. In general, lighting effects can be overcome and more edge and spatial information can be provided for indoor scene semantic segmentation by integrating RGB and depth images and taking advantage of depth channels. Unfortunately, the majority of existing semantic segmentation techniques either do not take into account the utilization of depth channels for context inference or use relatively single-source depth information that is solely utilized to create regional-level features. Additionally, the complexity of network training is increased by the filters' inability to leverage previous knowledge, such as the orientation and scale of indoor objects.

3. Methods. Fig.3.1 illustrates the general layout of the model used in this paper, which is centered on the convolution operation based on weighted Gabor directional filters, and combines the feature fusion and encoding-decoding frameworks in order to improve the model's ability to express the features abstractly. Specifically, the model in this paper takes RGB and depth images as inputs, extracts the depth features of the images through a novel wide residual Gabor convolutional network, and then pools the RGB and depth image features using a pyramid pooling module, respectively. The dual-stream features at different scales are acquired by this module to alleviate the object difference problem. In addition, RGB and depth image features of different scales are cascaded and fused. The fused multi-scale features are up-sampled and cascaded features in the form of decoding to obtain the fused features containing information about different scales, which are finally fed into the SoftMax classifier for classification.

Here, the role of the SoftMax classifier is particularly crucial.The SoftMax function converts the output of the model into a probability distribution, especially in multi-category classification tasks, where it is able to differentiate between features of different categories and assign a probability value to each category. In this way, the model is able to identify which category it belongs to from the multiscale fusion features extracted

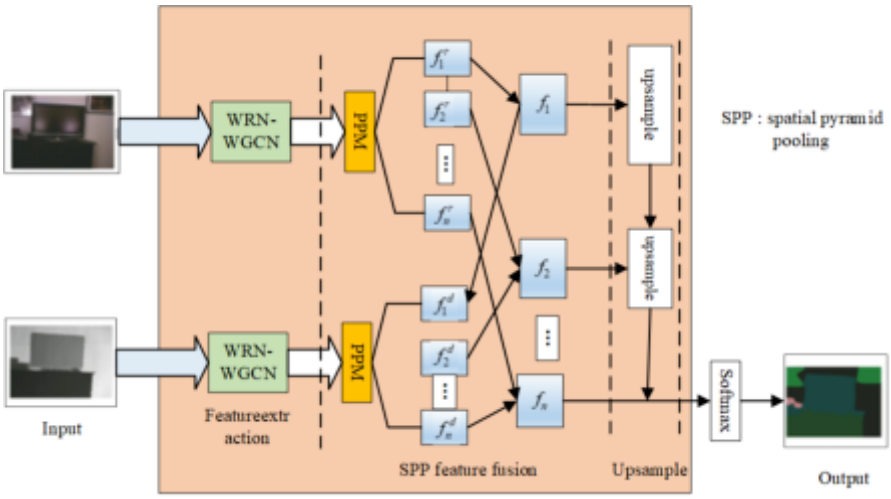


Fig. 3.1: Semantic segmentation of RGBD images fused by dual-stream weighted Gabor convolutional network.

from the RGB and depth features of the input image, thus accomplishing the multiclassification task.

3.1. Weighted Gabor directional filter. Deep convolutional neural networks have a long training period, high space complexity, difficulty adapting their extracted features to changes in direction and size, and an inability to dynamically update their own parameters based on feature differences. There are certain benefits to using traditional filters for feature extraction. Features that are invariant to spatial change are extracted through focused image processing, and feature redundancy is frequently lower than that of deep convolutional neural networks. The Abor filter exhibits good qualities in obtaining the target's local spatial-frequency information because it is a filter that closely resembles the basic cellular visual stimulus response. The Gabor filter resembles the convolutional neural network's shallow filter, according to the visualization results. A weighted Gabor direction filter is suggested, based on the developed Gabor direction filter. The convolution filter is modulated by the Gabor filter. The adaptability of the feature to the direction and scale is improved while lowering network parameters by modifying the convolution filter's feature extraction procedure. In order to highlight the differences between features in different directions, the method described in this paper involves the following specific processing steps: first, generate Gabor filters in different scales and directions; second, learn a weight coefficient for the filters; and finally, modulate the convolution filter. which, in order to make the output features flexible to the orientation and scale of the picture, generates a weighted Gabor orientation filter (Whoof) in each direction. Among them, Whoof is used as a filter with adjustable parameters, and the convolution filter is adjusted by the Gabor filter to enhance the expression of the feature map. The calculation process of the weighted Gabor direction filter is as follows: First, generate Gabor filters with U directions and V scales, weight each direction by learning a weight vector W, and then apply a learnable filter of size $N \times M \times M$. $M \times M$ represents the size of the two-dimensional filter and the modulation process is:

$$C_{i,u}^{v} = C_{i,o}^{o}[W.G(u,v)] \tag{3.1}$$

where u and v are the direction and scale indices, respectively; $G(u,v)$ is the corresponding Gabor filter; $C_{i,o}^{o}$ is the learnable filter; u is the dot product operation; $C_{i,u}^{v}$ is the modulation filter. Whoof can be expressed as:

$$C_{i}^{v} = \left(C_{i,1}^{v}, C_{i,2}^{v}, \cdots, C_{i,U}^{v}\right) \tag{3.2}$$

Since the Gabor filter has multiple directions, Whoof can be regarded as a three-dimensional convolutional filter, where the filter scale has different performances at different layers.

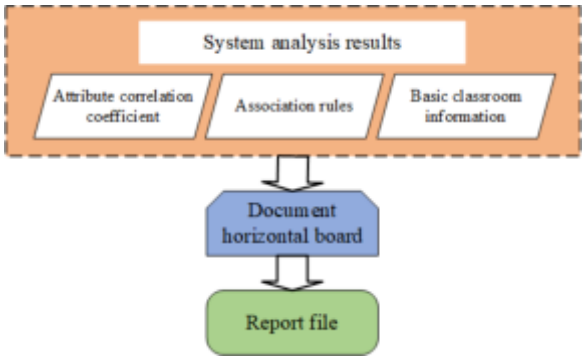


Fig. 3.2: Broad residual module. (a) initial residual module; (b) Extensive residual module L; (c) Extended residual module 2.

3.2. Wide Residual-Weighted Gabor Convolutional Network Module. The deep neural network model's layers are continuously deepened, which improves learning and allows for the extraction of richer characteristics. However, due to the influence of the gradient disappearing, the test accuracy of the model becomes harder to improve when the number of network layers reaches a certain point. Additionally, even as the training process goes on, the test accuracy gradually declines, and ultimately the loss function is unable to converge to the minimum value. To alleviate this problem, the residual module came into being.

The main idea is to use shortcut connections to skip multiple convolutional layers, and by continuously stacking modules, the network can continue to deepen without being affected by the disappearance of gradients. However, the deepening of the network makes the problem of decay feature reuse gradually prominent, resulting in only some parameters in the residual module participating in the update. A module-stacked network model is shallow and wide, so that a shallower model can be used to represent a deeper network. The background of the indoor scene is relatively complex, and there is interference from various illuminations, which makes feature extraction difficult. To construct features with high resolution, it is often necessary to use a deeper network model, and it is necessary to continuously fuse feature images between layers to alleviate the problem of excessive differences in multi-scale objects. To extract better features while building a lightweight network, this paper uses WRB to build a feature extraction network to extract RGB and depth image features respectively. The primary distinction between WRB and the standard residual module is the increase in the coefficient k and the quantity of convolution kernels. This guarantees the number of parameters while reducing the number of network layers, thereby fulfilling the goal of expediting the model training process.

Fig.3.2 displays a comparison of the structures of the various residual modules, with Fig.3.2a displaying the original residual module with two convolutional layers, batch normalisation layers, and REL layers. Figure 3(b) and Figure 3(c) represent two different WRBs, among them, X_1 and X_{1+1} are the input features of the lth layer and the two layer, respectively; FM is the feature learning with different number and width of convolution kernels. Compared with the original wide residual, which only adds different structure coefficients, WRB not only adds a different number of convolutional layers, but also adds different features.

The number of graphs, in turn, builds a wide and shallow network model. In this paper, the convolution filters used in WRB to build the network are all Woof's. On the one hand, the model is shallow by using WRB, and on the other hand, Woof's is used to extract the orientation and scale invariant features in the image, to better focus on the RGB image. It can improve the model's ability to represent information by extracting edge contour information from the depth image. To build a lightweight model and use a shallow neural network to achieve the same performance of a deep neural network, the wide residual-weighted Gabor convolutional network module in this paper has three broad residual groups and a network layer number of 13. The width k of each residual group is set to 4. In each residual group, the depth coefficient L determines the structure of the residual group. The first residual group GCConv2 and the second residual group GCConv3 adopt the structure of Fig.4.1.The specific structural parameters are shown in Table3.1. Fig.4.2 displays the schematic

Table 3.1: WRN-WGCN structural parameter configuration.

Group name	Output feature size	Block type
GCConv1	$N \times N$	[3×39]
GCConv2	$N \times N$	$\begin{bmatrix} 3\times 3 & 16\times k \\ 3\times 3 & 16\times k \end{bmatrix} \times L$
GCConv3	$N \times N$	$\begin{bmatrix} 3\times 3 & 16\times k \\ 3\times 3 & 16\times k \end{bmatrix} \times L$
GCConv4	$(N/2) \times (N/2)$	$\begin{bmatrix} 3\times 3 & 32\times k \\ 3\times 3 & 32\times k \end{bmatrix} \times L$

architecture of the WRN-WGCN module structure. First, the Gono is used to convolve the input image, and then the features are extracted through two wide residual groups, and finally the corresponding feature image is output.

4. Experiments.

4.1. Data set and experimental platform. NYUDv2 dataset the experiments in this paper are carried out on the mainstream semantic segmentation dataset NYUDv2, which contains 1449 densely annotated RGB and depth image pairs (image resolution is 640pixel×480pixel) and contains 44 scenes and 35064 images. target, with 894 target categories. Figure 8 shows an RGB image, a depth image, and a semantic label in the dataset. In the experiment, 40 types of semantic labels are used, and the dataset is divided according to the standard division strategy. There are 654 photos used for testing and 795 images utilised for training. During the training phase, the photos undergo flipping, translating, cropping, and colour jittering in order to address the issue of inadequate data.

4.2. Experimental results and discussion. This work presents a method comprising many functional modules, including a weighted Gabor direction filter, pyramid pooling feature fusion module, and wide residual convolution module. To validate the effectiveness of every module, the proposed method is extended to generate four alternative approaches. Variant model 1 should be set to Baseline in order to compare each module's effectiveness: Using an RGB image and a depth image as input, a conventional convolutional neural network is used to extract image features. Feature cascade is then used for fusing, and the network's convolution operation uses a conventional convolution filter. Variant approach 2 is configured as WRN-CNN in order to confirm the efficacy of the suggested wide residual feature extraction network: RGB and depth images make up the input, the wide residual network is used for feature extraction, and the two features are directly graded. Union and fusion, where the wide residual network consists of regular convolutional filters. In order to confirm the efficacy of the suggested weighted Gabor direction filter, variation method 3 is configured as WGCN: the weighted Gabor direction filter is utilised as the convolution filter with RGB and depth images as input, and the two resulting features are directly Fusion and cascading. Compared to variant model 1, this method replaces the weighted Gabor directional filter with a regular convolution filter. To verify the effectiveness of the feature fusion method proposed in this paper, the variant method 4 is set as PP-Fusion: taking RGB image and depth image as input, extracting features through conventional convolutional neural network, and utilising the fusion module's pyramid pooling capability for fusion.

In this paper, a training set is constructed based on the NYUDv2 data set, and the semantic segmentation model in the indoor scene is obtained by training, and then quantitative analysis is carried out on the test set. This work additionally tests the model and related techniques on the SUN-RGBD dataset, evaluating and visualising the results to confirm the model's generalisation performance. To verify the effectiveness of the design of each module, this paper designs ablation experiments, constructs different network models according to four variant methods, trains and tests the models, and obtains evaluation indicators. Additionally, employing FCN and Signee semantic segmentation, the approach in this work is contrasted with the current classical methods. Table4.1 displays the quantification results on the NYUDv2 dataset.

Based on the quantification results of the NYUDv2 data set, the network model based on the weighted Gabor

Table 4.1: Comparing the outcomes of various segmentation methods using the nyudv2 dataset.

Method	WRN-CNN	WGCN	PP-Fusion	Block typeA_{cc}%	mA_{cc}%	$mlon_{cc}$%	$Wmlon_{cc}$
Ours	.	.	.	60.4	50.9	40.2	53.2
Variant1				58.4	41.7	30.2	45.9
Variant2	.			58.7	42.5	31.8	45.4
Variant3	.			60.9	45.3	35.9	50.5
Variant4			.	63.3	45.9	36.5	46.7
FCN				65.5	45.2	34.4	48.7
Seg Net				56.3	47.7	35.2	50.2

Table 4.2: Comparison of various segmentation algorithms' output on the sun-robed dataset.

Method	WRN-CNN	WGCN	PP-Fusion	Block typeA_{cc}%	mA_{cc}%	$mlon_{cc}$%	$Wmlon_{cc}$
Ours	.	.	.	58.3	38.6	28.3	42.1
Variant1				45.3	33.8	21.9	38.5
Variant2	.			44.9	34.6	23.2	38.7
Variant3	.			54.7	35.2	27.4	37.8
Variant4			.	56.2	35.7	26.2	36.4
FCN				49.6	36.6	23.7	35.9
Seg Net				48.9	34.7	26.3	38.4

direction filter has greater advantages over the standard convolution filter network, and the four evaluation indicators are enhanced by 2.5% when compared to the baseline model 6.6%, 5.7%, and 4.6% demonstrate how the learnable convolution filter may be modulated with Gabor filters of various orientations and scales to efficiently extract features from the image, reducing the likelihood that orientation and scale variation interference will affect the segmentation results. To a certain extent, the segmentation performance of the network model can also be enhanced by the multi-scale fusion of RGB image data and depth image features. The accuracy of the proposed method has significantly improved when compared to the classic semantic segmentation models FCN and Signet. The Four index has also improved by 4.5% and 3.0%, respectively. This indicates that the method is effective when combined with dual-stream images and that multi-scale feature fusion and invariant feature extraction can enrich the information expression of images. The segmented images are labelled with different colours in order to intuitively depict the experimental results, and this yields the semantic results that are displayed in Fig.4.1. The figure shows that the traditional semantic segmentation techniques, FCN and Signet, are generally capable of segmenting a variety of objects. This suggests that the encoder-decoder structure may be able to recover certain features, but its fine-grained performance makes it unsuitable for small-scale objects. Even more so. Thanks to the benefits of the proposed pyramid pooling feature fusion module in multi-scale processing, the segmentation accuracy of some objects with smaller scales is rather good, and the segmentation of some objects' edges is more refined. In detail, the weighted Gabor direction filter-based model described in this study performs better in terms of both direction and scale invariance. This suggests that features linked to direction and scale invariance are extracted, and that this has an impact on the segmentation effect. WRB's dual-stream network model of colour and depth images fully leverages a range of various information representations in the image, offering definite advantages in object integrity and fine edge segmentation.Cross-dataset experimental results are a valuable tool in the research of picture semantic segmentation because they allow different approaches' generalisation performance to be tested. This paper trains the model on the NYUDv2 dataset and tests it on the SUN-RGBD dataset to validate the cross-dataset segmentation outcomes of the proposed technique. The SUN-RGBD dataset has 37 semantic categories and 10335 registered colour and depth image pairs. Only 37 common labels are quantified since the label set in the SUN-RGBD dataset is a subset of the label set in the NYUDv2 dataset. Table 4.2 displays the quantification outcomes of the various techniques.

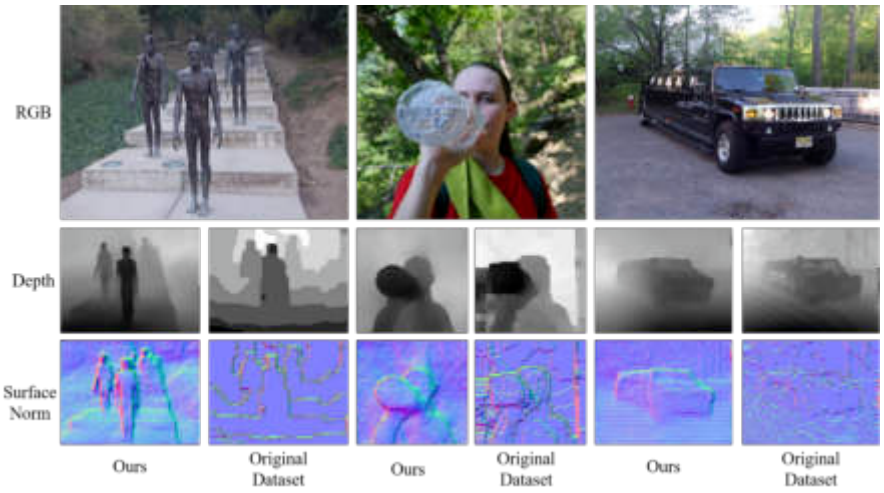


Fig. 4.1: Results of semantic segmentation using several techniques on the nyudv2 dataset. (a)RGB; (b) depth; (c)GT.

Although both NYUDv2 and SUN-RGBD datasets are derived from indoor scenes, they still have some differences in scene distribution, object categories, and labeling accuracy. Therefore, experiments on different datasets can comprehensively verify the generalization ability and robustness of the models.

Compared with the classical encoding-decoding semantic segmentation frameworks (e.g., FCN and Signet), the semantic segmentation method proposed in this study demonstrates stronger competitiveness and achieves certain performance improvement. By conducting ablation experiments on different modules, it can be found that each module contributes to the overall performance, especially the weighted Gabor direction filter and pyramid pooling feature fusion module, which have significant advantages in improving segmentation accuracy and model adaptation.

Fig. 4.2 shows the visualization results of different methods in the semantic segmentation task, and the colors represent different semantic categories. The experimental results show that the method in this study has better performance in detail portrayal, can accurately segment targets with large scale differences, and at the same time, effectively reduces the impact of environmental lighting changes and demonstrates strong scene adaptation ability. This indicates that the method has high practical value in semantic segmentation tasks in complex indoor environments.

4.3. Model complexity assessment. In order to create a relatively light network model, the approach suggested in this research takes into account both the model's complexity and the accuracy of semantic segmentation. Table4.3 compares the space complexity and minus value of different algorithm models. Considering the complexity of model space, Signet adopts pooling index for nonlinear up-sampling, and does not need to perform parameter learning in the up-sampling process, so the number of parameters is much less than that of the FCN method, but the segmentation performance of the two methods is comparable. This paper's model has a low complexity since it uses a large residual network for feature extraction, which drastically lowers the number of parameters. There are also less parameters when incorporating pyramid pooling. The Gabor directional filter also helps the network stay lightweight, which enables a more basic network to pick up complicated feature representations. When taking into account the model's temporal complexity, the network built using the conventional convolution filter has a longer inference time, whereas the network built using the broad residual network is shallower and offers some benefits during the inference process. When used in conjunction with the Gabor direction filter, this can efficiently shorten the model inference time by extracting the direction and scale features.

Taken together, this research method effectively reduces the model complexity and shortens the inference

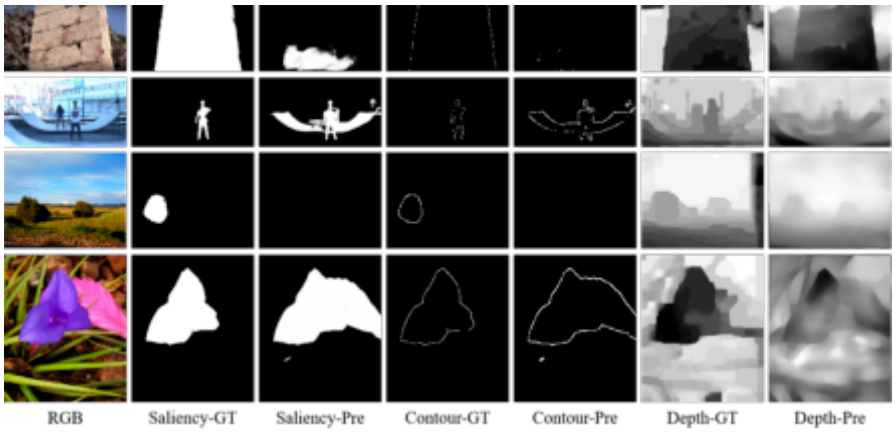


Fig. 4.2: Results of different techniques for semantic segmentation on the SUN-RGBD dataset.

Table 4.3: Comparison of the space complexity and reasoning time of several algorithms.

Method	WRN-CNN	WGCN	PP-Fusion	Model size/MB	Reasoning time/MS
Ours	.	.	.	118	43
Variant1				382	77
Variant2	.			116	36
Variant3	.			189	49
Variant4			.	246	52
FCN				548	44
Seg Net				127	59

time while ensuring high-quality semantic segmentation performance by adopting strategies such as wide residual network, large-scale residual feature extraction, Gabor direction filter and pyramid pooling, showing better lightweight characteristics and computational efficiency.

5. Conclusion. Our proposed approach addresses the challenge of color language expression in animation for film and television by introducing a semantic segmentation method for RGB-D images, leveraging a two-stream weighted Gabor convolutional network. The network first extracts image features, which are subsequently processed across multiple scales through pyramid pooling to capture both RGB and depth image features. The extracted dual-stream features undergo deep fusion at multiple scales via the pyramid pooling feature fusion module. To generate multi-scale fused features, the fused outputs are upsampled and cascaded in a decoding structure before being fed into a SoftMax classifier for final segmentation. Experimental results demonstrate that the proposed method achieves high-quality segmentation of object characteristics and edge contours, effectively adapts to variations in scale and orientation, and performs well in the semantic segmentation of indoor scenes.

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

REFERENCES

[1] Mirzakarimova, Z. D. *The acquisition of a subjective color as a result of the conversion of the meaning of the word.* ACADEMICIA An International Multidisciplinary Research Journal, 11(2),(2021) 1404-1413.

[2] Yanmin Xu, Yitao Tao, Chunjiong Zhang, Mingxing Xie, Wengang Li, Jianjiang Tai, *"Review of Digital Economy Research in China: A Framework Analysis Based on Bibliometrics"*, Computational Intelligence and Neuroscience, vol. 2022, Article ID 2427034, 11 pages, 2022.
[3] TU, Xing; WANG, DongRong; YANG, Qian. *Emotional Analysis in Animated Films Using Big Data and IoT: An In-Depth Study of'Krek'. In: Proceedings of the 2024 8th International Conference on Big Data and Internet of Things.* 2024. p. 175-182.
[4] Zeng, R. *Research on the application of computer digital animation technology in film and television.* Journal of Physics: Conference Series, 1915(3),(2021) 032047 (6pp).
[5] Xu, L. *Fast modelling algorithm for realistic three-dimensional human face for film and television animation.* Complexity, 2021(2), 1-10.
[6] Wang, S., Xu, Q., & Liu, Y. *Research on the creation of film and tv works based on virtual reality technology.* Journal of Physics Conference Series, 1744(3),(2021) 032015.
[7] Pan, Y. *Application of computer visual art in digital media art.* Journal of Physics: Conference Series, 1961(1)(2021), 012059 (6pp).
[8] Zhang, B., & Teng, Y. *A practical exploration of "ideological and political course" in film and television art education—take the "project training of 2d animation creation" as an example.* Open Journal of Social Sciences, 08(9)(2020), 229-236.
[9] Liu, X., & Pan, H. *The path of film and television animation creation using virtual reality technology under the artificial intelligence.* Scientific Programming, 2022, 1-8.
[10] Martin, A. E., & Baggio, G. *Modelling meaning composition from formalism to mechanism.* Philosophical Transactions of The Royal Society B Biological Sciences, 375(1791),(2020) 20190298.
[11] Avval, A. M., Hosseininejad, S. R., & Zahraei, S. H. *The most productive suffix - [i] of the persian language in the meaning of grammatical person and the ways of its expression in russian.* Bulletin of Udmurt University Series History and Philology, 30(3),(2020) 428-433.
[12] Stol, M. *The meaning of color in ancient mesopotamia. by shiyanthi thavapalan.* culture and history of the ancient near east 104. leiden: brill, 2020. pp. xiii + 509 + 30 plates. $163.00 (cloth). Journal of Near Eastern Studies, 80(1),(2021) 195-199.
[13] Kim, B.*A study on the content of eonmunjamo and the meaning of korean language education.* Korean Historical Linguistics, 32,(2021)33-80.
[14] Xu, L. *Fast modelling algorithm for realistic three-dimensional human face for film and television animation.* Complexity, 2021(2), 1-10.
[15] Pan, Y. *Application of computer visual art in digital media art.* Journal of Physics: Conference Series, 1961(1),(2021) 012059 (6pp).
[16] Arslan, G., & Gkearslan, A. *The use of clay animation in television advertisements.* e-Journal of New World Sciences Academy, 15(1),(2020)52-70.
[17] Sene, R. *The socio-historical factor behind change in meaning: the case of old french.* Taikomoji Kalbotyra, 15,(2021) 26-36.
[18] Cao, J. *Research on the application of color language in computer graphic design.* Journal of Physics Conference Series, 1915(4), (2021)042033.

Edited by: Ashish Bagwari
Special issue on: Adaptive AI-ML Technique for 6G/ Emerging Wireless Networks
Received: Sep 6, 2024
Accepted: Feb 25, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 1985–2001, DOI 10.12694/scpe.v26i5.4748

ENHANCING SCALABLE USER EXPERIENCE IN SMART HOME SYSTEMS WITH UBIQUITOUS VIRTUAL REALITY INTERFACES

NATESH MAHADEV*, SOWMYA V L†, SHANKAR R‡, ANITHA PREMKUMAR§, SYED SALIM¶, AND RAJESH NATARAJAN‖

Abstract. The popularity of smart home systems has greatly increased over the past several years, which provide convenience, automation, and control over various areas of daily life at home. However, these technologies still have issues with scalable user experience, mainly because there aren't any engaging and intuitive user interfaces. Incorporating pervasive virtual reality (VR) interfaces into smart home systems is the unique strategy proposed in this study to improve scalable user experience. The study's goal is to use VR technology to develop engaging and simple user interfaces that seamlessly integrate with the actual surroundings of a smart home. Users can engage with their smart home appliances and services through a mixed-reality experience in which virtual items and information are seamlessly incorporated into their environment by donning portable VR headsets. This application considers various difficulties related to developing Ambient Assisted Living (AAL) solutions, including unique characteristics of each end user, appliance, technology, deployment, and data-sharing problems. The Smart Home Systems take advantage of Semantic Web technologies integration abilities and their capacity facing represent significant information into legal models. A virtual reality application called Smart Home Systems allows for residential settings' setup and customization in AAL systems. Additionally, it effectively uses VR technology to streamline the creation of specialized AAL settings. The application and underlying framework were evaluated for each through two scenarios: designing a home setting specifically for specific scalable user categories.

Key words: virtual reality (VR) interfaces, Ambient Assisted Living (AAL), Smart Home Systems, users, customized design, smart home appliances, scalable user experience

1. Introduction.

Incorporating ubiquitous virtual reality (VR) interfaces into smart home systems raises comfort and immersion in our daily lives. This state-of-the-art technology allows people to interact seamlessly with their homes and environs, managing numerous gadgets and getting information intuitively and immersive. Imagine entering your living room and being able to change the lighting, temperature, and audio-visual settings with a single gesture to create the ideal mood. Scalability users using VR interfaces may explore virtual versions of their homes, interact virtually with furnishings security systems, and even remotely manage energy use [1]. These systems are linked gadgets and appliances operated remotely by smartphone and voice commands. Scalable computing solutions for smart home technology enables customers to easily manage their homes, from adjusting the thermostat and shutting off the lights to monitoring security cameras and controlling entertainment systems. Smart homes adapt to our tastes and save energy while improving the quality of our lives with features like automatic routines and customized settings. Smart Home Systems have become a crucial component of contemporary living, enabling a more connected and intelligent way of life, whether boosting security, maximizing energy use, or creating a seamless entertainment experience [2]. A key factor

*Department of Computer science and Engineering, Vidyavardhaka college of engineering, Mysore, Karnataka, India (natesh.m@vvce.ac.in)
†Department of Artificial Intelligence & Machine Learning, BMS Institute of Technology and Management, Bengaluru 560064, India (soumyavl@bmsit.in)
‡Department of Computer Science and Engineering, BMS Institute of Technology and Management, Bengaluru 560064, India (shankar@bmsit.in)
§Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Bengaluru 576104, India, (Corresponding author, p.anitha@manipal.edu)
¶Department of Computer Science and Design, School of Engineering, Mysore University, Manasa Gangothiri, Mysuru, Karnataka 570006, India (prof.syedsalim@gmail.com)
‖Information Technology Department, College of Computing and Information Sciences, University of Technology and Applied Sciences-Shinas, Al-Aqr, Shinas 324, Oman (rajesh.natarajan@utas.edu.om)

that influences how people interact with their living spaces is user experience (UX), which is a key component of smart home systems. These intelligent solutions attempt to improve and simplify our residential settings by seamlessly fusing technology and daily living. The scalable user is at the center of a well-thought-out smart home system, which ensures the scalable user interface is simple to use, open to everyone, and attentive to their specific requirements [3].

Residents can easily control and personalize their settings thanks to the scalability user experience of smart home systems, which range from simple voice commands to intelligent smartphone apps. These technologies anticipate human preferences through sophisticated automation and thoughtful design, resulting in a seamless and individualized living experience. Their connection with homes has changed due to the widespread adoption of VR technologies. These systems offer intuitive and engaging experiences by utilizing VR technology, converting our homes into intelligent spaces. Users may easily access and operate various smart home features with the help of pervasive VR interfaces [4]. The options are boundless, from controlling temperature and lighting to working security and entertainment systems. These user interfaces offer more interaction, enabling users to see and change their houses in real-time. Users may easily customize their smart home settings and make educated selections using accurate simulations and virtual representations. Smart home solutions using ubiquitous VR interfaces improve comfort, convenience, and customization while altering our relationship to living surroundings [5].

Additionally, the immersive aspect of VR might pose possible health hazards, such as motion sickness or visual discomfort, particularly for people predisposed to such conditions. Additionally, those with physical limitations or those who are uneasy with technology may use interfaces that rely on VR. Further, the requirement for a steady internet connection for efficient functioning may be a drawback in places with low connectivity or during network outages [6]. Finally, when smart home systems collect and process personal data connected to pervasive VR interfaces, privacy and security problems may surface, increasing the danger of unauthorized access or data breaches [7]. As scalable smart home systems gain popularity for their convenience and automation capabilities, it faces significant challenges in scalable user experience due to lack of engaging and intuitive interfaces The study aims to create integrating pervasive VR interfaces into smart home systems, creating a mixed-reality environment that enhances user interaction with appliances and services. However, developing these Ambient Assisted Living (AAL) solutions requires addressing diverse user characteristics, appliance interactions and data, sharing complexities. The proposed VR application seeks to streamline the customization of smart home settings, yet its effectiveness and usability in real-time scenarios remain to be thoroughly evaluated. This study's novel approach to enhancing scalable user experience involves integrating pervasive VR interfaces into smart home systems.

Key contributions of this research:

1. The scalable user experience and functionality of smart home systems can be improved by including ubiquitous Virtual Reality (VR) interfaces.
2. Improved handling and interaction and pervasive VR interfaces may offer a more immersive and intuitive approach to handling scalability smart home appliances.
3. To build personalized and context-aware experiences within scalable smart homes, VR interfaces may take advantage of user preferences and contextual data.

2. Related work. The study [8] provided a unique Augmented Reality (AR) decision support system (STARE) that enhances the user's focused objects with collections of semantically relevant Internet of Things (IoT) data and matching suggestions. Research [9] determined End users may use the smartphone camera to frame the required sensor or item through the produced prototype, then receive the current automation, amend their definition, build new ones, and watch the automation involving the full present environment. The study [10] intended to give references on effective setup of interaction systems for various purposes, throw light on future development directions in this field, and aid researchers in better understanding the function of cross-device Interaction in digitalizing a complex smart home system. The study [11] investigated the possibility of automation replacing normal interactions in Ambient Assisted Living (AAL) or Smart Home settings, generally dependent on the users' location. They particularly recommend the cutting-edge strategy of using the indoor location capabilities of Augmented Reality (AR) head-mounted displays (HMD) to detect, monitor, and identify occupants to automatically manage different Internet of Things (IoT) devices in Smart Homes [12]. The study

[13] created a vision-based detection pipeline and integrated AR device to track 3D edges and identify touch interactions between fingers and edges. The human-centered approach, although technical research into smart homes, has grown quickly, focusing on home automation, gadgets, software, and protocols. Providing organic, practical, cozy, welcoming, and secure user experiences in the smart home requires a human-centered [14]. The foundation for establishing smart manufacturing shop floor systems based on ubiquitous augmented reality technology (SSUAR) was described [15]. The study [16] used a smart home environment to include eye-tracking with a brain-computer interface. The paper [17] uses the interactive functionality of a smart home environment with the passive haptics of SR, offering a unique strategy we term Smart Substitutional Reality (SSR). The study [18] investigated pervasive augmented reality in the smart home setting, allowing users to pick a control by pointing at a specific appliance, streamlining the interaction process.

The paper [19] introduced IoT, a well-defined technique that combines self-sustaining predictions and simulations of trigger-condition-action rules to determine when these rules will activate and what state changes linked smart home entities will experience. The study [20] provided an AR framework for Internet of Things applications involving telemetry and home automation. The paper [21] suggested a gesture-based safe interaction system for IoT health devices in smart homes to serve elderly or special needs individuals. The study [22] planned for a career and a virtual adviser to co-construct the procedure. The virtual adviser's knowledge about smart homes, daily routines, and help is organized in an ontology. The study [23] described the usage of a mobile application as a user interface that makes it easy for users to manage a home appliance remotely using two different tasks: voice command and graphical user interface. The study [24] used smart home technology for regular environment-integrated blending. For instance, head-mounted displays (HMD) are an augmented reality (AR) that gives users a non-intrusive method to interact with gadgets and access privacy-related data. Research [25] evaluated the buttonless device, which produces both an infrared (IR) and a visible (VIS) laser beam, is made be utilized consistently and universally for a wide range of appliances and equipment in private homes.

Article addressed the limitations of traditional intrusion prevention systems in autonomous systems by proposing the Gannet optimized-mutated k-nearest Neighbour (GO-MKNN) approach [26]. The method leverages MKNN for improved detection accuracy and adaptability, ensuring the reliability and safety of autonomous vehicles. Study explored various methods for fall identification, emphasizing the importance of mobile phone sensors in fall detection [27]. It evaluations, discussed key challenges, innovation methods and advancements in this critical area of health monitoring technology. Scalable smart door access systems utilize traditional biometric methods like fingerprint and iris recognition, but face speed and accuracy challenges [28]. Advancements in computer vision, particularly AI, improved face detection and gesture recognition, enhancing security and scalable user experience.

Limitations While ubiquitous VR interfaces significantly enhance scalable user experience in smart home systems by providing immersive and intuitive interactions, there are notable drawbacks. High costs associated with VR technology limit accessibility for some users. The need for advanced hardware can create compatibility issues. Scalable users also experience discomfort or motion sickness during prolonged use. Furthermore, the massive data collecting needed for tailored experiences raises privacy issues. Finally, the learning curve associated with new interfaces could hinder user adoption, particularly among less tech-savvy individuals. To overcome scalable user experience challenges in smart home systems, this study suggests integrating pervasive VR interfaces. By creating intuitive, mixed-reality interactions tailored to individual preferences, leveraging Semantic Web technologies for enhanced data sharing, and conducting user feedback evaluations, the approach aims to deliver engaging and accessible scalable computing of smart home environments.

3. Smart Home Systems.

3.1. Domestic Environments and Universal Design. Traditional product and solution design is frequently focused on standard individuals and are fictionalized representations of actual people. This scalable method must consider the numerous factors about existing end scalable users, including their skills, knowledge, social connections, and needs. The Universal Design paradigm, also called create for All, is a strategy to surpass this restriction. It focuses on providing solutions that can adjust to each scalability user's demands while considering the different characteristics that identify actual human users. The fundamental concepts of this model are also included in AAL, which deals with problems about improving people's Quality of Life (QoL) at all

phases of life and giving them assistive devices for an independent existence appropriate to their capacities. To assist elderly or disabled persons in maintaining an independent and autonomous lifestyle, AAL systems search for effective solutions. Although numerous attempts have been made to build AAL solutions, most generated methods and apparatuses must consider the requirements and ignore their users' practical issues and personal contact. The initiative strives to incorporate Universal Design principles into inclusive and domotic settings.

Regarding various daily activities (ADL), the surroundings, called scalable smart homes, should be able to anticipate and adapt to the demands. Throughout its entire existence, designing a scalable smart home for families, seniors, people with disabilities, and individuals with impairments requires a variety of tools. Additionally, it necessitates paying close attention to the unique people that utilize the provided services and technologies and how they employ these functions. The huge quantity of heterogeneous data that these interactions might produce must thus be managed logically and effectively to make it accessible to both end users and distant users. With a focus on physical activity, the following subsections explore the requirements and potential for making it compatible with one with the household's equipment and tools.

3.2. Requirements and Needs of Users. As indicated above, an expected definition of a smart home is a residential building equipped with a collection of gadgets commonly referred to as scalable smart objects. The scalable Smart Home's equipment ensures occupants have comfortable and customized living arrangements. The numerous pieces of equipment used in traditional homes can carry out their functions independently and separately. In a scalable smart home, however, the appliances must cooperate dependably and predictably and gather, manage, exchange, and transmit information about the residents to achieve their comfort and well-being. Devices must work with other appliances, maybe specialized and multivendor, to increase dwellers' assistance for bettering, hygienizing, and securing their daily lives. However, these days, only home environments where appliances are installed using the same protocol stack and communication style can ensure this synergy among distributed equipment. Along with data and information models, services offered, and discovery procedures, targeted appliance interoperability should go beyond only communication interoperability. Reaching this degree of interoperability is a difficult endeavor that takes effort.

3.3. Cooperation and Interoperability among Scalable Smart Home Appliances. As indicated above, an expected definition of a smart home is a residential building equipped with a collection of gadgets commonly referred to as scalable smart objects. The scalable Smart Home's equipment ensures occupants have comfortable and customized living arrangements. The numerous pieces of equipment used in traditional homes can carry out their functions independently and separately. In a scalable smart home, however, the appliances must cooperate dependably and predictably and gather, manage, exchange, and transmit information about the residents to achieve their comfort and well-being. Devices must work with other appliances, maybe specialized and multivendor, to increase dwellers' assistance for bettering, hygienizing, and securing their daily lives. However, these days, only home environments where appliances are installed using the same protocol stack and communication style can ensure this synergy among distributed equipment. Along with data and information models, services offered, and discovery procedures, targeted appliance interoperability should go beyond only communication interoperability. Reaching this degree of interoperability is a difficult endeavor that takes effort.

3.4. Project's Methodology and System for Smart Homes. The framework's objective is to offer services and products that can accommodate weak and impaired users, enabling them to carry out tasks that would otherwise be difficult or impossible. The system provides services to all house residents and is not restricted to any particular scalable user demographic. The construction of the dwellers' home allows for the analysis of the interoperable appliances' behavior and the satisfaction of the individual user's physiological needs before their hard deployment in the actual house. Standard nomenclature for health-related features was used in the reintroduction of international classification functioning (ICF) framework, which was created to correctly measure users' health.

To facilitate communication between health stakeholders and provide a consistent, comparable depiction of people's operational experiences across the globe, ICF views an individual's functioning as a dynamic interplay between their personal, environmental, and health conditions. Due to its simple terminology that even non-clinical experts readily comprehend, many health sectors can utilize the categorization, reintegrating injured

Table 3.1: An illustration of the ICF's construction

b <<Body function>>	Component
b2<<Sensory functions and pain>>	Chapter-first level item
b210 << Seeing function >>	Second Level item
b2102 <<Quality of vision >>	Third level item
b21020 << Light sensitivity >>	Fourth level item

persons into the workforce. The categorization is split into functioning and disability is the first of two major parts describing the elements. Activities and Participation, Body Functions, and Body Structures; the second, Contextual Factors, provides a way to identify the influence of both environmental and personal factors. The relationship among a person's health and the environment in which they behave then describes how that person functions. A letter denotes each element, and numbers may add further information (Table 1). The appropriate category and its accompanying code can be chosen, together with a qualifier, to assess a person's functionality and disability.

In addition to the users' ICF categorization, the users' cardiorespiratory fitness (CRF) was considered to further our understanding of their physical state. When characterizing the user and creating Smart Home, which is made to also care for a scalable a user's health and well-being, it is important to consider the user's general physical health because it is essential in preventing the emergence of many illnesses. CRF describes a person's physical fitness by describing their capacity to engage in dynamic, moderate-to-intense exercise for an extended amount of time. Additionally, studies have indicated a negative correlation between it and the chance of dying and cardiovascular, pulmonary, and coronary diseases. As a result, CRF serves as a reliable measure of a subject's overall health condition and may be used to track the improvement or deterioration of the occupants of Smart Homes over time. Formal and shared explanations of ideas and relationships were created using semantic web technologies, including modelling domain knowledge into ontologies, residences, and appliances. Semantic web technologies may describe that devices work and services they offer, improving the semantic interoperability between them. Additionally, it is possible to set off certain inferences using "Semantic Web Rule Language (SWRL)" rules, such as deploying particular services and appliances to meet the demands of users. The Smart Home System (SHS) that created to simplify a-priori review and evaluate designing integrated appliance systems that provide AAL services to customers and their services. This program uses a virtual depiction of a home and its furnishings to enable domestic environment designers to simulate and create a home. Before actually hard-deploying the customized AAL solution, SHS makes it possible to create an accurate and thorough picture of the situation of aided users, including their daily surroundings. This is accomplished by letting the designer alter each device's features using the Semantic Web technologies' descriptions. The result is a house simulation that is as close to reality as possible, allowing the designer to arrange the appliances in response to changes in the occupants and their surroundings. The SHS creates an integrated and compiled picture of pertinent knowledge about the household environment through Virtual Reality and Semantic Web. This linkage makes it possible for real-time interaction and coexistence of virtual and physical objects, improving setup capabilities of appliances in practical settings. The digital equivalent of the actual domestic environment may also be defined and expressed formally using Semantic Web technologies. Fig.3.1 shows the scalable smart home system's function

3.5. The System Architecture. A program called SHS utilizes the stated integrated service-oriented platform. The major goals of this platform are to control information about the user's and home domains while enabling data flow across various equipment. The Virtual Home Framework is a framework that is built on four foundations:

1. The initial information foundation Home layer semantic, an ontology collection that explains important information in the above mentioned domains.
2. The second pillar is the virtualization scalable of a home environment, which enables the virtualization of devices like sensors and appliances.
3. Integration Services ensure data synchronization among the system's physical and virtual components.

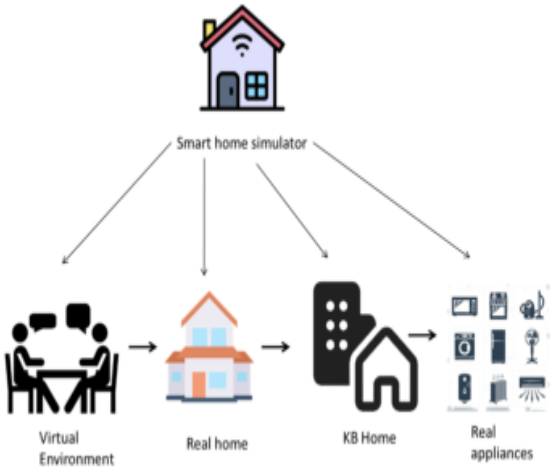


Fig. 3.1: The Scalable Smart Home System's function

This strategy is explored, and in this setting, apps may easily communicate with one another while information is integrated, distributed, and gathered using methods that are invisible to the clients.To facilitate and improve semantic synchronization generated by various home appliances and sensors:

(a) They make information gathering from any device possible.
(b) They make it possible to organize, analyze, and control the data received.
(c) When a device requests them or the required information becomes available, they communicate and dispatch it.

4. The collaboration between the three pillars of architecture results in Real Home, which displays proposed solutions using its equipment.

The SHS performs its role to enable designers to set up and test AAL solutions tailored to particular users.

4. Knowledge Base at Home. An ontology set called KBHome models, physical status information, scalable and information about the living spaces that make up the home and the installed and deployable appliances. The Web Ontology Language (OWL) and Resource Description Framework (RDF), along with Semantic Web Rule Language (SWRL), were chosen as the implementation languages for the knowledge base. Pellet was used as the reasoner to carry out reasoning operations on the ontologies. The four primary ontologies that makeup KBHome each target a different domain.

4.1. User's Cardiorespiratory Fitness and Health Ontology. Starting with his registry entries, this domain ontology describes ideas related to the scalable user's domain. The ICF codes and qualifiers used to describe the user's health state are the most important aspect of this ontology. This goal necessitated some re-engineering of the ICF ontology, which is accessible to the public via the Bio Portal: The particular ICF codes, which were first depicted as persons, were changed into datatype characteristics to enable the modeling of several health conditions using a single ICF code. As shown in the following calculation, the user's UP_2 maxvalue may be utilized to evaluate the customized workload of an activity.

$$WL(Cycle - ergometere) = (((UP_2max.0.65) - r))/n \tag{4.1}$$

The user must enter the data into an interpolation program to determine the angular coefficient m and user-dependent intercept q. According to ACMS, 65% of the recommended training intensity is the lowest, which might still increase users' CRF. To compute loads of exercise that are applied to a cycle ergometer, Equ.4.1 may easily be transformed into SWRL: As shown in Fig.4.1, it is feasible to express pertinent information about the user and some aspects of their physiological condition in this manner.

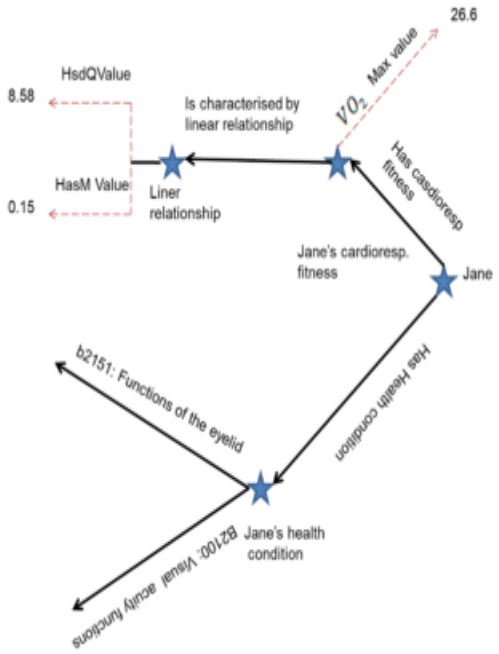


Fig. 4.1: An illustration of user characterization in the KBHome

4.2. Domestic Environment Ontology and Appliances. By describing the appliances and connecting them to the environment where they are used, this ontology intends to provide information about the equipment. The first module gives a straightforward depiction of atypical residential setting, complete with classes and people. This description is accomplished by utilizing the HicMO language, a collection of XML attributes that may describe the features of any device. HicMo details were transformed into datatype and object properties following the NeOn methodology's principles to develop XML descriptor semantics. This makesit feasible to offer that ID cardany equipment now installed and deployable in a home. Each program is listed individually and connected to one or more appliances in a submodule integrated with a scalable smart object's description. The third module enables describing these measures because each appliance can give one. Measurements that describe a user's physiological condition utilize a largely re-engineered that describes household items, and devices used, providing a unit ontology trustworthy explanation. Fig.4.2 shows a sensor that measures a user's body temperature.

4.3. Service coordination in the house. This ontology explains the actions brought about by a single or a combination of environmental or user-related conditions. It is possible to specify the circumstances under which a certain action is activated by using SWRL rules. This ontology included scenarios similar to the one above, outlining the events under which appliances can provide the necessary services. A complicated issue entails adopting tailored services for certain user categories with disabilities.

4.4. Arrangement of Living Environments. Appliances may provide useful services to their users, helping them to manage their limitations when doing regular activities, depending on their characteristics and the programs they contain. This ontology enables the classification of appliances into different varieties. Therefore, using the groups "Cognitive Impairment Appliance, Hearing Impairment Appliance, Motor Impairment Appliance, and Visual Impairment Appliance," household equipment is also categorized according to their appropriateness for specific users with disabilities. A structure project may be modelled by utilizing these aspects. With a distinct ID number and customized style, each task may be customized with the appliances and development. It is also possible to specify that project is meant for Fig.4.3. According to the programs offered for

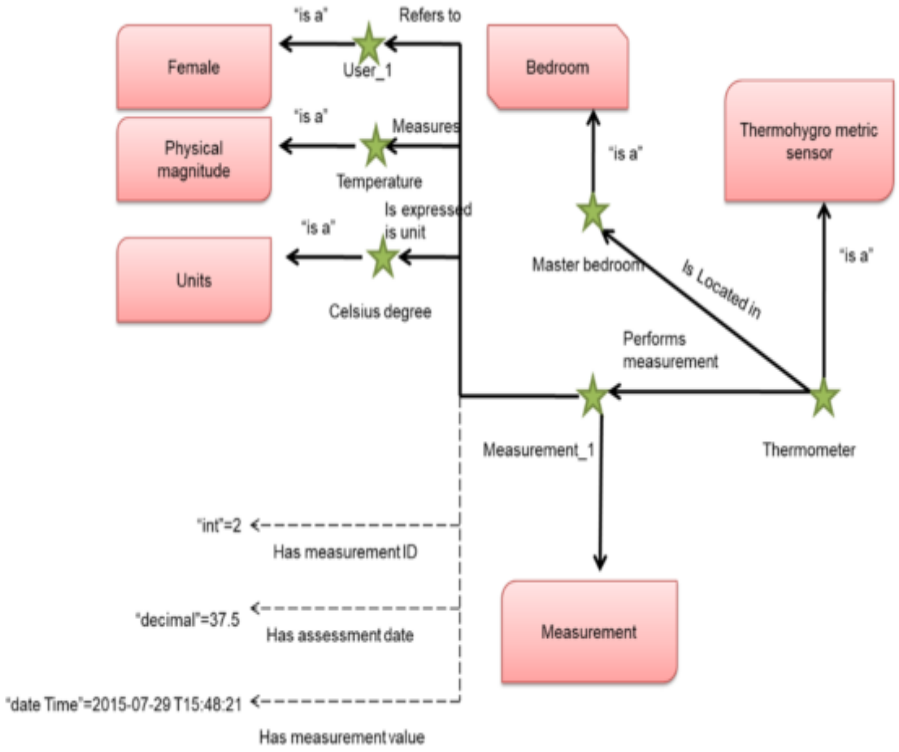


Fig. 4.2: A model of a KBHome application, as an example

the application, ontology also enables automatic inference of whether a device is acceptable for a certain type of disabled user.

In this method, a designer may pick from a list of devices that are assumed to be capable of coping with a user's handicap, and the results of this ontology's reasoning process can be used as a framework for supporting decisions throughout the living environment's design phase.

5. Scalability Smart Home Services Personalization. The offered solution must be customized if a scalable smart home is to be created to provide its residents with the tools and services to improve their autonomy and well-being in ADL. This process involves gathering the subjects' demands, assessing their health, and designing a unique solution that can assist users in controlling their disabilities and enhancing their class of life.

5.1. Identification of User. The physicians' involvement is necessary to complete the ICF-based module to assess the health state of the home occupant(s). The healthcare staff must evaluate all employing validated assessments and clinical scales created with a focus on a certain area, the scalable user's impairments, and general health state of older adults. Alternative methods call for estimating the CRF indirectly after measuring the subject's "Heart Rate (HR)." Additionally, they believe every one of a certain age has the same maximum heart frequency. Using a multistage workout strategy, the CRF of the Smart Home inhabitant may be found after specifying the tools and equipment required to assure subject's safety throughout test.

The selected protocol calls for a low-intensity warm-up phase that lasts for three minutes, following which the workload will rise by 20 W. Each element causing the increase in effort is described in full in Fig.5.1 and Table.5.1.

Since the ultimate goal of a test is subject's greatest attempt to push themselves over their physiological limits, trained clinical staff must continually check the subject's health during the test. As a result, the exercise

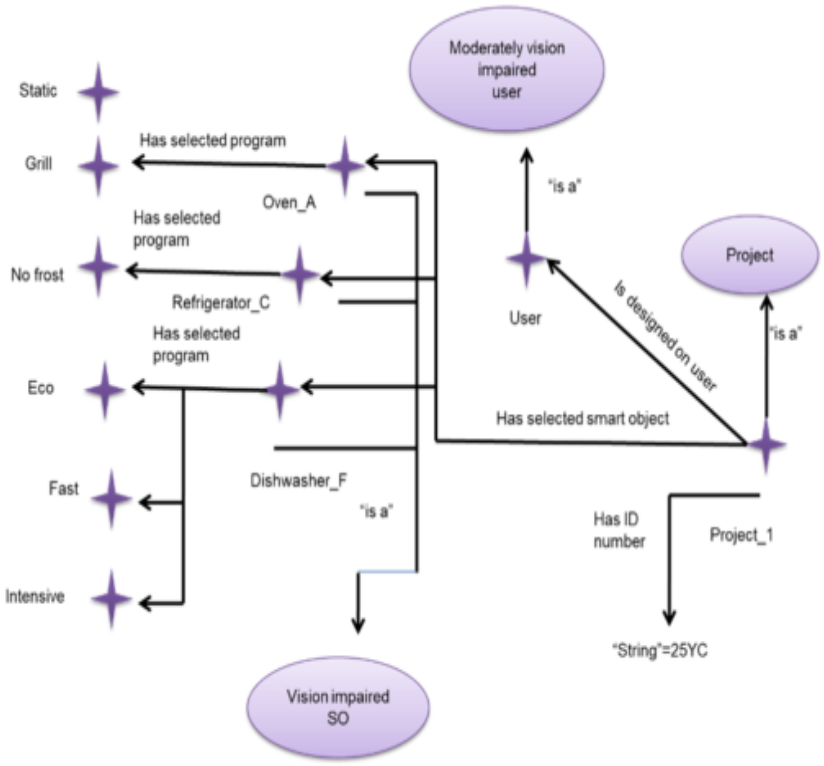


Fig. 4.3: Project designed for a user with visual impairment using the KBHome

Table 5.1: Test variables for the maximum amount of cycling activity

HR warm-up(bpm)	Ma ∗ WL(W)	Δ WL(W)	Starting WL
≤ 90	161	21	101
81≤ HR ≤ 91	142	21	81
91 ≤ HR ≤ 100	101	21	65
≥ 100	88	21	45

must end promptly if any of the circumstances indicated in Table.5.2 materialize. Additionally, if practice is interrupted, a person needs medical assistance until equilibrium and normal levels are achieved.

After completing these two evaluation rounds, each subject receives a detailed report on their health state, according to the international ICF standard, accurate information about their CRF, and as a result, their physical capacities. The latter would allow for personalizing each person's daily workout regimen. Because it is safer than the treadmill, the only other piece of exercise equipment that can directly manage the workload, a smart home arrangement, a cycle ergometer is kept as a training tool. Using a bike ergometer is acceptable for people with significant motor impairments or postural control problems. These critical issues are identified at first clinical evaluation, which is formalized by the completion of the ICF, and they need to be treated using specific methods, such as an arm ergometer.

5.2. The scalable smart home's user's configuration and test. The scalable Smart Home designer may create a mixed reality setting that can provide the finest answers in response to the user's needs by starting with the informational bits obtained during the user assessment phase. In this situation, employing mixed reality and virtual reality has several benefits. As was already indicated, the first is the capability of

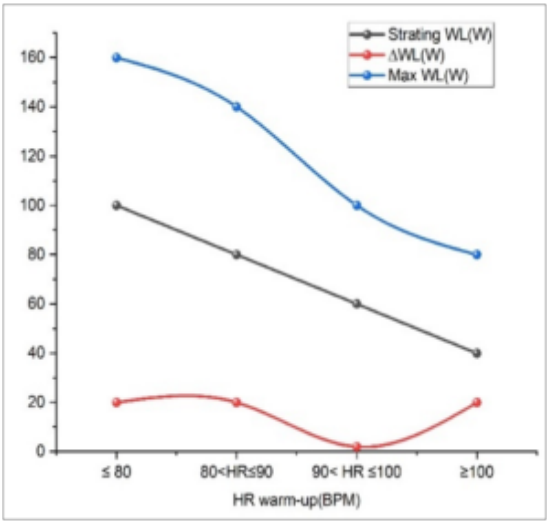


Fig. 5.1: Testing parameters for the highest level of cycling exercise

Table 5.2: Conditions for stopping the bicycle exercise test at maximum intensity

Conditions for test interruption
Loss of limb coordination
Sudden pallor
decreased blood pressure
Dizziness, faintness, or confusion
Severe desaturation (SpO2 < 80%)
Hypertension (HT)
ECG alterations indicating ischemia or ectopy
Chest pain suggests ischemia
Dyspnoea
Heart rate is higher than 85% of the highest estimate of HR

testing the different equipment's communication in real-time. The second option is offering an end-user chance to encounter the VEs first-hand with dual goals of familiarising themselves with their newly configured scalable Smart Homes and providing feedback to the designer that may enhance a final home design. Finally, a designer of a smart home and final user may assess that sensor and appliance performance without actually owning it thanks to the usage of virtual reality and semantic models, saving time and money; resulting numbers represent the impact on PMV value a scalable home environment's variable elements alter and make sense. As shown in Fig.5.2 and Table.5.3, the impact of temperature and relative humidity on the PMV value is calculated.

5.3. Customized scalable Smart Home Inauguration. The designer can model the ultimate users' existing home using an authoring tool in this step after receiving the blueprint and reconstructing a digital model of the interior spaces. The creator can enhance the digital representation of the area with additional devices and appliances house using this program by selecting them from the ones portrayed in the semantic catalog specified. The user's preferences and peculiarities were put in a semantic repository and queried via SPARQL to update the catalog in real-time inside the VE. Following the scalable architecture mentioned above, which enables data flow between the physical and virtual worlds, a designer might include actual hardware and digital illustration of a scalable smart home with sensors. The designer may save and preserve the project for future adjustments after creating customized surroundings and validating communication between the gadgets,

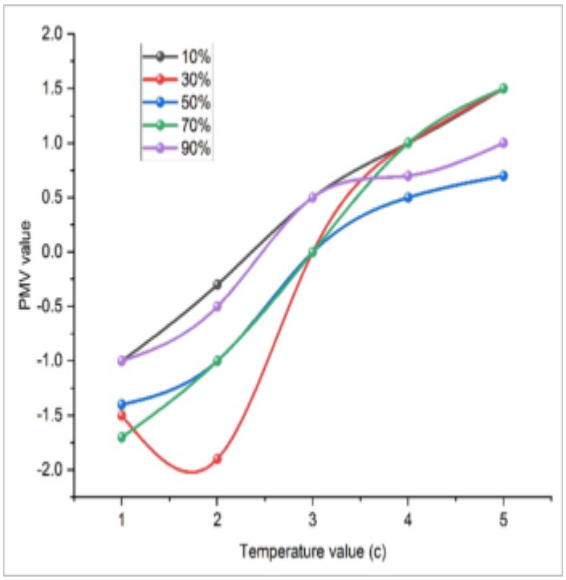


Fig. 5.2: The impact of temperature and relative humidity on the PMV value

Table 5.3: Temperature and humidity affect PMV

Temperature value (c)	PMV value				
	10%	30%	50%	70%	90%
1	-1	-1.5	-1.4	-1.7	-1
2	-0.3	-1.9	-1	-1	-0.5
3	0.5	0	0	0	0.5
4	1	1	0.5	1	0.7
5	1.5	1.5	0.7	1.5	1

sensors, and human users.

5.3.1. The designed solution is being tested. The scalable smart home system allows the end user to evaluate a solution that was particularly built following their characteristics and demands after the design phase. The explanation could implement various stages contingent intended user's features and kind of setting. Because of the greater sensation of presence they evoke and the more natural engagement they frequently offer, immersive and semi-immersive experiences undoubtedly represent the most promising way to test various scenarios and understand that Smart Home services operate. The target user must be carefully considered when selecting VR technology, nevertheless, since there is no denying that utilizing head-mounted displays poses a danger of unpleasant events and illness for fetus elderlies or severely cognitively handicapped individuals, as shown in Fig.5.3 and Table.5.4 scalable Smart home security has grown in value recently. Revenue is predicted to reach $5 billion in 2020 and 8.2 billion in 2023.

Therefore, other solutions ought to be chosen even in return for a diminished sensation of presence. When using real appliances or equipment that are part of the final solution, it is best to utilize non-immersive surroundings. If the interaction is accurately recreated in the virtual world, it might become easier and more manageable. Once the best solution has been found, it may be tested again in the configurator before being installed by the end user at their real location.

6. Real-World Use Case Validation. Two actual use cases served as a basis for SHS of proposed framework; in contrast to the first, second concerns the design of a kitchen for a user that is visually challenged

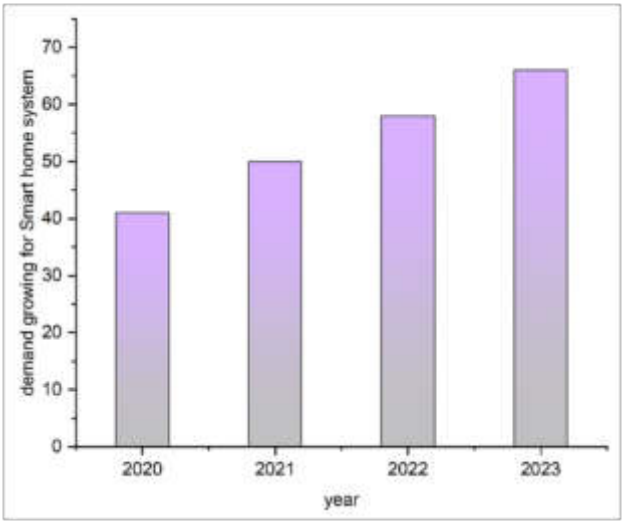


Fig. 5.3: Demand for smart home systems is increasing

Table 5.4: Scalable Smart home demand rising

year	demand growing for Smart home system
2020	41
2021	50
2022	58
2023	66

and addresses the topic of active aging by imagining an older user participating in physical activity performed at home.

6.1. Establishment Inside a Real Environment: The Kitchen. A kitchen designer in this case study must utilize the SHS to design a kitchen while choosing the best equipment; a person with a slight vision impairment and hyposmia is the intended user. The designer builds a virtual representation of the kitchen using the user's kitchen plans. In this approach, user chooses the appliances and sensors to a list provided by the KBHome that is suitable for their impairment(s). The KBHome gathers information for this user.

1. Two induction cooktops: one with textured button surfaces that may provide high contrast, other having a black cover, controls with a high degree of difference and a lit digital display.
2. Two convection ovens have a set of control lights, a digital display, and a backlit display.
3. A total of four dishwashing models featuring backlit, high-contrast displays.
4. Two refrigerator types featuring digital, backlit interfaces that show the current interior temperature and light.

6.2. Assessment of the Application. Following the formative assessment process, the configuration part validation is carried out. This kind of validation entails watching and evaluating empirically how layout of the kitchen is the job that causes representative users to engage with the VEs. This review uses qualitative and quantitative data to pinpoint usability issues and gauge user learning curves and task performance. Five designers will participate in the study to do this; during the initial training phase, which will last around 15 minutes, they will be given a situation and a scalable Smart Home System program. A computerized model of an empty kitchen will also be made available to them as a starting point for their work to make it easier and quicker. Timing issues, mistakes, possible software flaws, and user feedback will all be recorded throughout the configuration step. For the System Usability Score (SUS),each participant will get a questionnaire after the

Table 6.1: Conditions for stopping the user's cycling exercise test during the domestic activity

Conditions for test interruption
"Severe" pain
Decreased blood pressure
"Very strong" perceived e=ort
Hypertension (Ht)
Heart rate is higher than 85% of the highest estimate of HR

experiment. Additionally, open feedback will be gathered and considered to modify the program to suit the intended audience's requirements better.

6.3. Environment Setup for Physical Activities in the Home. In the first use case, a fragile 72-year-old woman must exercise daily on a bicycle-ergometer set up in her bedroom, a common scenario for older individuals. It was discovered that her CRF was 18.3. The actress is provided a suitable physical exercise to complete while keeping track of her outcomes and conditions. Using a tablet, the user is provided with performance-related information. She is asked to go into the bedroom to complete her daily physical exercise, and a presence sensor there recognizes her. It next expected to set up the bedroom for the workout by venting it and waiting until it reaches the right temperature. The user may put on the sensors and start the activity once the temperature and air quality are appropriate. The exercise's workload will be determined based on the user's health situation. The physiological condition is watched as she engages in the activity. The information on the physical activity and the physiological measures identified is kept and made available to the carers after the exercise session. The scalable user receives all the instructions and may utilize a virtual tablet to evaluate her work in real-time and even get other notifications depending on sensor data.

The designer creates a virtual depiction domestic setting with an actuator to allow automated window opening and closing, incorporating actual sensors into the environment utilizing SHS characteristics. A presence sensor, an ambient temperature, and a sensor for air quality, which can gauge room's CO2 content, are all included in the virtual environment. These sensors may all be used to determine whether the user is in the space. The KBHome contains information on user's CRF and state of health. The user may navigate the virtual environment and touch the tablet to receive instructions to prepare the space for her activity and click the exercise icon. The bike ergometer is in the bedroom, and she is first told to enter. She may be detected by the presence sensor, which also activates the room's air quality sensor. It determines the amount of CO2 present in the space and stores the information in the KBHome, where it is analyzed.

6.4. Validation on Elderly Subjects for the Physical Activity Configured Environment. The same procedure is used to validate older adults. A modest sample of target customers is engaged in a first-pilot study to examine possible software concerns in a designed scenario.The enlisted subjects must include the following requirements: age 65 and older, with a mild-to-moderate impairment of total physical stamina, and be deemed by a physician to benefit from light daily physical activity. Severe cognitive or eyesight impairments and incapacity to give informed consent are exclusion criteria. Each participant should receive a clinical CRF assessment after enrollment. CRF results will determine the target workload after the test.

Attempts are made to closely resemble the specified circumstance in the setting to verify an elderly-focused scenario. Therefore, a whole room simulates the situation, including a cycle ergometer, blood pressure monitor, pulse oximeter, tablet, presence sensor, automated window, air quality sensor, and virtual temperature. A wall projector will show the intended user all of the virtual items. Each topic will get instruction from knowledgeable staff who will provide an overview of each component and the system's goal and functionality. The participants can engage with the system when comfortable with the setup. Still, they will remain under continual clinical and technical staff monitoring. They will follow the instructions provided on the tablet. After opening and closing each subject's (virtual) window and setting the workload following CRF, each issue will cycle for 20 minutes. The measured parameters go beyond the limits in Table.6.1 to assure the safety of the training at all times. Collecting impartial quantitative data will validate this scenario and the subject interviews using a specially created questionnaire. Semi-structured interviews were deemed the best method for gathering

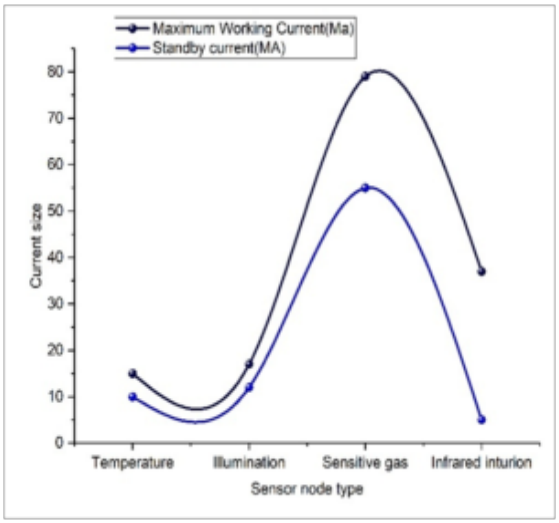


Fig. 7.1: Sensor node type

Table 7.1: The typical sensor node operating current

Sensor node type	Current size	
	Maximum Working Current(Ma)	Standby current(MA)
Temperature	15	10
Illumination	17	12
Sensitive gas	79	55
Infrared intrusion	37	5

qualitative information needed to evaluate the acceptability and utility of the proposed system according to the population's characteristics.

7. Analysis of a scalable smart home system for virtual reality.

7.1. The scalable smart home node's power consumption test. The survey's intelligent home technology collects data in battery-powered mode; therefore, the terminal node's service life depends on how much power it uses. This part primarily evaluates the ZigBee network node's power usage, which helps predict the node's battery life. For data gathering, two dry cells power sensor nodes. More than 2.5V must be present in the battery. The node is unable to ensure regular functioning otherwise. The data collecting nodes for temperature and humidity, lighting, and power consumption test principally uses high-sensitivity gas detection and an infrared intruder detection checkpoint, as shown in Fig.7.1 and Table.7.1.

Formula 7.1 provides the computation for an average working current, assuming that current in node's operating mode isc, the active time is J_j, and current in sleep mode is J_i, and the sleep duration is S_j.

$$J_avg = \frac{(c * S_j + J_i * S_i)}{(S_j + S_j)} \tag{7.1}$$

The average operating current (Iavg) of all tested sensor nodes is calculated using formula (2). The second design makes use of a 2000mAh battery. Calculating the working period of the sensor node using formula (3) is possible, given that the average battery discharge depth is around 60%.

$$S = 2000x0.6/J_avg(day) \tag{7.2}$$

The formulas (2) and (3) may be used to determine the duration of sensor node operation using test data. Calculate the ZigBee terminal's temperature, humidity, and light intensity data collecting nodes utilizing the test results. These batteries completely fulfill the specifications for low power usage and can operate in an actual system. They can function constantly for approximately 4 months when just one battery power supply is used.

Perspective Projection. Perspective projection converts 3D virtual scene points to 2D screen coordinates. The perspective projection matrix can be shown as code for copy

$$o = [e/aspect_ratio000][0e00][00(z_far + z_near)1][00 - (2 * z_far * z_near)0] \tag{7.3}$$

where z_near and Z_far are the near and far clipping planes, respectively, and f is virtual camera's focal length. Aspect ratio is the ratio of the screen's width to height.

Transformations. Objects are positioned and oriented in the virtual scene via transformations. Scaling, rotation, and translation procedures are used in this. These transformations can be represented by a 4 X 4 transformation matrix called the model-view matrix (M):

$$N = [QS][0^S1] \tag{7.4}$$

T is a 3D translation vector, 0T is a zero vector, and R is a 3x3 rotation matrix. One uses the model-view matrix to translate 3D object coordinates to the camera's coordinate system.

Homogeneous Coordinates. Computer graphics uses homogeneous coordinates to make transformations simpler. The following equations can be used to convert a point (x, y, z) in 3D space to a homogeneous coordinate (x', y', z', w'):

$$W' = W/XZ' = Z/XY' = Y/XX' = 1 \tag{7.5}$$

Binocular Disparity. By giving each eye slightly distinct images, binocular disparity creates a stereoscopic perspective. The interocular distance and the distance between the item and the viewer both affect how much difference there is. Given the object's depth (z), the formula for scalable computing the disparity (d) is as follows:

$$d = \frac{(2 * eye_distance * focal_length)}{z} \tag{7.6}$$

where $eye_distance$ is the space between the observer's eyes, and $focal_length$ is the virtual camera's focal length. Please be aware that these equations only offer a simplified picture and do not account for all the complexity inherent in virtual reality systems. An in-depth grasp of computer graphics, computer vision, and signal processing is necessary since VR technology involves various methods and methodologies.

8. Conclusion. This article introduced scalable Smart Home System as an AAL application that simulates the setting of a home utilizing Semantic Web and Virtual Reality technology. The health of the residents, which is periodically evaluated by medical staff, is modelled as a semantic knowledge base (KBHome), which enables automatic inference of various devices that allow the inhabitants to complete several everyday duties without assistance. By giving a formal account of the users' physiological condition, cardiorespiratory fitness, measurements for ambient comfort, gadgets, and behaviour, the ontologies open the door to providing the inhabitants with specialized services. Reasoning results permit, on the one hand, determination of range about devices assists residents with daily tasks, allowing them to live freely and cope with their disabilities; However, KB Home may provide instructions for residents to establish a daily physical exercise schedule. One of the scalable smart home system's most important aspects is the language for clinical and nonclinical practitioners in areas connected to health. The scalable smart home system personalizes support by assessing users' health through heart rate and cardiovascular fitness (CRF), a subject with CRF engages in tailored exercises, monitored using sensors for safety and efficiency during cycling sessions. The results of the discussion process are input into a VR-based application that can imitate a genuine household environment so that the appliances may be chosen and installed by smart home designers so that their behaviour and potential for accommodating people

with disabilities may be studied advantage of this virtual setting procedure is that it is entirely suited to user's actual expectations and restrictions and considerably saving setup costs and time by testing the appliances in a virtual environment before setting up a scalable smart home. In addition to their many advantages, smart home systems and commonplace virtual reality (VR) interfaces have certain drawbacks. Smart home systems scalable user experience will continue to incorporate the use of virtual reality environments in the future including developing interfaces that adapt to the user over time, enhancing cross-device communication and compatibility, incorporating Artificial Intelligence (AI) for the customization of space environments, fancier haptic devices to enhance the control scalable experience and make it more natural and cleaner which in turn focuses on developing better and scalable smarter homes.

REFERENCES

[1] Liao, L., Liang, Y., Li, H., Ye, Y. and Wu, G., 2023. A systematic review of global research on natural user interface for smart home system. International Journal of Industrial Ergonomics, 95, p.103445.

[2] Santos, B.S., 2019, August. Configuration and Use of Pervasive Augmented Reality Interfaces in a Smart Home Context: A Prototype. In Human Systems Engineering and Design II: Proceedings of the 2nd International Conference on Human Systems Engineering and Design (IHSED2019): Future Trends and Applications, September 16-18, 2019, Universität der Bundeswehr München, Munich, Germany (Vol. 1026, p. 96). Springer.

[3] Schenkluhn, M., Peukert, C. and Weinhardt, C., 2023, April. It is augmented Reality-based Indoor Positioning for Smart Home automation. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

[4] Liu, X., Shi, Y., Yu, C., Gao, C., Yang, T., Liang, C. and Shi, Y., 2023. It is understanding In-Situ Programming for Smart Home Automation. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 7(2), pp.1-31.

[5] Lyu, Z., Yang, J., Lam, M.S. and Landay, J.A., 2022, October. HomeView: Automatically Building Smart Home Digital Twins With Augmented Reality Headsets. In Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (pp. 1-6).

[6] Ariano, R., Manca, M., Paternò, F. and Santoro, C., 2023. Smartphone-based augmented reality for end-user creation of home automation. Behavior & Information Technology, 42(1), pp.124-140.

[7] Iliev, Y. and Ilieva, G., 2022. A Framework for Smart Home System with Voice Control Using NLP Methods. Electronics, 12(1), p.116.

[8] Zheng, M., Pan, X., Bermeo, N.V., Thomas, R.J., Coyle, D., O'hare, G.M. and Campbell, A.G., 2022. Stare: Augmented reality data visualization for explainable decision support in smart environments. IEEE Access, 10, pp.29543-29557.

[9] Ariano, R., Manca, M., Paternò, F. and Santoro, C., 2023. Smartphone-based augmented reality for end-user creation of home automation. Behavior & Information Technology, 42(1), pp.124-140.

[10] Mao, C. and Chang, D., 2023. Review of cross-device Interaction for facilitating digital transformation in smart home context: A user-centric perspective. Advanced Engineering Informatics, 57, p.102087.

[11] Schenkluhn, M., Peukert, C. and Weinhardt, C., 2023, April. It is augmented Reality-based Indoor Positioning for Smart Home automation. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

[12] Flick, C.D., Harris, C.J., Yonkers, N.T., Norouzi, N., Erickson, A., Choudhary, Z., Gottsacker, M., Bruder, G. and Welch, G., 2021, November. Trade-offs in augmented reality user interface for controlling a smart environment. In Proceedings of the 2021 ACM Symposium on Spatial User Interaction (pp. 1-11).

[13] He, F., Hu, X., Shi, J., Qian, X., Wang, T. and Ramani, K., 2023, April. Ubi Edge: Authoring Edge-Based Opportunistic Tangible User Interfaces in Augmented Reality. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

[14] Yao, Y., Huang, L., He, Y., Ma, Z., Xu, X. and Mi, H., 2023, April. Reviewing and Reflecting on Smart Home Research from the Human-Centered Perspective. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-21).

[15] Wang, X., Yew, A.W.W., Ong, S.K. and Nee, A.Y., 2020. Enhancing smart shop floor management with ubiquitous augmented reality. International Journal of Production Research, 58(8), pp.2352-2367.

[16] Putze, F., Weiß, D., Vortmann, L.M. and Schultz, T., 2019, October. Augmented reality interface for smart home control using SSVEP-BCI and eye gaze. In 2019 IEEE international conference on Systems, Man, and Cybernetics (SMC) (pp. 2812-2817). IEEE.

[17] Eckstein, B., Krapp, E., Elsässer, A. and Lugrin, B., 2019. Smart substitutional reality: Integrating the smart home into virtual reality. Entertainment Computing, 31, p.100306.

[18] Marques, B., Dias, P., Alves, J. and Santos, B.S., 2020. Adaptive augmented reality user interfaces using face recognition for smart home control. In Human Systems Engineering and Design II: Proceedings of the 2nd International Conference on Human Systems Engineering and Design (IHSED2019): Future Trends and Applications, September 16-18, 2019, Universität der Bundeswehr München, Munich, Germany (pp. 15-19). Springer International Publishing.

[19] Coppers, S., Vanacken, D. and Luyten, K., 2020. Fortnite: Intelligible predictions to improve user understanding of smart home behavior. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(4), pp.1-24.

[20] Mishra, A., Karmakar, S., Bose, A. and Dutta, A., 2020. Design and development of IoT-based latency-optimized augmented

reality framework in home automation and telemetry for smart lifestyle. Journal of Reliable Intelligent Environments, 6, pp.169-187.

[21] Rahman, M.A., Abualsaud, K., Barnes, S., Rashid, M. and Abdullah, S.M., 2020, February. A natural user interface and blockchain-based in-home smart health monitoring system. In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT) (pp. 262-266). IEEE.

[22] Haidon, C., Pigot, H. and Giroux, S., 2020. Joining semantic and augmented reality to design smart homes for assistance. Journal of Rehabilitation and Assistive Technologies Engineering, 7, p.2055668320964121.

[23] Suesaowaluk, P., 2020, June. Home automation system-based mobile application. In 2020 2nd World Symposium on Artificial Intelligence (WSAI) (pp. 97-102). IEEE.

[24] Knutzen, K., Weidner, F. and Broll, W., 2021, October. They are exploring augmented reality privacy icons for smart home devices and their effect on users' privacy awareness. In 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 409-414). IEEE.

[25] Ruser, H., Vorwerg, S., Eicher, C., Pfeifer, F., Piela, F., Kaltenbach, A. and Mechold, L., 2021. Evaluating the accuracy and user experience of a gesture-based infrared remote control in smart homes. In Human-Computer Interaction. Interaction Techniques and Novel Applications: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II 23 (pp. 89-108). Springer International Publishing.

[26] Sujatha Krishna, & Paryati. (2024). Advancing Cyber Resilience for Autonomous Systems with Novel AI-based Intrusion Prevention Model. International Journal of Data Informatics and Intelligent Computing, 3(3), 1–7. https://doi.org/10.59461/ijdiic.v3i3.121

[27] Asma Abdallah Nasser Al-Risi, Shamsa Salim Mattar Albadi, Shima Hamdan Said Almaamari, Saleem Raja Abdul Samad, & Pradeepa Ganesan. (2024). Automated Fall Detection for Disabled Individuals Using Mobile Phone Sensors and Machine Learning: A Survey. International Journal of Data Informatics and Intelligent Computing, 3(2), 27–33. https://doi.org/10.59461/ijdiic.v3i2.106

[28] Dutta, C., & Ritu Maity. (2024). Intelligent door unlock system using AI MediaPipe. International Journal of Data Informatics and Intelligent Computing, 3(1), 36–43. https://doi.org/10.59461/ijdiic.v3i1.96

Edited by: Dhilip Kumar V
Special issue on: Unleashing the power of Edge AI for Scalable Image and Video Processing
Received: Aug 27, 2024
Accepted: Nov 28, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2002–2015, DOI 10.12694/scpe.v26i5.4818

ENHANCED PRE-PROCESSING STRATEGIES FOR ACCURATE DIABETES PREDICTION IN HEALTHCARE USING NOVAL METHOD: ANN+LDA

SOUMYA K N* AND RAJA PRAVEEN N†

Abstract. In the recent past, so many chronic diseases have been emerging and spreading in the world and even in the developing and advanced countries as well. One of such serious chronic diseases is Diabetes Mellitus that covers and impacts the health of people from early age. Nevertheless, the available Machine Learning (ML) and Deep Learning (DL) approaches are unable to provide good predictions in patients relating to diabetes. In addition, this study evaluated the proposed pre-processing procedure on large datasets for diabetes prediction that contained outliner detection and removal, missing values imputation, and standardization, to improve diabetes ascertainment. This research evaluated the proposed pre-processing procedure on a large set of data by outlier identification and removal, missing values imputation and data standardization were done to improve diabetes forecast. To ensure rapid and accurate classification of diabetes, the researchers employed and initialized an Artificial Neural Network (ANN). Data was gathered from the PIMA Dataset and North California State University (NCSU). Following this, Bivariate filter was applied to sort out features which were relevant. The selected features were subsequently subjected to Pearson correlation towards feature set refinement considering a threshold below which features were eliminated and only the most effective features selected. From the results it was evident that the proposed approach was significantly better than the existing methods in terms of accuracy as it achieved a classification accuracy of around 93% as opposed to the other methods

Key words: Artificial Neural Network, Bivariate filter, Diabetes mellitus, Pearson correlation, Standardization.

1. Introduction. Diabetes is a chronic condition that may soon threaten the health of the entire population. The International Diabetes Federation reported that currently there are 382 million diabetics across the world. Unfortunately, the forecast is for this number to rise to 592 million by the year 2035, and essentially doubling. It is a disease in which blood glucose levels are high, making it a concern for much of the population as well as healthcare systems [1]. Early prediction of diseases like diabetes can play a crucial role in controlling and potentially saving human lives. This research aims to predict diabetes by analyzing various disease-related attributes. The study utilizes the Pima Indian Diabetes Dataset and applies various machine learning (ML) classification and ensemble techniques. These methods aim to accurately predict diabetes occurrence, facilitating early intervention and improving healthcare outcomes [2]. Machine Learning (ML) is an explicit training method used to efficiently gather knowledge by building various classification and ensemble models from large datasets. This approach can be employed to develop predictive models for diabetes using extensive data repositories [3]. These learning models generate insights from large datasets, which can then be applied to new data for prediction and analysis. In recent years, ML has gained significant attention in healthcare, particularly for tasks such as disease diagnosis, including diabetes prediction.

In the context of disease diagnosis, ML enables the development of systems that can effectively assist physicians in identifying diseases [4]. Rapid advancements in Artificial Intelligence (AI), especially in ML and computer vision, have led to applications that automate complex, intelligence-demanding tasks in healthcare. These technological developments are particularly valuable for tasks such as diabetes prediction and diagnosis, where analysis of large datasets and recognition of subtle patterns are crucial.

This research focused on training an ML model to predict the progression from pre-diabetes to diabetes. The model utilized Electronic Medical Records (EMR), incorporating both historical and current patient data. The researchers thoroughly described the model's development and validation process, using data from The

*School of Computer Science and Engineering, JAIN (Deemed to be University) Bangalore, Karnataka, India (Soumya.kn16@gmail.com)

†Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, 562112, India (p.raja@jainuniversity.ac.in)

Health Improvement Network (THIN) database for both internal and external validation [5]. This approach aligns with the broader ML process for diabetes prediction, which involves gathering pertinent data from EMR systems like THIN, including comprehensive medical records and lifestyle factors. From the collected EMR data, researchers extract relevant features and train an ML algorithm to build a predictive model for diabetes progression. The resulting model then analyzes new patient data to evaluate the risk of diabetes development based on specific input variables [6]. This approach equips healthcare systems with a powerful tool for early diabetes risk assessment, seamlessly integrating software engineering technology with ML techniques to support proactive patient care. By providing timely risk evaluations, the system enables healthcare providers to implement targeted interventions and personalized management strategies, potentially improving patient outcomes and reducing the overall burden of diabetes on healthcare systems.

The diabetes management system aims to determine optimal nutrition requirements and provide tailored meal recommendations for patients. The system also sends timely medication reminders, enhancing overall healthcare management and support [7]. Beyond these specific features, ML plays a crucial role in broader diabetes care. It assists healthcare professionals in early detection and intervention for high-risk individuals, facilitating the development of personalized treatment plans and lifestyle adjustments. Furthermore, ML is instrumental in developing decision support systems for diabetes management, providing real-time insights and recommendations to healthcare providers. This comprehensive approach, combining personalized patient support with advanced clinical decision-making tools, has the potential to significantly improve diabetes outcomes and quality of life for patients.

Additionally, these predictive models are essential for population health studies, as they enable the identification of risk factors and the implementation of preventive measures on a broader scale [8]. Benefits of using ML for diabetes prediction include early detection, personalized risk assessment, optimized treatment planning, reduced healthcare costs, improved patient management, enhanced quality of life, and the potential to prevent complications. ML algorithms can process vast amounts of data, identify patterns, and deliver accurate predictions, enabling timely interventions and better outcomes for those at risk of developing diabetes [9]. However, drawbacks include the possibility of false positives or negatives and dependence on accurate data input.The models have limited interpretability and pose a risk of over-reliance on technology without sufficient clinical judgment. Overcoming these challenges demands careful planning, stakeholder involvement, and a well-executed transition strategy to effectively leverage the benefits of ML while minimizing disruptions to clinical workflows [10-13]. The suggested method changed the feature scale led to varying coefficients, rendering the magnitude coefficient an unsuitable choice for determining feature importance in the model.

Isfafuzzaman Tasin et al. [14] implemented a six ML approaches, including decision tree, SVM, Random Forest, Logistic Regression, KNN, and various ensemble techniques, to achieve the most accurate diabetes prediction. It also utilized a semi-supervised model with extreme gradient boosting to predict insulin features from a private dataset. To address the issue of class imbalance, SMOTE and ADASYN techniques was implemented. Moreover, it developed a user-friendly Android smartphone application and a suggested framework that allows users to input diverse features for instantaneous diabetes prediction. However, the suggested method, need to further enhance by incorporating fuzzy logic techniques and optimization approaches into the ML models.

Annamalai R and Nedunchelian [15] have developed an Optimal Weighted based Deep Artificial Neural Network (OWDANN) algorithm, for predicting diabetes mellitus disease and estimating its severity level. The system comprises two distinct phases: disease prediction and severity level estimation. In the disease prediction phase, the Pima dataset undergoes preprocessing to enhance data quality. Subsequently, relevant features were extracted from the preprocessed data, and the classification step employs the OWDANN algorithm. This approach effectively addresses noise and efficiently restores corrupted data, leading to improved prediction accuracy. However, OWDANN requires further modifications and need to enhance a wide range of scenarios and provide more comprehensive predictions for diabetes-related complications.

The proposed method aims to address the limitations found in existing diabetes prediction approaches. By enhancing data quality and eliminating inconsistencies, the suggested approach aims to significantly improve the accuracy and reliability of prediction models. As a result, the diabetes predictions generated are expected to be more robust and effective.

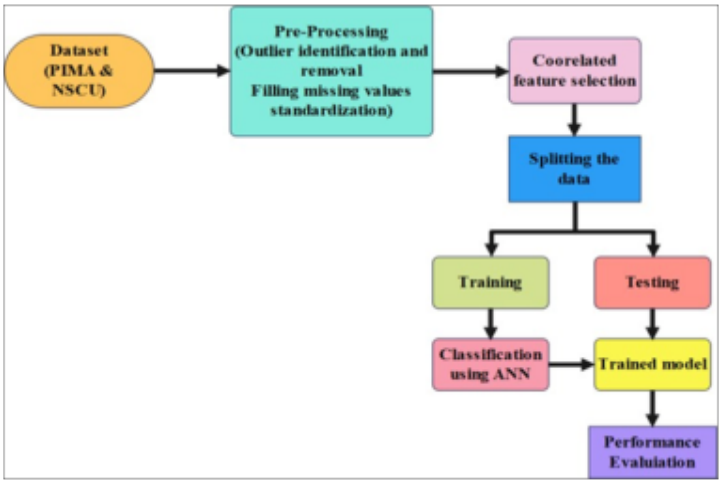


Fig. 2.1: Flow diagram of the proposed method

2. Methodology. The purpose of the ITSOA has to improve the accuracy of EEG based emotion classification. The process involved in various stages of classifying diabetes are data acquisition, pre- processing, feature selection and classification of diabetes.

The initial step involves acquiring data from a publicly available dataset, followed by a pre- processing phase to eliminate irrelevant or inappropriate features. Subsequently, a feature selection process is applied to choose relevant and non-redundant features.

Finally, an efficient classification is conducted using an ANN classifier to achieve accurate predictions. These steps aim to enhance the accuracy and reliability of the prediction models by improving data quality and removing inconsistencies, leading to more robust and effective diabetes predictions. The block diagram of the suggested method is illustrated in Fig. 2.1.

2.1. Data collection. In this research, the raw data is obtained from two publicly available datasets such as PIMA Dataset [16] and North California State University (NCSU) [17] dataset. The description of the mentioned dataset is mentioned as follows:

PIMA: The Pima Indian Diabetes dataset has been a standard benchmark for diabetes classification research due to its binary outcome variable, making it suitable for supervised learning, especially logistic regression. However, researchers have explored various ML algorithms to build classification models using this dataset, allowing for diversity and avoiding reliance on a single type of model.

NCSU: The diabetes prediction dataset used in this study was obtained from NC State University and consists of 442 instances with 10 attributes. The feature set comprises age and sex, and for the analysis, one feature set was selected.

2.2. Data Pre-processing. After the stage of data collection, in this research, the preprocessing step of the proposed framework aims to transform the data into a processed format without complexities. It involves outlier identification and removal to eliminate extreme data points, filling missing values to ensure complete datasets, and standardization to normalize variables. The process of the proposed method is briefly outlined as follows;

Outlier identification and removal. The purpose of outlier identification and removal using pre-processing for diabetes prediction is to improve model accuracy by eliminating extreme data points by improving the model's performance, ensuring more reliable predictions for better outcomes. It is a crucial step in data preparation to enhance the accuracy of predictive models. Outliers are data points that significantly deviate from the majority, potentially distorting the model's performance. Through the application of pre-processing techniques, outliers can be detected and effectively eliminated from the dataset. In the context of diabetes prediction, this

procedure enhances the model's robustness and generalization capabilities by reducing erroneous data. The conventional approach for identifying and removing outliers in multivariate data analysis involves measuring the distance of each observation using Mahalanobis distance, as depicted in Equ.2.1.

$$D_M = (X - \bar{X})^T S^{-1} (X - \bar{X}) \tag{2.1}$$

where D_M=Scaler matrix, X=Vector, $\bar{X}$=Mean of matfrix X, S^{-1}=Inverse of matrix. The observations associated with large values of DM are classified as outliers and then discarded. The Mahalanobis distance can be related to the principal components: it can be shown, in fact, that the sum of squares of the PC, standardized by the eigenvalue size, equals the Mahalanobis distance for observation I was illustrated in Equ.2.2.

$$\sum_{k=1}^{Q} \frac{Z_i k^2}{l_k} = \frac{Z_i 1^2}{l_1} + \frac{Z_i 2^2}{l_2} + ... + \frac{Z_i Q^2}{l_Q} = D_{M,i} \tag{2.2}$$

where, Q =upper limit, Z_{i1}^2=Specific instance of Z, $D_{M,i}$=Results of summerization, Z_{ik}^2=weights.

In high-dimensional datasets, some outliers may not be apparent when examining individual dimensions, making them undetectable through univariate analysis. Consequently, a multivariate approach is necessary. In this regard, Principal Component Analysis (PCA) is an excellent tool for effectively identifying and removing outlier observations.

Filling missing values. The goal of filling in missing values during pre-processing for diabetes prediction is to create a complete dataset that is ready for analysis, as missing data can adversely affect the performance of predictive models. Imputing missing values enables the model to make more accurate predictions, leading to better diabetes diagnosis and treatment. Missing data can occur for various reasons, including data entry errors or incomplete information. To ensure the integrity and accuracy of the predictive model, these gaps are filled with estimated or imputed values using different techniques.Common techniques for handling missing values include mean, median, or mode imputation, which use the central tendency of the available data to replace missing entries. More advanced methods, such as regression imputation or K-Nearest Neighbors (KNN) imputation, leverage the relationships between variables in neighboring data points. In the proposed framework, missing or null values were filled using the mean values of the attributes instead of being discarded, as shown in Equ.2.3. Mean imputation is advantageous because it fills continuous data without introducing outliers.

$$(x) = f(x) = \{(mean(x), if x = null/missedx, otherwise \tag{2.3}$$

where x is the instances of the feature vector that lies in n-dimensional space, $x \epsilon R$.

Standardization. Standardization is a key preprocessing step to ensure that features are on a comparable scale, which can improve the performance of machine learning models. We standardize these datasets:

Steps for Standardization:

1. Calculate the Mean and Standard Deviation: For each feature in the dataset, compute the mean and standard deviation.

$$Mean(\mu) = \frac{1}{N} \sum_{i=0}^{N} xi \tag{2.4}$$

$$Standard\ deviation = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [x_i - \mu]^2} \tag{2.5}$$

2. Transform the Features: Subtract the mean from each feature value and divide by the standard deviation:

$$Standardized\ Value = \frac{x_i - \mu}{\sigma} \tag{2.6}$$

The Pima Indians Diabetes dataset often includes features like glucose levels, blood pressure, skin thickness, insulin levels, and BMI. Here's a brief outline of standardization for this dataset:

1. Compute the mean and standard deviation for each feature
2. Apply the transformation to each feature to ensure they have a mean of 0 and a standard deviation of 1.

2.3. Feature selection. The first pre-processing step yields the pre-processed output, which is then used as input for the feature selection stage that follows. By choosing the most pertinent features that make the diabetes classification process easier, feature selection seeks to increase classification accuracy. The present study employs the Bivariate statistics technique for feature selection, which effectively picks pertinent and suitable characteristics from extensive datasets such as PIMA and NCSU [18-24]. The Bivariate filter is employed for feature extraction, effectively integrating heterogeneous data layers to address uncertainties in the input data. Additionally, this filter utilizes a certainty factor to identify relevant features, and its evaluation is carried out using the next equation:

$$CF = \frac{PP_a - PP_s}{PP_a(1 - PP_s)}, if PP_a \geq PP_s else \frac{PP_a - PP_s}{PP_s(1 - PP_a)}, if |PP_a < PP_s| \tag{2.7}$$

where the conditional probability of CF is denoted as PP_a and the prior probability of the selected features are represented as PP_s. The value of PP_a and PP_s is evaluated using the Equ.2.8 and Equ.2.9 respectively.

$$PP_a = PS|B \tag{2.8}$$

$$PP_s = PS \tag{2.9}$$

where the conditional probability unit of **B** is represented as $PS|B$. Positive results indicate an increase in the certainty value, whereas negative results signify a decrease in the certainty value. Furthermore, the features are extracted using the Weight of Evidence (WoE) based on the Bayesian probability approach, utilizing weights to determine their significance. The positive and negative weights of the features are evaluated using two parameters, $W+$ and $W-$, as represented in Equ.2.10 and Equ.2.11.

$$W^+ = ln\frac{PB|A}{PB|\bar{A}} \tag{2.10}$$

$$W^- = ln\frac{P\bar{B}|A}{P\bar{B}|\bar{A}} \tag{2.11}$$

The logarithm and probability values are represented as P and ln respectively. The features selected using the bivariate filter method are then used as input for Pearson correlation, which identifies effective features based on a specified threshold value.

Pearson Correlation. To enhance the relationship between Pearson correlation and diabetic characteristics, the parameters are optimized to remove redundant information. The Pearson correlation coefficient is used to assess linear relationships between random variables. Equ.2.12 illustrates the linear correlation between two continuous variables

$$r_{xy} = \frac{\sum(x_i - \bar{x})\sum(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y}2)}} \tag{2.12}$$

If rxy= 1, x and y are a totally positive correlation, If rxy = 0, the linear correlation between x and y is not obvious and when rxy = -1, x and y are a totally negative correlation.

The Pearson correlation, with an R-value of 0.12, exhibits a less significant effect on Diabetes. The relationship between certain features and diabetes is found to be moderate (r = 0.33, r = - 0.42, r = 0.23). It is important to note that correlation does not imply causation. Utilizing this information, relevant features strongly associated with diabetes are identified and selected as input variables for precise disease classification. The output of Pearson correlation is then fed into the classification process to identify cases of type II diabetes.

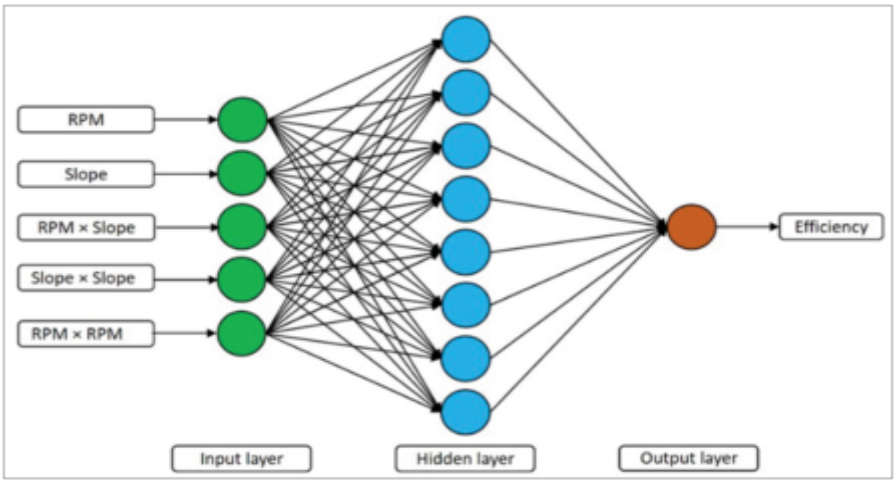


Fig. 2.2: Architecture of ANN model

2.4. Classification using ANN. Following the feature selection phase, the PIMA and NCSU datasets were used for the classification. Artificial Neural Networks (ANN), one of the machine learning (ML) algorithms used for classification, are renowned for producing results that are more accurate than those obtained using other methods. An ANN is made up of one or more hidden layers where information is processed by neurons. In order to get better outcomes, every node serves as an activation node that classes the output of artificial neurons. Numerous factors, such as the choice of hidden units for each layer, can be tuned in order to reduce the training error. These units determine the name and size of the layers, enabling users to customize the neural network's structure. The ANN algorithm is adept at finding minima, controlling variance, and subsequently updating the model's parameters, as expressed in next equation.

$$\theta = \theta - \eta * De; taJ(\theta) \tag{2.13}$$

The learning rate is another critical parameter in ANN, responsible for adjusting the weights at each step and playing a crucial role in the model's learning process. It must be carefully chosen, as a learning rate that is too high may hinder the selection of minima, while one that is too low can slow down the learning speed. Commonly selected learning rate values are in the power of 10, such as 0.001, 0.01, 0.1, and 1. In this model, the learning rate is set to 0.1. Figure 2.2 represents the architecture of ANN model.

The ANN-LDA algorithm in shown in Algorithm 1.

Start with new weights each time to perform training. Next, the LDA is used to minimize the dimensionality of the input characteristics, simplify the operations of the neural network, and optimize the classes. The weights are continuously adjusted by the model based on the error ascertained by the error Calculation. Following that, the ANN learns to decrease error by modifying the rule weights. The process is repeated until the error no longer decreases dramatically; at that point, the model is considered to have learned the relationships discovered in the data. After training each LDA-transformed feature via the ANN, apply the activation function to the weighted sum to provide an output for each neuron.

Determine the efficient result of the output layer, which gives the predicted risk of diabetes. Find the gradient of the loss function with respect to each network weight. Include the LDA-generated probabilities in the weight update process and adjust the weights to emphasize the components that LDA identified.

The network converges when weight changes in certain threshold when the maximum number of iterations is reached. Update the weights through the whole training dataset many times. Make use of both forward and backward propagation methods to adjust weights based on the difference between the actual and labels. This method effectively combines the dimensionality reduction and class separation strengths of LDA with the nonlinear modelling capabilities of an ANN.

Algorithm 1 ANN-LDA algorithm

1. **Step 1:** Data Preprocessing
   Input: Dataset (X, y)
   Output: Processed data ($X_processed$)
   (a) Normalize the dataset X (mean = 0, variance = 1)
   (b) Handle missing values (imputation or removal)
   (c) Convert categorical variables to numerical (one-hot encoding)
2. **Step 2:** Dimensionality Reduction using LDA
   Input: Processed data ($X_processed, y$)
   Output: Reduced dimension data (X_LDA)
   (a) Apply LDA to $X_processed$ with target labels y
   (b) Compute the linear discriminants for class separation
   (c) Project $X_processed$ onto the LDA components to get X_LDA
3. **Step 3:** Initialize the ANN Model
   Input: X_LDA, hyperparameters ($learning_rate$, epochs, $hidden_layers$, etc.)
   Output: Initialized ANN model (weights, biases)
   (a) Define the ANN structure with input layer size = X_LDA dimensions
   (b) Initialize weights and biases for all layers
   (c) Choose activation function (e.g., ReLU, Sigmoid)
   (d) Define the loss function (e.g., Cross-Entropy Loss)
4. **Step 4:** Train the ANN Model
   Input:X_LDA, y, ANN model
   Output: Trained ANN model (optimized weights and biases)
   (a) For each epoch in range(epochs):
   For each batch in the training data:
   Forward Pass:
   - Compute the output of each layer (activations)
   - Calculate the predicted output using the final layer
   Compute Loss:
   - Compare the predicted output with the true labels (y)
   -Calculate the loss using the chosen loss function
   Backward Pass (Backpropagation):
   - Calculate the gradient of the loss with respect to weights and biases
   - Update the weights and biases using the learning rate
   End For
5. **Step 5:** Model Evaluation
   Input: Test data (X_test, y_test), Trained ANN model
   Output: Performance metrics (accuracy, precision, recall, F-score)
   (a) Apply LDA transformation to X_test
   (b) Forward pass X_test_LDA through the trained ANN model
   (c) Calculate the predicted output for test data
   (d) Compare predictions with true labels y_test
   (e) Compute performance metrics (accuracy, precision, recall, F-score)
6. **Step 6:** Output the Model and Performance Metrics
   Return: Trained ANN-LDA model, Performance metrics

Table 3.1: Performance comparison of classifiers for PIMA dataset

Classifiers	Accuracy (%)	F-measure (%)	Recall (%)	Precision (%)
LR	78	75	77	74
SVM	87	78	85	73
KNN	76	72	73	72
ANN	82	81	86	77
DT	76	74	74	74
ANN-LDA	93	84	88	81

3. Result Analysis. In this section, the results obtained from the proposed is evaluated to obtain the results based on diabetes classification. The result section is sub-sectioned to performance analysis and the comparative analysis. The performance analysis involves assessing the efficiency of the proposed approach on two distinct datasets, namely PIMA and NCSU. For the comparative analysis, the proposed approach's effectiveness is evaluated against existing approaches documented in the literature. The evaluation metrics encompass accuracy, precision, recall, and f-measure, which are computed using the equations (12-15) as provided below.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

$$F1measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.4}$$

where the TP,FP,TN,FN is the True Positive, False Positive, True Negative and False Negative respectively.

3.1. Experimental setup. The proposed ANN-LDA classification approach was implemented on a system with the following specifications: Anaconda Navigator 3.5.2.0 (64-bit), Python 3.7 software, Windows 10 (64-bit) operating system, Intel Core i7 processor, and 16 GB of random-access memory.

3.2. Performance analysis of PIMA Dataset. The performance analysis of the PIMA dataset includes evaluating ML models for diabetes prediction, exploring pre-processing, Feature selection and Classification models for effectiveness, thus enabling valuable insights for healthcare applications and improving diabetes diagnosis and management [25-27]. Figures 3.1 to 3.4 presents the different performance measures. Table 3.1 shows Performance comparison of classifiers for PIMA dataset.

3.3. Performance analysis of PIMA Dataset for Pre-processing. In this subsection, we evaluate the performance of the proposed approach using various classifiers, including Naïve Bayes, KNN, Support Vector Machine (SVM), and Random Forest. The evaluation is conducted on the PIMA dataset, and the results are presented in Table 3.1 and Table 3.2.

Table 3.1 displays the results obtained from the proposed method on the PIMA dataset without applying any pre-processing techniques, while Table 3.2 shows the results after applying pre- processing techniques. Additionally, Figure 3.5 provides a graphical representation of the performance analysis for the PIMA dataset. Table3.3 presents Performance analysis for after pre-processing techniques.

The results from Table 3.1 and Table 3.3 demonstrate that the proposed method serves as an excellent classifier for distinguishing diabetic patients in the PIMA dataset. The performance of the proposed classification approach outperforms existing methods in terms of overall metrics, particularly in accuracy.

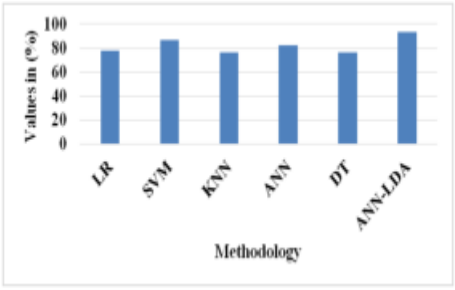


Fig. 3.1: Graphical representation of classification performance of accuracy for PIMA dataset

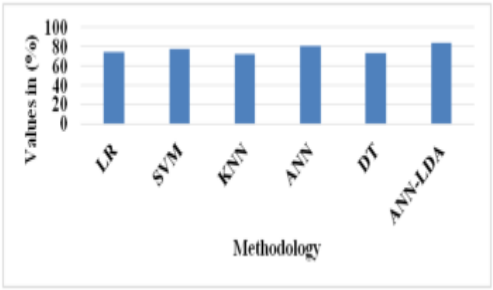


Fig. 3.2: Graphical representation of classification performance of F1-measure for PIMA dataset

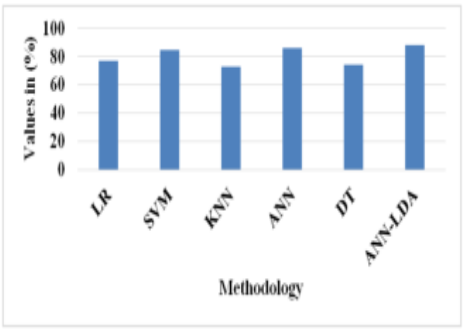


Fig. 3.3: Graphical representation of classification performance of Recall for PIMA dataset

Table 3.2: PIMA dataset for without pre-processing techniques

Methods	Accuracy (%)		
	Mean	Median	Most Frequent
Naïve Bayes	74.21	68.03	74.52
SVM	75.30	74.25	76.35
Random Forest	76.32	74.54	73.21
Proposed	70.16	58.23	61.98

3.4. Performance analysis of PIMA Dataset for feature selection. Table 3.4 presents the results obtained from the proposed method applied to the PIMA dataset using various feature selection techniques. The

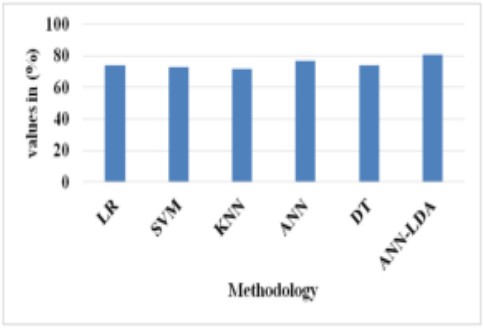


Fig. 3.4: Graphical representation of classification performance of precision for PIMA dataset

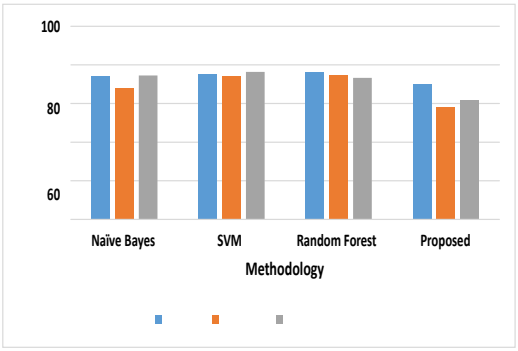


Fig. 3.5: Graphical representation of PIMA dataset for without pre-processing techniques

Table 3.3: Performance analysis for after pre-processing techniques

Missing value strategy	Z- Score	Minmax Scalar
Mean	74.65	83.45
Median	60.19	81.10
Most Frequent	64.15	80.26

Table 3.4: Performance analysis of Feature selection for PIMA Dataset

Classifier	Accuracy for Testing (%)	Accuracy for Validation (%)
SVM	75.37	81.41
Random Forest	77.23	82.89
Correlated function	78.25	85.21

dataset is divided into training and test sets in a ratio of 70% and 30%, respectively. The training set consists of 70% of the data randomly chosen, while the remaining 30% is allocated to the testing set. This specific split ratio was determined after exploring various combinations and has proven to be efficient in achieving better results.

According to the data presented in Table 3.4, the correlated function outperforms the SVM and Random Forest classifiers in terms of training and testing accuracy, following data pre- processing. Furthermore, both classifiers achieve similar validation accuracy. Notably, the correlated function demonstrates a substantially

Table 3.5: Comparing the performance of the classifiers for PIMA dataset

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
KNN	75.11	71.61	72.32	72.64
LR	76.25	73.84	76.74	74.45
DT	75.25	73.16	73.45	73.73
SVM	86.59	72.18	84.27	83.38
ANN	90.66	80.32	87.73	85.35

Table 3.6: NCSU dataset for without pre-processing techniques

Methods	Accuracy (%)		
	Mean	Median	Most Frequent
Naïve Bayes	73.31	69.06	73.11
SVM	74.36	73.35	75.22
Random Forest	76.66	73.36	72.35
Proposed	70.13	57.22	60.46

Table 3.7: Performance analysis for NCSU dataset after pre-processing techniques

Missing value strategy	Z- Score	Minmax Scalar
Mean	73.21	82.47
Median	60.32	80.65
Most Frequent	63.51	80.51

higher true negative rate, implying a more accurate prediction capability.

3.4.1. Performance analysis of PIMA Dataset for Classification. The performance evaluation of mentioned classifiers was conducted using the PIMA dataset, as shown in Table 3.5. Additionally, Table 3.5 displays the results obtained from the proposed approach for the same PIMA dataset.

Table 3.5 demonstrates that the proposed ANN serves as a highly effective classifier for distinguishing diabetic patients within the PIMA dataset. The proposed classification approach outperforms existing methods across various metrics. Additionally, the classification accuracy of the proposed ANN reaches an impressive 90.66%, surpassing the accuracies of other classifiers, such as KNN of 75.11%, LR of 76.25%, DT of 75.25%, and SVM of 86.59% respectively.

3.5. Performance Analysis of NCSU dataset. The performance analysis of the NCSU dataset involves assessing various ML models for predicting diabetes. It also encompasses exploring the effectiveness of pre-processing, feature selection, and classification models. These findings provide valuable insights for healthcare applications, leading to enhancements in diabetes diagnosis and management.

3.5.1. Performance analysis of NCSU Dataset for Pre-processing. Within this sub-section, the proposed approach's performance is assessed using various classifiers, including Naïve Bayes, KNN, SVM, and Random Forest. The evaluation is based on the NCSU datasets, and the results are presented in Table 3.6 and Table 3.7. These tables illustrate the outcomes of the proposed method on the NCSU dataset, both without and after employing pre-processing techniques. Furthermore, a graphical representation of the performance analysis for the NCSU dataset is provided in Figure 3.6.

The results from Table 3.6 and Table 3.7 demonstrate that the proposed method serves as an outstanding classifier in accurately identifying diabetic patients within the NCSU dataset. When compared with existing classification methods, the proposed approach achieves superior results in overall metrics, particularly in terms of accuracy.

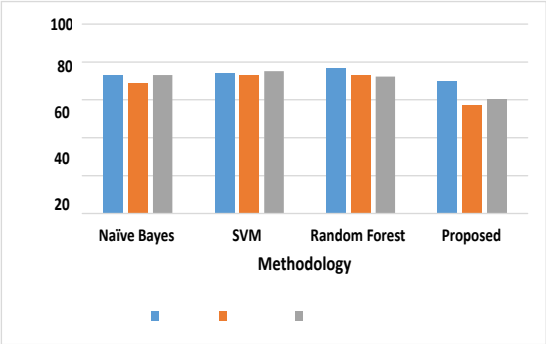


Fig. 3.6: Graphical representation of the NCSU for without pre-processing techniques

Table 3.8: Performance analysis of Feature selection for NCSU Dataset

Classifier	Accuracy for Testing (%)	Accuracy for Validation (%)
SVM	74.32	80.25
Random Forest	76.45	81.94
Correlated function	78.86	85.37

Table 3.9: Comparing the performance of the classifiers for NCSU dataset

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
KNN	74.36	72.94	71.33	71.62
LR	75.65	74.34	75.42	73.31
DT	74.12	74.16	73.97	72.30
SVM	85.10	71.88	83.55	82.11
ANN	90.37	80.34	88.52	86.19

3.5.2. Performance analysis of NCSU Dataset for feature selection. The Table 3.8 presents the results obtained from the proposed method applied to the NCSU dataset using various feature selection techniques. The dataset is divided into training and test sets at a ratio of 70% and 30%, respectively. This split is determined after exploring various combinations, proving its efficiency in achieving optimal performance.

Table 3.8 reveals that after data pre-processing, the training accuracy and testing accuracy of the correlated function surpass those of the SVM and Random Forest classifiers. Additionally, both classifiers achieve similar validation accuracy. These results indicate that the correlated function exhibits a significantly higher true negative rate, highlighting its superior correctness in predictions.

3.5.3. Performance analysis of NCSU Dataset for Classification. The performance evaluation of the recommend classifiers was conducted using the NCSU datasets, as depicted in Table 3.9. Additionally, Table 3.9 presents the results obtained from the proposed approach for the NCSU dataset.

Table 3.9 demonstrates that the proposed ANN serves as an outstanding classifier for accurately classifying diabetic patients within the NCSU dataset. The proposed classification approach achieves superior results in overall metrics compared to existing classification methods. Notably, the classification accuracy of the proposed ANN reaches 90.37%, which is significantly higher than the accuracies of existing classifiers, such as KNN of 74.36%, LR of 75.65%, DT of 74.12%, and SVM of 85.10%.

3.6. Comparative analysis. Comparative analysis refers to the Comparison of data to identify similarities and differences for meaningful insights or decision-making. In this subsection, the classification approach's performance is assessed by comparing it with existing approaches listed in related works. Evaluation is based

Table 3.10: Comparative analysis of various classifier for PIMA dataset

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
2GDNN [12]	97.93	98.11	97.23	97.95
Six- ML [14]	88.05	82.02	80.56	81.21
OWDANN [15]	98.97	97.02	93.84	94.04
Proposed	98.99	98.15	97.25	97.96

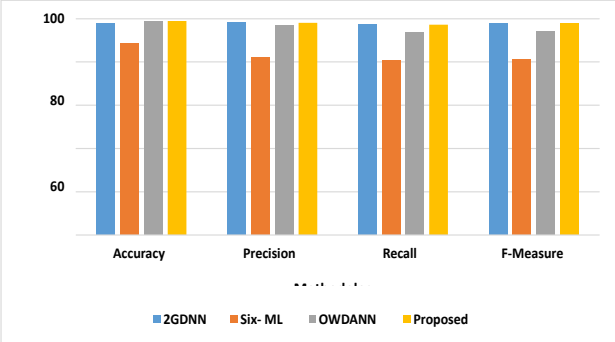


Fig. 3.7: Graphical representation of comparative analysis for PIMA dataset

on performance metrics such as accuracy, precision, recall, and F-measure score. The results obtained from evaluating the proposed approach for the PIMA dataset are presented in Table 3.10.

The graphical representation of the comparative analysis for PIMA dataset was illustrated in figure 3.7.

Table 3.9 and Figure 3.2 demonstrate that the proposed classification approach outperformed other methods in overall performance metrics. The accuracy achieved by the proposed approach is 98.99%, significantly higher than the Twice Growth Deep Neural Network (2GDNN) (97.93%), Six ML methods (88.05%), and Optimal Weighted based Deep Artificial Neural Network (OWDANN) (98.97%).

4. Conclusion. The research introduces a pre-processing approach involving outlier identification, missing value filling, and standardization to enhance diabetes Mellitus prediction accuracy. The proposed method utilizes an ANN with optimized weight initialization for effective diabetes classification. The approach's performance is evaluated on both PIMA and NCSU datasets using accuracy, precision, recall, and F-measure metrics. Following the Bivariate filter-based feature selection stage, relevant features are selected, and the chosen features undergo Pearson correlation analysis using a threshold value. The resulting effective features are then utilized as input for the ANN classifier, performing the final classification. The proposed approach outperforms existing methods in overall metrics, with an accuracy of 98.99%, surpassing 2GDNN, Six ML methods, and OWDANN of 7.93%, 88.05%, and 98.97% respectively. Future work can explore incorporating meta-heuristic algorithms to further enhance accuracy by selecting appropriate features.

REFERENCES

[1] Rani, K.J., 2020. Diabetes prediction using machine learning. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 6, pp.294-305.
[2] Soni, M. and Varma, S., 2020. Diabetes prediction using machine learning techniques. International Journal of Engineering Research & Technology (Ijert) Volume, 9.
[3] Kaul, S. and Kumar, Y., 2020. Artificial intelligence-based learning techniques for diabetes prediction: challenges and systematic review. SN Computer Science, 1(6), p.322.
[4] Assegie, T.A. and Nair, P.S., 2020. The performance of different machine learning models on diabetes prediction. International journal of scientific & technology research, 9(01).
[5] Cahn, A., Shoshan, A., Sagiv, T., Yesharim, R., Goshen, R., Shalev, V. and Raz, I., 2020. Prediction of progression from

pre-diabetes to diabetes: Development and validation of a machine learning model. Diabetes/metabolism research and reviews, 36(2), p.e3252.

[6] Sowah, R.A., Bampoe-Addo, A.A., Armoo, S.K., Saalia, F.K., Gatsi, F. and Sarkodie-Mensah, B., 2020. Design and development of diabetes management system using machine learning. International journal of telemedicine and applications, 2020.

[7] Vehí, J., Contreras, I., Oviedo, S., Biagi, L. and Bertachi, A., 2020. Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. Health informatics journal, 26(1), pp.703-718.

[8] Nibareke, T. and Laassiri, J., 2020. Using Big Data-machine learning models for diabetes prediction and flight delays analytics. Journal of Big Data, 7, pp.1-18.

[9] Jaiswal, V., Negi, A. and Pal, T., 2021. A review on current advances in machine learning based diabetes prediction. Primary Care Diabetes, 15(3), pp.435-443.

[10] Ramesh, J., Aburukba, R. and Sagahyroon, A., 2021. A remote healthcare monitoring framework for diabetes prediction using machine learning. Healthcare Technology Letters, 8(3), pp.45-57.

[11] Chang, V., Bailey, J., Xu, Q.A. and Sun, Z., 2022. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, pp.1-17.

[12] Olisah, C.C., Smith, L. and Smith, M., 2022. Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. Computer Methods and Programs in Biomedicine, 220, p.106773.

[13] Kibria, H.B., Nahiduzzaman, M., Goni, M.O.F., Ahsan, M. and Haider, J., 2022. An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI. Sensors, 22(19), p.7268.

[14] Tasin, I., Nabil, T.U., Islam, S. and Khan, R., 2023. Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters, 10(1-2), pp.1-10.

[15] Annamalai, R. and Nedunchelian, R., 2021. Diabetes mellitus prediction and severity level estimation using OWDANN algorithm. Computational Intelligence and Neuroscience, 2021.

[16] Chang, V., Bailey, J., Xu, Q.A. and Sun, Z., 2022. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, pp.1-17.

[17] SVKR Rajeswari, V., 2021. Prediction of diabetes mellitus using machine learning algorithm. Annals of the Romanian Society for Cell Biology, pp.5655-5662.

[18] Parente, A. and Sutherland, J.C., 2013. Principal component analysis of turbulent combustion data: Data pre- processing and manifold sensitivity. Combustion and flame, 160(2), pp.340-350.

[19] Hasan, M.K., Alam, M.A., Das, D., Hossain, E. and Hasan, M., 2020. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, pp.76516-76531.

[20] Naz, H. and Ahuja, S., 2020. Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes & Metabolic Disorders, 19, pp.391-403.

[21] Sannasi Chakravarthy, S.R., Bharanidharan, N., Vinothini, C. et al. Adaptive Mish activation and ranger optimizer-based SEA-ResNet50 model with explainable AI for multiclass classification of COVID-19 chest X-ray images. BMC Med Imaging 24, 206 (2024). https://doi.org/10.1186/s12880-024-01394-2

[22] Musthafa, M.M., T R, M., V, V.K. et al. Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification. BMC Med Imaging 24, 201 (2024). https://doi.org/10.1186/s12880-024-01356-8

[23] Mahesh T R, Muskan Gupta, Anupama T A, Vinoth Kumar V, Oana Geman, Dhilip Kumar V, An XAI-Enhanced EfficientNetB0 Framework for Precision Brain Tumor Detection in MRI Imaging,Journal of Neuroscience Methods, 2024,110227,ISSN 0165-0270,https://doi.org/10.1016/j.jneumeth.2024.110227.

[24] Kumaran S, Y., Jeya, J.J., R, M.T. et al. Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. BMC Med Imaging 24, 176 (2024). https://doi.org/10.1186/s12880-024-01345-x

[25] Kudithi, T., Balajee, J., Sivakami, R. et al. Hybridized deep learning goniometry for improved precision in Ehlers-Danlos Syndrome (EDS) evaluation. BMC Med Inform Decis Mak 24, 196 (2024). https://doi.org/10.1186/s12911-024-02601-4

[26] Mahesh TR, Surbhi Bhatia Khan, A. Balajee, Ahlam Almusharraf, Thippa Reddy Gadekallu, Eid Albalawi, Vinoth Kumar; Water quality level estimation using IoT sensors and probabilistic machine learning model. Hydrology Research 2024; nh2024048. doi: https://doi.org/10.2166/nh.2024.048

[27] Natarajan, K., Vinoth Kumar, V., Mahesh, T.R. et al. Efficient Heart Disease Classification Through Stacked Ensemble with Optimized Firefly Feature Selection. Int J Comput Intell Syst 17, 174 (2024). https://doi.org/10.1007/s44196-024-00538-0

Edited by: Dhilip Kumar V
Special issue on: Unleashing the power of Edge AI for Scalable Image and Video Processing
Received: Sep 2, 2024
Accepted: Nov 8, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2016–2026, DOI 10.12694/scpe.v26i5.4851

RESEARCH ON THE PATH OF EMPOWERING RURAL REVITALIZATION WITH E-COMMERCE UNDER THE BACKGROUND OF DIGITAL ECONOMY

ZHENYA ZHAO* AND XIAOLAN FENG †

Abstract. Effective utilization of marketing data can enhance market mobility and competitiveness, thus increasing the number of users. In order to help enterprises proactively deal with competition in the Shaanxi Province market, this study validates and products a game theory-assisted machine learning scheme based on the theory of game decision-making and conditional relaxation. The technique helps enterprises identify users' churn tendencies and adopt differentiated handling strategies through machine learning models. Specifically, it enables enterprises to formulate targeted migration strategies and accurately identify "abnormal" users that may flow in or out. According to the experimental results, the program has been successfully transformed into a product, which greatly improves the marketing effect of enterprises. For example, in one company, the number of users increased by about 50%. In addition, this study analyzes the interest preferences of moviegoers on several dimensions by comparing the seed sets of the Movie Consumer Network and the Yahoo Answer Network.

Key words: Increase in users; Game theory; Machine learning; Marketing data; Comparative analysis

1. Introduction. With the development of Internet technology, the marketing medium has become more diverse, encompassing the emerging Internet, social networks, and mobile applications in addition to the more traditional media of radio, television, newspapers, and magazines. The new communication media also bring new marketing ideas and a range of innovative marketing techniques[1]. Two of these characteristic marketing techniques are particularly highlighted in this essay: viral marketing, which is centered on social networking sites, and online marketing, which connects well-known companies with the Internet.

Viral marketing leverages existing social networking platforms or other technology to distribute content in a way similar to a virus replicating and spreading itself in order to promote a brand or other marketing objectives [2]. Viral marketing has garnered significant interest from both industry and academia due to its often low cost and ability to disseminate information quickly and widely. Mathematically, viral marketing can be expressed as an influence maximization problem, or more precisely, how to select a subset of nodes of size K in a social network so that its influence is maximized (activating the most nodes in the network) [3]. In the real world, we might seek to influence as many nodes as we can, as well as to be influenced by as many different nodes as we can, in order to increase the likelihood of reaching the target audience and reduce marketing risk.

Meituan and Dianping in the catering and entertainment industry, Didi Express in the taxi industry, Ctrip and Qunar in the travel and tourism industry, and Soufun and chain home real estate in the real estate industry are just a few examples of traditional industries that have been seriously affected by the Internet in recent years [4]. When it comes to consumer consumption and review data, trajectory data or real estate market data, the Internet process makes it easier to collect and obtain data from traditional industries [5]. We can improve user experience and make marketing more intelligent and targeted by mining this data.

We acquired transaction data, such as dates, costs, specifics of property qualities, and pertinent geographic location information, for a commercial real estate company in Shaanxi Province. We made an effort to estimate the time needed for the house's sale using the information supplied [6]. By helping sellers set fair asking prices for properties that sell at the right time and helping buyers ascertain the property's popularity and the most effective bidding strategies, these pieces of information can assist both buyers and sellers in making decisions.

*School of Economics and Management, Gansu Vocational College of Communications,Lanzhou,Gansu, 730207, China; School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China (Corresponding author, jilly_ya@163.com)

†School of Information Engineering, Xi'an Translation Institute, Xi'an, Shaanxi, 710105, China (15389050978@163.com).

2. Game-theoretic based decision algorithms.

2.1. Game theoretic algorithms. Game theory has been effectively utilized in various domains, such as analyzing transaction selection, pricing competition in sales, and fraud penalties [7]. Building on this success, we propose applying a game theory approach to strategically evaluate the offensive and defensive marketing behaviors of mobile network operators (MNOs). This framework is designed to optimize the marketing decisions made by enterprises operating within the mobile network operations mode, considering the complex interplay between competitors and market conditions. In this context, MNOs must make critical marketing choices, including pricing strategies, customer acquisition efforts, and retention tactics, all while anticipating the reactions of competing companies. By modeling these decisions through game theory, operators can predict the outcomes of different strategies and select the optimal course of action. Specifically, the offensive strategies focus on gaining market share and attracting new customers, while defensive strategies aim at retaining existing users and minimizing churn. This game-theoretic framework allows MNOs to account for both short-term and long-term market dynamics. For example, it helps businesses evaluate the effectiveness of price cuts, promotional offers, and loyalty programs while anticipating potential counteractions from competitors. By understanding the strategic responses of other players in the market, MNOs can better position themselves to maximize profitability and reduce the risk of adverse outcomes, such as customer defections or revenue loss. Moreover, this approach offers the flexibility to incorporate various market variables, such as customer preferences, competitor behaviors, and external factors like regulatory changes. Ultimately, the proposed game theory-based marketing strategy enables MNOs to navigate the competitive landscape with a more informed and proactive approach, leading to better decision-making and improved financial performance. Fig.2.1 depicts the workflow of the suggested game theory-based offensive and defensive marketing strategy approach, which entails feature extraction of the most important business data, thorough scoring of the worth of the company, a market game with rivals, and the production of the optimal decision strategy. For Chinese mobile network operators, a decision algorithm based on game theory has been created to maximize the effectiveness of mobile network operations [8]. The algorithm dynamically outputs the best current marketing decision, including maximum carry-in ($\max_{in}$), minimum carry-out ($\min_{out}$) or a combination of maximum carry-in and minimum carry-out $\left(\max_{in/out}\right)$, depending on the current marketing objectives and market conditions.

Algorithm 1 gives the algorithmic process of the game in the MNP business of the firm. First, the important features $X_i, i = 1, 2, ..., n$ of the firm's public and/or private business data are selected as inputs and the relative importance matrix β between X_i and X_j is calculated as follows:

$$\beta_{i,j} = \begin{cases} 1, X_i \text{ is more important than } \mathrm{X}_j \\ 0, X_i \text{ is not as important as } \mathrm{X}_j \end{cases} \tag{2.1}$$

The score of feature importance a_i and weight ω_i can be calculated using Eq.2.2 andEq.2.3:

$$a_i = \sum_{j=1}^{n} \beta_{i,j} \tag{2.2}$$

$$\omega_i = \frac{a_i}{\sum\limits_{j=1}^{n} a_i} \tag{2.3}$$

Assuming a marketing objective of $\tilde{x}_i^v$ for feature i and marketing strategy v in the MNP business, where v denotes $\max_{in}, \min_{out}$ or $\max_{in/out}$, the baseline score for the marketing objective is shown in Table 2.1, $\tilde{s}_i^v$ is the baseline score for $\tilde{x}_i^v$ and the combined value of marketing strategy v is:

$$Z_v = \sum_{i=1}^{n} \omega_i \tilde{s}_i^v \tag{2.4}$$

V represents the combination value of marketing strategies, used to measure the effectiveness of different marketing strategies. S represents the benchmark score for features, used to evaluate marketing objectives for

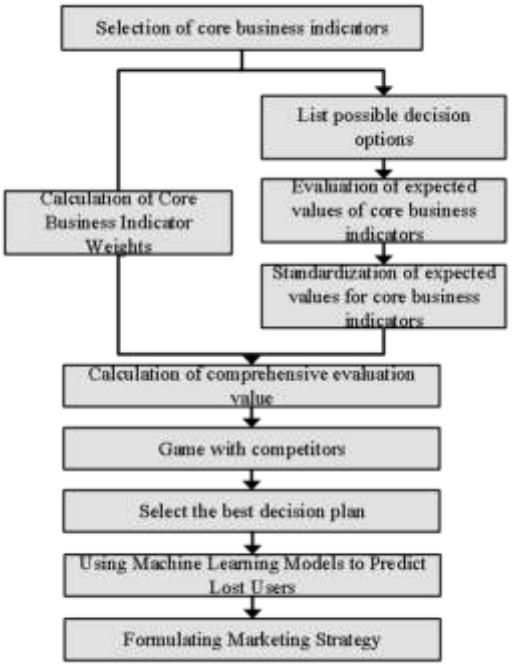


Fig. 2.1: Workflow of Attack and Defense Marketing Decision Evaluation Method Based on Game Theory.

Algorithm 2 Game algorithm for Chinese enterprises to port their numbers to other networks

Selecting core indicators $X_i, i = 1, 2, ..., n$.
The parameters $\beta_{i,j}, a_i$ and $\omega_i for v \in \left\{\max_{in}, \min_{out}, \max_{in/out}\right\} do$ are calculated by Eq.2.1 to Eq.2.3 respectively.
For i^{th} characteristics and v^{th} marketing strategy, $i = 1, 2, ..., n$,set operational target expectations $\widetilde{x}_i^v$.
The scores were calculated using Table 2.1, $\tilde{s}_i^v, i = 1, 2, ..., n$ to calculate the overall evaluation value $Z_v = \sum_{i=1}^{n} \omega_i \tilde{s}_i^v$.
For All $v \in \left\{\max_{in}, \min_{out}, \max_{in/out}\right\} do$.
For All $o \in \{A, B, C\}$ *do*.
Calculate the score by Eq.2.5: $S_{v(i)}^{o(m)}$.
Calculate mathematical expectations for all $v \in \left\{\max_{in}, \min_{out}, \max_{in/out}\right\}$ and $o \in \{A, B, C\}$ all $E\left\{S_{v(i)}^{o(m)}\right\}$.
Output the best i for all $o \in \{A, B, C\}$.

different features. J may represent an index of a specific marketing strategy or market segment, and its specific role needs to be further explained.

Due to the fact that the three major telecommunications companies A, B, C in China have three possible marketing strategies in actual games, namely $\max_{in}, \min_{out}$ and $\max_{in/out}$, and any two companies will play games with each other, there are 27 possible permutations and combinations of their game strategies. In order to obtain the optimal solution of each player, all possible game theory results are calculated by Eq.2.5.

$$S_{v(i)}^{o(m)} = 2Z_{v(i)}^{o(m)} - Z_{v(j)}^{o(n)} - Z_{v(k)}^{o(l)} \tag{2.5}$$

Among them, $S_{v(i)}^{o(m)}$ is the final score of Enterprise $o\,(m)$ after selecting Strategy $v\,(i)$ and playing games with the other two enterprises. Z_v^o is the comprehensive score of enterprise o marketing strategy v calculated by Eq.2.4, where the values of $o \in \{A, B, C\}\, v \in \left\{\max_{in}, \min_{out}, \max_{in/out}\right\}, i, j, k, m, n, l$ are 1, 2, or 3, and

Table 2.1: Scoring benchmarks.

Characteristic	$Score \geq 5$	$Score \geq 4$	$Score < 4$	$Score \geq 3$	$Score < 3$	$Score \geq 2$	$Score < 2$	$Score \geq 1$	$Score < 1$	$Score = 0$
$\tilde{x}_1$	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}
$\tilde{x}_2$	S_{20}	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}
$\tilde{x}_3$	S_{30}	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
					$\vdots$					
$\tilde{x}_n$	S_{n0}	S_{n1}	S_{n2}	S_{n3}	S_{n4}	S_{n5}	S_{n6}	S_{n7}	S_{n8}	S_{n9}

$m \neq n \neq l$.

By calculating the expected value $E\left\{S_{v(i)}^{o(m)}\right\}$ for all v and o, the optimal strategy for enterprise $o\,(m)$ can be obtained.

$$v\,\{i\} \leftarrow \max\left[E\left\{S_{v(i)}^{o(m)}\right\}\right] \tag{2.6}$$

All $v \in \left\{\max_{in}, \min_{out}, \max_{in/out}\right\}$ and all $o \in \{A, B, C\}$ are considered here.

2.2. Subscriber churn prediction algorithm. In the Chinese telecommunications industry, subscribers are limited to three major service providers: China Mobile, China Unicom, and China Telecom. With the implementation of the Mobile Number Portability (MNP) policy [9], subscribers are now able to switch between these providers while retaining their original phone numbers. This policy introduces new challenges for telecom companies, as they must now proactively anticipate and manage subscriber turnover in order to maximize their revenue. To stay competitive, these businesses must develop strategies that not only attract new customers but also reduce churn by understanding the factors that influence customers' decisions to switch providers. By leveraging data analytics and predictive modeling, telecom companies can optimize their marketing efforts and improve customer retention rates.

Assuming that user i decides to move from firm m to firm n, the transfer decision $p_{m,n}^{i}$ can be expressed as:

$$p_{m,n}^{i} = \sum_{k} \delta_{m,k} p_{k,n}^{i} \tag{2.7}$$

where $\delta_{m,k}$ is the Dirac Delta formula. The transfer matrix can be constructed from Eq.2.7. For any given user, his choice is deterministic and unique. Thus, the user's transfer decision $P_{m,n}$ can be written as:

$$P_{m,n} = \begin{cases} 1, \text{Usertransitionfromton} \\ 0, \text{other} \end{cases} \tag{2.8}$$

From Eq.2.8, it can be seen that only one element of the transfer matrix is 1 and all other elements are 0. Assuming that there are N subscribers in the entire communications market, at time $t+1$ (in months), the total number of subscribers carrying into s firm is:

$$in\,(m, t+1) = \sum_{i=1}^{N} \sum_{n \neq m} p_{nm}^{i}\,(t) \tag{2.9}$$

and the total number of users carried out from the enterprise m:

$$out\,(m, t+1) = \sum_{i=1}^{N} \sum_{n \neq m} p_{nm}^{i}\,(t) \tag{2.10}$$

In order to attract users to bring in and reduce users to bring out, the transfer decision matrix of any user i needs to be known in advance M^i. According to previous research, there are many factors that influence a user's transfer decision, including education level, package bundle status, portability costs, call quality, monthly fees and call duration [10]. Therefore, a user's transfer decision will depend on the attributes of the user, i.e.:

$$p^i_{m,n}(t) = p^i_{m,n}\left(t, D^i\right) \tag{2.11}$$

where D^i is the characteristic of user i.

The challenge now is to obtain $p^i_{m,n}$ from $\left(t, D^i\right)$ in Eq.2.11. Recently, some studies have proposed to predict $p^i_{m,n}$ using Markov chains, logistic, social network analysis and some other statistical methods [11]. Given that consumers have three options—remaining with the original firm and moving to the other two firms—the transfer decision problem is essentially a triple classification problem. In order to achieve great prediction performance and interpretability, users' transfer decisions are predicted in this research using a stacked neural network model (SNN) [12, 13]. Additionally, some additional models were employed for comparisons, including Random Forest (RF), Decision Tree (DT), Long Short Term Memory Network (LSTM), Gradient Boosted Tree (GBDT), XG Boost, Extra Tree, and Fully Connected Neural Network (NN) [14, 15]. Due to the relatively small number of user carry-outs, resampling and down-sampling techniques were used to address the problem of data imbalance. After a user churn prediction model is constructed, the expected return for each strategy can be expressed as:

$$\max\left\{in\left(m, t+1\right)\right\} = \max\left\{\sum_{i=1}^{N}\sum_{n\neq m}\left(p^i_{nm}\right)^*\left(t, D^i\right)\right\} \tag{2.12}$$

$$\min\left\{out\left(m, t+1\right)\right\} = \min\left\{\sum_{i=1}^{N}\sum_{n\neq m}\left(p^i_{nm}\right)^*\left(t, D^i\right)\right\} \tag{2.13}$$

$$\max\left\{\frac{in\left(m, t+1\right)}{out\left(m, t+1\right)}\right\} = \max\left\{\frac{\sum\limits_{i=1}^{N}\sum\limits_{n\neq m}\left(p^i_{n,m}\right)^*\left(t, D^i\right)}{\sum\limits_{i=1}^{N}\sum\limits_{n\neq m}\left(p^i_{m,n}\right)^*\left(t, D^i\right)}\right\} \tag{2.14}$$

where p^* is the machine learning prediction of user transfer decisions.

Using this approach, whatever the strategy obtained from game theory, the transfer decision matrix first needs to be obtained by a machine learning model. For strategies $\max_{in}, \min_{out}$ and $\max_{in/out}$, the machine learning models for predicting the propensity of users to shift are the same. After identifying the users that will be lost, companies can target their marketing activities. At the same time, the results of the machine learning model can be used as a basis for game-theoretic decisions.

3. Condition loosening and expansion. According to the general framework mentioned earlier, we need to compute μ^s_j (i.e., the probability that node j is activated when the seed set is S) to calculate the diversity of the affected population. However, a common way to compute μ^s_j is to run Monte Carlo simulations enough times (e.g., 20,000 times), which is very time-consuming, even for medium-sized networks (e.g., a graph with 10k nodes and 100k edges. In fact, some studies give methods on how to estimate μ^s_j, however, most of these studies address special cases, such as assuming that influence propagates along the shortest path, or that nodes have a small probability of influencing each other [16]. Furthermore, most of these methods do not have theoretical performance guarantees.

This motivates us to define a direct diversity measure that can be effectively computed. To simplify the problem, we focus on optimizing the diversity of the seed set nodes instead of directly optimizing the diversity of the affected population. This approach is inspired by the concept of "homogeneity" in social networks [17], where the spread of influence often results in a concentration of similar behaviors or opinions. By targeting the

seed set, we aim to maximize the potential for diversity within the broader network, thus indirectly achieving the desired diversity in the affected population. This approach is referred to as a relaxation problem, where the original, more complex problem is simplified to facilitate computational efficiency while maintaining the core objectives.

$$\begin{aligned} &\max F_s(S) = (1-\gamma)\frac{\sigma(S)}{\overline{\sigma}} + \gamma\frac{D(S)}{\overline{D}}, \\ &s.t. \quad S \subseteq V, |S| = K \end{aligned} \tag{3.1}$$

where $D(S)$ is the value of the diversity of the seed set S.

In particular, we choose two forms of $D(S)$:

$$D(S) = \sum_{i=1}^{C} f\left(\sum_{j \in S} w_{ji} * 1\right) \tag{3.2}$$

$$D(S) = \sum_{i=1}^{C} f\left(\sum_{j \in S} w_{ji}\sigma(\{j\})\right) \tag{3.3}$$

We call Eq.3.2 a consistent diversity metric because each node's contribution to diversity is the same.

We call Eq.3.3 a weighted diversity metric because the contribution of seed node j to diversity is weighted by the influence of that node.

However, the time complexity of Eq.3.1 is still high, as the calculation of $\sigma(S)$ still requires a Monte Carlo simulation. Nevertheless, this relaxed thinking can guide us to incorporate diversity into some impact maximization heuristics. Next, we present specific methods using the degree centrality metric (degree centrality) and PageRank as examples.

The degree centrality approach is very simple: the node with the largest K degrees is selected, while PageRank selects the node with the largest K PageRank values. These heuristics can also be seen as optimizing specific submodular functions. For example, the degree centrality approach can be seen as finding a set S of nodes of size K to maximize $\sum_{i \in S} \deg(i)$, where $\deg(i)$ is the degree of node i. Thus, we can easily incorporate diversity into the degree centrality approach using this relaxed form (Eq.3.1):

$$\begin{aligned} &\max F_d(S) = (1-\gamma)\frac{\sum_{i \in S} \deg}{\overline{\deg}} + \gamma\frac{D(S)}{\overline{D}}, \\ &s.t. \quad S \subseteq V, |S| = K \end{aligned} \tag{3.4}$$

Similarly, we have consistent indicators of diversity:

$$D(S) = \sum_{j \in S}^{C} f\left(\sum_{j \in S} w_{ji} * 1\right) \tag{3.5}$$

and weighted diversity indicators:

$$D(S) = \sum_{i=1}^{C} f\left(\sum_{j \in S} w_{ji} \deg(j)\right) \tag{3.6}$$

Replacing the degree of a node with the PageRank value of the node, we formalize the diverse PageRank heuristic as follows:

$$\begin{aligned} &\max F_P(S) = (1-\gamma)\frac{\sum_{i \in S} PR(i)}{\overline{PR}} + \gamma\frac{D(S)}{\overline{D}}, \\ &s.t. \quad S \subseteq V, |S| = K \end{aligned} \tag{3.7}$$

Table 4.1: Some statistics for the graph data.

Figure	Number of nodes	Number of edges	Number of categories
Film consumption network	2,113	276,676	20
Yahoo Q&A Network	42,043	194,619	31

where $PR(i)$ is the PageRank value of node i. Similarly, we have consistent diversity metrics:

$$D(S) = \sum_{i=1}^{C} f\left(\sum_{j \in S} w_{ji} * 1\right) \tag{3.8}$$

and weighted diversity indicators:

$$D(S) = \sum_{i=1}^{C} f\left(\sum_{j \in S} w_{ji} PR(j)\right) \tag{3.9}$$

4. Analysis of experimental results. Movie Lens and Yahoo Answers data were included in the experimental dataset we used. The benchmarking technique we employed and the evaluation metrics are then presented. Finally, the approaches' performance comparison and efficiency comparison findings are shown.

4.1. Experimental datasets. We use two real datasets, one from Movie Lens * and the other from Yahoo! Answer). Especially, we have built a movie consumption network based on the Movie Lens dataset and a Yahoo!

In a movie consumption network, each node represents a movie. If a user evaluates movie A and then movie B, we insert a directed edge $(A \to B)$ between A and B. We then replace parallel edges (heavy edges) with edges with weights, which are the number of parallel edges. Then, because these edges are probably noisy, we filter away edges with weights below a predetermined threshold (in the trials, this threshold was set at 10). Then, we multiply the incoming edges of every node by a weighting factor such that the sum of the incoming edge weights for every node is a number between 0 and 1 (representing the probability of the node being activated, which in our experiments was set to 0.1), so the edge weights can be thought of as activation probabilities. So an edge with weight P $(A \to B)$ means that the probability of a user watching movie A and then going to see movie B is P.

In the Yahoo Answers network, we selected the 30 categories with the most responses as our experimental data. Nodes represent users, both askers and answerers. The category of a node is the category of the question asked by that user. If a user A answers a question asked by user B, we insert a directed edge $(A \to B)$ between A and B. These edges are then assigned weights in a similar way to the movie consumption network and are not repeated.

We assume that the distribution of nodes across the tagged categories is uniformly collinear due to the lack of detailed information regarding the exact affiliation of nodes to categories. For instance, a movie may be labeled under both "love" and "war" categories, but it is unclear how the movie aligns with each category or to what degree. In the context of films, the category is defined by the genre of the film; for Yahoo Answers data, the category corresponds to the topic area in which the user asked the question, such as "technology," "health," or "entertainment." This uniform assumption simplifies the analysis but also limits the accuracy of category-specific inferences. Table 4.1 presents the basic details and properties of the two datasets used in this study, providing further insights into their structure and the relationships between nodes and categories. By considering the assumptions about the uniform distribution and categorization, we can attempt to better understand the patterns in user behavior, content interaction, and their impact on network dynamics. However, future work could explore incorporating more granular data on category affiliation to improve model precision and offer a deeper understanding of cross-category influence.

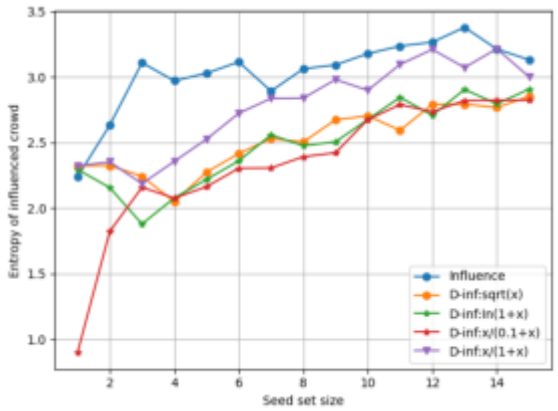


Fig. 4.1: Influence diversity on the diversity of affected populations in two datasets (Movies).

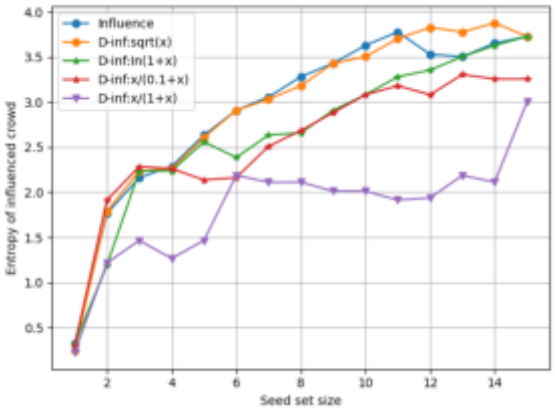


Fig. 4.2: Influence diversity on the diversity of affected populations in two datasets (Yahoo).

4.2. Baseline methods for experiments. In order to evaluate the effectiveness of our proposed diverse impact maximization approach, we chose the following baseline methodology:

Influence: Basic greedy algorithm for maximizing influence (Eq.3.1).

PageRank: greedy algorithm for maximizing the sum of the PageRank values of the set of seed nodes.

Degree: greedy algorithm for maximizing the sum of the degrees of the set of seed nodes (degree centrality method).

The methods we proposed in the previous subsections are summarized as follows:

D-mf: greedy algorithm for maximizing diversity impact (Eq.2.14).

Seed-DU(W): greedy algorithm for maximizing the diversity and influence of the set of seed nodes with a consistent (weighted) diversity measure (Eq.3.1).

Deg-DU(W): a greedy algorithm for maximizing the sum of the diversity of the set of seed nodes with a consistent (weighted) diversity measure (Eq.3.4).

PR-DU(W): a greedy algorithm to maximize the sum of the diversity PageRank values of the set of seed nodes with a consistent (weighted) diversity measure (Eq.3.7).

Although our formalism is general for both the independent cascade model and the linear threshold model, the implementations of $In-fluence, D-Inf$ and $Seed-DU\,(W)$ are different under different influence models. We chose the independent cascade model for the experimental validation of our proposed strategy because it is more widely used and shares similarities with the linear threshold model, such as the fact that they both have equivalent live-edge graph models and their generalised versions are equivalent.

4.3. Evaluation indicators. We chose the widely used Shannon entropy to measure the diversity of the influenced population. After obtaining the seed set using each method, we run 20,000 Monte Carlo simulations under an independent cascade model to estimate the influence of the seed set $\sigma\,(S)$ and a vector consisting of the probability of each node being activated μ^S. The Shannon entropy of the influenced population is defined as:

$$Shannonentropy = \sum_{i=1}^{C} -p_i \log_2 p_i \tag{4.1}$$

where $p_i = \frac{\sum_{j=1}^{|V|} w_{ji}\mu_i}{\sum_{j=1}^{|V|} \mu_i} p_i$, has an intuitive physical meaning, i.e. the number of people activated in category i as a proportion of all activated people. Please refer to Fig.4.1, Fig.4.2, Fig.4.3 and Fig.4.4 for details.

4.4. Diversification of the seed set. Fig.4.5 to Fig.4.12 respectively show the results of seed set diversification on the movie consumption network and Yahoo Q&A network.

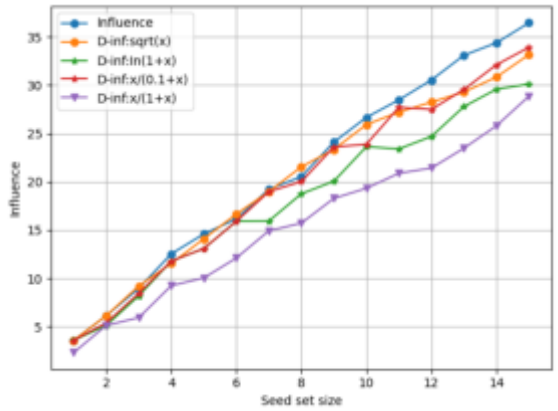


Fig. 4.3: The impact of diversification on two datasets (Movie).

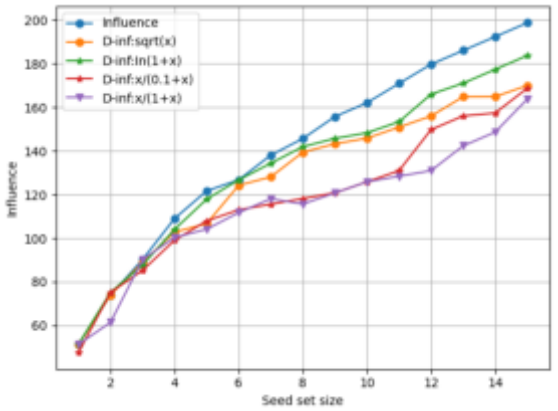


Fig. 4.4: The impact of diversification on two datasets (Yahoo).

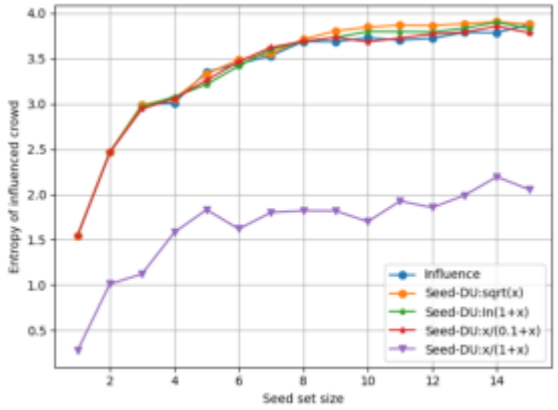


Fig. 4.5: Seed diversification affects the diversity of the affected population in the film consumption network (Unified).

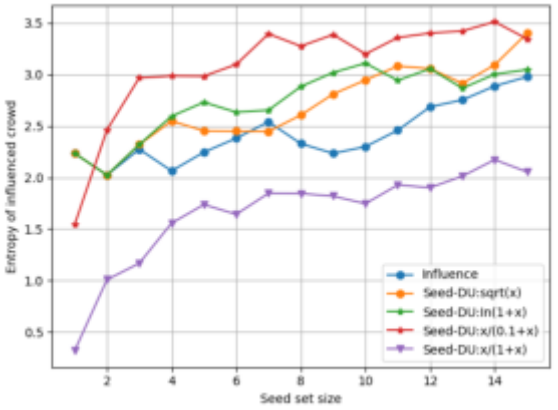


Fig. 4.6: Seed diversification affects the diversity of the affected population in the film consumption network (Weighted).

In both figures, γ is also set to 1, as in the previous subsection. Two things can be seen from the experimental results: first, diversifying the seed set (i.e., maximizing $F_s(S)$ in Eq.3.1) also leads to a more diverse population, which confirms the validity and reasonableness of the model. Second, we can see that in the $\gamma = 1$ setting, the consistent diversity indicator generally corresponds to higher diversity (and lower influence) than the weighted diversity indicator. Since $\gamma = 1$ corresponds to the maximum value of diversity we can obtain, these plots show that the consistent diversity indicator is more tunable in terms of balancing diversity and influence.

5. Conclusion. In response to the increasing market rivalry and the need for businesses to adapt quickly, this study introduces a novel framework that combines game theory and machine learning to enhance marketing strategies. By applying decision-making principles from game theory and incorporating conditional relaxation techniques, the proposed framework helps businesses proactively identify user behavior patterns, such as the likelihood of user churn or migration. Using a machine learning algorithm, the framework allows businesses to tailor their marketing strategies to different user segments, improving engagement and retention. This framework has been successfully validated through experimental data and subsequently productized, leading

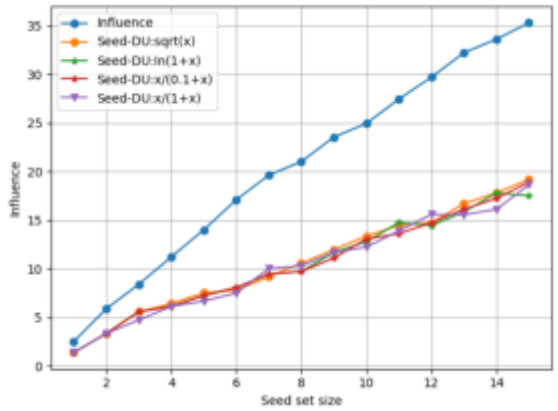


Fig. 4.7: Seed diversification affects the diversity of the affected population in the film consumption network (Unified).

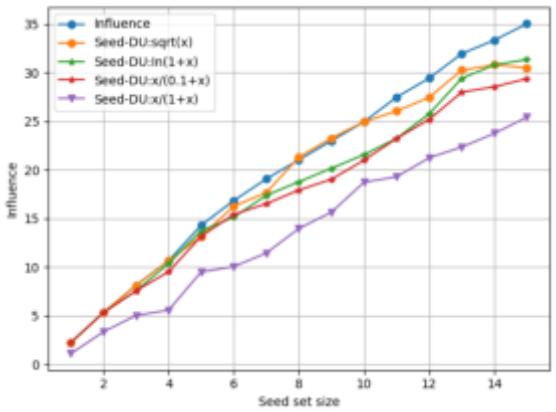


Fig. 4.8: The impact of seed diversification on the film consumption network (Weighted).

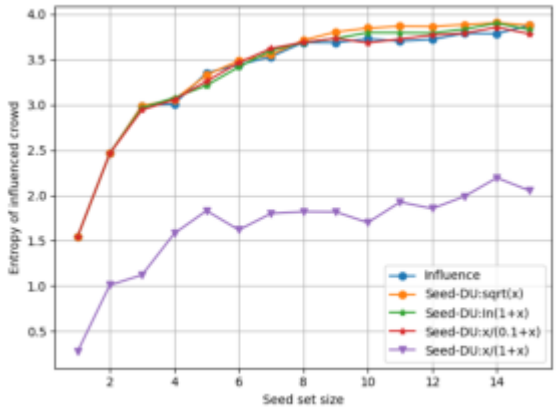


Fig. 4.9: Diversity of Seed Sets and the Affected Population of Yahoo Q&A Network (Unified).

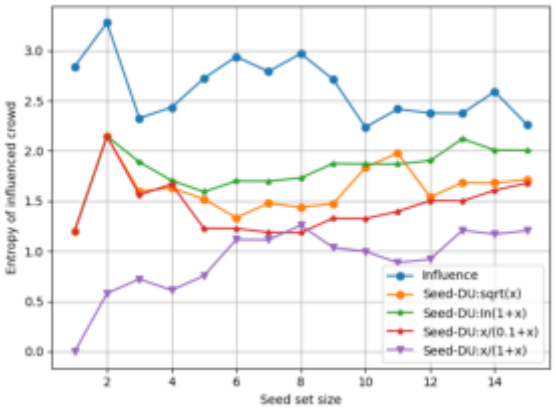


Fig. 4.10: Diversity of Seed Sets and the Affected Population of Yahoo Q&A Network (Weighted).

to a significant boost in marketing effectiveness. For instance, one large company implemented the solution and saw a 50% increase in its subscriber base, demonstrating the practical benefits of applying these advanced techniques to real-world marketing challenges.

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

Funding Statement. This work was supported by Provincial Youth Talent Project of Gansu Province (2024QNGR32) and Research on the Path of Empowering Rural Revitalization with E-commerce in Shaanxi Province under the Background of Digital Economy (2024ZD462).

REFERENCES

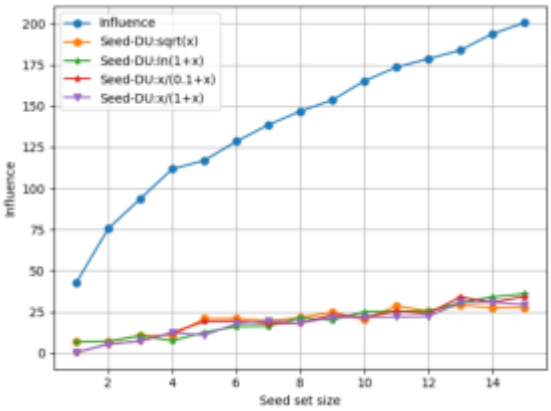


Fig. 4.11: The impact of diversified seed sets on Yahoo Q&A network (Uniform).

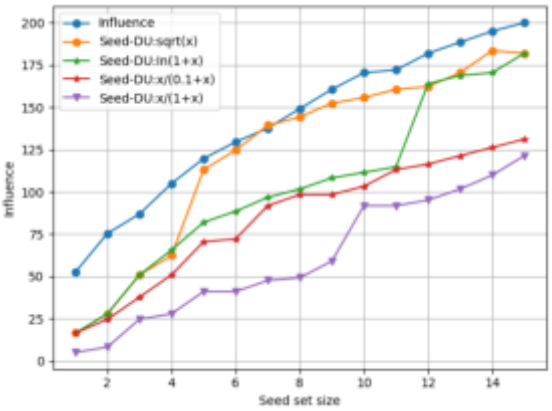


Fig. 4.12: The impact of diversified seed sets on Yahoo Q&A network (Weighted).

[1] Abreu, M., Ferreira, F., & Silva, J. *To be or not to be sustainable in an emerging market? conjoint analysis of customers' behavior in purchasing denim jeans.* Journal of Fashion Marketing and Management: An International Journal, (2022), 26(3), 452-472.

[2] Ying, K., & Zhong, W. *A new data-driven robust optimization approach to multi-item newsboy problems.* Journal of Industrial and Management Optimization, (2023), 19(1), 197-223.

[3] Han, W., & Bai, B. *Pricing research in hospitality and tourism and marketing literature: a systematic review and research agenda.* International journal of contemporary hospitality management, (2022), (5), 34.

[4] Petersen, J. A., & Schmid, F. *Leveraging stakeholder networks with outside-in marketing.* Industrial Marketing Management, (2021), 92(Suppl), 72-75.

[5] Ibrahimov, Z., Hajiyeva, S., Nazarov, V., Qasimova, L., & Ahadov, V. *Bank efficiency analysis of financial innovations: dea model application for the institutional concept.* Marketing & Management of Innovations, (2021), (1), 290-303.

[6] Pollak, F., Vavrek, R., J Váchal, Markovi, P., & Konen, M. *Analysis of digital customer communities in terms of their interactions during the first wave of the covid-19 pandemic.* Management & Marketing: Challenges for the Knowledge Society, (2021), 16(2), 134-151.

[7] Almuraqab, N., & Andleeb, N. *Special issue 2, 2021 1 marketing management and strategic planning.* Academy of Strategic Management Journal, (2021), 20(2), 1-19.

[8] Jiang, H., & Cheng, Y. *Customer–brand relationship in the era of artificial intelligence: understanding the role of chatbot marketing efforts.* Journal of Product & Brand Management, (2022), 31(2), 252-264.

[9] Le, L. H., Kim, J., & Min, J. E. *Impacts of brand familiarity and brand responses on perceived brand credibility, similarity, and blog recommendation intention: a study of corporate blogs.* Journal of Fashion Marketing and Management: An International Journal, (2022), 26(2), 328-343.

[10] Rossi, M., Festa, G., Carli, M. R., & Kolte, A. *Envisioning the challenges of the pharmaceutical sector in the indian health-care industry: a scenario analysis.* Journal of Business & Industrial Marketing, (2022), 37(8), 1662-1674.

[11] Shafiee, M. M. *Knowledge-based marketing and competitive advantage: developing new scales using mixed method approach.* Journal of modelling in management, (2021), (4), 16.

[12] Irfan, I., Chan, F., Khurshid, F., & Sumbal, M.*Toward a resilient supply chain model: critical role of knowledge management and dynamic capabilities.* Industrial Management & Data Systems, (2022), 122(5), 1153-1182.

[13] Urbanski, M. *Company global competitive strategy and employee's awareness.* Marketing and Management of Innovations, (2021), 5(2), 49-64.

[14] Smith, C., Ogutu, M., Munjuri, M., & Kagwe, J. *The effects of foreign market entry strategies on financial performance of listed multinational firms in kenya.* European Journal of Business Management and Research, (2021), 6(3), 216-225.

[15] Acharya, S., Tajane, V., & Prof. Shubhangi. *Novel technologies for processing mushrooms and its marketing strategies.* International Journal of Engineering and Management Research, (2021), 11(1), 93-96.

[16] Abou-Shouk, M., & Soliman, M.*The impact of gamification adoption intention on brand awareness and loyalty in tourism: the mediating effect of customer engagement.* Journal of Destination Marketing and Management, (2021), 20(2), 100559.

[17] Crick, J. M., & Crick, D.*Rising up to the challenge of our rivals: unpacking the drivers and outcomes of coopetition activities.* Industrial Marketing Management, (2021), 96(3), 71-85.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: Sep 9, 2024
Accepted: Feb 27, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2027–2035, DOI 10.12694/scpe.v26i5.4711

THE APPLICATION SYSTEM OF INTELLIGENT WEARABLE DEVICES IN PHYSICAL EDUCATION

YONGLIANG WANG*

Abstract. To address the challenge of real-time data collection and analysis in human motion data mining, the author proposes a system that integrates intelligent wearable devices into physical education. Initially, motion data is gathered through these smart devices, then transformed into binary format. This data undergoes cleaning and supplementation processes before being clustered. The Firefly Algorithm is employed to enhance the K-means clustering technique, which is then applied to the processed data. Experimental results indicate that this refined approach achieves an average recall rate of 97.12% and an average data mining accuracy of 98.42%, thus offering a valuable foundation for the real-time monitoring and assessment of students' physiological metrics. This algorithm can be applied in student physical condition assessment, sports injury and fatigue monitoring, sports posture and movement assessment, personalized training and rehabilitation program development, scientific decision-making and management, and other aspects.

Key words: Smart wearable devices, Human movement, Data mining, Data cleaning, K-means clustering algorithm, data acquisition

1. Introduction. With the rapid development of technology, smart wearable devices have become an indispensable part of people's lives [1]. These devices provide new possibilities for health management and lifestyle improvement by monitoring users' physiological and behavioral data [2]. Especially among the student population, smart wearable devices have received widespread attention due to their interactivity and fun. Student life is a critical stage for physical and psychological development, and exercise not only contributes to physical health, but also has a positive impact on social skills and mental health [3]. Therefore, studying the application of smart wearable devices in student sports is of great significance for promoting students' comprehensive development [4]. There are various types of smart wearable devices, including smartwatches, fitness trackers, exercise monitoring belts, etc. These devices enable the continuous tracking of different physiological metrics, including heart rate, step count, caloric burn, and sleep quality. In recent years, with the advancement of technology, the functions of smart wearable devices have become increasingly powerful, and the user interface has become more user-friendly, resulting in a rapidly growing market trend [5].

Intelligent wearable sports equipment can not only obtain exercise data during the exercise process for fitness tracking and data recording, real-time monitoring of physical indicators, but also help users with exercise analysis and provide targeted fitness, dietary advice, and scientific training plans [6]. Based on the fact that most students only engage in sports activities unilaterally without forming a complete exercise prescription based on their own physical data indicators. In this context, the development and demand for intelligent sports wearables for students have also attracted people's attention. Smart wearable devices have great development prospects and innovation space; Sports smart wearables bring us convenient functions and real-time monitoring of physical health in sports, while advocating for more college students to develop healthy lifestyles and good sports habits. In this context, studying the application system of smart wearable devices in physical education has important theoretical and practical significance. Through research, we can explore how to better integrate these advanced devices with the modern physical education teaching system, solve the pain points in current physical education teaching, and promote the comprehensive development of students' physical literacy [7].

2. Literature Review. In recent years, physical education has developed rapidly, and sports training is the key to improving physical education performance. Therefore, it is of great significance to obtain real-time

*Police Skills and Tactics Training Department Criminal Investigation Police University of China, Shenyang, 110035, Liaoning, China (wangyongliang714@126.com)

data on sports training in order to provide scientific basis for formulating targeted technical and physical training plans, thereby improving students' physical performance and enhancing the guidance and training level of sports training. Therefore, the optimization of real-time data collection methods for sports training has become a hot topic of concern for researchers. In previous studies, a large number of researchers have provided various data mining algorithms, but there are still some shortcomings in the mining of human motion data. Jiang et al. developed an innovative real-time 3D interactive system that leverages smart clothing. This system employs compact sensor modules to gather human motion data and incorporates a dual-stream fusion network with pulse neural units for motion classification and recognition, facilitating seamless real-time interaction between users and the sensors [8]. Dong et al. developed a system for analyzing and evaluating physical education steps using deep learning techniques. The system employs smart wearable devices to continuously track students' exercise steps and heart rate during class, creating a dataset of physical education activities. They utilized a transformer-based deep model to analyze temporal step sequences, enabling the assessment of motion effects [9]. Ma, Z. et al. introduced an affordable solution designed to assist teenagers in practicing basketball dribbling independently. This solution incorporates a smart wearable headband along with a dribbling aid called DribbleAid. The system monitors typical posture problems during dribbling sessions and delivers targeted feedback to help users overcome common challenges [10].

To address these challenges, the author presents a human motion data mining algorithm leveraging intelligent wearable devices. Initially, these devices gather motion data, which is then refined through data cleaning and clustering processes. To enhance clustering accuracy, the algorithm employs the Firefly Algorithm to optimize the K-means clustering approach. The experimental results show that the algorithm designed by the author can quickly and accurately mine human motion data, provide decision-making information for sports training, and promote the development of sports training technology.

3. Method.

3.1. Human motion data mining algorithm . In response to the problem that current human motion data mining algorithms cannot collect and analyze real-time data, resulting in low accuracy and long time for human motion data mining, the author proposes a human motion data mining algorithm based on intelligent wearable devices [11]. Specifically, by analyzing and mining the collected data, the following objectives can be achieved:

1. Assessing students' physical condition and exercise performance: By analyzing data such as heart rate, steps, speed, muscle activity, etc., students' physical condition and exercise performance can be evaluated to determine training intensity and frequency, and improve training effectiveness [12].
2. Monitoring and warning of sports injuries and fatigue: By analyzing data on students' muscle activity and energy expenditure, sports injuries and fatigue can be monitored and warned to avoid overtraining and sports injuries.
3. Optimizing training and competition strategies: By analyzing students' posture, gait, movements, and other data, training and competition strategies can be optimized to improve students' competitive level and performance.
4. Personalized training and rehabilitation programs: By analyzing students' physical condition and exercise performance, personalized training and rehabilitation programs can be developed for different students to meet their needs and requirements.
5. Scientific decision-making and management: By mining and analyzing a large amount of sports data, it can provide coaches, doctors, and managers with scientific decision-making and management basis to improve the quality and effectiveness of training and competition.

From this, it can be seen that the purpose of human motion data mining based on smart wearable devices is to improve students' competitive level and physical health status, optimize training and competition strategies, achieve personalized training and rehabilitation plans, and provide a basis for scientific decision-making and management [13,14].

3.2. Data collection of smart wearable devices. In the process of traditional sports data analysis research, the inability to obtain accurate real-time data results in low reliability of the analysis results. In this study, smart wearable devices were used as collection devices for human motion data, which can obtain

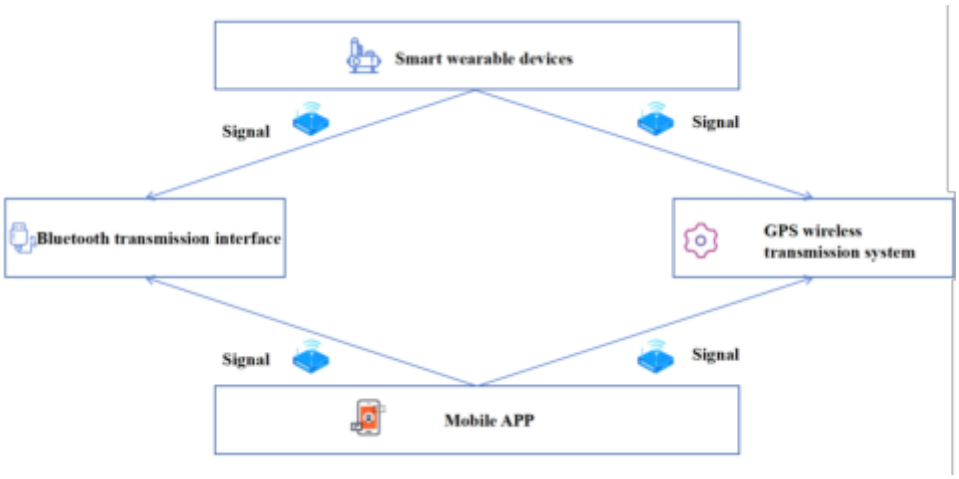


Fig. 3.1: Architecture of Intelligent Wearable Devices

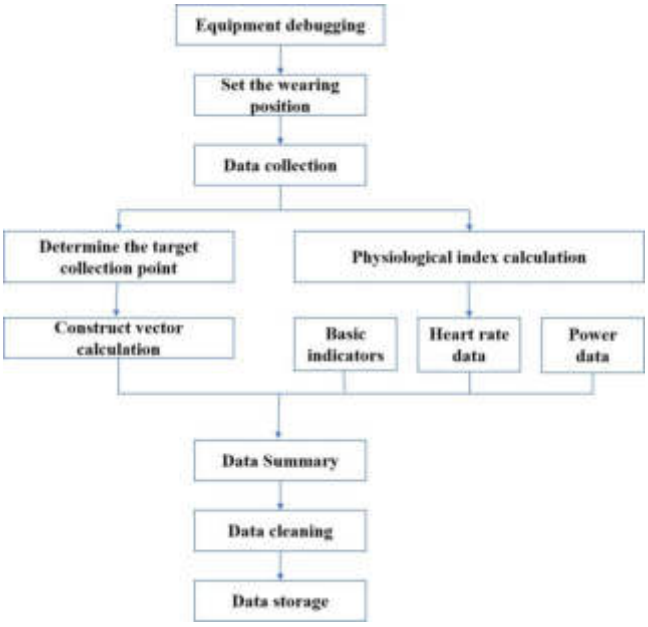


Fig. 3.2: Data Collection Process

real-time motion data of students [15]. The architecture of smart wearable devices is shown in Figure 3.1.

As shown in Figure 3.1, the architecture of smart wearable devices consists of smart wearable devices, Bluetooth transmission ports, GPS wireless transmission systems, and mobile apps. Install smart wearable devices into the elastic sportswear worn by students to collect human motion signals, and transmit the signals to the mobile app through Bluetooth transmission ports and GPS wireless transmission systems to ensure the quality and efficiency of data collection. The process of collecting human motion data for smart wearable devices is shown in Figure 3.2.

As an intelligent motion monitoring device, smart wearable devices can obtain various physiological indicators of students during their exercise process, providing a data foundation for subsequent data mining. During

the data collection process on this device, corresponding settings need to be made [16]. Assuming that the smart wearable device is worn on the limbs of the human body and used as a data collection point, this point can be represented as $q(x, y, z)$. During the data collection process, it is necessary to import its constructed vector into the data collection results, which can be represented as:

$$\begin{cases} \overline{o_1} = q_1 - q_0 \\ \overline{o_2} = q_2 - q_1 \\ \overline{o_3} = q - q_3 \\ \overline{o} = q - q_4 \end{cases} \tag{3.1}$$

In the formula, q_0, q_1, q_2, q_3 and q_4 represent the wearing points of smart wearable devices, and $\overline{o}, \overline{o_1}, \overline{o_2}, \overline{o_3}$ represents different construction vectors.

Using this formula, basic physiological indicators during human movement can be collected. Some physiological indicators cannot be reflected in the form of data and need to be converted into data information in order to complete subsequent analysis work. Therefore, the following indicators were selected for conversion processing in the study:

1) Real time heart rate.

$$g = \frac{\alpha}{r_1 - r_2} = \frac{(e_1 - e_2)}{2} \tag{3.2}$$

In the formula, r_1 and r_2 represent the heartbeat data of the previous time node and the current period, α represents the standard coefficient for heart rate calculation, and e_1 and e_2 represent the blood pressure data of the previous time node and the current period, respectively [17].

2) Total power of exercise. Assuming that the real-time physiological index data fitting function $f(a)$ of students is continuous within a certain constant interval, the total exercise power of students can be obtained based on this fitting function. The specific calculation result can be expressed as

$$\int_j^i f(a) = f(j) - f(i) \tag{3.3}$$

In the formula, $f(i)$ represents the fitting function calculation result of interval endpoint 1, and $f(j)$ represents the fitting function calculation result of interval endpoint 2.

According to the above formula, after sorting and analyzing these two physiological indicators, they will be imported into the same database as the results collected by smart wearable devices, providing a foundation for subsequent human motion data mining [18].

3.3. Organizing and Analyzing Human Movement Data. Due to the diverse categories and different dimensions of the raw data collected by smart wearable devices, in order to better complete the data mining process, it is necessary to preprocess the collected data and integrate it into binary data format. Assuming the collected raw data is $E = (1, 2, 3, \cdots, n)$, the binary digit β_v of physiological indicator data can be represented as

$$\beta_v = (\beta_{v1}, \beta_{v2}, \cdots, \beta_{vn}) \tag{3.4}$$

In the formula, n represents the number of physiological indicators. After completing this part of the processing, set H as the candidate dataset for mining data 3, and calculate the basic support for each physiological indicator. This includes:

$$sup(H) = \sum_{i=1}^{n} N_i \tag{3.5}$$

In the formula, the calculation formula for the support coefficient N_i is as follows:

$$N_i = \begin{cases} 1, \beta_v \cap H = H \\ 0, \beta_v \cap H \neq H \end{cases} \tag{3.6}$$

According to the above formula, integrate the data into binary data format and import it into the database. The raw data collected by smart wearable devices may contain abnormal data, which needs to be cleaned and subjected to secondary analysis. Assuming the data is complete, then $\beta_v = 1$; If the data is abnormal, then $\beta_v = 0$. After discovering abnormal data, use interpolation method to supplement the data. After completing the analysis of all data in the database, conduct human motion data mining.

3.4. Human motion data mining. On the basis of the previous text, the author cited the K-means clustering analysis method in data mining technology to mine and process the data, and constructed a human motion data mining algorithm [19]. Assuming that this part of the data can be divided into k subclasses, and multiple iterative calculations are required during the subclass partitioning process, the variation between each subclass can be expressed as

$$U = \sum_{i=1}^{n} \sum_{w \in L,i} dist(w, l_i)^2 \tag{3.7}$$

In the formula, U represents the deterioration of data category, l_i represents the data cluster code, w represents the centroid of each data cluster, and (w, l_i) represents the Euclidean distance between the observed data and the centroid of the data cluster. Based on this formula, construct a human motion data mining algorithm. Enhance and refine sports data according to its specific features. Given the substantial volume of collected human motion data, extended computation times during processing are an unavoidable challenge. Therefore, in this study, the cocoon firefly algorithm was used to optimize it. Assuming that the core information of each raw data is C, there exists:

$$C_i = f(A_i) \tag{3.8}$$

In the formula, A_i represents the basic information value, and $f(A_i)$ represents the objective function of sports information data mining. According to this formula, the attractiveness between the original data can be determined as follows:

$$\mu = \mu_0 * exp(-vp_{ij}) \tag{3.9}$$

In the formula, μ_0 represents the initial attraction between data, p_{ij} represents the Euclidean distance between two raw data, and v represents the absorption coefficient of attraction between different data [20]. When the basic information of two data sets is different, the clustering mining process of the data can be represented as

$$g_{i+1} = g_i + \mu(g_j - f_i) + \eta(rand - \frac{1}{2}) \tag{3.10}$$

In the formula, g_i and g_j represent the spatial positions of two sports data, μ represents the attractiveness of the data, η represents a random constant, and rand represents any integer from 0 to the maximum random number. Combine formula 3.10 with formula 3.7 to complete the data mining process. Thus, the design of a human motion data mining algorithm based on smart wearable devices has been completed.

3.5. Experimental verification.

3.5.1. Dataset. The author chose the Actions dataset and the measured dataset as the data sources. The Actions dataset contains 7 types of sports data, including diving, skiing, snowboarding, platform diving, racewalking, artistic gymnastics, and trampoline, with 10000 pieces of data. The measured dataset is a spatiotemporal data detection dataset, which comes from human motion measurement data of 30 volunteers. The data can be divided into 6 categories of human movements, totaling 5000 pieces of data. Merge the data from the two datasets into the test data and randomly divide them into 10 groups, each containing data from both datasets. The division results are shown in Table 3.1.

Select test groups XC1-XC5 as the training group and XC6XC10 as the testing group. Apply the proposed algorithm to analyze the data through mining techniques to assess the effectiveness of the algorithm in practice. The types of human motion data collected by smart wearable devices include the following aspects:

Table 3.1: Example Test Data Group

Group number	Data volume	Group number	Data volume
XC1	1500	XC6	1500
XC2	1500	XC7	1500
XC3	1500	XC8	1500
XC4	1500	XC9	1500
XC5	1500	XC10	1500

(1) Heart rate: By measuring electrocardiogram (ECG) signals, students' heart rate data can be obtained to evaluate their exercise status and physical health. The data size is 15.61GB.
(2) Step count and step frequency: Through sensors such as accelerometers and gyroscopes, students' step count and step frequency data can be obtained to evaluate their gait and walking posture, with a data volume of 10.36GB.
(3) Distance and speed: Through sensors such as GPS and accelerometers, students' distance and speed can be obtained to evaluate their exercise intensity and speed, with a data volume of 7.89GB.
(4) Posture and motion trajectory: Through sensors such as gyroscopes and accelerometers, students' posture and motion trajectory data can be obtained to evaluate whether their posture and actions are correct. The data size is 8.96GB.
(5) Muscle activity and energy consumption: Through technologies such as electromyography (EMG) sensors, students' muscle activity and energy consumption data can be obtained to evaluate their muscle state and exercise intensity, with a data volume of 5.96GB.
(6) Blood oxygen and blood pressure: Through technologies such as blood oxygen and blood pressure sensors, students' blood oxygen and blood pressure data can be obtained to evaluate their physical condition and exercise load, with a data volume of 11.34GB.

3.5.2. Experimental indicators. The experiment selected algorithms such as big data analysis based motion wearable smart devices (Algorithm 1), data mining algorithm based motion wearable smart devices (Algorithm 2), information entropy based optimization algorithm (Algorithm 3), improved mean and LSTM based algorithm (Algorithm 4), and full entropy based anomaly data mining algorithm (Algorithm 5) to conduct data mining analysis with the proposed methods. Assess the strengths and weaknesses of different algorithms based on predefined evaluation criteria to confirm that the proposed algorithm meets the design specifications. The evaluation metrics include three main categories: recall rate, accuracy rate, and precision of human motion data collection. By analyzing these indicators in combination, determine the computational performance of the proposed algorithm.

The recall rate calculation formula for human motion data mining:

$$D_1 = \frac{S_0}{S_{all}} \times 100\% \tag{3.11}$$

In the formula, S_{all} represents the data that should be recalled, and S_0 represents the actual recall data. Formula for calculating the accuracy of human motion data mining:

$$D_2 = \frac{S_1}{S_0} \times 100\% \tag{3.12}$$

In the formula, S1 represents the data that has been correctly mined and processed in the recall data. Formula for calculating the accuracy of human motion data collection:

$$D_3 = \frac{S_1}{S_{all}} \times 100\% \tag{3.13}$$

Table 4.1: Comparison Results of Data Recall Rates

Group number	Proposed algorithm	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
XC6	98.874	96.85	95.06	96.14	96.52	95.74
XC7	95.46	95.03	96.03	96.24	96.58	96.01
XC8	96.89	96.03	95.70	96.36	95.10	96.07
XC9	97.26	95.48	96.63	96.56	95.02	96.03
XC10	97.78	95.04	96.80	96.37	96.73	96.23
average value	97.20	95.67	96.05	96.33	96.09	96.02

Table 4.2: Comparison Results of Data Accuracy

Group number	Proposed algorithm	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
XC6	98.12	95.89	95.88	94.18	94.01	95.01
XC7	98.68	94.88	95.64	95.36	94.33	95.22
XC8	98.02	94.07	94.67	94.55	95.13	94.16
XC9	98.16	94.46	94.34	95.57	95.44	94.55
XC10	98.64	94.31	95.44	95.16	94.71	94.49
average value	98.31	94.71	95.29	95.05	94.72	94.65

Table 4.3: Comparison Results of Data Collection Accuracy

Group number	Proposed algorithm	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
XC6	98.49	94.61	93.02	94.06	94.05	93.14
XC7	98.00	94.05	93.55	93.04	94.17	93.03
XC8	97.87	94.04	93.06	93.16	93.81	94.91
XC9	97.57	94.88	95.21	94.17	93.75	94.08
XC10	98.04	95.03	94.43	93.24	93.76	94.25
average value	98.05	94.45	93.83	93.51	94.00	93.82

4. Results and Discussion. Table 4.1 compares the recall rates of various algorithms for human motion data mining. Analysis of the data reveals that the recall rate of the proposed algorithm is comparable to those of the other five algorithms, with any differences being no greater than 4.0%. Through analysis, it can be seen that the average recall rate of the algorithm proposed by the author is 97.31%, which is 1.53%, 1.15%, 0.87%, 1.21%, and 1.18% higher than Algorithm 1, Algorithm 2, Algorithm 3, Algorithm 4, and Algorithm 5, respectively. The algorithm proposed in the experiment has a relatively high recall rate, overall stability, and relatively stable data analysis and processing capabilities, demonstrating a certain degree of feasibility.

From the analysis of the data in Table 4.2, it can be seen that there are differences in the accuracy of data mining between the algorithm proposed by the author and the other five algorithms. Among them, the average data mining accuracy of the algorithm proposed by the author is 98.31%, which is 3.5%, 3.01%, 3.25%, 3.48%, and 3.55% higher than algorithms 1, 2, 3, 4, and 5, respectively. The proposed algorithm demonstrates notably superior accuracy in data mining compared to other methods, highlighting its strong capabilities in data analysis. While the other five algorithms meet the accuracy standards for human motion data, their overall performance in this aspect is relatively lower, suggesting a need for further optimization.

The analysis presented in Table 4.3 reveals that the proposed algorithm significantly outperforms the other five algorithms in terms of human motion data collection accuracy. Specifically, the proposed algorithm achieves an average accuracy of 98.04%, which surpasses the accuracy of algorithms 1, 2, 3, 4, and 5 by 3.3%, 4.01%, 4.22%, 4.04%, and 4.01%, respectively. This demonstrates that the proposed algorithm excels in accurately collecting human motion data compared to the alternatives.

5. Conclusion. In this paper, we put forward a study on the application of intelligent wearing equipment in physical education. In response to the problems that arise in the current sports data analysis research process, smart wearable devices are applied to obtain raw data and a new data mining algorithm is set up. The results of the example show that the proposed method is of great practical value. This algorithm can be applied in student physical condition assessment, sports injury and fatigue monitoring, sports posture and action assessment, personalized training and rehabilitation program development, scientific decision-making and management, etc., to improve students' physical health status, optimize training and competition strategies, implement personalized training and rehabilitation programs, and provide a basis for scientific decision-making and management. Although the research has achieved certain results, this design only focuses on optimizing real-time data collection problems and has not improved other issues in the data mining process. In future research, it is necessary to optimize other aspects of sports data analysis to ensure that the application effect of this algorithm meets the requirements of data analysis.

Funding. This work has received support from 2023 University-level Teaching and Research Project for Criminal Investigation Police University of China, Project Number 3JYQN05.

REFERENCES

[1] Liu, W., Zhu, T., Xie, F., Zhan, B., & Xu, M. (2022). A smart sensing module aims at reduction of motion artifact for wearable tonometry blood pressure device. Journal of Hypertension, 40(Suppl 1), 87.
[2] Zhu, Y., Lu, Y., Gupta, S., Wang, J., & Hu, P. (2023). Promoting smart wearable devices in the health-ai market: the role of health consciousness and privacy protection. Journal of Research in Interactive Marketing, 17(2), 257-272.
[3] Alattar, A. E., & Mohsen, S. (2023). A survey on smart wearable devices for healthcare applications. Wireless Personal Communications, 132(1), 775-783.
[4] Liu, M. Y., Liu, M. Y., Hang, C. Z., Hang, C. Z., Wu, X. Y., & Wu, X. Y., et al. (2022). Investigation of stretchable strain sensor based on cnt/agnw applied in smart wearable devices. Nanotechnology, 33(25), 255501.
[5] Secara, I. A., & Hordiiuk, D. (2024). Personalized health monitoring systems: integrating wearable and ai. Journal of Intelligent Learning Systems and Applications, 16(2), 9.
[6] Zovko, K., Seric, L., & Solic, B. P. (2023). Iot and health monitoring wearable devices as enabling technologies for sustainable enhancement of life quality in smart environments. Journal of cleaner production, 413(Aug.10), 1-12.
[7] Arya, M. B. S., emailprotected, Emailprotected, E., Arya, S., Sandeep Arya * Sandeep AryaDepartment of Physics, University of Jammu, Jammu, Jammu and Kashmir, India *Email: emailprotectedMore by Sandeep Aryahttps://orcid.org/—, & ,and, et al. (2023). Recent developments in wearable nems/mems-based smart infrared sensors for healthcare applications. ACS Applied Electronic Materials, 5(10), 5386-5411.
[8] Jiang, M., Tian, Z., Chenyu, Y. U., Shi, Y., Liu, L., & Peng, T., et al. (2024). Intelligent 3d garment system of the human body based on deep spiking neural network. Virtual Reality and Intelligent Hardware (both in Chinese and English)(001), 006.
[9] Dong, A. (2023). Analysis on the steps of physical education teaching based on deep learning. International journal of distributed systems and technologies, 14.
[10] Ma, Z., & Hao, Q. (2022). Posture monitoring of basketball training based on intelligent wearable device. Journal of healthcare engineering, 2022, 4121104.
[11] Siemon, D., & Wessels, J. (2023). Performance prediction of basketball players using automated personality mining with twitter data. Sport, Business and Management: An International Journal, 13(2), 228-247.
[12] Shu, Y. U., Qiu, C., & Yang, R. (2022). Spatial characteristics of sports tourism destination system based on data fusion and data mining. Mobile information systems, 2022(Pt.11), 1-12.
[13] Junwei, F. (2023). Designing an artificial intelligence-based sport management system using big data. Soft computing: A fusion of foundations, methodologies and applications, 27(21), 16331-16352.
[14] Jianwattanapaisarn, N., & Sumi, K. (2022). Investigation of real-time emotional data collection of human gaits using smart glasses. Journal of Robotics, Networking and Artificial Life, 9(2), 159-170.
[15] Francisco Tomás.González-Fernández, Sarmento, H., Sixto.González-Víllora, Pastor-Vicedo, J. C., Luis Manuel.Martínez-Aranda, & Clemente, F. M. (2022). Cognitive and physical effects of warm-up on young soccer players. Motor control, 26(3), 334-352.
[16] Nayrolu, S., Ylmaz, A. K., Silva, A. F., Silva, R., Nobari, H., & Clemente, F. M. (2022). Effects of small-sided games and running-based high-intensity interval training on body composition and physical fitness in under-19 female soccer players. BMC sports science, medicine & rehabilitation, 14(1), 119.
[17] Zhen, T., Zheng, H., & Yan, L. (2022). Human motion mode recognition based on multi-parameter fusion of wearable inertial module unit and flexible pressure sensor. Sensors and materials: An International Journal on Sensor Technology, 34(3 Pt.2), 1017-1031.
[18] Xia, M. B. H., Xia, H., Hui Xia Hui XiaCollege of Control Science and Engineering, China University of Petroleum East China, Qingdao , ChinaMore by Hui Xia, ,and, Zhang, M. B. H., & Zhang, H., et al. (2023). High-sensitivity wearable

flexible pressure sensor based on mxene and polyaniline for human motion detection. ACS Applied Polymer Materials, 5(12), 10386-10394.
[19] Zhang, T., Zhang, W., & Jiang, Y. T. (2023). High-efficient flexible pressure sensor based on nanofibers and carbon nanotubes for artificial electronic skin and human motion monitoring. Journal of porous materials, 30(6), 1797-1806.
[20] Chen, K., Guo, Y., & Sun, X. (2023). Intelligent fiber optic integrated sensing system for human motion monitoring. Optical fiber technology, 81(Dec.), 1-11.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Aug 24, 2024
Accepted: Oct 10, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2036–2044, DOI 10.12694/scpe.v26i5.4714

NEAR INFRARED SPECTROSCOPY DETECTION OF SOYBEAN FATTY ACID CONTENT BASED ON NEURAL NETWORK COMBINED WITH GENETIC ALGORITHM

LI MIN[*], QI XIAOCUI[†], LI RONGYAO[‡], WANG RUFENG[§] AND YU DANDAN[¶]

Abstract. In order to solve the problem of poor performance in rapid analysis of soybean fatty acid content using traditional methods, the author proposes a near-infrared spectroscopy detection method for soybean fatty acid content based on neural network combined with genetic algorithm. The author first collected sample spectra and preprocessed them using mean centralization and Savitzky Golay smoothing differentiation method. Then, the optimal band was selected through segmented combination modeling, and genetic algorithm was used to further screen wavelength points sensitive to content prediction modeling. Finally, the spectral data was decomposed using principal component analysis (PCA), and the score matrix was input into a 3-layer neural network for training. The optimal model was established through parameter optimization. The experimental results show that the calibration relative analysis error RPD of the model based on genetic algorithm combined with neural network is, indicating that the model has good prediction accuracy and stability. To further validate the model's reliability in practical applications, the correlation coefficient R between the predicted values and the standard values within the model's prediction range was found to be 0.986, indicating a strong linear relationship between them. The relative deviation of each value is less than 2%, and the standard deviation of the difference is 0.0091. Paired t-test =1.2706 was performed on both methods, which is less than =2.354. This indicates that there is no significant difference between the analysis results of this method and the standard method, verifying the strong predictive ability of the model in practical applications. The model designed by the author can accurately and efficiently complete near-infrared spectroscopy detection.

Key words: Neural network, Genetic algorithm, near-infrared spectroscopy, Soybeans, fatty acid

1. Introduction. Soybean Glycine max (L.) Merr. is an important source of plant protein and oil for humans, with an oil content of around [1]. It is the main raw material for the plant oil processing industry. Soybean oil contains a large amount of unsaturated fatty acids, which are beneficial to human health and have attracted widespread attention. However, due to the generally high content of linolenic acid in soybean oil, reducing its storage period, and improving the fatty acid composition from a breeding perspective is an effective method. Therefore, improving the fatty acid composition has become an important part of soybean quality breeding [2]. In the past, gas chromatography and liquid chromatography techniques were mainly used for the analysis of soybean fatty acids, which not only consumed a lot of manpower and material resources, but also caused time waste and reduced breeding efficiency. It can be seen that how to quickly and accurately determine the fatty acid composition in soybean seeds to accelerate soybean quality breeding is an urgent problem to be solved [3].

Near infrared spectroscopy analysis technology is an emerging qualitative and quantitative analysis technique that uses sample spectral data. The rapid and non-destructive measurement method has made this technology increasingly widely used in the determination of physicochemical properties and state monitoring of soybean fatty acids. Near infrared spectroscopy mainly refers to the harmonic and combined frequency absorption generated by the vibration of hydrogen containing groups in molecules. Due to the complex composition of lubricating oil, the spectral bands formed in the near-infrared spectral wavelength range overlap severely, contain a large amount of non target information, and are subject to external signal interference such as background and instrument noise. As a result, spectral preprocessing and calibration techniques greatly

[*]Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China.
[†]Admission and Employment Office, Cangzhou Medical College, HeBei, 061001, China.
[‡]Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China.
[§]Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China.
[¶]Department of Basic Medicine, Cangzhou Medical College, HeBei, 061001, China (Corresponding author, dan5507803@126.com).

influence the model's predictive performance. Commonly used wavelength selection methods include the correlation coefficient method, interval partial least squares (iPLS), the successive projections algorithm (SPA), and genetic algorithm (GA). The genetic algorithm, a global optimization-based wavelength selection method, is particularly effective for selecting wavelengths in multi-component systems and can be integrated with various calibration methods to enhance and optimize the model's predictive performance [4,5,6].

2. Literature Review. Near-infrared spectroscopy (NIRS) is an innovative technique that leverages the vibrational absorption characteristics of organic substances in the near-infrared spectrum to quickly measure the content of various chemical components in samples. At present, there have been some research reports abroad on using NIRS technology to determine fatty acid composition in different crops. Deng, Z. et al. investigated the use of fatty acid composition and near-infrared spectroscopy for tracing the geographical origin of camellia oil in Hainan. They measured the relative content of 16 fatty acids in camellia oil samples using gas chromatography and gathered spectral data through near-infrared spectroscopy. The study demonstrated that both fatty acid composition and near-infrared spectroscopy are effective tools for accurately determining the geographical origin of camellia oil [7]. Zhe, L. I. et al. devised a method using near-infrared reflectance spectroscopy (NIRS) to rapidly and approximately assess the nutritional composition of Pacific oysters (Crassostrea gigas) [8]. Tonolini, M. et al. explored the application of near-infrared (NIR) spectroscopy combined with chemometrics to monitor the degumming process by measuring diglyceride levels in soybean oil in both offline and online environments. To develop an NIR spectroscopy prediction model, they studied 15 lab-scale batches of enzyme degumming, utilizing various water, acid, and enzyme doses from the final screening design. The findings suggest that glycerol diester content serves as a reliable indicator of enzyme effectiveness, and that NIR spectroscopy is an appropriate analytical technique for real-time quantification of glycerol diesters [9].

The author studied the use of genetic algorithm for wavelength screening in the determination of soybean fatty acid content by near-infrared spectroscopy, and combined it with a highly nonlinear prediction ability backpropagation neural network to establish an analysis and correction model, effectively improving the accuracy and adaptability of the model prediction.

3. Method.

3.1. Basic Framework. The author presents a convolutional neural network (CNN) learning method enhanced by genetic algorithm optimization, as illustrated in Figure 3.1. The CNN framework includes input layers, convolutional layers, pooling (sampling) layers, fully connected layers, and output layers. The number of convolutional and pooling layers is not fixed, but rather adjusted based on computational tasks and available resources [10]. A significant challenge with traditional CNNs is that their learning performance heavily depends on the initial weight settings of the convolutional and fully connected layers. To address this issue, the author employed a genetic algorithm to optimize the weights of each convolutional and fully connected layer during the training process. The basic idea is to construct the chromosomes of the genetic algorithm based on these weights, and obtain the optimal solution through selection, crossover, and mutation operations.

3.2. Configuration and Training of Convolutional Neural Networks.

(1) Configuration of Convolutional Neural Networks. Convolutional neural network classifiers typically consist of two parts: convolutional neural networks used for feature extraction and classifiers used for feature classification. The hierarchical framework of convolutional neural networks is shown in Figure 3.2. Assuming that the feature dimension of each sample is d and the total number of samples in the training set is m, storing the features of each sample in rows, where each row represents a sample, the input layer size of the convolutional neural network is $m \times d$ [11].

During the training process, continuously learn convolutional masks to identify the most discriminative local patterns in the classification process. Due to the fact that in the input layer, the features of each d-dimensional sample are stored in rows, the convolution mask only uses linear convolution masks in the horizontal direction [12]. To minimize the complexity of pattern recognition and enhance computational efficiency, the author applied a convolution mask in the initial convolutional layer. In the following convolutional layers, the number of convolution masks sequentially doubles—for instance, the second convolutional layer has 2 masks, the third layer has 4 masks, and so forth.

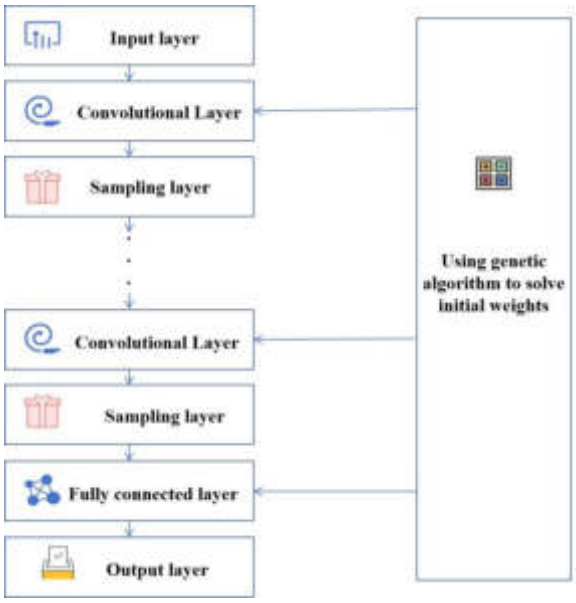


Fig. 3.1: Convolutional Neural Network Learning Framework Optimized with Genetic Algorithm

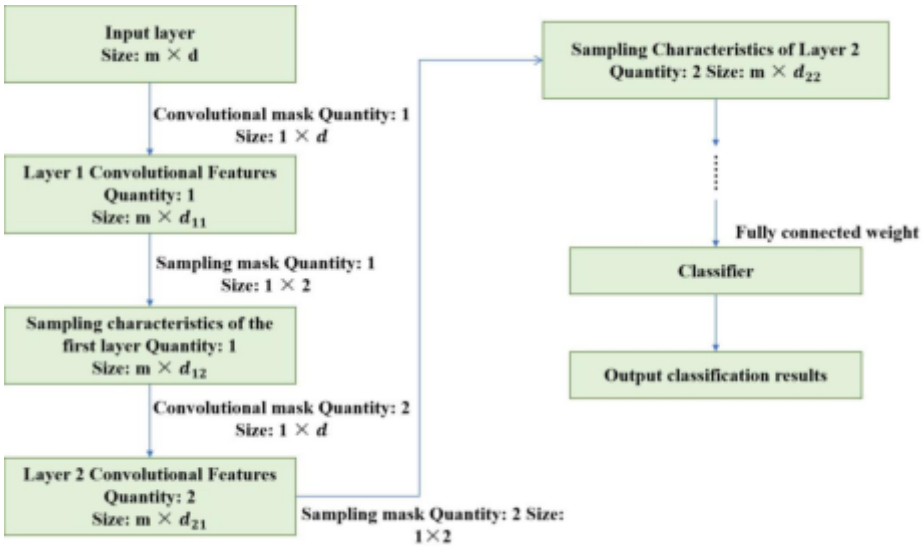


Fig. 3.2: Convolutional Neural Network Layered Framework

In addition, for ease of calculation, each convolutional mask uses the same mask size, denoted as $1 \times d'$. The sampling layer contains the same number of masks as the preceding convolutional layer, with each mask having a size of 1 * 2. After passing through the first convolutional layer, the dimension of the features is

$$d_{11} = d - d' + 1 \tag{3.1}$$

The feature dimension after the first sampling layer is

$$d_{12} = \frac{d - d' + 1}{2} \tag{3.2}$$

Similarly, the feature dimension after passing through the second convolutional layer and sampling layer is

$$d_{22} = \frac{d_{12} - d' + 1}{2} = \frac{1}{4}d - \frac{3}{4}d' + \frac{3}{4} \tag{3.3}$$

If the classifier to be trained only uses two convolutional layers and two sampling layers to extract features, then the dimension d' of each convolutional mask satisfies the condition:

$$\frac{1}{4}d - \frac{3}{4}d' + \frac{3}{4} > 0 \tag{3.4}$$

The final layer of the convolutional neural network classifier is the fully connected layer, which uses a classifier to categorize the features extracted by the convolutional and sampling layers. The author uses a commonly used logistic regression classifier [13,14].

(2) Training of Convolutional Neural Networks. The main problem with using the steepest descent method to train convolutional neural networks is that they are prone to getting stuck in local optima. To address this issue, the author employs a genetic algorithm, with the core concept being to determine the initial optimal weights for the convolutional neural network classifier using the genetic algorithm. Specifically, the weights of the convolutional and fully connected layers in the network are treated as the population for the genetic algorithm, and the weight combinations for each group are binary encoded to form the chromosomes of the genetic algorithm [15].

Assuming the number of convolutional layers in the convolutional neural network classifier to be trained is k, according to the aforementioned convolutional neural network configuration, the number of convolutional masks in the i-th convolutional layer is 2^{i-1}, the total number of convolutional masks in the convolutional neural network is $\sum_{i=1}^{k} 2^{i-1}$, the total number of weights in the convolutional layers is $d' \times \sum_{i=1}^{k} 2^{i-1}$, and the weight of a fully connected layer is added. Therefore, the number of bits of the genetic algorithm chromosome generated by binary encoding is

$$B = d' \times \sum_{i=1}^{k} 2^{i-1} + 1 \tag{3.5}$$

In this way, the first B-1 bit of the chromosome is used to encode the convolutional mask, and the last bit is used to initialize the fully connected layer [16]. Conventional convolutional neural networks typically employ the steepest descent algorithm for training, and their learning performance is significantly influenced by the choice of initial weights in the network. The author uses genetic algorithms to generate different initial weights, so that the optimal weight obtained through selection, crossover, and mutation operations is significantly better than the initial weight obtained through random selection.

3.3. Genetic algorithm solution for initial weights. Genetic algorithm is an adaptive heuristic search method, which originates from the natural law of biological evolution, namely survival of the fittest. In this method, each chromosome in the population undergoes a series of selection, crossover, and mutation changes to obtain a new chromosome with stronger adaptability. The implementation process of this algorithm is shown in Figure 3.3.

The author treats the weights of both the convolutional and fully connected layers in the neural network as the genetic algorithm's population. Initially, several sets of weights are chosen, and each set is encoded into chromosomes [17]. By performing selection, crossover, and mutation operations on these chromosomes, various weight combinations are produced. Then, the fitness value of chromosomes is calculated, which is the classification accuracy of the convolutional neural network under various weight combination methods.

Based on this, the final population, that is, the optimal weight combination, is selected.

The solving process of genetic algorithm is shown in Figure 3.3, and the specific steps will not be repeated by the author. The author focuses on introducing the fusion method of genetic algorithm and convolutional neural network. Correspondingly, the core of the above algorithm is the encoding of chromosomes and the solution of chromosome fitness values. The encoding method of chromosomes has been explained in the previous section. Here, we will specifically introduce the steps used by the author to solve the chromosome fitness values, which are as follows:

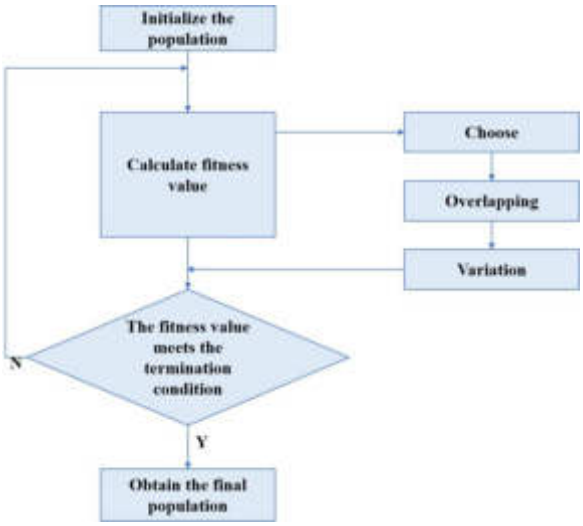


Fig. 3.3: Process of Genetic Algorithm

Step 1: Decode the encoded weight set to retrieve the initial weights;
Step 2: Apply these initial weights to the corresponding convolutional and fully connected layers of the neural network;
Step 3: Train the convolutional neural network classifier using the PL steepest descent algorithm;
Step 4: Assess the classification accuracy of the trained network and use this accuracy as the fitness score for the corresponding chromosome.

Among them, the parameter pl involved in step 3 refers to the number of iterations for training the convolutional neural network. In order to avoid overfitting of the data, the number of iterations should not be set too large [18].

The use of genetic algorithms can label many local valleys, which facilitates the steepest descent algorithm to quickly find local optimal values from these valleys. For a given population size n, after executing multiple rounds of genetic algorithm, the final population can be obtained, which contains n sets of initial weights. These n initial weights can be used to train n convolutional neural network classifiers. Below is an introduction on how to combine these n convolutional neural network classifiers for feature classification.

3.4. Joint classifier. Assuming the output value of a classifier is $o_1, o_2, \cdots, o_c$, where c is the total number of categories. Generally, the output value is binary, which means that for $i = 1, 2, \cdots, c$, there are

$$o_i = \begin{cases} 1, \text{The sample belongs to category i} \\ 0, \text{The sample does not belong to category i} \end{cases} \tag{3.6}$$

Moreover, assuming that each input sample belongs to only one of the c categories, that is

$$\sum_{i=1}^{c} o_i = 1 \tag{3.7}$$

For any test sample, the output category after classification by the classifier is

$$p = arg \underset{1 \leqslant i \leqslant s}{max} o_i \tag{3.8}$$

For the n classifiers generated earlier, the author used the maximum value criterion to fuse the classification results of multiple classifiers and searched for the binary decoding output in the joint classifier model.

Specifically, assuming that the output value of the i-th classifier in the joint classifier model is $o_{1i}, o_{2i}, \cdots, o_d$, the jth output value of the joint classifier model can be represented as

$$f_j = \max_{1 \leqslant i \leqslant n} o_j \tag{3.9}$$

For any test sample x, if its set of output values in the joint classifier model is denoted as $\{f_j | j = 1, 2, \cdots, c\}$, then the category of the test sample should be

$$p = arg \max_{1 \leqslant j \leqslant i} f_j \tag{3.10}$$

In this way, a joint classifier composed of multiple classifiers can achieve higher classification accuracy [19].

3.5. Experimental verification.

3.5.1. Instrument device. The soybean fatty acid content detection device used in the experiment is the Perten8620 near-infrared spectrometer produced by the Swedish company Borton. This instrument has 20 filters corresponding to 20 wavelengths, and is equipped with NIR software program. The instrument used for chemical determination of soybean fatty acid content is the GC9A gas chromatograph produced by Shimadzu, Japan.

3.5.2. Near infrared spectroscopy acquisition. The near-infrared spectroscopy of the sample is collected after the instrument is preheated for 1 hour and the working state is stable. The instrument sample cell is a 5mm quartz flow colorimetric dish. Before spectral collection, it should be cleaned with n-hexane at least three times and air dried. Using air as a reference, perform sample spectral scanning in the 1000-1800nm wavelength range, scan 10 times, and take the average spectrum for modeling.

3.5.3. Spectral preprocessing and calibration methods. A total of 50 samples were taken, and 40 samples were selected as the calibration set and 10 samples as the validation set using the Kemand Stone classification method. In order to reduce random noise and background interference in the original spectrum, mean centering and Savitzky Golay smoothing differentiation method were used for spectral preprocessing. To identify spectral bands with significant informational content for modeling, the spectrum is divided into several subintervals for band selection, and a genetic algorithm is employed to refine the choice of wavelength points. Within the chosen wavelength range, a backpropagation neural network (BP-ANN) is then utilized to develop a calibration model, optimizing parameters such as the number of hidden layer nodes, learning rate, momentum factor, and the number of training iterations. The constructed model is evaluated and optimized through calibration set correlation coefficient (R_c), validation set correlation coefficient (R_p), cross validation root mean square error (SEC), and prediction root mean square error (SEP). Spectral analysis modeling is implemented using Unscrambler 10.4 and Matlab 7.0 software [20].

4. Results and Discussion. The mean centrality method was used to reduce background interference in the near-infrared spectra of the collected soybean fatty acid samples in use, reflecting the information of spectral differences. The author uses the S-G smooth derivative method to filter out spectral noise and eliminate the influence of baseline drift. The window width for smooth differentiation is optimized between 13 and 21 points, with an interval of 2. The modeling results SEC and SEP are used for evaluation, as shown in Figure 4.1. The smooth window width is set to 17 points, and the modeling effect is ideal when taking the first-order derivative of the spectrum.

The genetic algorithm, which selects wavelengths based on optimizing the objective function, offers superior global optimization capabilities compared to the correlation coefficient method. In addition, the determination of the threshold in the correlation coefficient method generally relies on experience and is highly subjective. Using genetic algorithm to screen wavelengths is beneficial for improving the prediction accuracy and stability of the model. Select wavelengths within the measurement range of 1000-1800nm for soybean fatty acid near-infrared spectroscopy. First, divide the entire spectrum into multiple sub intervals to investigate the effectiveness of different interval combinations in modeling, and screen for band combinations that are sensitive to predicting kinematic viscosity. Then, genetic algorithm is used to further optimize the selected bands and enhance the

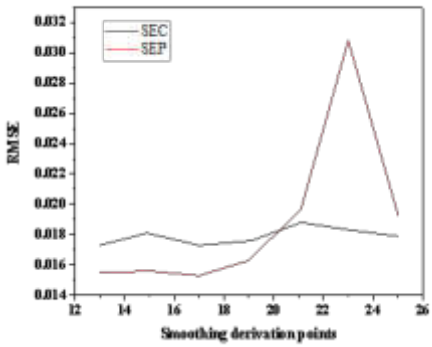


Fig. 4.1: Optimization of Smooth Derivative Points

Table 4.1: Modeling wavelength optimization for multi interval combinations

Sub-region Number	Main Factor	Selected Regions	SEC	SEP
21	5	[1 4 5 17]	0.02756	0.02645
23	7	[1 6 16 18]	0.02347	0.02424
25	6	[1 5 8 18]	0.02157	0.02201
27	7	[1 4 6 20]	0.02341	0.02406
29	7	[1 7 15 21]	0.02545	0.02447

Table 4.2: Modeling effect of wavelength selection using genetic algorithm and correlation coefficient method

Wavelength Selection Method	Wavelength Points	R_c	R_p	SEC	SEP
Genetic Algorithm	87	0.96114	0.97368	0.01712	0.01508
Correlation Coefficient	267	0.93871	0.92617	0.01981	0.0206

predictive ability of the model. The full spectrum is divided into 21-29 sub intervals, with band combinations of 3 and 4. SEC and SEP are used as evaluation criteria to preliminarily optimize the modeled wavelengths. The results are shown in Table 4.1.

According to Table 4.1, when the spectrum is divided into 25 sub intervals, the selected spectral bands are 1000-1031nm, 1128-1159nm, 1224-1255nm, and 1544-1575nm, respectively, resulting in the best modeling effect. Adopting R The improved genetic algorithm proposed by Leardi screens 128 wavelength points in the combined spectral regions of the four bands mentioned above, removing wavelength points that are insensitive to viscosity prediction and further optimizing the modeling wavelength. The genetic algorithm's objective function is defined as the root mean square error from cross-validation, with the algorithm configured to an initial population size of 32, a crossover rate of 0.05, a mutation rate of 0.01, and a maximum of 100 evolutionary generations. Table 4.2 displays the selected wavelength points and model evaluation metrics for wavelength modeling using both the genetic algorithm and the correlation coefficient method.

The author developed a 3-layer neural network model using a combination of principal component analysis (PCA) and backpropagation (BP). They first applied principal component decomposition to the preprocessed spectrum, then selected the principal components that accounted for 99% of the cumulative contribution. The resulting principal component scores were used as input for the BP network modeling. The GA algorithm selects 88 input elements for spectral data, which are reduced to 10 after principal component analysis, reducing the

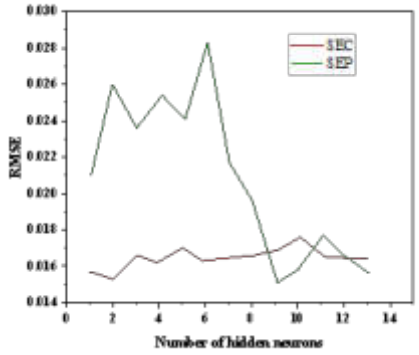


Fig. 4.2: The influence of the number of hidden layer nodes on modeling performance

Table 4.3: Evaluation of Correction Method Analysis Results

Correction Algorithm	R_c	R_p	SEC	SEP
PLS	0.94701	0.91212	0.01701	0.01625
BP-ANN	0.96014	0.97268	0.01612	0.01408

complexity of the training process. The number of hidden layer nodes is optimized based on the root mean square error of the modeling results, ranging from 1 to 13 points, as shown in Figure 4.2.

As observed in the figure above, the SEC and SEP values stabilize at lower levels when the number of hidden layer nodes reaches 9. Consequently, the optimal number of hidden layer nodes is determined to be 9. Additionally, the model uses the tansig function as the transfer function for the hidden layer and the purelin function for the output layer. When the number of training iterations reached 33, the training errors of the modeling set and prediction set reached the lowest value, and the predictive ability of the model tended to stabilize, establishing the optimal prediction model. The comparison between the constructed model and the results of linear partial least squares (PLS) modeling is shown in Table 4.3.

Table 4.3 reveals that the correlation coefficients for both the calibration and validation sets of the BP-ANN model are high and show minimal variation. This indicates that the model has strong prediction accuracy and stability. Compared with PLS method, the BP-ANN method effectively integrates the nonlinear information present in the spectrum, improving the predictive performance of the model. The calibration relative analysis error RPD of the model based on genetic algorithm combined with neural network is 5.01>3, indicating that the model has good prediction accuracy and stability. In order to further verify the reliability of the model in practical applications, within the prediction range of the model, the correlation coefficient R between the predicted value and the standard value is 0.986, indicating a good linear relationship between the two. The relative deviation S_d of each value is less than 2%, and the standard deviation of the difference is 0.0091. Paired t-test t_{solid}=1.2706 was performed on both methods, which is less than $t_{0.05,7}$=2.354. This indicates that there is no significant difference between the analysis results of this method and the standard method, verifying the strong predictive ability of the model in practical applications.

5. Conclusion. The author introduces a method for detecting soybean fatty acid content using near-infrared spectroscopy, enhanced by a neural network and genetic algorithm. This approach employs a genetic algorithm-based wavelength selection technique to identify optimal wavelengths for modeling, thereby maximizing the use of spectral information while reducing the number of variables in the model. By integrating the BP neural network, the method effectively addresses the nonlinear relationship between spectral data and fatty acid content, significantly boosting the accuracy and reliability of predictions. The analysis of actual samples

using optimization models yielded ideal prediction results, further verifying the reliability of the method in practical applications.

Acknowledgements. This study is supported by 2021's Cangzhou science and technology planning and self-financing project Research on Cangzhou Nutritional Food Informalization Evaluation under the ""Internet +"" Background. Cangzhou science and technology planning and self-financing project research result (213206018).

REFERENCES

[1] Anai, T. (2023). Improvement of soybean seed oil quality using fatty acid mutants. Nippon Shokuhin Kagaku Kogaku Kaishi, 70(1), 47-51.
[2] Bukowski, M. R., & Goslee, S. (2024). Climate-based variability in the essential fatty acid composition of soybean oil. The American journal of clinical nutrition., 119(1), 58-68.
[3] Song, H., Taylor, D. C., & Zhang, M. (2023). Bioengineering of soybean oil and its impact on agronomic traits. International Journal of Molecular Sciences, 24(3), 2256.
[4] Sim, J., Kuwabara, C., & Yamada, A. T. (2023). Recent advances in the improvement of soybean seed traits by genome editing. Plant Biotechnology, 40(3), 193-200.
[5] Johnson, J. B., Thani, P. R., & Naiker, M. (2023). Through?container detection of tea tree oil adulteration using near?infrared spectroscopy (nirs). Chemical papers, 77(4), 2009-2017.
[6] Palou, A., Patricia Jiménez, Casals, J., & Ingrid Masaló. (2023). Evaluation of the near infrared spectroscopy (nirs) to predict chemical composition inulva ohnoi. Journal of Applied Phycology, 35(5), 2007-2015.
[7] Deng, Z., Fu, J., Yang, M., Zhang, W., Y.-H., Y., & Zhang, L. (2024). Geographical origin identification of hainan camellia oil based on fatty acid composition and near infrared spectroscopy combined with chemometrics. Journal of Food Composition and Analysis, 125(1), 342-351.
[8] Zhe, L. I., Haigang, Q. I., Ying, Y. U., Liu, C., Cong, R., & Li, L. I., et al. (2023). Near-infrared spectroscopy method for rapid proximate quantitative analysis of nutrient composition in pacific oyster crassostrea gigas. Journal of Oceanology and Limnology(1), 150(3), 1-19.
[9] Tonolini, M., Wawrzynczyk, J., & Nielsen, M. E. S. B. (2023). On-line monitoring of enzymatic degumming of soybean oil using near-infrared spectroscopy. Applied Spectroscopy: Society for Applied Spectroscopy, 77(12), 1333-1343.
[10] Tonolini, M., Wawrzynczyk, J., & Nielsen, M. E. S. B. (2023). On-line monitoring of enzymatic degumming of soybean oil using near-infrared spectroscopy. Applied Spectroscopy: Society for Applied Spectroscopy, 77(12), 1333-1343.
[11] Liang, Y., Lin, H., & Kang, H. C. Q. (2023). Application of colorimetric sensor array coupled with machine-learning approaches for the discrimination of grains based on freshness. Journal of the Science of Food and Agriculture, 103(14), 6790-6799.
[12] Zhang, Q., Pizzi, A., Lei, H., Xi, X., Cao, M., & Cao, L. (2023). Maldi tof investigation of the reaction of soy protein isolate with glutaraldehyde for wood adhesives, 11(3), 12.
[13] Dai, C., Xu, X., & Huang, R. M. H. (2023). Monitoring of critical parameters in thermophilic solid-state fermentation process of soybean meal using nir spectroscopy and chemometrics. Journal of Food Measurement and Characterization, 17(1), 576-585.
[14] Bartalne-Berceli, M., Izso, E., & Gergely, S. S. A. (2022). Monitoring of soybean germination process by near-infrared spectroscopy. Acta Alimentaria: An International Journal of Food Science, 51(2), 204-217.
[15] Da Costa, G. B., De Fernandes, D. D. S., & Veras, GermanoDiniz, Paulo Henrique Goncalves DiasGondim, Amanda Duarte. (2024). Combining nir spectroscopy with dd-simca for authentication and ispa-pls-da for discrimination of ethyl route and oil feedstocks of biodiesels in biodiesel/diesel blends. Journal of the American Oil Chemists' Society, 101(2), 187-196.
[16] Kuang, J., Luo, N., Hao, Z., Xu, J., He, X., & Shi, J. (2022). Ni-raman spectroscopy combined with bp-adaboost neural network for adulteration detection of soybean oil in camellia oil. Journal of Food Measurement and Characterization, 16(4), 3208-3215.
[17] Liu, W., Sun, S., Xia, Y., Zhao, P., Liu, C., & Zheng, L. (2022). Rapid determination of benzo(a)pyrene concentration in soybean oil by terahertz transmission spectroscopy with chemometrics. Journal of Infrared, Millimeter, and Terahertz Waves, 43(7), 695-708.
[18] Santos, C. S., Clara, S., & Bagheri, M. S. M. W. (2024). Predicting iron deficiency and oxidative stress in glycine max through fourier transform infrared spectroscopy in a time-course experiment. Plant and Soil, 496(1/2), 161-177.
[19] Wang, J., He, J., Yang, J., & Gai, J. (2023). A novel procedure for identifying a hybrid qtl-allele system for hybrid-vigor improvement, with a case study in soybean (glycine max) yield. The Crop Journal, 11(1), 177-188.
[20] Petravicius, D., Binelo, M. O., & Binelo, Marcia de F. B.Faoro, Vanessada Silva, Jose A. G. (2023). Genetic algorithm in the design of soybean silos for airflow homogenization 1. revista brasileira de engenharia agricola e ambiental, 27(7), 531-538.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Aug 25, 2024
Accepted: Oct 5, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2045–2052, DOI 10.12694/scpe.v26i5.4716

THE CONSTRUCTION OF SMART CITIES AND BIG DATA GOVERNANCE STRATEGIES BASED ON ARTIFICIAL INTELLIGENCE

YINPING LI*

Abstract. To tackle the challenge of sluggish resource scheduling in smart city management, the author introduces a study focused on smart city development leveraging artificial intelligence and big data governance strategies. The research uses the Cloud Management Module to monitor the various hardware equipment, and establishes the target functionality of the Cloud Computing Resource Scheduling. Through the application of PSO to this goal function, a optimal scheme is created for the Cloud Resource Scheduling. Experimental results demonstrate that the system achieves optimal resource utilization, nearing 100%, whereas the other two systems have lower utilization rates, both under 90%. This indicates that, compared to similar systems, this approach offers superior overall utilization. The system is capable of efficiently managing smart cities and providing real-time monitoring of urban conditions. When scheduling resources, it achieves shorter task completion times and higher system efficiency, ensuring maximum resource utilization.

Key words: Cloud computing, smart city, Management system, Cultural particle swarm, Resource scheduling

1. Introduction. In recent years, the swift advancement of the Internet of Things (IoT), artificial intelligence, and large-scale model technologies has led to transformative shifts in the development of smart cities. As a fundamental component of smart city infrastructure, the IoT enables seamless connections and integration among people, machines, and objects through the deployment of numerous sensors that gather real-time data on urban operations. However, the increasing number of IoT devices has resulted in an exponential rise in the volume of data being generated [1,2]. How to effectively analyze and utilize this data has become a major challenge in the construction of smart cities. Traditional data processing and analysis methods are no longer able to cope with such large-scale, high-speed, and diverse data. Therefore, utilizing artificial intelligence technology for deep mining and value extraction of IoT data has become an effective way to solve this problem [3]. A smart city represents the future of urban development, combining and surpassing the elements of a digital city, knowledge-based city, ecological city, and creative city into a comprehensive urban system. It is built on an Internet of Things (IoT) infrastructure, where intelligent sensors embedded in urban objects are interconnected through the internet to provide real-time awareness of the physical city. By employing computer technology to process and analyze the gathered information, smart cities integrate the online "digital city" with IoT systems. This enables intelligent responses and decision-making to meet various needs such as governance, public welfare, environmental management, public safety, urban services, and commercial activities [4].

The smart city IoT perception system is an important construction content under the digital national strategy, and it is also a key infrastructure for achieving fine and precise management of smart cities. Through the use of state-of-the-art technologies like the Internet of Things, Cloud Computing, Big Data Analytics, and Artificial Intelligence, cities can achieve real time monitoring and insight into many fields. This provides urban managers with precise and comprehensive data, supporting more informed and rational decision-making. As a result, these technologies drive the modernization and intelligent evolution of urban environments. However, in the process of building smart cities, how to effectively govern big data has become an important challenge. The diversity, complexity, privacy, and security issues of data make big data governance particularly important. Making data safe, reliable and transparent is one of the most urgent problems in developing intelligent cities. Therefore, the establishment of a strong Large Data Management Framework is crucial to support Smart Cities' long term development and sustainable development. The author aims to explore the overall framework of building smart cities based on artificial intelligence, and focus on analyzing the strategies and methods of

*Anhui Wenda University of Information Engineering, Hefei, 231201, China (liyinping6541@163.com)

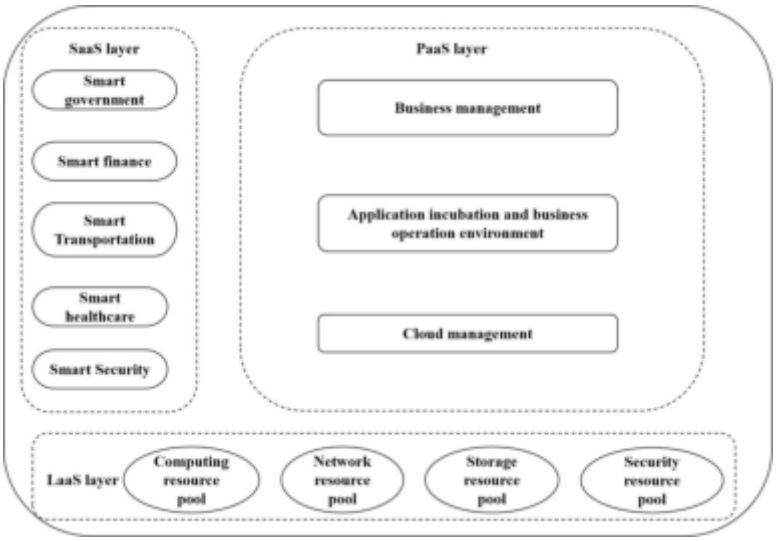


Fig. 3.1: Overall System Structure

big data governance, providing theoretical and practical support for the construction of smart cities [5,6].

2. Literature Review. Smart City was originally conceived in the 1990s, and it was used to describe the process of technological, innovative, and globalized urban development[7]. Liu, X. et al. examined the current state of smart cities, exploring specific needs and broader developmental trends. Their work has been thoroughly analyzed, and they have created a Intelligent Urban Management System with Java Programming Language and Browser/Server (B/S) architecture [8]. Li et al. investigated issues such as insufficient planning in the development of national smart city information service systems. To address these challenges, they proposed building a smart city information service system aligned with the goals of big data. They employed data clustering algorithms to effectively integrate and manage smart city information within a big data environment [9]. Liu, Y. et al. studied the frameworks for smart city development and services, focusing on the applications of cloud computing and the Internet of Things. Their research included extensive testing of data mining algorithms, demonstrating their impact on enhancing computational efficiency and optimization within smart city public management systems [10].

Based on the above issues, the author used cloud computing platforms as the foundation and designed an intelligent city management system to meet the requirements of modern urban development.

3. Method.

3.1. Overall structure of smart city management system based on cloud computing platform. Taking into account the system's overarching objectives and management principles, an evaluation of the smart city system's operational functions was performed. Following this analysis, a detailed framework for managing smart cities through a cloud computing platform was designed. The outcomes of this framework are depicted in Figure 3.1.

The basic architecture of the system consists of an IaaS (Infrastructure as a Service) resource level, with the primary focus on the promotion of the Platform as a Service Platform. This level deals with the administration of the system and is characterized by developing the Cloud Management and Business Management Modules, ensuring the smooth operation of both the system and its associated business processes.

The IaaS layer, serving as a unified resource foundation, is primarily tasked with supporting the PaaS platform layer. The system consists of the Net Resource Pool, the Memory Resource Pool, the Computation Resource Pool, and the Safety Resource Pool. This layer delivers essential hardware resources to both the SaaS and PaaS layers. Additionally, it manages and maintains the health of various resource pools and physical

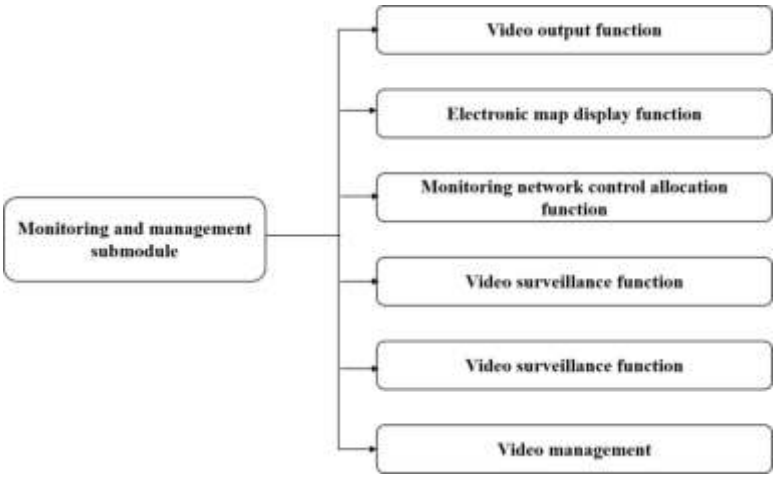


Fig. 3.2: Structure of Monitoring Management submodule

devices, overseeing the allocation and regulation of resources.

The PaaS layer delivers standardized, shared cloud services to various applications. It encompasses the business operation environment and application development infrastructure, and is divided into three main functional components: the cloud service engine, middleware, and data management components. The cloud service engine handles tasks such as service scheduling, resource management, monitoring, routing, and authentication. Middleware manages dynamic resource sharing and provides unified management capabilities. The data management component oversees dynamic resource allocation and centralized database administration. In this layer, the cloud management module oversees the overall cloud operations for the smart city management system, including resource management and operation of various sub-modules. Meanwhile, the business management module focuses on the cloud platform's business operations, ensuring the smooth functioning of service management and other related sub-modules.

The SaaS (Software as a Service) layer is designed to support a range of applications across seven key domains within smart cities. This layer primarily consists of national-level management applications as well as those tailored for provincial and municipal levels, with applications categorized according to industry and geographical region. The way to showcase applications is through cloud platforms.

3.2. Design of Monitoring and Management Module. To efficiently monitor and assess the city's condition, the smart city management system depends significantly on the cloud platform's monitoring and management submodule. The configuration of this submodule is shown in Figure 3.2. It includes six essential functions and utilizes various equipment such as audio and video capture devices, transmission media, control systems, and terminal monitoring units. These components work together within the camera system to fulfill the six primary functions of the monitoring control submodule. By strategically deploying surveillance cameras throughout the city, real-time monitoring capabilities are established [11].

3.3. Cloud computing resource scheduling strategy.

3.3.1. Scheduling principle. In the cloud, there is no direct one-to-one relation between computational resources and tasks. Rather, a task is initially allocated to a resource, which is then accessible by the respective physical device. The Map/Reduce programming model, introduced by Google, is widely utilized in cloud computing platforms. This model can be described using a five-tuple to represent the resource scheduling framework of cloud computing:

$$S = \{T, V, D, M_{TV}, M_{VD}\} \tag{3.1}$$

In the formula, $V = \{v_1, v_2, \cdots, v_m\}$,$D = \{d_1, d_2, \cdots, d_m\}$ and $T = \{t_1, t_2, \cdots, t_m\}$ represents the resource

set, physical device set, and task set [12]. Equation 3.2 reflects the total duration needed for device dk to complete the execution of the assigned task:

$$\begin{cases} Sum(d_k) = \sum_{i=1}^{m} c_{ik} Finish(t_i M_{TV}, d_k) \\ c_{ik} = \begin{cases} 1 & t_i M_{TV} = d_k \\ 0 & t_i M_{TV} \neq d_k \end{cases} \end{cases} \tag{3.2}$$

In the formula, represents that physical device dk is ultimately mapped and executed by task ti. The total execution time of all tasks $T = \{t_1, t_2, \cdots, t_m\}$ is represented by equation 3.3:

$$total(T) = \sum_{k=1}^{n} Sum(d_k) \tag{3.3}$$

The main goal of Cloud Computing Resource Scheduling is to minimize the value expressed in Formula 3.3. Thus, formula 3.4 is used as an objective function for optimization of resource scheduling in cloud computing.

$$Goal(T) = min \sum_{k=1}^{n} Sum(d_k) \tag{3.4}$$

3.3.2. Cloud computing resource scheduling strategy based on cultural particle swarm algorithm. Traditional particle swarm optimization algorithms utilize rule-based binary encoding, which is not well-suited for cloud computing resource scheduling. The main aim in cloud computing resource scheduling is to minimize task completion times. To achieve this, it is crucial to ensure high particle quality, which is reflected by a high fitness value, in order to derive an optimal scheduling solution. Equation 3.5 specifies the fitness function for evaluating particles [13,14].

$$fit = \frac{1}{min \sum_{k=1}^{n} sum(d_k)} \tag{3.5}$$

The common way for particle swarm optimization algorithms to generate initial particle swarm is random, which can easily lead to particle concentration in a certain local area, resulting in non-uniform feasible solutions. In this study, the initial particle swarm was generated using a uniform approach, ensuring that the particles were evenly distributed at the outset.

Figure 3.3 illustrates the flow of Cloud Computing Resource Scheduling with Cultural PSO. First, we initialise the knowledge space and the initial particle swarm. Then, the iterative count is set, and the knowledge space is updated as the fitness of the particles is evaluated. Once the knowledge space update is complete, the group undergoes self-evolution to verify if the evolution meets the termination criteria. If not, the evolution process is repeated. If the criteria are met, the influence operation is applied to the particle swarm space. The fitness of the particles is recalculated, and both individual and global best values are updated accordingly. Simultaneously, the location and speed of the particle are adjusted to check whether the end condition is satisfied [15]. If this condition is not satisfied, the procedure will proceed for a specified number of iterations. When these conditions are satisfied, the best particle location can be found, and then the ultimate optimum solution can be obtained.

3.4. Construction strategy of smart city management platform based on urban information model. After extensive research and development, the framework of smart cities has evolved into a four-tiered system, organized from the ground up as follows: the Perception layer, the Transmission layer, the Data Processing layer, and the Application layer (see Table 3.1).

Smart city management focuses on serving the public by enhancing and streamlining social resources to deliver high-quality and efficient urban services. The development of a new smart urban management platform should begin with the creation of a unified service system, incorporating medical, educational, dietary, transportation, consumer, and administrative services [16]. Aiming to establish a high-standard service platform, this approach relies on the urban information model as a crucial step in building the smart management

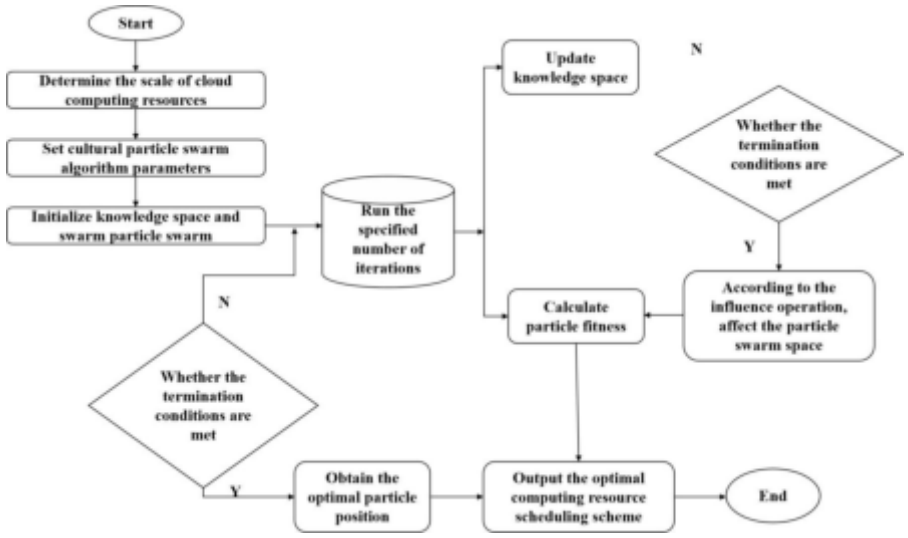
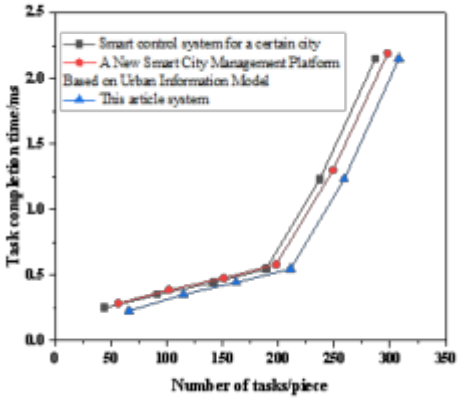


Fig. 3.3: Cloud computing resource scheduling process

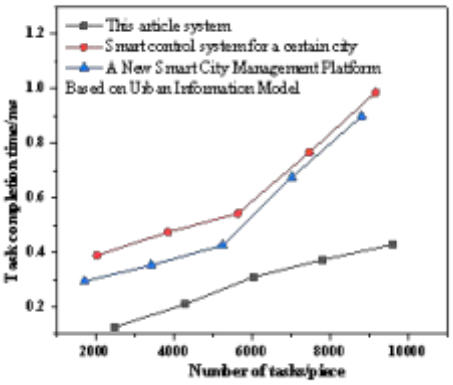
Table 3.1: Structural System of Smart Cities

Level	Function	Related research directions
Perception layer	Responsible for collecting various data parameters in the city, such as humidity, temperature, pressure, light, etc	Various sensing devices, such as Zigbee, Bluetooth and Radio Frequency Identification (RFID) sensors, cameras, and Global Positioning System (GPS) terminals
Transport layer	Collect and summarize data obtained from sensors through various communication networks	Network transmission, such as 3G, 4G Long Term Evolution (LTE), 5G, and Low Power Wide Area Network (LP-WAN)
Data Processing Layer	Perform functions such as data processing, organization, analysis, storage, and decision analysis	Data integration, such as the fusion of BIM and GIS data, the fusion of sensing equipment and environmental data, etc
Application layer	The top-level of ”smart city” is the functional design layer guided by practical applications	Smart economy, smart citizens, smart government, smart travel, smart life, smart environment, etc

platform. This facilitates the integration of advanced information technology with urban services and infrastructure, promoting coordinated urban development. In urban development and intelligent management, it is essential to establish clear objectives that emphasize enhancing convenience and fostering a high-end smart economy to achieve effective urban management. Utilizing CIM technology as a foundation and focusing on network security, a robust framework for smart city development should be built, leading to the creation of a smart city characterized by its omnipresence, intelligence, and integration. For instance, in developing a smart city management platform, resident ID cards can be utilized to consolidate various functions such as urban public administration, social security, micro-payments, utility services, and identity verification. By leveraging these multifunctional cards, the level of urban intelligence and service efficiency can be significantly enhanced. Additionally, residents can install mobile apps on their smartphones, allowing them to use their ID cards for streamlined services including medical insurance, public transportation, and digital payments. An



(a) When the number of tasks is small



(b) When there are a large number of tasks

Fig. 4.1: Completion time of resource scheduling for different task scenarios

intelligent management platform is crucial for urban development. It drives urban construction by prioritizing intelligence, overcoming previous limitations, and ultimately achieving the goal of enhancing and simplifying life for residents [17].

Building a connected and multi-dimensional urban smart management platform, centered around urban development, aims to enhance and refine city services and administration. The goal of smart urban management should be to advance smart city initiatives by gathering and optimizing extensive data, establishing large-scale computing servers, and supplying robust data support for networked systems. This platform should facilitate coordinated development, allowing residents to access improved online services and thereby significantly boost operational efficiency.

4. Results and Discussion. To evaluate the efficiency of resource scheduling, two different task volumes were assigned to each cloud computing node: fewer than 300 tasks and over 10,000 tasks. This paper compares the Intelligent Control System of an urban area with that of a newly developed Intelligent Urban Management Platform. The comparison is illustrated in Figure 4.1. As demonstrated in Figure 4.1 (a), when the task count is low, there is no noticeable effect on the resource scheduling completion time, with all systems displaying nearly identical task completion times and minimal differences. However, as shown in Figure 4.1 (b), with a larger volume of tasks, the time required for resource scheduling increases across all systems. Despite this increase, the new smart city management platform consistently demonstrates the shortest scheduling time, indicating a clear advantage in resource scheduling efficiency compared to the other systems [18].

Also, the amount of Cloud Computing Resource Nodes affects the efficiency of Resource Scheduling. In this evaluation, 5000 tasks were distributed across 60 cloud computing nodes, and a comparison was made between the three systems. The results of this comparison are presented in Figure 4.2.

In the smart city management system, the resource utilization rate serves as a key metric for assessing resource scheduling efficiency within the cloud computing platform, reflecting the system's resource availability. The primary objective of the cloud computing platform is to optimize resource utilization and ensure efficient resource sharing. The comparison of overall utilization rates among the three systems is illustrated in Figure 4.3. According to Figure 4.3, this system achieves the highest overall utilization rate, nearing 100%, whereas the other two systems fall below 90%. This indicates that, in comparison to similar systems, this system demonstrates superior resource utilization [19,20].

5. Conclusion. The author suggests a study focused on developing smart cities and big data governance strategies using artificial intelligence. The study involves designing a comprehensive system architecture and

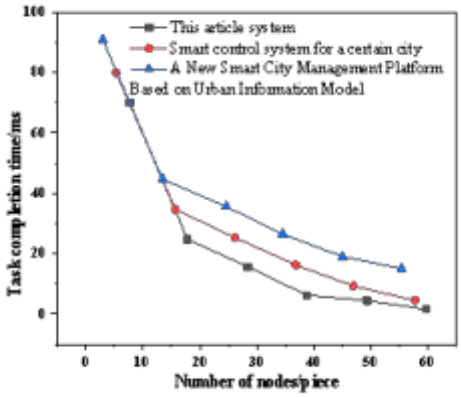


Fig. 4.2: Number of Nodes and Task Completion Time

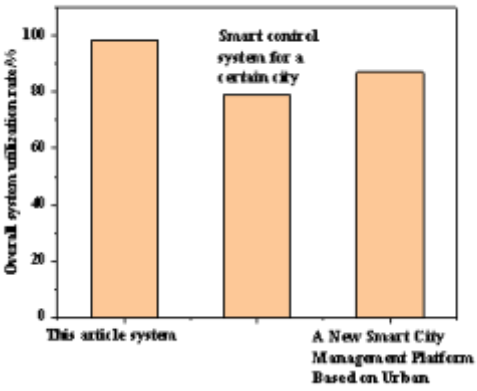


Fig. 4.3: Overall utilization rate of the system

implementing cloud computing resource scheduling through a cultural particle swarm algorithm. This approach enables efficient smart city management and real-time urban monitoring, with notable effectiveness in resource scheduling. Even with a high volume of scheduling tasks, the system maintains rapid performance and demonstrates superior resource scheduling outcomes compared to similar systems.

Acknowledgement. Research on smart city construction and Big Data governance strategy based on artificial intelligence (2024AH050596).

REFERENCES

[1] Lu, J. (2022). Building a smart city planning system integrating multidimensional spatiotemporal features. Scientific programming, 2022(Pt.16), 1-9.

[2] Wilson, D., & Wyly, E. (2023). Toward a dracula urbanism: smart city building in flint and jakarta. Dialogues in Urban Research, 1(2), 135-154.

[3] Majdi, A., Dwijendra, N. K. A., & Muda I.Chetthamrongchai P.Sivaraman R.Hammid A.T. (2022). A smart building with integrated energy management: steps toward the creation of a smart city. Sustainable Energy Technologies and Assessments, 53(Oct. Pt.C), 1-8.

[4] Arief, A., Sensuse, D. I., & Mursanto, P. (2022). Smart city's trends, architectures, components, and challenges: a systematic review and building an initial model for indonesia. engineering and applied science research, 49(2), 259-271.
[5] Liu, C., & Ke, L. (2023). Cloud assisted internet of things intelligent transportation system and the traffic control system in the smart city. Journal of Control and Decision, 10(2), 174-187.
[6] Zouhri, A., Ezzahout, A., Chakouk, S., & Mallahi, M. E. (2024). A numerical analysis based internet of things (iot) and big data analytics to minimize energy consumption in smart buildings. Journal of Automation, Mobile Robotics and Intelligent Systems, 18(2), 46-56.
[7] Zeng, Q., Lv, Z., & Li, C. S. G. (2023). Fedprols: federated learning for iot perception data prediction. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 53(3), 3563-3575.
[8] Liu, X., Liu, G., & Wan, H. D. (2023). Innovation of smart city management system based on computer application technology. Springer, Singapore, 14(7 Pt.1), 8811-8826.
[9] Li, C. (2023). The construction of smart city information service system in the era of big data. International Journal of Data Science, 35(Sep.), 1-9.
[10] Liu, Y., & Sun, G. (2022). Analysis of government public management information service and computer model construction based on smart city construction. Mathematical Problems in Engineering: Theory, Methods and Applications(Pt.8), 2022.
[11] Dimitrov, W., Spasov, K., & Syarova, T. S. (2022). Complexity assessment of research space for smart city cybersecurity*. ifac papersonline, 55(11), 1-6.
[12] Liao, R., & Chen, L. (2022). An evolutionary note on smart city development in china, 23(6), 9.
[13] Gelmez, E., & Zceylan, E. (2023). Evaluation of the smart cities listed in smart city index 2021 by using entropy based copras and aras methodology. Foundations of Computing and Decision Sciences, 48(2), 153-180.
[14] Ghosh, S., & Sk., J. (2024). Impact of changing urban form and production-living-ecological space on changing ecosystem services of a smart city of eastern india, durgapur. Environmental Monitoring and Assessment, 196(9), 1-21.
[15] Senthamilarasi, N., Kumar, V. D. A., & Robin, C. R. R. (2023). Deep learning framework for attack and malicious node detection for smart city protection applications. Journal of environmental protection and ecology, 24(7), 2361-2372.
[16] Lipianina-Honcharenko, K., Komar, M., Melnyk, N., & Komarnytsky, R. (2024). Sustainable information system for enhancing virtual company resilience through machine learning in smart city socio-economic scenarios. ECONOMICS, 12(2), 69-96.
[17] Yu, X. (2022). Smart city economic management prediction model based on information analysis system. Mobile information systems, 2022(Pt.14), 1-10.
[18] Li, Y., Jin, H., & Wang, H. X. (2023). Research on intelligent city traffic management system based on webgis. International Journal of Nanotechnology, 20(1/4), 410-420.
[19] Silva, F. A., Iure Fé, Silva, F., & Nguyen, T. A. (2024). Quantifying the impact of resource redundancy on smart city system dependability: a model-driven approach. Cluster Computing, 27(5), 6059-6079.
[20] Li, W., Stidsen, C., & Adam, T. (2023). A blockchain-assisted security management framework for collaborative intrusion detection in smart cities. Computers and Electrical Engineering, 111(Pt.A), 1-13.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Aug 26, 2024
Accepted: Oct 11, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2053–2060, DOI 10.12694/scpe.v26i5.4857

HUMAN BEHAVIOR RECOGNITION IN COMPLEX SCENES BASED ON DEEP LEARNING

RUIFENG HONG*

Abstract. In order to overcome some of the problems that have been encountered in the past, for example, due to their reliance on manual feature extraction and limitation of model generalization, this paper proposes a new method to identify people's behaviour in complicated situations. Based on Convolutional Neural Networks (CNN), this method has been proposed for the automatic extraction of a large number of datasets. In addition, the Long Short Term Memory (LSTM) network is used to capture the long-run dependence in time order. Lastly, we use the soft max classifier to classify the various actions of people. Experiments show that the CLT network is able to achieve a high performance of 97.5% over 13 different types of people, outperforming CNN alone, LSTM, and BP models on the DaLiAc dataset, demonstrating superior performance in human behavior recognition and classification. The accuracy, recall, and F1 score evaluation indicators of the CLT net model are the highest, while all indicators of the BP model are the lowest, indicating that the CLT net model has good stability and reliability in recognizing and classifying different human behaviors.

Key words: Human behavior recognition, Deep learning, Convolutional neural network, Long Short Term Memory Network

1. Introduction. With the continuous breakthroughs in computer vision research and computer hardware performance, the idea of machines possessing partial visual abilities of the human eye has become possible. As a core topic in the field of computer vision, human behavior recognition has been active at the forefront of research, playing a key role in intelligent security, medical assistance, smart education, human-computer interaction, and gradually changing people's lives [1]. The traditional video surveillance methods mainly rely on manual monitoring, which inevitably results in a significant waste of manpower and material resources. Due to the limitations of the human body, the monitor cannot maintain a high level of attention at all times during work. When the monitor becomes tired or briefly leaves the monitoring screen, it is difficult to handle emergencies in a timely manner, ultimately leading to missed or false detections in the monitoring process [2]. Secondly, the video streams generated by traditional monitoring methods only include simple storage and playback functions, and there is still a possibility of false or missed detections when manually reviewing historical events [3].

With the fast developing of smart hardware and computing techniques, it is becoming more and more important to substitute for conventional human surveillance with artificial intelligence, complete the recognition and analysis of human behavior and actions in the monitoring area, and combine the advantages of computer vision technology such as real-time, efficiency, and accuracy with video surveillance to build intelligent video surveillance has become an inevitable development trend [4]. Deep learning based human behavior recognition in complex scenarios mainly includes two aspects: motion object detection and human behavior recognition. By collecting video data of human behavior actions through cameras, using object detection methods to identify and detect moving targets, extracting behavioral action features of moving targets and completing classification, the current human behavior state can be determined, providing timely and effective information for staff [5-6]. However, despite the enormous potential of deep learning in behavior recognition, it still faces many challenges. In complicated situations, for instance, the influence of environmental noise and disturbance on the recognition capability of the model will be influenced. Furthermore, the behavior patterns of each person are not uniform enough, so that it is difficult to generalize the model. This paper is intended to investigate the approach of human behaviour identification in complicated situations, especially for complicated background, illumination, multi-object occlusions, etc.

*Guangzhou Civil Aviation College, Guangdong, China (hongrf@qq.com)

2. Literature Review. Deep learning based human behavior recognition is a process in which machines obtain human limb movement information from videos and complete action category judgment. It is a research hotspot in the field of computer vision and a key technology in video understanding. It has huge application prospects in smart homes, surveillance security, smart healthcare, education, and other fields [7]. Although many algorithms have emerged in this field, there are still problems such as difficulty in determining the starting time of actions, poor recognition performance in cases of occlusion or low resolution, and insufficient real-time recognition. Wang et al. developed a meta-learning framework named MetaTTE, which uses a well-designed DED (a data-pre-processing module and a coder decoding net module) to continually deliver an accurate journey time estimation [8]. Fuentes et al. proposed an approach to tracking the behaviour of an individual cow by using the image data as an input to identify the movement. In this paper, we use an image sequence as an input, which is used to recognize the hierarchy of activities that are divided into parts and single actions. These areas of concern are then entered into the tracing and identifying mechanism so that the system can continually follow every individual on the spot and give them a unique identifier. Through this approach, the behaviour of cows can be continually monitored and statistically analyzed to assess behavioral changes in time [9]. Jiao et al. presented a new approach to detect the variation of facial features in movement, and made a comparison between them and the former one. Experiments indicate that the algorithm is more accurate by 1. 68%, and the precision of the modified moving historical image is raised by 14.8%. The proposed approach has been successfully applied to the identification of human motion [10].

In order to improve the precision of human behaviour identification, we propose a new approach, called CLT network, which is based on temporal and spatial characteristics fusion. Using Convolutional Neural Networks (CNN), this approach utilizes Long Short Term Memory (LSTM) net to capture the intrinsic temporal relations of the data. The classification of human behaviors is then done using a softmax classifier within this framework.

3. Method.

3.1. CNN Model. Convolutional neural networks have achieved remarkable success in many fields such as object detection, facial recognition, and speech recognition, there is still no widely accepted architecture for their use in sequence signal classification [11]. To address this gap, the author developed a CNN model tailored for human behavior recognition, inspired by the LeNet-5 architecture. Unlike LeNet-5, this CNN model is designed to handle sequential data as input. After each max pooling layer, additional batch normalization and activation layers (using Leaky ReLU) are incorporated [12]. The CNN structure includes a number of critical elements as shown in Figure 3.1: Sequential Entry Level, Collapse Level, 3 CNN Characteristic Extracting Levels (Convolution, Max Pooling, Batch Normalization, and Leaky ReLU Activation Layer), Unfolding, Leveling, Full Connection, and Soft Max Classifying Layer [13]. The feature extraction layers are crucial for the model, with the convolutional layers extracting essential features from human behavior data [14]. The max pooling layers serve to compress the data and reduce dimensionality, while the batch normalization layers ensure that the extracted features are normalized. The Leaky ReLU activation layers introduce non-linearity, aiding in the effective mapping of features post-normalization. Lastly, the full connection level reduces the loss of the data in the process of extracting the characteristic, and makes the ultimate classification of the person's actions by soft max.

The proposed method can not only increase the rate of convergence, but also reduce the "gradient dispersion" and improve the stability of the training model [15]. The Leaky Relu Trigger Function is used to solve the problem that Relu has a negative input and a constant output of zero, whereas the first derivative is also zero, which leads to no change in neural parameters and no learning. The definition is shown in equation 3.1:

$$f(x) = \begin{cases} x, x \geqslant 0 \\ s \times x, x < 0 \end{cases} \tag{3.1}$$

Among them, s is a non-negative number not smaller than 1. The Leaky Relu activation function becomes Relu when s is set to zero. The softmax classification layer is shown in equation 3.2:

$$s(x_i) = \frac{e^{x_j}}{\sum_{j=1}^{K} e^{x_j}}, i = 1, 2, \cdots, K \tag{3.2}$$

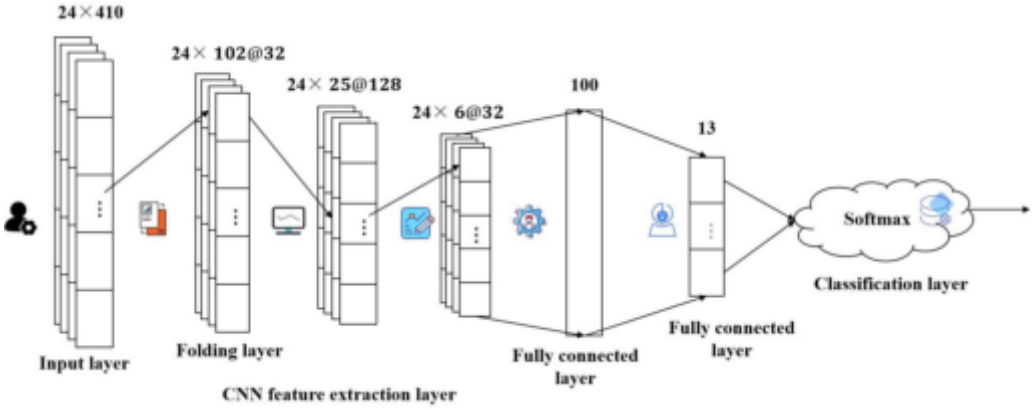


Fig. 3.1: Structure of CNN Model

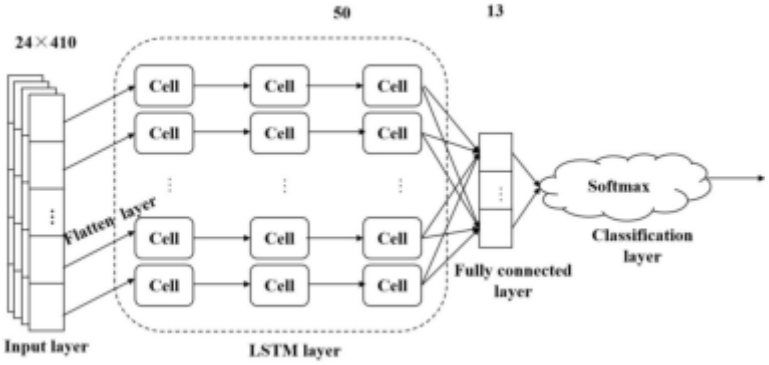


Fig. 3.2: Structure of LSTM Model

Among them, Xi is the character order of the extracted behavioral data, and K is the quantity of human behaviour. The Soft Max Function's classification result indicates the probability that an input sample will be categorized into different classes, and the total probability is 1 [16].

3.2. LSTM Model. LSTM was developed by SCHMIDHUBER and HOCHREITER in 1997 to improve Recurrent Neural Network (RNN). The key parts of LSTM networks include sequential entry and LSTM. The sequential entry level is capable of inputting sequential or temporal information into the net, and LSTM level is able to study the long term dependence among sequential data time steps, which efficiently resolves the RNN gradient disappearing. Because LSTM is an effective way to deal with and predict time signals, the LSTM is used as a characteristic filter in CLT network model [17]. The LSTM model is illustrated in Figure 3.2, which consists of sequential input, flat, LSTM, full connection, and soft max.

From Figure 3.2, we can see that the sequence entry level has a sample size of 24x410 x 1. Then, the multi-dimension data is smoothed out as an input to LSTM level. The LSTM level has 50 hidden cells, and the full connection level has 13 hidden cells. At last, we apply soft max classifier to classify the various actions of people. The LSTM layer's units provide temporal dependence and temporal properties of the input data [18]. The LSTM network achieves long-term control of a unit, which is then used for classification and prediction of temporal signals. Cellular function is primarily implemented by logical gates, input gates, and output gates. Figure 3.3 illustrates the inner architecture of the LSTM layer cells.

The LSTM layer is able to learn the weights such as the weight of the input, the value of the recurrence and the value of the error. Matrices W, R, and B represent a sequence of input weights, recurrence weights,

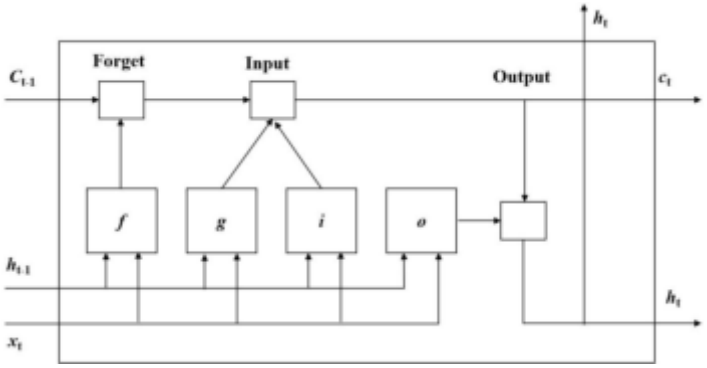


Fig. 3.3: Internal structure of LSTM cells

and deviations, as illustrated in formula 3.3:

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \tag{3.3}$$

The output of the cell status and the output of the hidden state at the time t is given by the expressions 3.4 and 3.5:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{3.4}$$

$$h_t = o_t \odot \sigma_c(c_t) \tag{3.5}$$

Among them, $\odot$ is Hadamard product (multiplication of vector elements); σ_c is the hyperbolic tangent function(tanh) state activation function.

At time t in Figure 3.3, forget f_t, activate input i_t, output o_t, activate candidate unit input g_t as shown in Equation 3.6 and Equation 3.9:

$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \tag{3.6}$$

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \tag{3.7}$$

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \tag{3.8}$$

$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \tag{3.9}$$

Use h_{t-1} and x_t as input information for the current time step in network training. After passing through the gate activation function , these pieces of information ultimately result in an output value between [19].

The larger the forget gate activation ft is, the less the previous cell state output ct−1 is forgotten. Conversely, the larger the input gate activation it is, the more the candidate input gt is represented, allowing more information to be written into the cell state at the current time. Together, the forget gate ft and input gate it determine how much of the new input information is incorporated into the current cell state output ct. Additionally, the activation of the output gate ot dictates the current hidden state output ht. This combined control mechanism enables the model to capture long-term dependencies in human behavior data over time steps.

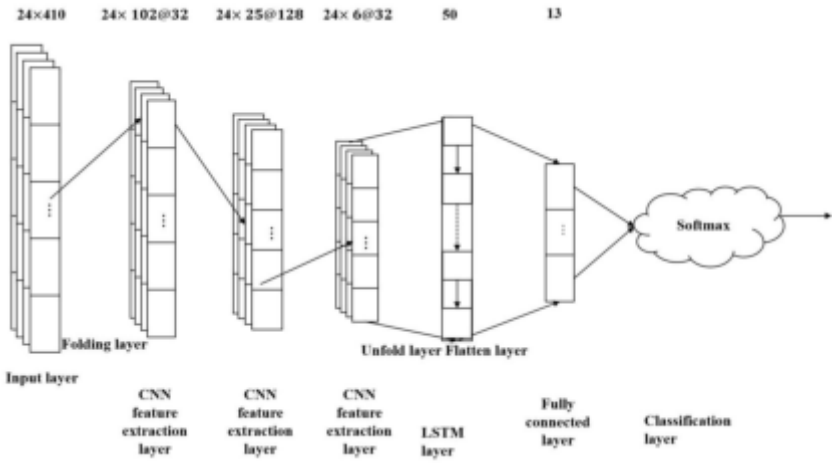


Fig. 3.4: CLT net model structure

3.3. CLT net human behavior recognition model. Since the INS can be viewed as a time sequence, the conventional SVM based on artificial characteristics might result in sub-optimal information usage and insufficient recognition of complicated human activities. In order to solve this problem, we present a novel CLT Net, which combines spatial-temporal and spatial characteristics to identify people's behaviour. The CLT Net model is a combination of CNNs that are good at automatic extraction from data, and LSTMs that can effectively capture time dependence in time sequence data. The proposed model uses the same architecture as CNN, except that the LSTM layer is used instead of a full connection layer. The CLT Net Model's parameters and functional specs match those of the CNN and LSTM modules. The CLT Net architecture, as shown in Figure 3.4, consists of a sequential entry level, a folding level, a CNN characteristic extracting layer (consisting of convolutions, Max Pooling, Batch Normalization, Leaky ReLU), unfolding, flat, LSTM, full connection, and soft max [20].

A CLT network model is employed to classify a person's behaviour in this way. Firstly, a CNN module is used to extract the input person's behavioral data series. Then, the 2D data characteristics are compressed to one dimension and transferred to the LSTM layer, where temporal feature filtering takes place. Next, a fully connected layer maps the selected features to the label space using a weight matrix. Lastly, we use soft max level to classify and select the class that has the best forecast probability as the forecast tag of the input data sample. During training, the model adjusts its parameters by comparing the predicted categories from forward propagation with the true labels of the samples. The error is used to backpropagate through the network, with the loss function and optimizer continually refining the weights and bias terms, allowing the model to improve and ultimately achieve optimal performance.

3.4. Experimental verification.

3.4.1. Experimental Dataset. The author conducted research on human behavior recognition based on wearable sensor data, and the experiment used the publicly available DaLiAc (Daily Life Activities) dataset. The dataset was collected by placing four 6-axis inertial sensor nodes on the subjects' right buttock, chest, right wrist, and left ankle. Each sensor node includes a three-axis accelerometer and a three-axis gyroscope. The accelerometers have a range of ±6g, and the gyroscopes, particularly the one on the wrist, chest, and hip sensor nodes is $\pm 500(\circ)/s$, the gyroscope range of the ankle sensor node is $\pm 2000(\circ)/s$, and the data sampling frequency is $204.7Hz^2$. A total of 19 healthy subjects participated in the data collection experiment (8 females, 11 males, age 26 ± 8 years, height $177 \pm 11cm$, weight $75.2 \pm 14.2kg$, deviation ± mean), and a total of 13 activities were collected. The activities and corresponding tags are shown in Table 3.1.

Table 3.1: Activities and Corresponding Tags

Activity Description	Label
Sit still	1
Lie flat	2
Stand	3
Wash dishes	4
Vacuum cleaner	5
Sweep the floor	6
Walk	7
Go upstairs	8
Go downstairs	9
Treadmill running	10
Test bike riding (50W)	11
Test bike riding (100W)	12
Skipping rope	13

Table 3.2: Experimental Parameter Settings

Parameter	Set up
Initialization of CNN layer weight coefficients	Kaiming method
Initialization of LSTM layer weight coefficients	Orthogonal method
Initialization of fully connected layer weight coefficients	Kaiming method
Optimizer	Adam optimizer
Loss function	Cross baking
Initial learning rate	0.001
Sample sequence size	24× 410
Number of samples in the training set	20088
Number of test set samples	2232
Training rounds	20
Batch size	500
Leaky Relu factor	0.1

3.4.2. Experimental operating environment. All of the author's models were trained and tested on a computer equipped with a Core i5-6500U CPU @ 3.20 GHz and 16 GB of RAM. The system ran Windows 10 Professional 64-bit, and the models were developed using the Matlab 2020b Deep Learning Toolbox framework.

3.4.3. Experimental Parameters. Firstly, the human behavior data is divided into samples with a sliding window length of 410 (rounded to twice the sampling frequency), and there is 50% data overlap between adjacent windows. As a result, each sample sequence has a size of 24 × 410 (corresponding to 4 sensors with 6-axis data each). After segmenting the data, the samples were sorted, with the top 90% used as the training set and the remaining 10% as the testing set. The experimental parameters are detailed in Table 3.2. During the simulation experiments, all models were initialized with the same configuration to ensure a fair comparison, allowing for a more accurate assessment of the true performance of the CNN models, LSTM models, and CLT net models.

The CNN layers and fully connected layers utilize Kaiming initialization for their weights to speed up model convergence. For the LSTM layers, the weights are initialized using the orthogonal method. All models are optimized using the Adam algorithm, an adaptive moment estimation method that offers quicker convergence and lower memory usage. This approach also eliminates the need for a validation set during training.

4. Results and Discussion. To effectively showcase the generalization capability of the CLT net model, we calculated the macro precision, macro recall, and macro F1 score for the test results of the LSTM, CNN, CLT net, and traditional BP models. These metrics were computed by averaging the precision, recall, and

Table 4.1: Comparison of evaluation indicators for BP, LSTM, CNN, and CLT net models

Model	Macro precision	Macro recall rate	Macro F1 value
BP	0.5221	0.5127	0.5002
LSTM	0.7585	0.7117	0.7260
CNN	0.9510	0.9503	0.9501
CLT-net	0.9635	0.9605	0.9616

F1 score across all 13 categories of human behavior. The performance of the four models is summarized and compared in Table 4.1.

The overall average classification accuracy of BP, LSTM, and CNN models were 61.6%, 77.5%, and 96.3%, respectively. The author proposed the CLT net model, which achieved 97.5%, an improvement of 35.8, 20.1, and 1.1 percentage points, respectively. The CNN model can extract features of human behavior data, which represent the original human behavior data to the maximum extent possible. Using these features for human behavior recognition and classification has good performance. Compared to the LSTM model, the CNN model has a higher recognition rate. The LSTM model is only used for modeling temporal data to learn the correlation between data, and cannot achieve feature extraction. This also indicates that feature extraction is the key to classification and recognition, and the CNN feature extraction module is the most important component of the CLT net model. The precision assessment measure measures the proportion of the correct positive samples to the total expected positive, while the recall measure measures the proportion of correct recognized positive samples to the overall true positives. F1 is the harmonic mean of accuracy and recall. According to Table 4.1, the CLT net model exhibits the highest accuracy among the evaluated models, recall, and F1 score evaluation indicators, while all indicators of the BP model are the lowest, indicating that the CLT net model has good stability and reliability in recognizing and classifying different human behaviors.

5. Conclusion. The author presents a study on recognizing human behavior in complex environments using deep learning techniques, introducing a novel model called CLT net that leverages spatiotemporal feature fusion. The proposed approach combines Convolutional Neural Networks (CNNs) to extract the time dependence of time sequence data using Long Short-Term Memory (LSTM) networks, and uses the soft max classifier to classify the behavior. The experimental results on the DaLiAc dataset show that compared to LSTM, CNN, and BP models, the CLT net model converges faster and has better performance in human behavior recognition and classification. Subsequently, lightweight deep learning models will be constructed to optimize sensor based human behavior recognition methods and further improve feature recognition accuracy.

REFERENCES

[1] Dong, X. Q., Wang, X. C., Li, B. J., Wang, H. Y., & Chen, G. C. (2024). Mp-abr: a framework for intelligent recognition of abnormal behaviour in multi-person scenarios. Multimedia Tools and Applications, 83(18), 55605-55626.
[2] Wu, H., Han, Y., & Meng ZhangBihonegn Dianarose AbebeMolla Betelhem LegesseRuoyu Jin. (2023). Identifying unsafe behavior of construction workers: a dynamic approach combining skeleton information and spatiotemporal features. Journal of construction engineering and management, 149(11), 1-15.
[3] Ozdemir, C., Hoover, R. C., Caudle, K., & Braman, K. (2024). Tensor discriminant analysis on grassmann manifold with application to video based human action recognition. International Journal of Machine Learning and Cybernetics, 15(8), 3353-3365.
[4] Noor, T. H. (2023). Human action recognition-based iot services for emergency response management. Machine Learning and Knowledge Extraction, 5(1), 330-345.
[5] Reddy, G. V., Deepika, K., Malliga, L., Hemanand, D., Senthilkumar, C., & Gopalakrishnan, S., et al. (2023). Human action recognition using difference of gaussian and difference of wavelet. Big Data Mining and Analytics, 6(3), 336-346.
[6] Ye, Q., Tan, Z., & Zhang, Y. (2022). Human action recognition method based on motion excitation and temporal aggregation module. Heliyon, 8(11), 11401.
[7] Sun, B., Kong, D., Zhang, W., & Jia, W. (2022). Survey on human action recognition from depth maps, 27(6), 29.
[8] Wang, C., Zhao, F., Zhang, H., Luo, H., Qin, Y., & Fang, Y. (2022). Fine-grained trajectory-based travel time estimation for multi-city scenarios based on deep meta-learning. arXiv e-prints, 41(12), 3788-3817.
[9] Fuentes, A., Han, S., & Nasir, Muhammad FahadPark, JongbinYoon, SookPark, Dong Sun. (2023). Multiview monitoring

of individual cattle behavior based on action recognition in closed barns using deep learning. animals, 13(12), 76(9), 2667-2684.
[10] Jiao, C. (2022). Recognition of human body feature changes in sports health based on deep learning. Computational and Mathematical Methods in Medicine, 2022.
[11] Srihari, P., Harikiran, J., & Reddy, C. V. S. (2023). Effective framework for human action recognition in thermal images using capsnet technique. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 45(6), 11737-11755.
[12] Gutoski, M., Lazzaretti, A. E., & Lopes, H. S. (2023). Unsupervised open-world human action recognition. Pattern analysis and applications: PAA, 26(4), 1753-1770.
[13] Kumar, R., & Kumar, S. (2024). A survey on intelligent human action recognition techniques. Multimedia Tools and Applications, 83(17), 52653-52709.
[14] Sam Slade Li Zhang Yonghong Yu Chee Peng Lim. (2022). An evo1ving ensemb1e mode1 of mu1ti stream convo1utiona1 neura1 networks for human action recognition in sti11 images. Neural Computing and Applications, 34(11), 9205-9231.
[15] Basak, H., Kundu, R., Singh, P. K., Ijaz, M. F., Woniak, M., & Sarkar, R. (2022). A union of deep learning and swarm-based optimization for 3d human action recognition. Scientific Reports, 12(1), 1-17.
[16] Berlin, Jeba, S., John, & Mala. (2022). Light weight convolutional models with spiking neural network based human action recognition. Journal of intelligent & fuzzy systems: Applications in Engineering and Technology, 39(1), 961-973.
[17] Ghosh, S. K., Rashmi, M., Mohan, B. R., & Guddeti, R. M. R. (2022). Deep learning-based multi-view 3d-human action recognition using skeleton and depth data. Multimedia Tools and Applications, 82(13), 19829-19851.
[18] Wu, Q., Huang, Q., & Li, X. (2022). Multimodal human action recognition based on spatio-temporal action representation recognition model. Multimedia Tools and Applications, 82(11), 16409-16430.
[19] Islam, M. S., Bakhat, K., Khan, R., Naqvi, N., Islam, M. M., & Ye, Z. (2022). Applied human action recognition network based on snsp features. Neural Processing Letters, 54(3), 1481-1494.
[20] Yongfeng, Q., Jinlin, H., & Xiaoxu, L. P. (2023). Semantic-guided multi-scale human skeleton action recognition. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 53(9), 9763-9778.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Sep 9, 2024
Accepted: Oct 11, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2061–2069, DOI 10.12694/scpe.v26i5.4905

THE BLOCKCHAIN IN THE DESIGN OF ELECTRONIC MEDICAL RECORD SYSTEM SUPPORTING MULTIMEDIA COMMUNICATION TECHNOLOGY

XIONG JIAN*, RAJAMOHAN PARTHASARATHY†, YINQING TANG‡, AND BINWEN HUANG§

Abstract. This paper proposes the design of electronic medical record system supported by multimedia communication based on blockchain technology. Blockchain technology ensures the secure storage and sharing of patient information through distributed ledger and smart contract algorithm. In this system, smart contracts are used to automatically execute cross-institutional data access control and audit functions to ensure the transparency and compliance of data access. At the same time, this paper introduces multimedia communication technology to support the efficient transmission and sharing of medical data, especially in diagnostic images, videos and voice. Simulation results show that the electronic medical record system based on blockchain has significantly improved data processing efficiency, security and reliability compared with traditional systems.

Key words: Smart contract algorithm; Multimedia communication; Electronic medical record system; Blockchain.

1. Introduction. As an important part of modern medical informatization, electronic medical record system (EMR) is gradually replacing the traditional paper medical record management method. However, the existing electronic medical record system still faces many challenges in data security, privacy protection, cross-institutional data sharing and data integrity. Especially when multiple roles such as medical institutions, doctors, and patients are involved, the circulation and security management of data are particularly prominent. Therefore, how to build a safe, reliable and efficient electronic medical record system has become a key issue in the current development of medical informationization.

The introduction of smart contracts enables blockchain to automatically execute pre-defined logical rules without the intervention of external third parties, ensuring that the data sharing and access process is transparent and compliant. Therefore, the application of blockchain in medical information sharing and privacy protection shows broad prospects. Reference [1] proposed a distributed medical information sharing solution based on blockchain, which successfully solved the problem that data in traditional centralized systems are easily tampered with and leaked. The solution achieves multi-party information synchronization through distributed ledgers, ensuring data security and consistency. Reference [2] discussed the application of blockchain in electronic medical record systems and proposed an automated data access control mechanism based on smart contracts, which effectively reduced the potential risks of manual operations and enhanced the transparency of the system. Reference [3] designed a cross-institutional information sharing framework based on blockchain, which solved the trust problem in information flow between different medical institutions and strengthened the protection of patient privacy. Reference [4] discusses the application of multimedia communication technology in medical systems, solving the problem of insufficient timeliness of medical data during transmission, especially in the real-time transmission of diagnostic images, videos, and voice data.

This paper proposes a blockchain-based electronic medical record system design. The system not only supports multimedia communication technology, but also realizes automatic access and sharing management of data through smart contracts. The core goal of this system is to solve the trust barriers and privacy leakage risks in traditional electronic medical record systems in data sharing, and significantly improve the efficiency and security of information transmission.

*Modern Educational Technology Center, Hainan Medical University, Haikou, Hainan, 571199, China; Faculty of Engineering, Built Environment and Information Technology SEGi University, Kuala Lumpur, 47810, Malaysia
†Faculty of Engineering, Built Environment and Information Technology SEGi University, Kuala Lumpur, 47810, Malaysia
‡Modern Educational Technology Center, Hainan Medical University, Haikou, Hainan, 571199, China
§Modern Educational Technology Center, Hainan Medical University, Haikou, Hainan, 571199, China (Corresponding author, huangbinwen88@126.com)

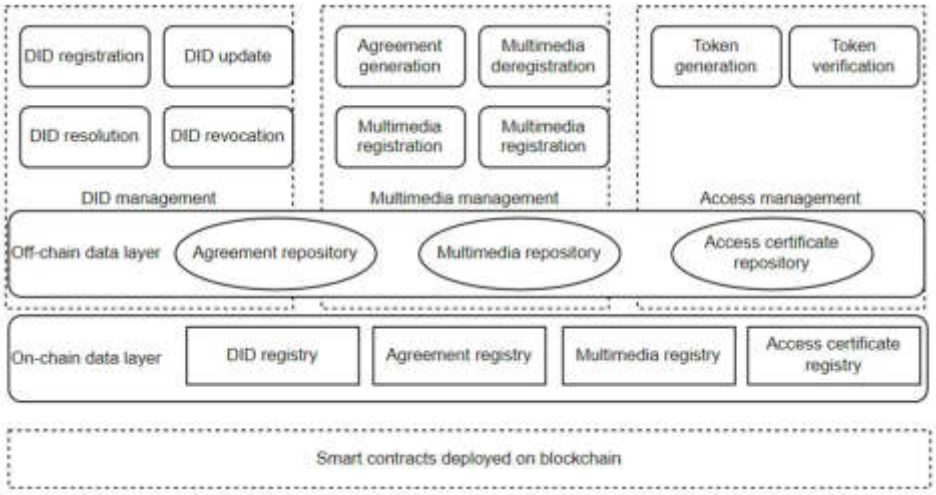


Fig. 2.1: Blockchain and multimedia communication integration architecture.

2. System architecture design.

2.1. Integration architecture of blockchain and multimedia communication. In order to solve the shortcomings of traditional electronic medical record systems, this paper proposes an electronic medical record system architecture based on blockchain and multimedia communication technology. The system architecture is shown in Figure 2.1.

The blockchain network layer is responsible for recording all data operations related to electronic medical records. Through the distributed ledger method, the data is tamper-proof and transparent. All operations, including the creation, access, and modification of medical records, will be recorded and executed through smart contracts.

The smart contract layer is used to realize automated medical record data management, especially the access control of data. Each query, sharing, or update of medical records will trigger the corresponding smart contract to ensure the legality and compliance of data access.

The multimedia communication layer is mainly used to process multimedia information in medical data. Through advanced communication protocols and technologies, this layer can efficiently transmit large-capacity data such as medical images and diagnostic videos, while ensuring the security of data transmission.

Due to the diversity and large volume of medical data, it is not feasible to rely solely on blockchain storage. Therefore, this system adopts a distributed storage solution to store data in various nodes and ensure the security of data through encryption mechanisms.

2.2. Application of smart contracts in medical record data management. Smart contracts play a key role in this system, especially in medical record data management. Through smart contracts, the system can realize automatic processing and access control of medical record data [5]. Smart contracts predefine the permissions of various roles, such as doctors, patients, insurance companies, etc. Only users who meet the permission requirements can access the corresponding medical record data through smart contracts.

In addition, smart contracts are also used to record logs of each data operation. Whenever a user queries or updates a medical record, the smart contract automatically records the time, operator, and content of the operation [6]. This logging mechanism not only ensures the transparency of data access, but also facilitates data auditing. Smart contracts can also realize multi-party data sharing. For example, in cross-institutional medical cooperation, different medical institutions can achieve secure data sharing through smart contracts without worrying about data leakage or tampering. At the same time, smart contracts can automatically execute data sharing agreements without the intervention of third parties, reducing the risk of human intervention.

2.3. Data encryption and storage. To ensure the security of data during transmission and storage, this system uses a variety of encryption and security algorithms, including elliptic curve encryption (ECC), hash algorithms, and distributed storage. The following are the main technologies used by the system and their mathematical descriptions.

2.3.1. Elliptic Curve Encryption. ECC is an efficient public key encryption technology, especially suitable for resource-constrained environments [7]. Compared with the traditional RSA algorithm, ECC requires a shorter key length while providing the same level of security, so it is more computationally efficient. The basic formula of ECC is as follows:

$$y^2 = x^3 + ax + b(\text{mod}) \tag{2.1}$$

a and b are the parameters of the elliptic curve, and p is a prime number. Through this formula, the system can generate public and private keys for encrypting and decrypting medical record data.

2.3.2. Hash algorithm. In the blockchain system, hash algorithms are widely used to verify the integrity of data. The system uses the SHA-256 hash function to generate a unique identifier for the data, ensuring that the medical record information remains intact during storage and transmission without any modification [8]:

$$H(m) = h_1\left(h_2\left(\ldots h_n(m)\right)\right) \tag{2.2}$$

m represents the input data, and $H($ m) is the generated hash value. The algorithm is calculated through multiple iterations to ensure that any small data change will lead to a significant change in the hash value.

2.3.3. Data encryption and decryption. In this system, each electronic medical record will be encrypted by ECC before storage. Assuming that the medical record data is D, its encryption process can be expressed as:

$$C = E(D, k) = D \cdot k(\text{modp}) \tag{2.3}$$

Among them, k is the encryption key, and C is the encrypted medical record data. The decryption process is:

$$D = C \cdot k^{-1}(\text{modp}) \tag{2.4}$$

This encryption and decryption mechanism ensures the security of medical record data. Even if the data is intercepted during transmission or storage, the attacker cannot decrypt the content [9].

2.3.4. Distributed storage mechanism. This system adopts a distributed storage mechanism [10]. Medical record data will not be stored directly on the blockchain, but will be stored through distributed nodes. The blockchain only saves the index and hash value of the data. The distribution of data in the distributed storage network is as follows:

$$D = (D_1, D_2, \ldots, D_n) \tag{2.5}$$

$D_1, D_2, \ldots, D_n$ represent different parts of the medical record data after being sharded. These fragments are distributed and stored in different nodes [11].

3. Application of blockchain technology.

3.1. The role of blockchain in the electronic medical record system. In the electronic medical record system, blockchain technology plays a vital role, mainly in the storage, security verification and sharing of information. First, the distributed ledger structure of blockchain ensures the security and tamper-proof nature of data storage. Whenever an electronic medical record is operated, a corresponding block is created, and a consensus mechanism is used to ensure that these operations are consistent and verifiable throughout the network. The distributed storage of data not only reduces the possibility of single point failures, but also enhances the redundancy and security protection of information [12].

Secondly, blockchain also plays a key role in the verification of medical record data. Each data operation generates a hash value, and the integrity and legitimacy of the operation are verified by a hash function. For example, assuming there is a medical record data D, its corresponding hash value $H(D)$ can be obtained by the following formula:

$$H(D) = \text{hash}(D) \tag{3.1}$$

This hash value is compared with the historical data records in the blockchain. If the hash values are consistent, the verification is passed, indicating that the data has not been tampered with. Otherwise, the system will prompt that the data may be tampered with or damaged.

3.2. Medical record transmission and verification under multimedia communication technology. In the medical environment, medical records are not just text data, but more often include complex multimedia information, such as medical images, videos, audio, etc. Therefore, how to efficiently and securely transmit these multimedia medical record data is one of the focuses of system design [13]. The combination of blockchain and multimedia communication technology provides an ideal solution for this. The distributed structure of blockchain can effectively protect multimedia data from tampering during transmission. Each piece of multimedia data will be encrypted before transmission and a unique hash value will be generated to ensure the integrity and consistency of the data. For example, for a medical image I, the encrypted ciphertext $E(I)$ and the corresponding hash value $H(E(I))$ can be expressed as follows:

$$\begin{aligned} E(I) &= \text{encrypt}(I, k) \\ H(E(I)) &= \text{hash}(E(I)) \end{aligned} \tag{3.2}$$

k is the encryption key, $E(I)$ is the encrypted image data, and $H(E(I))$ is the hash value of the data. After the data transmission is completed, the receiver can verify the integrity and consistency of the data through decryption and hash verification.

3.3. Design and implementation of smart contracts. The role of smart contracts in blockchain cannot be ignored, especially in electronic medical record systems, where its main functions are reflected in permission management, data auditing and automated processing.

Smart contracts allow the system to assign different access rights to different users. Taking doctors, patients and medical insurance companies as examples, different roles have different access rights [14]. For example, patients can view all their medical records, while doctors can only view relevant medical records. Every creation, modification or access operation of medical record data will be recorded by the smart contract on the blockchain to form an unalterable log. In this way, when a security incident occurs, the system administrator can use these logs to audit the data and track the source, time, and content of each data operation. Smart contracts can also automate the management of medical records. For example, when a doctor enters a new diagnosis, the system automatically activates the smart contract and sends the diagnosis information to the patient and related medical institutions without any human intervention. This automated operation significantly improves the efficiency of the system and reduces the potential risks of human intervention [15]. The execution efficiency and stability of smart contracts depend on the design quality of the underlying code. During the design phase, the system must ensure the accuracy of the contract logic to avoid security vulnerabilities and logical errors. The execution process of a smart contract is as follows:

$$T_i = C\left(x_i\right) \tag{3.3}$$

T_i represents the execution result, C is the logic function of the smart contract, and x_i is the input data. The contract processes the input data according to the preset rules and outputs the corresponding results.

3.4. Privacy protection mechanism of blockchain. Patients' medical information is very sensitive, so the system must have an effective privacy protection mechanism.

Zero-knowledge proof is a cryptographic technology that allows users to prove the legitimacy of their possession of certain data without disclosing specific information [16]. For example, a user can prove that he

Table 4.1: Experimental data set

Data type	medical records (n)	Data size (GB)	Data content
Plain text medical records	1000	0.5	Plain text information such as diagnosis reports and prescriptions
Image medical records	1000	5.2	MRI images, CT scan images
Video medical records	500	20	Ultrasound videos, dynamic image records
Voice medical records	500	1.5	Doctor's diagnosis voice records

has access to a medical record without revealing the specific contents of the medical record. The mathematical model of zero-knowledge proof is as follows:

$$P(x) \to V \tag{3.4}$$

Among them, $P(x)$ represents the secret data held by the prover, and V is the verifier. In the zeroknowledge proof process, the specific content of $P(x)$ will not be passed to V, but the verifier can still confirm the legitimacy of the prover. Homomorphic encryption allows encrypted data to be directly operated without decryption. In the process of medical data processing, homomorphic encryption can ensure that the data remains encrypted during processing and analysis, thereby protecting data privacy [17]. For example, suppose there are two encrypted medical data $E(A)$ and $E(B)$. Homomorphic encryption allows the encrypted data to be added, and the result is:

$$E(A+B) = E(A) \oplus E(B) \tag{3.5}$$

Combined with zero-knowledge proof and homomorphic encryption technology, the application of blockchain in medical record system can effectively protect patient privacy and ensure that data will not be leaked during storage, transmission and processing.

4. Experiment and performance analysis.

4.1. Design of experimental data set and multimedia medical record. In this experiment, electronic medical record data sets from multiple sources are selected, and multimedia medical record data are introduced for testing. Traditional electronic medical record data sets contain text records, diagnosis reports, drug prescriptions, etc., while multimedia medical records are extended to include medical images, video records, and voice diagnosis. In order to simulate the actual medical scenario, this paper obtained 3,000 medical records from the hospital system, including 1,000 MRI data with images, 500 dynamic ultrasound videos, and 1,500 medical records containing voice records [18]. To ensure the breadth and representativeness of the experiment, this paper divides the data set into three categories: plain text medical records, multimedia medical records (images and videos), and medical records combined with multimedia and text. Different types of medical records have different storage, transmission and processing requirements in the system. Through these multi-dimensional data experiments, the integration effect of blockchain and multimedia communication technology can be more comprehensively evaluated. The specific contents of the experimental data set are shown in Table 4.1.

4.2. Evaluation Indicators. In order to evaluate the performance of this system in the multimedia electronic medical record scenario, this paper selects system performance, storage efficiency, communication delay and security analysis as the main evaluation indicators. The results of the system performance evaluation are shown in Table 4.2.

The system has high performance when processing plain text medical records, while the processing efficiency of image and video medical records is relatively low due to the large data volume, but it can still meet the actual application needs.

Table 4.2: The results of the system performance evaluation.

Metrics	Plain text medical records	Image medical records	Video medical records	Voice medical records
Throughput (records/s)	1200	850	450	750
Response time (milliseconds)	200	400	600	350
Contract execution time (s)	0.8	1	1.5	1.1

Table 4.3: Comparison of the three consensus mechanisms in system performance.

Consensus Mechanism	Throughput records	Response time (ms)	Consensus time (s)
PoW	500	800	2.5
PoS	1000	300	1.2
PBFT	1200	200	0.8

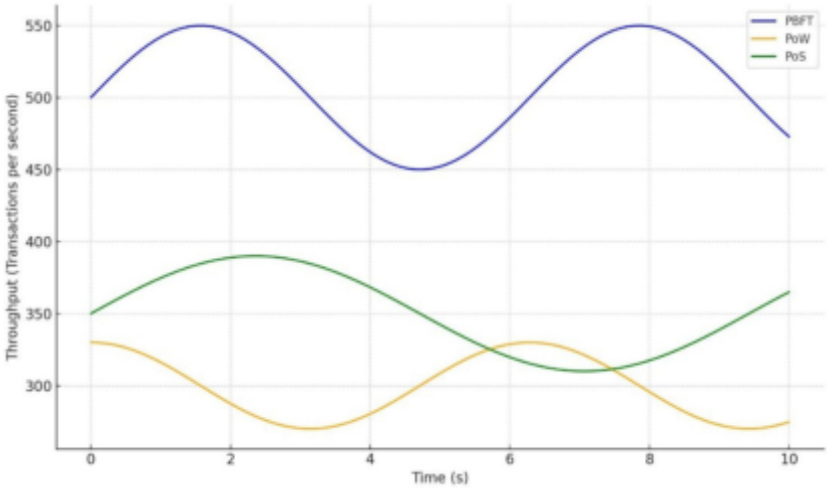


Fig. 4.1: Performance trends of different consensus mechanisms in terms of processing throughput.

4.3. Performance comparison analysis of blockchain systems. This paper compares the performance differences of proof of stake (PoS), practical Byzantine fault tolerance (PBFT) and proof of work (PoW) when processing electronic medical record data. Table 4.3 shows the comparison of the three consensus mechanisms in system performance.

PBFT has the best overall performance and is suitable for deployment in medical data environments that require high efficiency. Figure 4.1 shows the performance trends of different consensus mechanisms in terms of processing throughput.

4.4. Evaluation of the effectiveness of data encryption and distributed storage. In order to ensure the security and storage efficiency of medical record data, the system adopts a solution that combines elliptic curve cryptography (ECC) with distributed storage. In the experiment, this paper compares the changes in the storage efficiency of the system before and after using distributed storage, as well as the performance overhead caused by encryption during data transmission. First, the storage efficiency of the system in both encrypted and unencrypted conditions is evaluated. Table 4.4 shows the storage size comparison of various medical record data in encrypted and unencrypted states:

Table 4.4: The storage size comparison of various medical record data in encrypted and unencrypted states.

Medical record type	Original size (GB)	Encrypted size (GB)	Storage Optimization Ratio
Plain text medical record	0.5	0.6	-20%
Image medical record	5.2	4.9	6%
Video medical record	20	18.5	7.50%

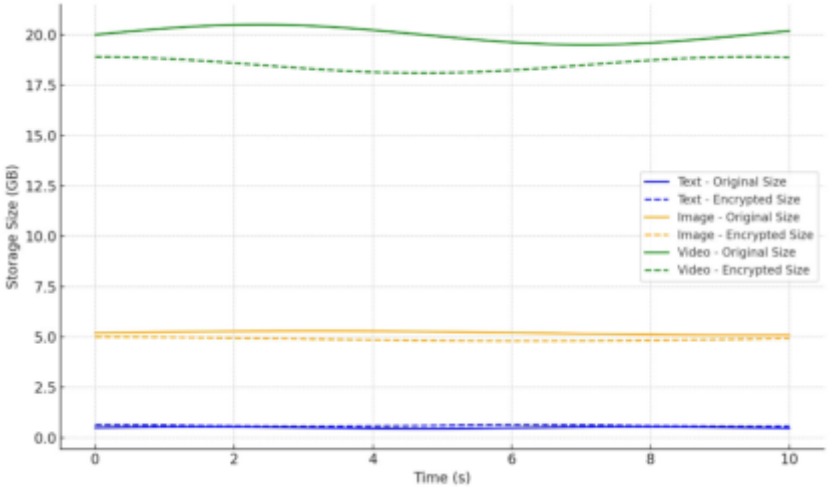


Fig. 4.2: Comparison of storage efficiency of different medical record type.

It can be seen that for multimedia data, the encryption process does not significantly increase the storage space requirements, but reduces the storage cost through the optimization of distributed storage. Figure 4.2 shows the storage efficiency comparison of different medical record types.

Secondly, the impact of the data encryption process on communication delay was evaluated. The experimental results show that although data encryption increases a certain amount of computational overhead, the delay increase is small for plain text data; while for video data, the delay caused by the encryption process is more obvious. Overall, the system can provide acceptable communication delay while ensuring data security. Figure 4.3 shows the impact of data encryption on communication delay of different medical record types.

5. Conclusion. This system can ensure the secure storage and access control of medical data. Smart contracts automatically execute permission management, ensure the transparency and compliance of data, and reduce the security risks caused by manual operations. At the same time, the application of multimedia communication technology improves the transmission efficiency of medical data, especially in supporting the transmission of large amounts of information such as diagnostic images and videos. The simulation results show that the electronic medical record system based on blockchain has significant improvements in security, reliability and data processing efficiency compared with traditional systems. The system improves the redundancy and anti-tampering ability of data storage through a distributed structure, while the smart contract algorithm effectively solves the trust problem in cross-institutional data sharing. Combined with multimedia communication technology, the system can meet the needs of modern medicine for efficient, convenient and reliable data sharing, especially in the application scenarios of telemedicine and multi-party collaboration.

6. Acknowledgements.

1. Research project on education and teaching reform of colleges and universities in Hainan Province: Application of medical data sharing in hospital virtual simulation teaching. (NO: Hnjg2019-53).

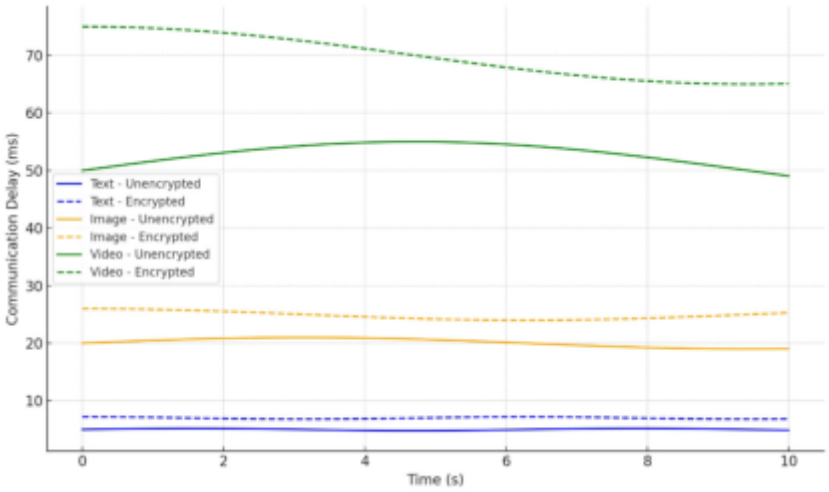


Fig. 4.3: The impact of data encryption on communication delays of different medical record types.

2. Major science and technology projects in Hainan Province: New intelligent multi-point trigger early warning and diagnosis technology for emerging and sudden infectious diseases. (NO:ZDKJ2021029).
3. Industry-university cooperation coordination education project of the Ministry of Education in 2022 (No.220604719293004).

REFERENCES

[1] Tith, D., Lee, J. S., Suzuki, H., Wijesundara, W. M. A. B., Taira, N., Obi, T., & Ohyama, N. (2020). Application of blockchain to maintaining patient records in electronic health record for enhanced privacy, scalability, and availability. *Healthcare informatics research*, *26*(1), 3-12.

[2] Mahajan, H. B., & Junnarkar, A. A. (2023). Smart healthcare system using integrated and lightweight ECC with private blockchain for multimedia medical data processing. *Multimedia Tools and Applications*, *82*(28), 44335-44358.

[3] Wu, H., Dwivedi, A. D., & Srivastava, G. (2021). Security and privacy of patient information in medical systems based on blockchain technology. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *17*(2s), 1-17.

[4] Mahajan, H. B., Rashid, A. S., Junnarkar, A. A., Uke, N., Deshpande, S. D., Futane, P. R., ... & Alhayani, B. (2023). RETRACTED ARTICLE: Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems. *Applied Nanoscience*, *13*(3), 2329-2342.

[5] Abbate, S., Centobelli, P., Cerchione, R., Oropallo, E., & Riccio, E. (2022). Blockchain technology for embracing healthcare 4.0. *IEEE Transactions on Engineering Management*, *70*(8), 2998-3009.

[6] Stafford, T. F., & Treiblmaier, H. (2020). Characteristics of a blockchain ecosystem for secure and sharable electronic medical records. *IEEE Transactions on Engineering Management*, *67*(4), 1340-1362.

[7] Reegu, F. A., Mohd, S., Hakami, Z., Reegu, K. K., & Alam, S. (2021). Towards trustworthiness of electronic health record system using blockchain. *Annals of the Romanian Society for Cell Biology*, *25*(6), 2425-2434.

[8] Sabu, S., Ramalingam, H. M., Vishaka, M., Swapna, H. R., & Hegde, S. (2021). Implementation of a secure and privacy-aware E-Health record and IoT data sharing using blockchain. *Global Transitions Proceedings*, *2*(2), 429-433.

[9] Liu, Y., Lu, Q., Zhu, C., & Yu, Q. (2021). A blockchain-based platform architecture for multimedia data management. *Multimedia Tools and Applications*, *80*(20), 30707-30723.

[10] Abbas, A., Alroobaea, R., Krichen, M., Rubaiee, S., Vimal, S., & Almansour, F. M. (2024). Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things. *Personal and ubiquitous computing*, *28*(1), 59-72.

[11] Mahajan, H. B. (2022). Emergence of healthcare 4.0 and blockchain into secure cloud-based electronic health records systems: solutions, challenges, and future roadmap. *Wireless Personal Communications*, *126*(3), 2425-2446.

[12] Wang, J., Fan, S., Alexandridis, A., Han, K., Jeon, G., Zilic, Z., & Pang, Y. (2021). A multistage blockchain-based secure and trustworthy smart healthcare system using ecg characteristic. *IEEE Internet of Things Magazine*, *4*(3), 48-58.

[13] Biswas, S., Sharif, K., Li, F., Latif, Z., Kanhere, S. S., & Mohanty, S. P. (2020). Interoperability and synchronization management of blockchain-based decentralized e-health systems. *IEEE Transactions on Engineering Management*, *67*(4), 1363-1376.

[14] Kim, H., Lee, S., Kwon, H., & Kim, E. (2021). Design and implementation of a personal health record platform based on patient-consent blockchain technology. *KSII Transactions on Internet and Information Systems (TIIS)*, *15*(12), 4400-4419.

[15] Shuaib, K., Abdella, J., Sallabi, F., & Serhani, M. A. (2022). Secure decentralized electronic health records sharing system based on blockchains. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5045-5058.

[16] Adegoke, A. (2023). Patients' reaction to online access to their electronic medical records: the case of diabetic patients in the US. *International Journal of Applied Sciences: Current and Future Research Trends*, *19*(1), 105-115.

[17] Tertulino, R., Antunes, N., & Morais, H. (2024). Privacy in electronic health records: a systematic mapping study. *Journal of Public Health*, *32*(3), 435-454.

[18] Lewis, A. E., Weiskopf, N., Abrams, Z. B., Foraker, R., Lai, A. M., Payne, P. R., & Gupta, A. (2023). Electronic health record data quality assessment and tools: a systematic review. *Journal of the American Medical Informatics Association*, *30*(10), 1730-1740.

[19] Landolsi, M. Y., Hlaoua, L., & Ben Romdhane, L. (2023). Information extraction from electronic medical documents: state of the art and future research directions. *Knowledge and Information Systems*, *65*(2), 463-516.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Sep 13, 2024
Accepted: Oct 14, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2070–2077, DOI 10.12694/scpe.v26i5.4911

INTERACTIVE DISPLAY SYSTEM INTEGRATING INFORMATION TECHNOLOGY IN ENVIRONMENTAL ART DESIGN

YUANYUAN YAO *AND ZHAOCHEN WANG †

Abstract. In order to solve the problems of dimensionality reduction and color display of spectral data after redundancy reduction in interactive display in environmental art design, the author proposes a research on an interactive display system that integrates information technology in environmental art design. The author first performs principal component transformation on the spectral data cube, assigning the first three components to the black and white channels, red and green channels, and yellow and blue channels in the color space. Then, after spatial transformation to sRGB space, the data is segmented and translated to the range of 0-1, mapped to 8-bit RGB, and single item evaluations of standard deviation, entropy, and average gradient are performed on each translated image. After all translations are completed, a comprehensive evaluation is performed on all evaluation values, and the interval with the highest comprehensive evaluation value is selected to output the mapping. The experimental results show that although this algorithm has a slightly longer time overhead, it is still far less than the acquisition time of 6.63 seconds for one spectral data cube, and is suitable for real-time detection applications of hyperspectral data. The fusion of images can maximize the energy, information, and clarity of the images, which is beneficial for the rapid recognition and judgment of the human eye.

Key words: Spectroscopy, Fusion display, Evaluation and selection, Environmental art design, Interactive Display

1. Introduction. With the development of the times, information technology has made rapid progress. Digital newspapers, online media, and self media platforms are all concrete manifestations of digital technology. Digital technology has accelerated the development speed of the design field, combining various advanced means such as computer technology and multimedia technology, subverting the traditional design ideas and paths, and making the previous environmental design methods more vivid and visual [1]. For example, digital images, 3D images, etc. are the result of combining traditional design methods with digital technology. In the digital age, the artistic nature of digital technology has become a new form and a key development trend in design and creation in the context of that era [2]. At the same time, the widespread application of digital technology has greatly improved the efficiency of environmental design. Environmental designers can also use various digital technological means to improve their creative efficiency, complete efficient creation of environmental art works, and enable all different forms of art to be combined with each other [3].

Environmental art and design is a comprehensive design discipline that integrates science and art into a theoretical system. The goal of environmental art design is to establish a sustainable humanistic and ecological environment, combine art and science for design, and better improve the human living environment. From the explanation of the concept of environmental art design at home and abroad, we can know that environmental art design is a systematic design project that includes multiple aspects, such as architectural design, garden design, urban community design, street design, public art design, public sculpture design, etc. It is a new discipline that includes art display and form display [4-6]. Environmental art design is a comprehensive art design process, to a certain extent, it is a holistic design that coordinates the relationship between humans and nature, and humans and society. This comprehensive design not only coordinates people's lifestyles, but also beautifies the natural environment, to a certain extent meeting people's spiritual needs and the need for harmonious coexistence between humans and nature. Modern interactive display design is a creative design activity that uses certain visual communication methods and exhibition facilities in the display space environment to display certain information and content in front of the public, and thus has a significant impact on the audience's psychology,

*College of Art and Design, Zhejiang Guangsha Vocational and Technical University of Construction, Jinhua, Dongyang, 322100, China.
†School of Economics and Management, Shangqiu Normal University, Shangqiu, Henan, 476000, China (Corresponding author, melody_wzc@163.com)

thoughts, and behavior. The author aims to explore how to effectively integrate information technology in environmental art design, construct innovative interactive display systems, and provide new ideas and practical guidance for the future development of environmental art design [7].

2. Literature Review. The fusion of imaging spectral data and pseudo color images in interactive display design of environmental art is one of the intuitive and effective ways to apply spectral data. The fusion processing provides pseudo color images that are easy for human eyes to observe and recognize. Spectral data fusion display first requires spectral data dimensionality reduction, and then the dimensionality reduction data is fused into a pseudo color image. The traditional dimensionality reduction data fusion method assigns the three images with the most information after dimensionality reduction to the RGB display model, which does not fully consider the characteristics of the dimensionality reduction data and the human eye's ability to distinguish different colors. Among all color spaces, the visual and psychological perception of color space and human eyes are relatively consistent [8]. Jiecheng, W. and others mainly analyze and explore the application of new media art in environmental art design to help us better understand the connotation of new media art [9]. Sang, Y. et al. introduced an algorithm called "Using Fuzzy Environmental Art Design" (EADF) to evaluate environmental standards and make better decisions. Initially, a fuzzy sequential preference technique similar to the Ideal Solution (FTOPSIS) was adopted in the design process, which considered multiple variables such as visual appeal, environmental impact, sustainability, and community audience engagement. Environmental art designers use the fuzzy TOPSIS method to evaluate artworks through multiple criteria [10]. Zhang et al. applied intelligent computer technology to environmental art design and combined it with intelligent computer technology to provide mathematical analysis tools required for precomputing light energy transmission technology. In addition, an environmental art design simulation system has been constructed. It is not difficult to see from the design evaluation results that the environmental art design system can effectively improve the effectiveness of environmental art design [11].

The author proposes a spectral data color fusion display algorithm based on multi interval translation mapping evaluation and optimization method to address the shortcomings of traditional pseudo color image display in environmental art interactive display design, which easily leads to mixed and difficult to distinguish targets. PCT dimension reduction data is allocated to the color space, and color data is mapped by inter partition translation for synchronous image evaluation. Finally, a comprehensive evaluation is conducted on all evaluation values, and the translation with the highest comprehensive evaluation is selected as the optimal translation [12]. The three optimal translation values are used to translate and map the color data of each channel, and the fused image with the best comprehensive energy, information, and clarity is obtained.

3. Method.

3.1. Principle of Hyperspectral Fusion Display. The spectral data cube undergoes redundancy reduction, noise reduction, and dimensionality reduction to obtain the dimensionality reduction data that best reflects the spectral features. It is then assigned to the corresponding color space and finally converted to the RGB space for storage and display. The principle is shown in Figure 3.1. The commonly used spectral data dimensionality reduction methods include band selection method, PCT method, independent component analysis (ICA), etc. Among them, PCT method has rigorous theoretical derivation, no iterative operation, and has received widespread attention. The author used PCT to achieve dimensionality reduction of spectral data.

The conversion of dimensionality reduction data to color space is the key to pseudo color display, which is related to whether the fusion result is consistent with human visual perception and whether the human eye can perceive various targets in the fusion result. The first three components of PCT are unrelated, with energy decreasing from strong to weak; The black and white channels O1, red and green channels O2, and yellow and blue channels O3 in the color space are also uncorrelated with the human eye, and their energy decreases from strong to weak. The first three components of PCT have similar characteristics to the color space of the human eye, and the three principal components of PCT can be assigned to 01, 02, and 03 [13,14]. It is not difficult to derive the conversion relationship from color space O1O2O3 to sRGB space based on the conversion

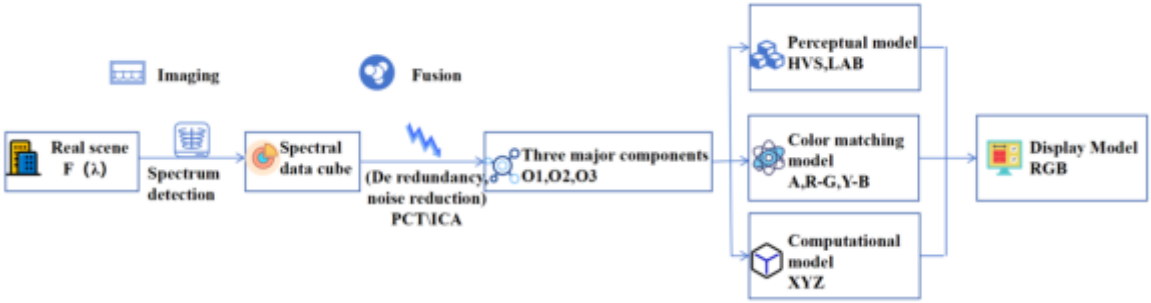


Fig. 3.1: Spectral imaging and fusion display process

relationship from color space to RGB image

$$[R_s G_s B_s] = \begin{bmatrix} 1.8063 & -5.8827 & -0.5766 \\ 0.6472 & 2.8929 & 0.1928 \\ 0.4103 & -0.0706 & 1.2418 \end{bmatrix} \begin{bmatrix} 01 \\ 02 \\ 03 \end{bmatrix} \tag{3.1}$$

Finally, a nonlinear transformation is performed to map the standard sRGB color space to an 8-bit RGB space, resulting in the final display result.

If $R_s, G_s, B_s \leqslant 0.00304$,

$$\begin{cases} R'_s = 12.92 \times R_s \\ G'_s = 12.92 \times G_s \\ B'_s = 12.92 \times B_s \end{cases} \tag{3.2}$$

If $R_s, G_s, B_s > 0.00304$,

$$\begin{cases} R'_s = 1.055 \times R_s^{1.0/2.4} - 0.055 \\ G'_s = 1.055 \times G_s^{1.0/2.4} - 0.055 \\ B'_s = 1.055 \times B_s^{1.0/2.4} - 0.055 \end{cases} \tag{3.3}$$

Finally, the non-linear $SR'G'B'$ value is converted into a digital encoded value according to the following equation:

$$\begin{cases} R_{8bit} = 255.0 \times R'_s \\ G_{8bit} = 255.0 \times G'_s \\ B_{8bit} = 255.0 \times B'_s \end{cases} \tag{3.4}$$

It is worth noting that in the above mapping process, only sRGB values within the [0,1] interval were captured, and values less than 0 and greater than 1 were not processed. This simple truncation mapping method sometimes cannot achieve the best visual display effect [15].

3.2. Multi information fusion translation mapping method. In order to map all values within the sRGB numerical range, a multi interval translation mapping method is used to map the sRGB standard color space to the RGB display space. That is, only a part of the sRGB numerical range is translated to the [0,1] interval each time, and the image quality after each translation mapping is comprehensively evaluated. Finally, the optimal interval of the image is selected for mapping. After transforming the color space 010203 to sRGB space, the minimum value of Rs is denoted as R_{min}, and the maximum value is denoted as R_{max} [16,17]. The

translation formula is

$$\begin{cases} R_n = (R - R_{min}) - \Delta R \times i \\ i = 0 \sim N_R \\ M_R = [(R_{max} - R_{min})/\Delta R] + 1 \end{cases} \tag{3.5}$$

In the formula: R_n is the translated value; ΔR is the translation step size; N_R is the total number of translation steps; i is the step count value.

Record translation amount:

$$M = R_{min} + \Delta R \times i \tag{3.6}$$

Then there are:

$$R_n = R - M \tag{3.7}$$

When the step count value i is 0, R takes the value $[R_{min}, R_{min} + 1]$, corresponding to R_n taking the value [0,1]. Substitute R_s into equations 3.2 and 3.4 for nonlinear transformation mapping, which is equivalent to translating the values within the $[R_{min}, R_{min} + 1]$ range; When i is 1, the range of R values $[R_{min} + \Delta R, R_{min} + \delta R + 1]$ is mapped by translation; When i is 2, the range of R values $[R_{min} + 2\Delta R, R_{min} + 2\Delta R + 1]$ is mapped by translation; Translating all values within the R range sequentially will result in segmented translation mapping with δR as the translation step size [18]. After each translation mapping, a fused image will be obtained. After R_s translation, N_R images will be obtained. Similarly, after G_s, B_s translation, N_G, N_B images will be obtained. In total, $N_R \times N_G \times N_b$ pseudo color images will be obtained. If manual discrimination of image quality is used, it will require a lot of work. The author objectively evaluates each pseudo color image and finally selects the image with the highest evaluation value as the final fused result image [19].

3.3. Comprehensive evaluation of fused images. At present, there are many objective quality evaluation methods for fused images, including entropy based on image information, cross entropy, correlation entropy, joint entropy, mean, standard deviation, variance, and covariance based on image energy distribution, signal-to-noise ratio and peak signal-to-noise ratio based on image signal quality, and various evaluation methods based on clarity, average gradient, and spatial frequency. These individual evaluation methods generally focus on one aspect and do not have a comprehensive effect; Some evaluation methods are difficult to operate and are not suitable for evaluating fused images; Some evaluation results are inconsistent with the visual effects of the human eye.

The author proposes a comprehensive evaluation method of "evaluation, scoring, and synthesis". After each translation mapping, the standard deviation V_{std}, entropy value V_{ent}, and average gradient value V_{avg} are calculated separately. After all translation mappings are completed, the individual evaluation values are scored, denoted as:

$$y = E(x) \tag{3.8}$$

The range of evaluation values for different types varies, making it difficult to comprehensively evaluate. Therefore, it is necessary to "score" again. The author uses a sorting scoring method, where the minimum score for evaluation value is 1 and the maximum score is the number of evaluation values. Finally, conduct a comprehensive evaluation:

$$E = E(V_{std}) + E(V_{ent}) + E(V_{avg}) \tag{3.9}$$

Table 3.1 is a schematic table of the comprehensive evaluation method used. The comprehensive evaluation value E with a translation M of 1.0 in the table is the maximum value of 13. By substituting the translation M at this time into equation 3.7, the optimal fusion mapping result graph can be obtained [20].

Table 3.1: Schematic Table of Comprehensive Evaluation

Serial Number	Transla--tion	standard deviation	entropy	average gradient	Standard deviation score	Entropy fraction	Gradient score	Comprehensive evaluation value
0	-2.0	80.4	0.70	30.4	1	2	1	4
1	-1.0	82.0	0.80	46.0	2	3	2	6
2	0.0	87.3	0.81	47.1	4	5	3	14
3	1.0	85.0	0.82	47.4	5	4	4	12
4	2.0	87.0	0.64	48.7	3	1	5	8

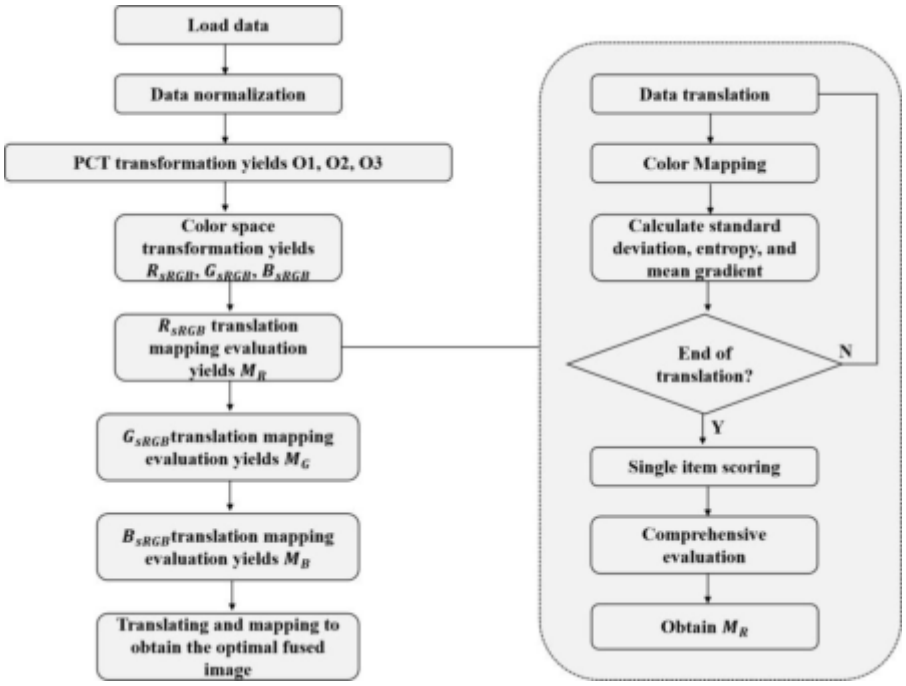


Fig. 3.2: Algorithm flowchart

3.4. Algorithm Implementation. The algorithm implementation process is shown in Figure 3.2. The specific implementation process is as follows:

1) Initialization: Set the translation step size $\Delta R = 0.5$, load hyperspectral data, and normalize it;
2) Spectral data dimensionality reduction: Perform principal component transformation on hyperspectral data to obtain the first three principal components after transformation;
3) Spatial variation: Assign the first principal component after PCT to the black and white channel 01 in the color space, assign the second principal component to the red and green channel 02 in the color space, and assign the third principal component to the yellow and blue channel 03 in the color space. Transform the color space data into sRGB space using equation 3.1 to obtain R_s, G_s, B_s;
4) Evaluate the translation mapping of the red component and obtain the optimal translation amount M_R for the red component; Use equation 3.5 to translate R_s, and after each translation, use equations 3.2 and 3.4 for mapping, and perform single item evaluations of standard deviation, entropy, and average gradient; After the translation is completed, use equations 3.8 and 3.9 for individual scoring and comprehensive evaluation, and obtain the red component translation amount M_R when the image quality is optimal through the optimal evaluation value E_R;
5) Green component translation mapping evaluation, obtaining the optimal translation amount for the green

Table 4.1: Performance Comparison of Four Mapping Methods

Mapping display method	R standard deviation	R entropy	R gradient	G standard deviation	G entropy	G gradient	B standard deviation	B entropy	B gradient
T. A. 2000	24.535	6.435	10.163	45.722	7.036	14.341	18.045	6.043	8. 875
J. S. T. 2003	47.432	7. 102	15.437	40.216	7. 002	16.166	30.071	6.627	10.702
Simple excerpt	92.027	1.027	16. 065	79.038	5.763	33.801	56.100	1. 601	4.8464
Author's Algorithm	113. 145	4.086	44.743	106.881	4.528	43.560	73.064	6.676	35.274

component, using the same method as step 4);

6) Evaluate the blue component translation mapping to obtain the optimal translation amount M_B for the blue component, using the same method as step 4);
7) Optimal image fusion mapping: Substitute M_R, M_G, and M_B into equation 3.7 to obtain translated R_s, G_s, and B_s; map the optimal fusion image using equations 3.2 and 3.4.

3.5. Experimental verification. The experimental equipment adopts a self-developed visible light imaging spectrometer, with a band range of $0.4\mu m \sim 0.9\mu m$, spatial resolution of 491 pixels (h) × 674 pixels (w), and a narrowband count of 61. The experimental subjects are 6 different $500mm \times 500mm$ camouflage target fabrics, with clear weather and an experimental distance of 100m; The collected spectral data cube is normalized and transformed by PCT.

4. Results and Discussion. In order to verify the author's fusion display algorithm, the author chose to compare it with Tirance Achalakul's color space mapping method proposed in 2000, J. Scott Tyo's HSV mapping method proposed in 2003, and a simple [0,1] interval mapping method. Table 4.1 is a comparison table of evaluation values for camouflage target plates.

A.2000 and J S. The two mapping methods of T.2003 have a soft overall color tone and can display the information content of all scenes in a summarized manner, but the contrast of the image is low and the details of the camouflage target board are difficult for the human eye to distinguish; Simple cropping and this algorithm have high image contrast and strong contrast between light and dark, reflecting the characteristics of the cropping mapping method; The simple truncation method only directly maps the [0,1] interval, but the information content of this part is not optimal and loses the details of the camouflage target; The algorithm used by the author traverses all possible values, selects the best among them, and the resulting image best reflects the details of the image. From the evaluation values of the camouflage target board in Table 4.1, it can be seen that compared to other methods, the evaluation values of this algorithm are significantly superior except for a slight decrease in entropy, indicating that the contrast and clarity of this algorithm are significantly better than other methods. T.Although the information content of the image is reduced, the human eye is more likely to distinguish the details of the camouflage target, which is consistent with the visual recognition effect of the human eye.

Table 4.2 lists the runtime models and testing times for four different schemes, where T_1 represents the PCA runtime; T_2 represents spatial transformation time; T_3 represents the digital image transformation time; T_4 represents the color space transformation time; T_5 represents the time of data interception; T_6 represents the translation mapping time. stay i5-3470@3.20GHz On a quad core CPU computer, the MatlabR2007a software platform was used to test the running time of four methods. As shown in Table 4.2, the running time of the first three methods is almost the same. Although this algorithm has a slightly longer time overhead, it is still far less than the acquisition time of one spectral data cube, which is 6.63s. It is suitable for real-time detection applications of hyperspectral data.

Table 4.3 lists the comprehensive evaluation values during the program running process. It can be seen from Table 4.3 that the comprehensive evaluation value is highest when the R channel translation step is 7, and the optimal translation amount for mapping the image is -4.1021.

(5) The ΔR translation step size of the equation affects the final mapping result. It is worth noting that when the translation step size is 1, there is no overlap in the translation intervals. When the translation step

Table 4.2: Comparison of Running Times for Four Mapping Methods

Mapping display method	Calculate time model	run time/s
T. A. 2000	T1+T2+T3	1.021
J. S. T. 2003	T1+T2+T3	1.01
Simple excerpt	T1+T4+T5+T3	1.037
Author's Algorithm	T1+T4+T6+T3	1.608

Table 4.3: Schematic Table of Comprehensive Evaluation of R Channel Part

Serial Number	Transla--tion	standard deviation	entropy	average gradient	Standard deviation score	Entropy fraction	Gradient score	Comprehensive evaluation value
4.0	-5.501	41.665	1.040	16.022	16.0	20.0	17.0	54.0
5.0	-5.001	77.563	3.040	34.407	18.0	26.0	26.0	72.0
6.0	-4.501	104.033	4. 213	45.853	24.0	31.0	31.0	84.0
7.0	-4.001	113.145	4.086	44.743	31.0	28.0	28.0	87.0
8.0	-3.501	112.220	3.018	37.031	28.0	27.0	27.0	84.0
9.0	-3.001	109.407	2.015	28.251	27.0	25.0	25.0	80.0
10.0	-2.501	107.063	1.382	22.604	23.0	24.0	24.0	76.0
...	...	...	...	...	...	...	...	...

Table 4.4: Performance Comparison of Different Translation Steps

Translation step size	thorough--fare	Serial Number	Transla--tion	standard deviation	entropy	average gradient
	R	4.0	-3.501	112.220	3.018	37.031
1	G	6.0	-0.301	106.881	4.528	43.560
	B	4.0	-0.386	68.303	5.568	17.252
0.8	R	4.0	-4.301	109.512	4.266	46.406
	G	7.0	-0.701	99.321	3.367	40.200
	B	5.0	-0.386	68.303	5.568	17.205
0.5	R	7.0	-4.001	113.145	4.086	44.743
	G	12.0	-0.301	106.881	4.528	43.561
	B	7.0	-0.886	73.064	6.676	35.274
	R	11.0	-4.201	111.222	4.232	46.207
0.3	G	20.0	-0.301	106.881	4.528	43.560
	B	12.0	-0.786	72.057	6.761	31.747

size is less than 1, there is overlap in the translation intervals. The smaller the translation step size, the more overlapping areas there are. However, this overlapping translation is indeed beneficial, because when converting the value of to RGB [0.255], grayscale values close to 0 and 255 are insensitive to human eye discrimination and recognition. When the translation step is 1, some optimal images will be lost. In theory, the smaller the translation step, the better, but this will cause a sharp increase in computational complexity. Table 4.4 presents a performance comparison table for different translation steps, and the numerical evaluation results are basically consistent with the human eye evaluation results.

5. Conclusion. The author proposes a research on an interactive display system that integrates information technology in environmental art design. Spectral data fusion and color display are key factors for real-time application of hyperspectral imaging sensors in environmental art interactive display design. The spectral data color fusion display algorithm proposed by the author ensures that the final mapped image quality is optimal within a selectable range through segmented translation, mapping, evaluation, and optimization methods. Not

only does it avoid the disadvantage of directly capturing the 0-1 part of color data, but it also avoids the problem that digital mapping can easily result in low-energy targets less than 0 and high-energy targets greater than 1 being unable to be distinguished and displayed in the fused image; It also solves the problem of overall compression of color data to the 0-1 range, resulting in low contrast in digital mapping images and difficulty in distinguishing details with the human eye. The experimental results show that the algorithm can significantly improve the energy and clarity of the fused image, with high contrast and clear distinction of details, which is conducive to rapid recognition and judgment by the human eye. The algorithm principle is simple and easy to implement, and has broad application prospects in real-time reconnaissance of hyperspectral imaging sensors.

REFERENCES

[1] Yu, Y., & Yu, D. (2023). Research on the visual media design system of brands integrating new media technology in the attention economy. Proceedings of the 2nd International Conference on Information, Control and Automation, ICICA 2022, December 2-4, 2022, Chongqing, China.
[2] Li, B., & Lu, W. (2024). Application of image processing technology in the digital media era in the design of integrated materials painting in installation art. Multimedia Tools and Applications, 83(18), 54211-54228.
[3] Derkach, S., Melnyk, M., Fisher, V., Krypchuk, M., & Sovhyra, T. (2023). Video mapping technologies as spatial augmented reality in the artistic process. Preservation, Digital Technology And Culture, 52(2), 59-68.
[4] Chen, Y. (2023). The path and value of integrating ecological aesthetic education into the teaching of environmental art design major. Journal of Contemporary Educational Research, 7(11), 190-196.
[5] Liao, X. (2024). Protecting the environment through art: the significance of environmental art design in raising awareness about environmental pollution among universities in china. Current Psychology, 43(32), 26548-26570.
[6] Sullivan, A., Gianotti, A. S., Scollins, A., Tornatore, L., Ge, B., & Briones, M. (2024). Undergraduate experiences with sustainability courses: insights for diversifying sustainability education. Journal of Environmental Studies and Sciences, 14(3), 548-567.
[7] He, J., & Wu, Y. (2022). Application of digital interactive display design based on computer technology in vr film. Mobile information systems, 2022(Pt.31), 1-7.
[8] Zheng, Y. (2023). The generalized dice similarity measures for comprehensive evaluation of graphic design effects based on color psychology with t-spherical fuzzy sets. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 45(4), 6413-6427.
[9] Jiecheng, W. (2023). Analyze the application of new media technology in environmental art design. Procedia Computer Science, 228, 907-913.
[10] Sang, Y. (2024). Application in environmental art design practice based on a fuzzy evaluation system. Scientific Reports, 14(1).
[11] Zhang, H. (2024). Intelligent computer technology and its application in environmental art design. International Journal of Information and Communication Technology, 24(2), 213-227.
[12] Grzelec, M. (2024). Application of attenuated total reflectance—fourier transform infrared spectroscopy—(atr-ftir) and principal component analysis (pca) in identification of copying pencils on different paper substrates. Heritage Science, 12(1), 1-16.
[13] Shang, H., Li, P., & Peng, X. (2024). Enhancing the quality of low-light printed circuit board images through hue, saturation, and value channel processing and improved multi-scale retinex. Journal of Computer and Communications, 12(1), 1-10.
[14] Gan, Y., Zhang, H., & Yi, C. (2024). Multiscale fusion transformer network for hyperspectral image classification. JOURNAL OF BEIJING INSTITUTE OF TECHNOLOGY, 33(3), 255.
[15] Falahatnejad, S., & Karami, A. (2023). Deep fusion of hyperspectral and lidar images using attention-based cnn. SN Computer Science, 4(1), 1-11.
[16] Tian, X., Zhang, W., Yu, D., & Ma, J. (2023). Sparse tensor prior for hyperspectral,multispectral,and panchromatic image fusion, 10(1), 284-286.
[17] Zhang, Y., Zhang, J., Xu, S., & Yang, L. (2024). Research on fracture prediction method based on multi-source information fusion. Journal of Geoscience and Environment Protection, 12(6), 14.
[18] Tange, R. (2023). A combinatorial translation principle and diagram combinatorics for the general linear group. Transformation groups, 28(4), 1687-1719.
[19] Han, Y., Lu, G., Zhang, S., Zhang, L., Zou, C., & Wen, G. (2024). A temporal knowledge graph embedding model based on variable translation. Tsinghua Science and Technology, 29(5), 1554-1565.
[20] Hung, B. T., & Thu, N. H. M. (2024). Novelty fused image and text models based on deep neural network and transformer for multimodal sentiment analysis. Multimedia Tools and Applications, 83(25), 66263-66281.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Sep 13, 2024
Accepted: Oct 21, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2078–2086, DOI 10.12694/scpe.v26i5.4949

INTELLIGENT ALGORITHMS FOR COLLEGE PHYSICAL EDUCATION ATHLETE TRAINING USING COMPUTER BIG DATA TECHNOLOGY

HANYANG CUI* AND XINYU YANG†

Abstract. In order to solve the problems of long training content mastery time, poor training effect, and high cost in traditional training systems, the author proposes the use of computer big data technology in the research of intelligent algorithms for athlete training in university physical education. The author begins by gathering sports data from athletes, then utilizes a virtual reality perception interaction model to feed the data into the virtual environment generation unit. Data fusion is performed under the supervision of the simulation management module. Finally, combined with the motion behavior interpretation in the database and image card of the sports simulation training data unit, simulation 3D modeling is carried out in the 3D model processor according to user settings. The experimental results demonstrated that athletes trained with this system (the third group) had a markedly better understanding of the training content compared to those trained with the other two systems (the first and second groups) after the first week, achieving a mastery rate of 73.875%. As the duration of the training increased, all groups showed improved mastery of the content. This system enhances athletes' training effectiveness while also reducing training costs, offering significant practical application value.

Key words: Virtual Reality Technology, Athletes, Simulation training, Sports data

1. Introduction. In recent years, the continuous development and advancement of computer technology have led to its widespread application across various aspects of production and daily life. It has become a crucial tool in supporting both work and productivity, offering numerous conveniences that enhance the quality of life and increase efficiency in various fields [1]. The continuous upgrading and innovation of computer technology have significantly improved the level of computer applications, gradually transforming from simple computer system applications to composite business systems that are widely used [2]. Computer technology has a long-standing history of application in sports competitions. Integrating it into sports training and events can enhance the efficiency of athletes' training and improve the fairness and precision of competitions. Based on this targeted analysis, the application of computer technology in sports training and competitions promotes the optimization of computer technology applications, which is of great significance for sports training and competitions, and also plays an important role in improving the quality of college students [3,4].

Sports training is a long-term and tedious activity, in which the application of computers can enhance the fun and diversity of training [5]. Computers have the characteristics of real-time, novelty, sharing, and random input, and have high repeatability, which can ensure that athletes can focus on training or competition for a long time and promote their performance improvement [6]. When teenagers engage in sports learning and exercise, they are influenced by factors such as their age and physical and mental development patterns, resulting in a higher probability of interest transfer and change, which can seriously affect the effectiveness of sports training [7]. Reasonably applying computer related technologies to sports training and teaching can ensure that students' nervous system excitability is fully stimulated, promote a more relaxed and interesting learning and training atmosphere, help enhance students' interest and focus, and thus promote their better mastery of sports related skills, which can help transform the boring problems existing in traditional teaching.

Modern information technology and multimedia technology have unique and outstanding functions, especially in terms of diversity and fun, which are of great significance. They can stimulate students' interest in learning, enhance their focus, promote learning effectiveness, and help solve problems that ordinary teaching

*College of Education, Polytechnic University of the Philippines, Manila 1000, Philippines (Corresponding author, cuihanyang1122@163.com)
†School of Physical Education, Emilio Aguinaldo Collage, Manila 1000, Philippines.

methods cannot solve. From the current situation of competitive sports, to some extent, competition in competitive sports belongs to technological competition, and computers are a key factor affecting competition. The application of computer technology in sports training and competitions can play an important role. If advanced computer technology cannot be effectively applied to assist and support in the current sports field, the training level will not be significantly improved, and athletes will have difficulty breaking through. The author aims to explore the combined application of computer big data technology and intelligent algorithms in athlete training in university physical education.

2. Literature Review. Training simulation, also known as simulation training, refers to the use of modern scientific and technological methods to construct virtual scenes or implement training under certain special conditions. It is characterized by a high level of realism, strong focus, enhanced safety, and improved training efficiency.

Simulation training is a key approach in athlete preparation, typically involving scenarios like real-life simulations, obstacle navigation, and athlete condition modeling. This method is used to enhance athletes' adaptability and responsiveness, with the goal of helping them adjust quickly to competition environments, maintain peak performance during events, and achieve better results [8]. Xiao, S. et al. plan to integrate rock climbing training with the traditional university sports curriculum, utilizing collaborative teaching methods to innovate and enhance the university sports system. By conducting a rock climbing teaching experiment with students at School A, and analyzing the data through literature review, surveys, and statistical analysis, they aim to evaluate the practical impact of incorporating physical rock climbing into college physical education programs [9]. Die, H. et al. optimized sports management based on clustering algorithms in the context of big data. They implemented all the functions of the algorithm using the popular Java language and demonstrated the correctness of the fuzzy clustering algorithm with some small examples [10]. Fu, C. et al. focus on developing a system for athlete selection and physical fitness evaluation using big data technology. The system is built around evaluation indicators, weight assignments for these indicators, and specific evaluation criteria. Following data mining principles, it incorporates the Apriori and FP algorithms, using physical data indexing and algorithmic frameworks to guide the process, the minimum support degree of association rules is determined to establish an association model between stereo energy testing level and individuals [11].

As an advanced product of modern Internet technology, virtual reality technology creates a virtual environment for users by constructing immersive virtual worlds. Building on this, the use of virtual reality in athlete simulation training has emerged as a key future trend in the development of simulation training systems for athletes. Therefore, designing computer big data technology in the research of intelligent algorithms for athlete training in university physical education enables diversified development of athlete virtual training.

3. Method.

3.1. Overall System Structure Design. The athlete simulation training system based on virtual reality technology consists of three parts: a motion simulation training input unit, a virtual environment generation unit, and a motion simulation training output unit. The detailed structure is illustrated in Figure 3.1. The input unit for sports simulation training consists of a position and direction sensing module, a conversion module, a data glove module, and a glove input control module, which primarily handle the collection of athlete motion data [12]. The virtual environment generation unit comprises a simulation management module, a user application module, a 3D model processing module, and a computer, which together handle the creation and management of the virtual environment. This unit serves as the central component of the athlete simulation training system utilizing virtual reality technology. The output unit, primarily consisting of an effects generator and a signal converter, is responsible for delivering the virtual training outcomes and facilitating interaction with users.

3.2. Virtual Reality Perception Interaction Model. The athlete simulation training system based on virtual reality technology can simulate the scenes during the athlete training process and record the athlete training movements and related data input by the motion simulation input unit. Through the analysis of the movement movements and related data, it can identify whether the athlete's movements during the training process meet the set standards, and also provide prompts for the athlete's movements that do not meet the standards during the training process, thereby improving the quality of athlete training [13,14]. To effectively

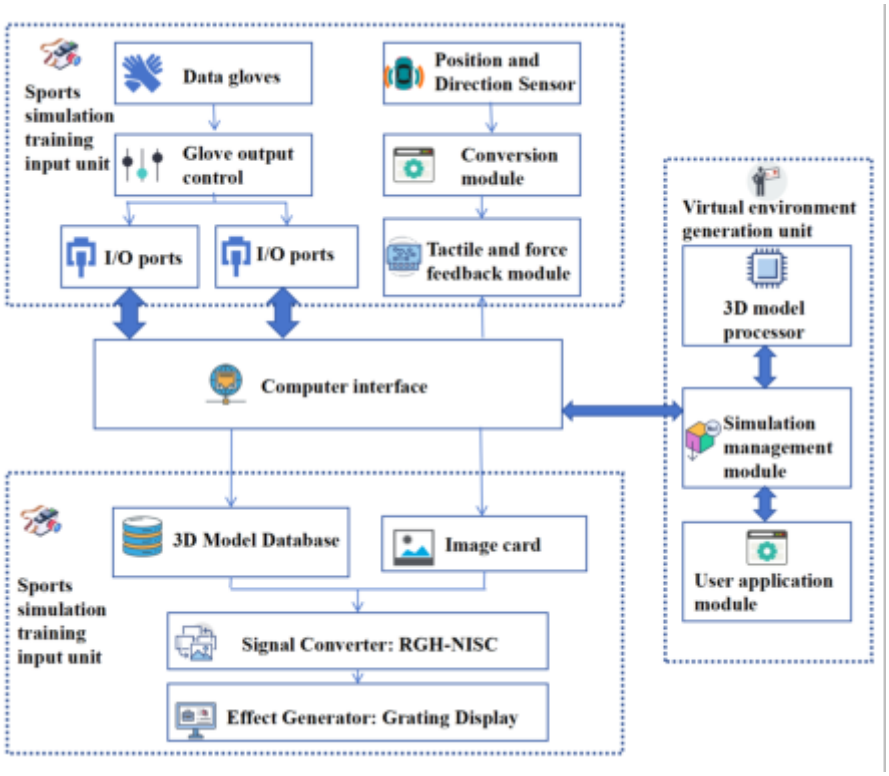


Fig. 3.1: Structure diagram of athlete simulation training system based on virtual reality technology

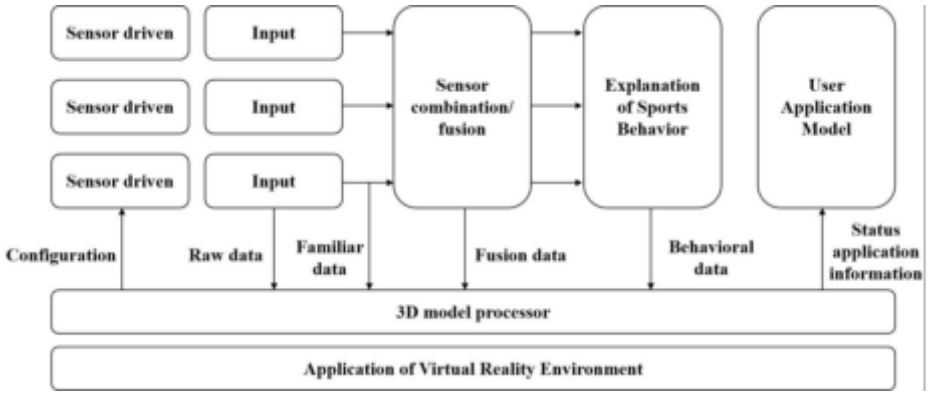


Fig. 3.2: Virtual Reality Perception Interaction Model

support data input and output functions, the athlete virtual training system must integrate with virtual reality technology. This integration ensures that athlete training movements are simulated with high fidelity, providing a realistic and dependable virtual environment for real-time sports training. Figure 3.2 illustrates the virtual reality perception interaction model within the simulation management module of the virtual environment generation unit.

The Simulated Management Module in the Virtual Environment Generating Unit is a 3-D Model Processor, which integrates the raw data of athlete sports training simulation collected by various sensors in the sports

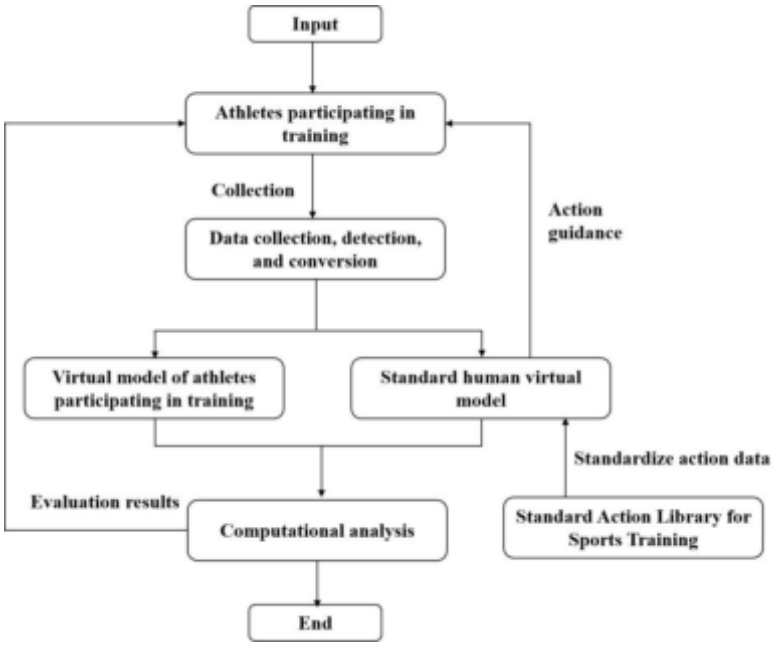


Fig. 3.3: System software operation process

simulation training input unit [15]. Combined with the interpretation of sports behavior, it simulates 3D modeling in the 3D model processor according to user settings, and finally displays the state application information on the display.

3.3. Three dimensional reconstruction of virtual athletes and construction of a standardized action library for sports. Design a virtual athlete 3D reconstruction software in an athlete simulation training system using the built-in scripting language Maxscript in 3D Studio MAX. The software aims to guide trained athletes in real-time and track the posture of their training parts. Throughout the athlete simulation training, virtual athletes are controlled according to standards to demonstrate training movements in real time. Concurrently, the user application module utilizes these virtual athletes to illustrate the movement trajectories of the participating athletes' training activities [16].

As a widely used 3D production software, 3D Studio MAX software includes various functions such as modeling, rendering, and animation roaming. In the 3D model processing module, the system calls internal programs to construct standard virtual humans based on standardized movements, and sets the time and angle of joint movements for a certain part to display guided standardized movements to participating athletes. The system collects data through functions such as Read Value to implement 3D reconstruction of intermediate document content, constructs a 3D virtual human of athletes participating in training, and real-time collects and reproduces the posture of the training parts of athletes participating in training in the real environment. In the design process using 3DStudio MAX software, a timer clock is added with a time interval of 20ms, which means that the intermediate document content is collected every 20ms to ensure that the sampling frequency of the same action position and direction tracker is the same, and to achieve the goal of real-time training of the posture of the trained parts of the athletes in the real environment. Athletes engaged in sports training follow real-time instructions based on standardized movements of virtual figures. The system uses a dialog box interface to offer feedback on their virtual training performance. Figure 3.3 depicts the software operation workflow of the athlete virtual training system [17,18].

To support athletes and coaches in tailoring training to various athletic scenarios and accessing specific training standards, it is essential to develop a virtual human sports movement library during the creation of a virtual reality-based athlete simulation training system. This library primarily consists of two components:

1. An athlete training record consisting of the names, ages, genders, training movements, and performance evaluations of athletes participating in sports training;
2. Including standardized movements for exercise training, such as movement number, type, position, and angle.

3.4. Simulation Application of Virtual Reality Technology. The advent of virtual reality technology has led to substantial advancements in athlete simulation training. Through technical analysis, the interactions between various positive feedback modules in virtual reality-based athlete simulation training are characterized by the following formula:

$$\frac{d \times L_{ev}}{t} = R_{\alpha} \tag{3.1}$$

In the formula, d and L_{ev} represent the time parameters of the simulation and the parameters of the virtual simulation, respectively; t and Ra indicate the number of simulated objects and the initial value of the simulation.

Optimize equation 3.1 using the differential equation shown in equation 3.2:

$$L_{ev} = L_{ev0} \cdot exp(c_0 \cdot t) \tag{3.2}$$

In the formula, L_{er0} and c_0 respectively represent the key parameters and scaling constants of the virtual simulation parameters.

In equation 3.2, the system state variables display dynamic variations, which can be used to compute relevant parameters for motion training simulations, providing crucial data for these simulations [19]. To meet the needs of virtual training simulations for athletes, the maximum value of feature roots max can be determined, as outlined in the following formula:

$$C = \frac{\lambda_{max} - n}{n - 1} \tag{3.3}$$

In the formula, C denotes the comparison matrix, and n represents the number of matrices used in the virtual simulation. Leveraging current experience with virtual training simulations, the comparison matrix helps quickly derive critical data related to virtual reality technology, leading to more precise data processing standards.

To meet the simulation needs of the athlete virtual training system, the consistency verification procedure must be applied to handle the simulation comparison matrix. The formula for this data verification process is given by:

$$C \cdot R = \frac{C \cdot I}{R \cdot I} \tag{3.4}$$

In the consistent verification, R and I. are the values of the matrix and the matrix parameters, respectively.

In virtual reality technology, the above process can be used to directly import important data of various sports training numbers, including athlete training plans and interdisciplinary results, during the recognition process of sports training data. Given these conditions, it is possible to determine the single-layer weight vector for sports training simulation. The details of this process are outlined in equation 3.5:

$$\overline{W_j} = \begin{bmatrix} W_1 \\ \vdots \\ W_N \end{bmatrix} j \tag{3.5}$$

In the weight vector identification process detailed in equation 3.5, $\overline{W}$ denotes the weight level, and j represents the parameter values of the single-layer fully connected weights. Following this weight vector identification requirement, the coach can utilize the system to assess the weight vector parameters for each layer. The formula for this evaluation process is as follows:

$$W_1 = \sum_{j=1}^{m} a_j b_{ij} \tag{3.6}$$

Table 4.1: Data Collection Error

	Lift your left arm up		Bending of waist	
sportsman number	Athlete's movements	System display action	Athlete's movements	System display action
12	45°00′	44°78′	32°45′	32°57′
19	44°21′	44°08′	30°06′	29°74′
26	41°44′	41°48′	42°27′	42°28′
38	17°46′	17°50′	24°22′	24°15′
54	57°70′	57°71′	53°24′	53°17′
60	70°03′	70°05′	42°01′	41°88′
66	45°78′	44°78′	63°03′	63°15′
71	96°26′	96°18′	19°06′	19°07′
77	80°25′	80°40′	77°36′	77°21′
89	74°04′	73°87′	15°25′	15°38′
105	22°60′	22°65′	95°30′	94°86′
113	19°30′	19°27′	82°00′	81°84′

$$W_2 = \sum_{j=1}^{m} a_j b_{2j} \tag{3.7}$$

$$W_3 = \sum_{j=1}^{m} a_j b_{nj} \tag{3.8}$$

In the formula, W signifies the actual vector values, while a and b denote the results from consistency tests and root analysis, respectively. By leveraging virtual reality technology, the performance of athlete simulation training systems can be optimized. This allows for the use of various simulation recognition and weight partitioning techniques to enhance the effectiveness of athlete training simulations.

3.5. Experimental verification. The experiment seeks to assess the real-world effectiveness of the athlete simulation training system developed using virtual reality technology. For this purpose, athletes from a specific school were chosen as subjects, with 120 participants from various sports training programs utilizing the system for their simulation training [20].

4. Results and Discussion. The author employs position and direction sensors to gather data on athlete movements. The error rates in data collection for this system are detailed in Table 4.1. Analysis of Table 4.1 reveals that the motion errors are generally limited to within 0° 20', demonstrating that the system effectively captures athletes' training data with high accuracy. This precision enhances the overall effectiveness of the sports training provided by the system.

To evaluate the training effectiveness for athletes, the mastery of training content was chosen as the benchmark. Out of 120 athletes with similar age and fitness levels, 15 were randomly assigned into three groups of five. The first group trained using a traditional artificial neural network-based system, the second group used a mechanical motion device-based system, and the third group utilized the newly designed system for comparison. Each group underwent eight weeks of training, and their progress was monitored and recorded weekly. The results, depicted in Figure 4.1, show that athletes using the new system (third group) demonstrated a notably better understanding of the training content after just one week, achieving a mastery rate of 73.75%. Over the eight weeks, all groups improved their content mastery, but the third group reached approximately 95% mastery by the end, compared to less than 80% for the other two groups. This indicates that the new system is more effective for athlete training and is better suited for enhancing training outcomes.

Table 4.2 presents the costs incurred by each group of athletes over the eight-week training period. During the first week, the cost of using the new system was reduced by 4000 yuan and 2000 yuan compared to the

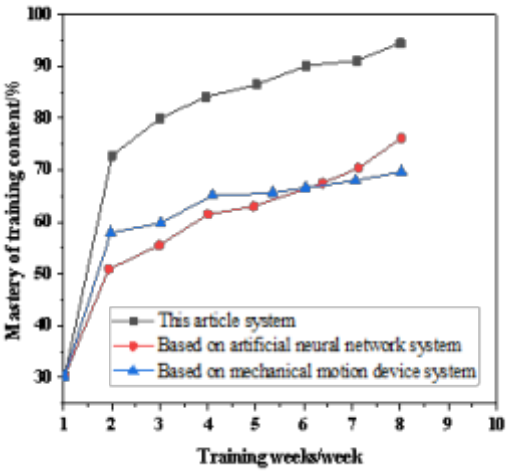


Fig. 4.1: Comparison results of mastery of training content

Table 4.2: Comparison Results of Training Costs (10000 yuan)

Training weeks/week	Traditional training system based on artificial neural network	Training system based on mechanical motion device	The system
1	1.3	1.1	0.8
2	1.5	2.3	0.8
3	1.7	3.5	0.8
4	2.1	4.7	0.8
5	2.1	6.1	0.8
6	2.3	7.1	0.8
7	2.5	8.3	0.8
8	2.7	9.5	0.8

traditional artificial neural network-based system and the mechanical motion device-based system, respectively. As the training continued, the cost of using the new system remained steady at 8000 yuan throughout the eight weeks. In contrast, the costs for the traditional and mechanical motion systems increased by 0.1 million yuan and 11,000 yuan per week, respectively. By the end of the eight weeks, the costs for the traditional and mechanical systems had risen to 27,000 yuan and 95,000 yuan, respectively. This demonstrates that the new system is more cost-effective and better suited for practical use in sports training.

Figure 4.2 illustrates the system's response speed during the comparative training effect test experiment. Analysis of this figure reveals that the response speed of the observation data calculation of the system basically meets the real-time requirements of virtual reality technology simulation analysis, which proves that the response speed of the system is guaranteed.

5. Conclusion. The author proposes that computer big data technology is an important component of future sports simulation applications in the field of intelligent algorithm research for athlete training in university physical education, as well as the design and development of athlete training simulation systems. The author's design leverages virtual reality technology to create an athlete simulation training system, with the goal of enhancing training effectiveness and lowering the costs associated with sports training. Due to time constraints, the performance testing content of the system is not comprehensive, and there are still certain defects within

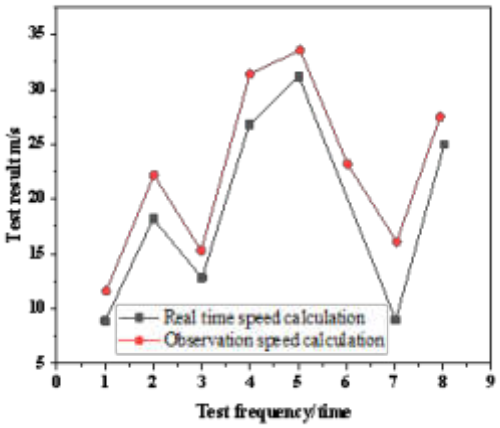


Fig. 4.2: System response speed test results

the system. These defects will be gradually optimized and resolved during the long-term testing process in the future.

REFERENCES

[1] Wu, M., Wang, R., & Wu, L. S. (2023). Invisible experience to real-time assessment in elite tennis athlete training: sport-specific movement classification based on wearable mems sensor data. Proceedings of the Institution of Mechanical Engineers, Part P. Journal of Sports Engineering and Technology, 237(4), 271-282.
[2] Zhang, Y., & Wei, W. (2023). Sports training correction based on 3d virtual image model. Mobile Networks and Applications, 28(5), 1687-1698.
[3] Luo, C., & Yang, H. (2024). Athlete training load monitoring using sensor-based technology and motion image analysis. Revista internacional de medicina y ciencias de la actividad fica y del deporte, 24(94), 322-339.
[4] Dan, J., Zheng, Y., & Hu, J. (2022). Research on sports training model based on intelligent data aggregation processing in internet of things. Cluster Computing, 25(1), 727-734.
[5] Zhang, Z., & Huang, X. (2024). Effectiveness assessment for the application of virtual reality technology in physical fitness training of engineering students. Revista internacional de medicina y ciencias de la actividad fica y del deporte, 24(94), 465-481.
[6] Luo, C., & Yang, H. (2024). Athlete training load monitoring using sensor-based technology and motion image analysis. Revista internacional de medicina y ciencias de la actividad fica y del deporte, 24(94), 322-339.
[7] Cheng, Y., Lu, H., & Guan, H. C. (2022). Monitoring simulation of athlete dynamic injury based on nosql database and localization algorithm. Mobile information systems, 2022(Pt.31), 1-10.
[8] Fang, L. (2024). The deployment of smart sharing stadium based on 5g and mobile edge computing. Wireless Networks, 30(5), 4121-4131.
[9] Xiao, S., Huang, H., & Wang, W. (2024). Research on the innovation system of physical climbing in college physical education under big data technology. Applied Mathematics and Nonlinear Sciences, 9(1).
[10] Die, H., & Lianhong, L. (2022). Research on the optimization of the physical education teaching mode based on cluster analysis under the background of big data. Scientific programming, 2022(Pt.11), 6340526.1-6340526.9.
[11] Fu, C. (2022). Physical fitness evaluation system for athlete selection based on big data technology. Int. J. Embed. Syst., 15, 249-258.
[12] Ma, Q., & Huo, P. (2022). Simulation analysis of sports training process optimisation based on motion biomechanical analysis. International Journal of Nanotechnology, 19(6/11), 999-1015.
[13] Jayanthi, N., Schley, S., Cumming, S. P., Myer, G. D., Saffel, H., & Hartwig, T., et al. (2022). Developmental training model for the sport specialized youth athlete: a dynamic strategy for individualizing load-response during maturation:. Sports Health: A Multidisciplinary Approach, 14(1), 142-153.
[14] Beaumont, P. L., Hoek, D. V. D., Holland, J., & Garrett, J. (2023). A scoping review of the physiological profiles of motorsport drivers: implications for athlete training. Strength and Conditioning Journal, 46(3), 257-278.
[15] Sansone, P., Rago, V., & Kellmann, M. A. P. E. (2023). Relationship between athlete-reported outcome measures and subsequent match performance in team sports: a systematic review. Journal of strength and conditioning research,

37(11), 2302-2313.
[16] Liu, Y., & Jing, H. (2022). A sports video behavior recognition using local spatiotemporal patterns. Mobile information systems, 2022(Pt.10), 1-10.
[17] Kechi, A. D., Lars, D., & John, K. (2023). Coaches' perceptions of factors driving training adaptation: an international survey. Sports medicine, 53(12), 2505-2512.
[18] Wu, M., Wang, R., & Wu, L. S. (2023). Invisible experience to real-time assessment in elite tennis athlete training: sport-specific movement classification based on wearable mems sensor data. Proceedings of the Institution of Mechanical Engineers, Part P. Journal of Sports Engineering and Technology, 237(4), 271-282.
[19] Zhang, Y., & Wei, W. (2023). Sports training correction based on 3d virtual image model. Mobile Networks and Applications, 28(5), 1687-1698.
[20] Zhu, C. (2022). Design of athlete's running information capture system in space-time domain based on virtual reality. Scientific programming, 2022(Pt.4), 1-10.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Sep 16, 2024
Accepted: Oct 21, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2087–2104, DOI 10.12694/scpe.v26i5.4639

OPTIMIZING EFFICIENTNETV2 MODEL WITH RANDAUGMENT DATA AUGMENTATION FOR DETECTING WHEAT DISEASES IN SMART FARMING

MANISHA SHARMA*, ALKA VERMA†, AND UMA RANI‡

Abstract. Wheat diseases threaten global food security, necessitating improved detection methods. In this paper, we integrate EfficientNetv2 model and RandAugment data augmentation to accurately and efficiently identify wheat diseases. EfficientNetv2, known for its optimal mix of accuracy and computing efficiency, is reinforced by RandAugment, a versatile data augmentation approach that randomly modifies training data. This augmentation method greatly enhances the model's generalisation and performance on new data. Our extensive experimentation reveals that this integrated technique improves model accuracy and robustness relative to baseline models. Proposed model gained the 96.73% accuracy on prescribed dataset. The results show that EfficientNetv2 and RandAugment can detect wheat illnesses on a large scale. This could change precision agriculture by enabling early and accurate disease management.

Key words: Deep Learning, Wheat Diseases, Detection, EfficientNetV2, Image Recognition, RandAugment

1. Introduction. Wheat is vital to global food security, yet various diseases endanger it. These diseases lower crop yields and quality, causing farmers and the food sector enormous problems. Understanding and treating these diseases is essential for wheat crop sustainability. Puccinia fungus causes wheat rust. The three most common rust species are Puccinia graminis, triticina, and striiformis. When conditions are right, many diseases spread quickly. Ug99 and other novel strains have panicked the world, and stem rust has destroyed crops. Powdery mildew, another major wheat disease, is caused by Blumeria graminis f. sp. tritici. This disease causes white, powdery fungal growth on wheat leaves, stems, and heads. Powdery mildew weakens plants, lowering photosynthesis and grain quality. Fungicides and resistant cultivars are often needed for management after resistance fails. Head blight (FHB) is another significant wheat disease caused by Fusarium species. FHB can deplete yields and contaminate grain with mycotoxin. Mycotoxin poisons humans and animals. The pathogen spreads quickly during wheat flowering because it likes warm, humid environments. Integrating management strategies is key to FHB control. Crop rotation, resistant cultivars, and timely fungal applications are essential. Wheat plants suffer from Septoria tritici blotch (STB) caused by Zymoseptoria tritici. The disease can reduce photosynthetic area and productivity. STB is difficult to control given to its genetic diversity and rising pesticide tolerance, even in many wheat-growing regions. Wheat can contract WSMV and BYDV. Insect-borne viruses limit growth, yellow leaves, and reduce grain production. Removal of disease-carrying insects and planting resistant crops are common disease management methods. Agronomy, environmental science, plant pathology, and genetics must work together to control wheat diseases. Wheat production and feeding a growing population depend on research on resistant cultivars, effective fungicides, and integrated pest management [1].

1.1. Problem Formulation. Temperature, humidity, soil moisture, light intensity, rainfall, and other environmental variables play a crucial role in smart farming disease detection for wheat. These factors impact plant health and disease development. For example, rust and powdery mildew are fungal infections that thrive at certain temperatures and humidity levels. Root rot and other soil-borne illnesses are affected by soil moisture levels, and signs of different diseases, such chlorosis or lesions, can be seen differently depending on the intensity

*Department of Electronics Communication and Engineering, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India (sharma9mani@gmail.com).

†Department of Electronics Communication and Engineering, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India (dralka.engineering@tmu.ac.in).

‡Department of Computer Science and Engineering, Dr. Akhilesh Das Gupta Institute of Professional Studies, Delhi, India (singh19uma@gmail.com).

of the light. Waterborne diseases can also spread due to changes in rainfall patterns. In order to better manage precision agricultural systems, it is important to keep an eye on these factors in addition to other contextual data such as crop type and development stage. This will allow for improved disease prediction and diagnostic models. Wheat gives the globe many calories, but diseases can diminish crop productivity and quality. Powdery mildew, rusts, and Fusarium head blight diminish wheat yield. Timely interventions and loss mitigation require precise sickness diagnosis and early identification. Traditional sickness detection methods like visual inspection and laboratory testing are laborious and error-prone. Thus, wheat disease prediction using deep learning and other cutting-edge technologies is growing. Deep learning models excel at image identification, making them excellent for plant photo disease diagnosis. These systems learn tiny patterns and features from massive databases to identify healthy and unhealthy plants. To maximize performance, deep learning models for wheat disease prediction need a big dataset, appropriate model topologies, and hyperparameter fine-tuning. Accurate forecasts require a sophisticated algorithm that can generalize across several settings and wheat kinds. A deep learning algorithm to detect wheat diseases starts with data collection. An exhaustive dataset should include photos of wheat plants in diverse settings, stages of development, and diseases. The dataset should include photos from multiple locations to make the model adaptable. Disease annotations are essential for supervised learning. The algorithm can learn to distinguish illnesses and improve predictions with well-labeled, high-quality data. CNNs recognize spatial hierarchies, making them suitable for image-based applications. ResNet, Inception, and EfficientNet have parameter efficiency, multi-scale feature extraction, and deeper skip-connected networks. To balance accuracy and computing efficiency, network depth and breadth must be changed [2-4].

Hyperparameter modification is key to deep learning model optimization. Learning rate, batch size, and epoch count greatly affect training duration and model accuracy. Bayesian optimization, grid search, and random search can explore hyperparameters. By using weight decay and dropout, the model can adapt to new data without overfitting. To optimize prediction, hyperparameters must be adjusted. The wheat disease prediction deep learning problem design requires data collection, model selection, and hyperparameter tweaking. An accurate, resilient, and generalizable optimized model requires many components. A novel deep learning disease detection system can improve wheat yield and food security. Wheat disease prediction manages production risk and stabilizes food supplies. Using machine learning, remote sensing, and pathogen tracking, predict wheat diseases [5]. These strategies provide more accurate and timely sickness prediction, enhancing treatment and intervention. Wheat disease prediction using machine learning is critical. Large database trends help these computers forecast illness transmission. Disease, weather, soil, and crop health data can teach neural networks, decision trees, and SVMs. This data lets models forecast sickness start in particular situations. Deep learning neural networks can handle complex, non-linear data.Wheat disease prediction requires remote sensing. This technology monitors crop health across wide areas using satellite or aerial photography. Remote sensing detects minor plant physiology changes before symptoms occur. Plant health and stress can be assessed with multispectral and hyperspectral imaging. Adding remote sensing data to machine learning models allows real-time, regionally precise crop condition predictions. Pathogen surveillance monitors viruses, fungi, and bacteria. This is possible using spore trapping, molecular diagnostics, and field surveys. New dangers and outbreaks can be found by monitoring pathogen populations and genetic changes. Temperature and humidity can improve disease surveillance predictions. Specific fungal spores in the air and correct weather can set off pandemic alarms. Wheat disease forecasting requires weather-based prediction models since many illnesses are climate-sensitive. Epidemiology Simulator (EPIDEM) simulates sickness progression using humidity, precipitation, and temperature. These models can warn farmers about weather. Agricultural growth models with weather data illustrate how plant development phases affect disease susceptibility [6-8].

Genomic approaches identify disease-resistant or vulnerable genetic markers to predict wheat illnesses. Genetic and marker-assisted selection breed disease-resistant wheat. Genetics helps researchers create disease-resistant crops. Genetic data enhances breeding and forecasts when paired with phenotypic and environmental data. Using IoT sensors in fields to capture real-time data on soil moisture, plant health, and environmental factors seems promising. For real-time sickness risk evaluations, these devices can send data to cloud platforms. Machine learning algorithms analyze data. A continual monitoring and prompt response to changing circumstances improves illness treatment. Combining weather predictions, genomic data, field surveys, and remote sensing requires big data analytics. Advanced analytics finds patterns and links that normal data analysis

cannot. Researchers enhance disease management and prediction with big data. To be useful, these methods must be integrated into user-friendly systems. Farmers may benefit from real-time data and prediction models in DSS. The systems can recommend fungicides, irrigation schedules, and other disease prevention strategies. Simple technologies are needed for general adoption and disease control. Last, wheat disease prediction requires machine learning, genetics, meteorology, big data analytics, IoT, decision support systems, and pathogen surveillance. Researchers can use these mechanisms to create disease prediction models to help farmers control disease risks and improve wheat yield and food security [9].

1.2. Research Contributions. This work has following research contributions as below:

- Integrated EfficientNetv2 with RandAugment to enhance wheat disease detection accuracy.
- Demonstrated superior performance of EfficientNetv2 with RandAugment over baseline model.
- Improved model generalization through diverse data augmentation techniques.
- Achieved efficient and scalable wheat disease detection suitable for large-scale applications.
- Contributed to precision agriculture by enabling early and accurate disease management.

This paper is designed with aiming of predicting wheat disease. Section 2 defines the various existing works carried out in this problem domain. Section 3 explains the datasets used for experimental purposes. In this paper, Wheat disease dataset is being taken. Section 3 demonstrated the modified EfficientNetv2 model. Next, section 4 illustrates the result, followed by the paper's conclusion in section 5.

2. Related Work. Using free remote sensing data, Pryzant et al. (2017) present an affordable, accurate, and scalable epidemic monitoring technique. Two ways our method beats the competition. Instead of remote sensing spectral characteristics, we use Convolutional and Long Short-Term Memory Network-generated automatically learned features. We merge data into broader regions. Our method outperforms others over nine years of agricultural output and is predictive. New agricultural disease surveillance methods may improve with time.

In 2020, WU et al. (2020) improved image processing with deep learning. Three wheat kinds, six backgrounds, and two image capturing methods with varied heights, angles, and grain numbers produced 1748 photos. All photo color spaces were rotated, flipped, and altered. Each dataset was divided into training, validation, and test sets after hand grain annotation. Faster Region-based Convolutional Neural Network Model was built with TensorFlow. Transfer learning improved wheat grain recognition and enumeration. Model precision averaged 0.91 and loss was less than 0.5. This model's grain counting error rate was under 3% and running time under 2 seconds, improving over previous methods. Image size, grain size, shooting angle, height, and grain crowding suit the model. It has wheat grain recognition and counting capabilities.

A one-shot learning-based wheat disease identification network was proposed by Mukhtar et al. (2021). A few photographs can teach this network new categories and types. Growers can test immediately after retraining the network with sick plant photos. MobileNetv3 extracts features quickly and accurately. The PlantVillage dataset fine-tuned this network. The final two thick layers were trained using CGIAR Crop Disease dataset and Google images plant photos of eleven wheat diseases. The One-shot network was trained with 440 photos, 40 each category. Siamese networks encode images. The absolute difference between encoded photos is calculated next. Similarity ratings use Sigmoid units. Similar photos get one point, while different ones get zero. Over 98% training and 96% validation for Mobilenetv3 and 92% accuracy, 84% precision, and 85 recall for the one-shot network. While traditional classification networks need retraining, our method simply needs a few image types.

Bukhari et al. (2021) realistically test Watershed, Grab Cut, and U2-Net segmentation methods. These methods divide the wheat stripe rust dataset into Watershed, GrabCut, and U2-Net. Segmentation's impact on classification accuracy is assessed using ResNet-18, a pre-trained deep learning model. Classification accuracy was best (96.196%) on U2-Net segmentation. To help researchers identify the optimum segmentation strategy, this study analyzes state-of-the-art techniques by correctness and classification accuracy. This study examines a neglected topic: segmentation and wheat stripe rust classification accuracy.

Haider et al. (2021) crowdsource agricultural experts, farmers, and growers. Next, data is processed to identify disorders. Farmers benefit from early crop disease diagnosis and control. Most academic disease management systems categorize agricultural illnesses using ML algorithms. Sadly, these systems cannot use static data because illnesses in different agricultural locations change constantly. The agricultural expert's experience is not considered while confirming classification results. We collected high-quality photos and text data from

farmers, domain experts, and users using a crowdsourcing platform to study wheat diseases' ever-changing nature. Augmentation improved training data. Modern general wheat disease diagnosis and classification utilizing Decision Trees (DT) and deep learning models is presented in this paper. Both algorithms improved decision tree accuracy by 28.5% and CNN accuracy by 4.3% (97.2% accuracy) and generated wheat disease knowledge-based decision rules after domain experts verified them.

Using UAV sensing, multispectral photography, vegetation segmentation, and deep learning, Su et al. (2021) monitor yellow rust infections U-Net. A DJI Matrice 100 drone and Red-Edge camera take multispectral aerial photos of winter wheat to test a yellow rust inoculum. Comparing high-resolution RGB images from the Parrot Anafi Drone reveals the calibrated and stitched multi-spectral orthomosaic for system evaluation. Drawing spectral-spatial information simultaneously outperforms a standard random forest-based spectral classifier. Three RGB bands and five spectral vegetation indices are tested using the wrapper algorithm's sequential forward selection.

Goyal et al. (2021) categorized wheat diseases differently. We classify 10 wheat illnesses using deep learning. A precise approach with 97.88% testing accuracy is suggested. It beats VGG16 and RESNET50 by 7.01% and 15.92%. The suggested method outperforms current methods in recall, f-score, and precision.

Deep learning is used to classify wheat varietal level (VLC) by Laabassi et al. (2021). Using grain photos, the Convolutional Neural Network recognized Simeto, Vitron, ARZ, and HD wheat cultivars. Five standard CNN architectures were trained using Transfer Learning to improve categorization. We compared the models using 31,606 single-grain photos from various Algerian locales captured using different scanners. The varietal level categorization accuracy was 85%–95.68%. The top three test accuracy rates were DensNet201 (95.68%), Inception V3 (95.62%), and MobileNet (95.49%). Therefore, the suggested approach delivers trustworthy and accurate results, making it worth attempting.

Visual wheat rust detection is the standard, however Ui Haq et al. (2022) believe it is inefficient and inadequate for broad agricultural regions. Experience and background determine farmers' monitoring reliability. Our AI-powered technology at the network's periphery classifies wheat leaves as healthy or corroded in real time. After assessing the dataset with multiple ML classifiers, Random Forest triumphed with 97.3% GLCM and 82.8% binary feature extraction accuracy. A Deep Convolution Neural Network (DCNN) model for rust and healthy leaf classification was 88.33% accurate after additional study. On the edge device, this trained DCNN model classifies wheat rust disease in real time. Wheat rot would be eliminated and technology favored over farming.

Chergui (2022) suggested data-augmentation to improve wheat output estimates using restricted data from two Algerian regions. We added features to each data set to provide dimension. Blending the sets increased their size. Three data sets—original, extra-featured, and merged—were tested. Support Vector Regression, Random Forest, Extreme Learning Machine, Artificial Neural Network, and Deep Neural Network were run to augment data. Our cross-validation showed that new data improved model performance.

Real-world plant pathologist datasets are compiled by Kundu et al. (2022). Deep learning is proposed to detect illnesses, predict severity, and estimate crop loss. K-Means clustering extracts the region of interest. A unique deep learning network called "MaizeNet" identifies diseases, forecasts severity, and calculates crop loss. Model accuracy peaks at 98.50%. Grad-CAM authors visualize features. The suggested model has an intuitive interface thanks to a web app. Plant pathology specialists benefit from the model's accuracy, limited number of parameters, fast training, and ability to extract critical information. Online application 'Maize-Disease-Detector' has copyright diary 17006/2021-CO/ SW.

Wang et al. (2022) used crop phenology, weather, and satellite images to create a machine learning model to predict European wheat mycotoxin risk. Deoxynivalenol, zearalenone, T-2, HT-2, fumonisins, aflatoxins, ochratoxin European wheat mycotoxin concentrations were monitored from 2010 to 2020. We linked this information to wheat phenology, weather, and satellite pictures by year and grid size (25 x 25 km). 80% of the 2010–2018 dataset was segregated from a 20% internal model validation set for training. For external validation, 2019 and 2020 data was used. Random forest (RF) was used on model training data. The model displays the low, medium, or high likelihood that wheat from a European grid has one of the six mycotoxins. The model did well in internal and external validation with 0.90-0.99 prediction accuracy. The Netherlands case study demonstrated satellite pictures improved model performance. The current strategy improves food security

and wheat-derived product safety by improving supply chain logistics and risk-based monitoring, including mycotoxin forecasts. Developing and improving models requires mycotoxin data with correct crop locations. Srivastav et al. (2023) developed an early detection model for leaf, stem, yellow, powdery, and septoria wheat crop fungal diseases. Disease may not spread to other plants with this model. The model was trained on 1972 Kaggle wheat fungal infestation images. The proposed model has 98.83% accuracy at epoch 12 after training and testing. First six epochs had different training and testing loss values, but as epoch values climbed, both phases' loss values declined. After numerous comparisons, the proposed model was most accurate. In 2024, Naik et al. (2024) used 18 CNN models to find lentils. Two-stage statistical study determined the optimum CNN model for Indian lentil recognition. The 18 CNN models are Alexnet, Darknet19, Darknet53, Densenet, EfficientNetB0, Google net, Inception, MobilenetV2, NasnetLarge, NasnetMobile, Resnet18, Resnet50, Resnet101, Squeezenet, Vgg16, Vgg19, and Two-stage statistical analysis used Wilcoxon signed-rank and Duncan's multiple range tests. Nine indicators—precision, sensitivity, accuracy, FPR, F1 Score, MCC, Kappa—were employed for statistical analysis. After 18 CNN models and two-stage statistical analysis, EfficientNetB0 identified lentils better than competitors. Table 2.1 demonstrates the summary of existing works.

2.1. Research gaps. The authors have revised the section to clearly identify and elaborate on the gaps in the existing literature, which motivate our study. Specifically, we have added the following points:

- There has been little investigation into using EfficientNetV2 in agricultural domains, especially for wheat disease detection, despite its potential in general computer vision applications. Previous research mostly used ResNet or Inception, two deep learning architectures that may not have taken full use of EfficientNetV2's enhanced performance and efficiency.
- The use of sophisticated data augmentation methods, such as RandAugment, is seldom ever discussed in the existing literature on wheat disease diagnosis. Models may be less resistant to changes in real-world data as many studies either employ conventional augmentation techniques or do not use any at all.
- It is not uncommon for some illnesses to be underrepresented in wheat disease databases. Data augmentation has the potential to increase model performance in all classes and even out the odds, although this has not been thoroughly investigated in the literature.
- Scalability, computing efficiency, and resilience to unforeseen data are some of the practical difficulties of deploying models for real-world settings, yet these issues have received little attention in the literature. Given this deficiency, it is clear that models such as EfficientNetV2 require a more thorough assessment of their accuracy and practicality for use in smart farming settings.
- Despite the fact that many different designs have been evaluated for wheat disease diagnosis, benchmarking comparisons utilizing cutting-edge architectures coupled with contemporary augmentation methods are still lacking. This disconnect makes it harder to get a complete picture of which approaches work best for this kind of work.

3. Material and Method.

3.1. Dataset. Wheat Plant Diseases Dataset is designed to empower researchers and developers in creating robust machine learning models for classifying various wheat plant diseases. It offers a collection of high-resolution images showcasing real-world wheat diseases without the use of artificial augmentation techniques [10]. There are 1,266 healthy and (Stip Rust, Septoria) diseased images in the dataset. This yielded 80% photos for training, 10% for validation, and 10% for testing. Data augmentation generates more photos to fit the model during training. Experimental results show that the suggested model detects Strip Rust and Septoria in wheat leaves. Experiments use VGG19, InceptionV3, MobileNet, and EfficientNet pretrained models. MobileNet is the best pretrained model and can categorize photos from a heterogeneous wheat farm with 90% accuracy. 902 healthy wheat leaves, 208 stripe rust disease, and 156 septoria-infected leaves are shown in Figure 1 and Figure 2. Fig 3.1 demonstrates a healthy leaf.

Fig 3.2 demonstrates various categories in this dataset.

To be compatible with EfficientNetV2, raw picture data was resized to 224x224 pixels. Normalizing the pixel values between 0 and 1 speeds up model convergence during training. Duplicate photos were eliminated to save data redundancy. We manually removed low-quality photos that were too blurry to detect illness.

Table 2.1: Review of existing works

Authors	Methods	Compared with	Dataset	Outcomes	Limitations
(Pryzant et al., 2017)	Convolutional and Long Short-Term Memory Networks	Convolutional Neural Network	Data into larger geospatial regions	AUC of 0.67	More expressive than traditional spectral indices
(WU et al., 2020)	Faster Region-based Convolutional Neural Network Model	Single Shot MultiBox Detector (SSD)	1748 images	Precision of 0.91	Many object features lost, and accuracy impaired
(Mukhtar et al., 2021)	MobileNetv3 network	Whole one-shot network	PlantVillage dataset	96% validation	Gives higher accuracy
(Bukhari et al., 2021)	ResNet-18 Model	U2-Net Model	Watershed segmented data	Accuracy of 96.19%	only been validated on the wheat stripe rust dataset
(Haider et al., 2021)	CNN model	Decision Trees	Wheat diseases image dataset	Accuracy of 97.2%	Less effective utilization of proposed approach for classifying wheat diseases.
(Su et al., 2021)	U-Net model CNN model	DJI Matrice	100 equipped with Red- Edge camera	Accuracy of 81.6%	Very noisy classification result
(Goyal et al., 2021)	RESNET50 model	VGG16 model	LWDCD2020 dataset	Accuracy of 97.88%	More computation requirement
(Laabassi et al., 2021)	DensNet201 model	MobileNet model	31,606 single-grain images collected	Accuracy of 95.68%	Less effective in classification
(Ui Haq et al., 2022)	Deep Convolution Neural Network (DCNN) model	Random Forest model	9232 diseased images	Accuracy of 88.33%	Required to improve accessibility and enhance data security
(Chergui, 2022)	Support Vector Regression	Random Forest model	Data sets of two distinct Provinces in Algeria	RMSE of 0.04 q/ha and R2 of 0.96	weather data taken at one are irrelevant
(Kundu et al., 2022)	K-Means clustering	Decision Trees	PlantVillage dataset	Accuracy of 94.60%	Not validated by the plant pathology experts
(Wang et al., 2022)	Random forest (RF) algorithm	Decision Trees	11 years of mycotoxin monitoring data (2010–2020)	0.90–0.99 prediction accuracy	more mycotoxin data with detailed locations of the cultivated crop are needed
(Zhang et al., 2022)	UNet model	SCA module	remote sensing dataset	Accuracy of 95.91%	requiring less labelled data is highly desirable
(Srivastav et al., 2023)	Convolutional Neural Networks (CNNs)	RCNN model	1972 images of wheat fungus diseases collected from Kaggle	Accuracy of 98.83%	lowest validation loss has been identified
(Naik et al., 2024)	Xception model	Resnet101, Squeezenet, Vgg16	Kaggle Wheat dataset	Accuracy of 94.94%	Less effective utilization of proposed approach

Data augmentation was used to strengthen the dataset and reduce overfitting. RandAugment was used as a fundamental augmentation approach to provide variety to the dataset by randomly rotating, brightness editing, and flipping. This balanced real-world image variants. Class distribution study also showed illness category imbalance. To guarantee balanced representation of all categories, training used minority class oversampling and weighted random sampling. These pre-processing processes enhanced data quality and diversity, improving smart farming wheat disease detection model generalization [16].

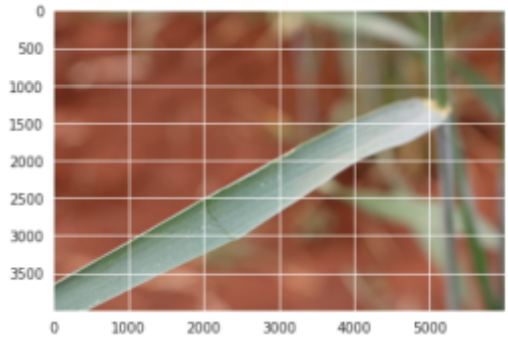


Fig. 3.1: Healthy Leaf

Fig. 3.2: Different categories of Wheat Leafs

3.2. Methods. In recent years, deep learning algorithms have helped farmers detect and predict wheat infections. These models have revolutionized sickness management. They automatically learn from data, handle complex patterns, and anticipate accurately. Early wheat crop disease diagnostic methods relied on arduous, error-prone manual examination. Due to deep learning, automation systems that can explore mountains of data for sickness patterns are a viable solution. One of the most prominent deep learning architectures for wheat disease prediction is CNNs. CNNs' image processing skills make them ideal for wheat disease diagnosis utilizing photographs of leaves, stems, or fields. The network trains on enormous datasets of labeled pictures to distinguish healthy and diseased crops. Early detection of wheat crop diseases like mildew, rust, and blight can be achieved using CNNs in real-time systems [17-20]. Other deep learning methods including LSTM networks and RNNs have been used to forecast wheat illnesses in addition to image analysis. These models can analyse time series, making them ideal for predicting disease spread using past data, weather, and other environmental factors. Farmers can predict disease outbreaks and take preventative measures using RNNs and LSTMs trained on this time series data. Multi-modal data is needed for deep learning wheat disease prediction. Besides the typical suspects, models can include image data, sensor data (soil moisture, temperature, and humidity), and genomic data (wheat kinds). Integrating diverse data sources improves the model's disease prediction accuracy. A model may estimate fungal infection onset using weather data, soil conditions, and early infection indications in pictures. Transfer learning advances deep learning for wheat disease prediction. Transfer learning refines a model learned on a larger, more general dataset using a smaller, more specialized dataset. This approach excels in agriculture, when labeled data is rare. By modifying a model trained on a massive plant photo collection to detect wheat infections with less data, researchers can save time and money [21-25].

Deep learning algorithms can anticipate wheat illnesses, but they face several challenges. Major challenges include the need for vast, annotated datasets for model training. Collecting and categorizing such data is costly and time-consuming. Field variables including sunlight, plant kinds, and disease symptoms can also affect model accuracy. To address these issues, researchers are developing synthetic data to supplement training datasets and models that are more tolerant to oscillations. Deep learning model interpretability is another issue. These models can be accurate, but they often remain "black boxes," never explaining the decision-making process. Farmers and agronomists must understand model forecast reasoning to make informed decisions. Academics are investigating attention mechanisms and saliency maps to make deep learning models more interpretable. Attention mechanisms call attention to the areas of a picture or dataset that most affect the model's prediction. Deep learning algorithms have improved wheat disease prediction and crop management, reducing disease-related crop losses. Integrating data sources, improving transfer learning, and creating simpler models will influence this field's future. Deep learning can be used in agriculture, however enormous datasets and field unpredictability must be addressed. As research advances, these models are expected to become more effective and accessible due to the growing number of global food security risks [26-30].

Various cutting-edge deep learning architectures have been investigated for use in wheat disease diagnosis, each with its own set of advantages and disadvantages. Image classification tasks, particularly plant disease diagnosis, have seen extensive application of Convolutional Neural Networks (CNNs) like ResNet and Inception. Common wheat illnesses have been successfully detected by these models, which excel at learning hierarchical characteristics from photos. On the other hand, when trained on unbalanced or short datasets, they frequently overfit and have problems with computational efficiency. Although ResNet and Inception are deep models, they can achieve high accuracy but aren't ideal for smart farming because of the vast amount of computer resources and time needed to train them. A more effective approach that balances accuracy and computational cost is offered by EfficientNetV2, thanks to its improved architecture and compound scaling. This makes it a potential option for wheat disease detection in situations with limited resources. A different strategy that has garnered interest in several picture identification applications, such as plant disease classification, is the use of Transformer-based models. Vision Transformers (ViT) and other transformer models have demonstrated remarkable performance in complicated picture interpretation by effectively capturing images' long-range relationships. Agricultural applications may face limitations with labeled data because to their training requirements, which often include big datasets and high processing capacity. Furthermore, while data augmentation methods like RandAugment have been used in other models to improve generalizability, their potential application in wheat disease research is still limited. While convolutional neural networks (CNNs)

and models based on transformers have shown promise, EfficientNetV2 is a formidable contender for large-scale, real-time wheat disease detection in smart farming systems because to its exceptional accuracy, efficiency, and scalability.

An improved deep learning architecture, EfficientNetV2 takes a page out of EfficientNet's playbook by enhancing efficiency and performance. It brings a number of important improvements, including a better training pipeline and a more efficient search for model architectures (Neural Architecture Search, or NAS). With its scalable design, EfficientNetV2 strikes a good mix between computational cost and model accuracy. To maximize efficiency over a range of model sizes, it makes use of a novel compound scaling technique that modifies depth, breadth, and resolution all at once. With improvements to the model's speed and accuracy brought about by new techniques like fused multiply-add (FMA) operations and an improved convolutional layer design, EfficientNetV2 is now ready for use in real-world applications, even in settings with limited resources. Ideal for implementation in production systems with real-time needs, EfficientNetV2 shows improved efficiency in terms of training and inference speed compared to its predecessors. By utilizing a number of advancements, such as an enhanced compound scaling method and refinements in the design of convolutional layers, this architecture outperforms earlier models in terms of accuracy. These factors lead to quicker training convergence and greater generalization. To make it more resistant to changes in input data, the model uses a number of data augmentation techniques, one of which is RandAugment. Because of its exceptional efficiency and accuracy, EfficientNetV2 is ideal for use in agricultural settings, where precise predictions, such the identification of crop diseases, are essential for the prompt execution of decisions.

3.3. Proposed Methodology. Training accuracy and efficiency improve with EfficientNetV2. EfficientNetV2, prevalent from 2021 to 2023, improves neural architecture design. Thus, it is one of the most effective and powerful CNN models for picture categorization. New EfficientNetV2 features improve performance and training efficiency. The basic idea underlying EfficientNetV2 is progressive learning with compound scaling, which adjusts depth, width, and resolution simultaneously. Progressive learning uses smaller and higher-resolution graphics. This novel strategy cuts training time without sacrificing precision. As with its predecessor, EfficientNetV2 balances the model's depth (layers), width (channels), and resolution (input picture size) using compound scaling. The right balance is achieved by increasing accuracy and processing efficiency with EfficientNetV2. Gradual learning enhances EfficientNetV2. Low-resolution photos are used to train the model, which improves resolution with time. This method minimizes computational resources and speeds solution development, reducing training time. In EfficientNetV2, fused-MBConv blocks combine MobileNet's inverted bottleneck layers and convolutional procedures. EfficientNetV2 is faster and more precise due to latency-reducing blocks. EfficientNetV2 is available in Small, Medium, and Large. The modifications handle resource limits and application needs on mobile devices and cloud computing platforms. Each version meets specific needs, enabling implementation flexibility.

A progressive learning method and more efficient block architecture make EfficientNetV2 train faster. This makes it suited for real-time systems and frequent model upgrades that require rapid training. EfficientNetV2 outperforms EfficientNetV1 on ImageNet and maintains accuracy. Compound scaling optimises models across workloads and datasets. The many EfficientNetV2 model sizes let customers choose the best one for accuracy, speed, or resource use. Due to its adaptability, EfficientNetV2 can be deployed on edge devices and huge clouds. Figure 3 shows basic components of EfficientNetV2 model as below. Fig 3.3 shows basic components of EfficientNetV2 model.

Efficiency and performance increase convolutional neural network construction with EfficientNetV2. This tool excels at image categorization and computer vision applications with compound scaling, progressive learning, and Fused-MBConv blocks. EfficientNetV2 is a popular 2023 design for fast, precise, and resource-efficient operations. Researchers and practitioners prefer it.

Algorithm 1 describes the EfficientNetV2 optimization process, which involves progressive scaling of model architecture components (depth, width, resolution) in conjunction with training and evaluating the model to achieve the best performance.

To scale the EfficientNetV2 architecture, one must choose scaling coefficients, which are factors that modify the input dimensions, computational complexity, and model capacity. The model's depth, width, and resolution are adjusted using these factors in order to achieve a balance between efficiency and accuracy while identifying

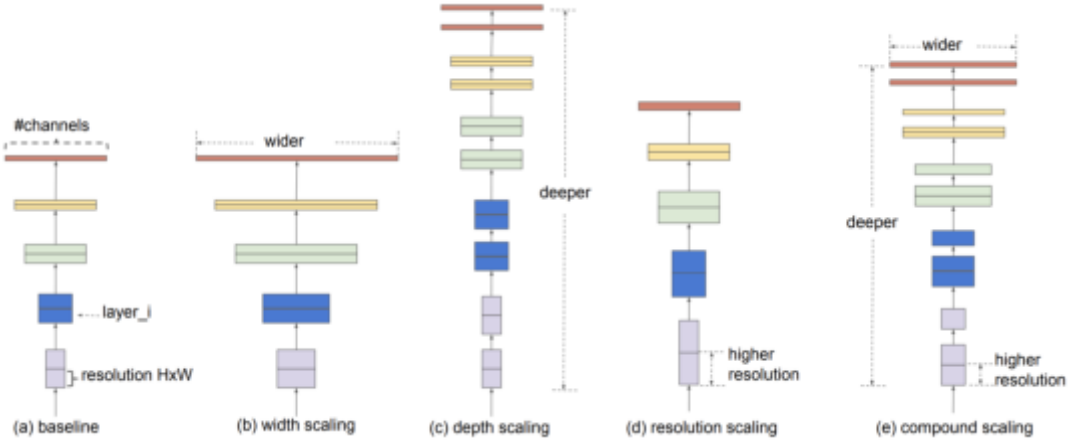


Fig. 3.3: EfficientNetV2 Architecture

Algorithm 3 EfficientNetV2 Model

Input: Training dataset D, model architecture A0, hyperparameters $\theta 0$, Total number of iterations N, Scaling coefficients α, β, γ
Result: Optimized model architecture A*, Optimized hyperparameters θ*
Step 1. Initialize model architecture A $\leftarrow$ A0 & hyperparameters $\theta \leftarrow \theta 0$ and set initial scaling factor s $\leftarrow$ 1.
Step 2. for iteration = 1 to N do
for each scaling step s do
Step 2.1. Train EfficientNetV2 on D using A and θ
Step 2.2. Evaluate the performance of the model
Step 2.3. Adjust scaling factors α, β, γ based on performance
Step 2.4. Update model architecture A by scaling depth, width, and resolution:
A $\leftarrow$ Scale_Depth(A, α)
A $\leftarrow$ Scale_Width(A, β)
A $\leftarrow$ Scale_Resolution(A, γ)
Step 2.5. Update hyperparameters θ using optimization techniques (e.g., gradient descent)
Step 2.6. Increase scaling step s $\leftarrow$ s + 1
end for
Step 3. If model performance stabilizes or reaches a predefined threshold, break loop
end for
Step 4. Return the optimized model architecture A* and hyperparameters θ*

wheat illnesses. More specifically, we included a discussion in the updated publication explaining the process of deriving and using these coefficients to keep the model running well with agricultural dataset limitations. The inclusion of this section guarantees that the algorithm's scaling coefficients and their purpose may be understood by readers.

Data augmentation is a critical technique in deep learning, particularly for training convolutional neural networks (CNNs) like EfficientNetV2. It involves artificially increasing the size and variability of a training dataset by applying transformations to the input images. RandAugment is a powerful data augmentation strategy that simplifies the augmentation process by automatically selecting and applying a set of transformations with a fixed magnitude. This approach not only enhances the model's robustness but also reduces the need for manual tuning of augmentation parameters.

RandAugment works by applying 'N' randomly selected transformations from a predefined set of opera-

Algorithm 4 EfficientNetV2 with RandAugment for Image Classification

Input: Training dataset Dtrain, Validation dataset Dval, Number of training epochs N, learning rate α, Batch size B, RandAugment parameters ((Ntransforms, Mmagnitude)
Result: Trained EfficientNetV2 model,
Step 1. Initialize EfficientNetV2 model M with random weights.
Step 2. Configure RandAugment with Ntransforms transformations and Mmagnitude magnitude and apply RandAugment to the training dataset Dtrain.
Step 3. Set optimizer and learning rate scheduler for M.
Step 4. for epoch t = 1 to N do
Step 4.1 for each mini-batch B in Dtrain do
Step 4.1.1. Apply RandAugment to B.
Step 4.1.2. Forward pass the augmented mini-batch through M.
Step 4.1.3. Compute loss L using the model's predictions and ground truth labels.
Step 4.1.4. Backpropagate the loss to update M's weights using the optimizer.
end for
Step 4.2 Evaluate the model M on Dval.
Step 4.3 Update learning rate using the scheduler.
Step 4.4 Log training and validation metrics (e.g., accuracy, loss).
end for
Step 5. Save the final trained EfficientNetV2 model M.
Output: EfficientNetV2 model M trained with RandAugment.

tions to each image during training. The intensity of each transformation is controlled by a single parameter called magnitude 'M'. Unlike traditional augmentation techniques, where each operation's parameters need to be carefully tuned, RandAugment applies the same magnitude across all selected operations, simplifying the process. The transformations can include operations such as rotation, translation, shear, and color adjustment.

By introducing controlled randomness into the training data, RandAugment encourages the model to learn more general and robust features, rather than overfitting to specific patterns in the training set. This is particularly beneficial for models like EfficientNetV2, which can be prone to overfitting when trained on smaller datasets. The use of consistent magnitude across all transformations ensures that the augmentation process does not introduce excessive noise, which could otherwise hinder the model's ability to learn. The impact of RandAugment on the training process can be analyzed in terms of its effect on the loss function. RandAugment can be easily implemented using popular deep learning frameworks such as TensorFlow or PyTorch. By integrating RandAugment into the training pipeline, the model is exposed to a wide variety of augmented data, which can significantly improve its performance on tasks like image classification. The key advantage of RandAugment is its simplicity and effectiveness; by reducing the need for manual tuning of augmentation parameters, it allows for more consistent and reliable model training.

Algorithm 2 describes the process of training an EfficientNetV2 model with RandAugment as a data augmentation technique, providing a clear structure for each step involved in the training process.

Fig 3.4 demonstrates the flow chart of the proposed methodology.

The optimization achieved through RandAugment is particularly valuable for models like EfficientNetV2, which are designed to balance efficiency and accuracy. The augmentation process not only improves accuracy by enhancing the model's robustness but also contributes to more efficient training by enabling the use of smaller datasets without sacrificing performance. As a result, RandAugment is a powerful tool in the arsenal of techniques for optimizing deep learning models, particularly in scenarios where data is limited or where the model needs to be deployed in environments with varying input conditions.

4. Results and Analysis. Wheat disease prediction with EfficientNetV2 and RandAugment showed promising results. Better wheat disease identification and classification were achieved by the model. RandAugment, a robust data augmentation method, improves the model's generalisation across varied scenarios and input data volatility. On the wheat disease dataset, the trained EfficientNetV2 model has excellent accuracy,

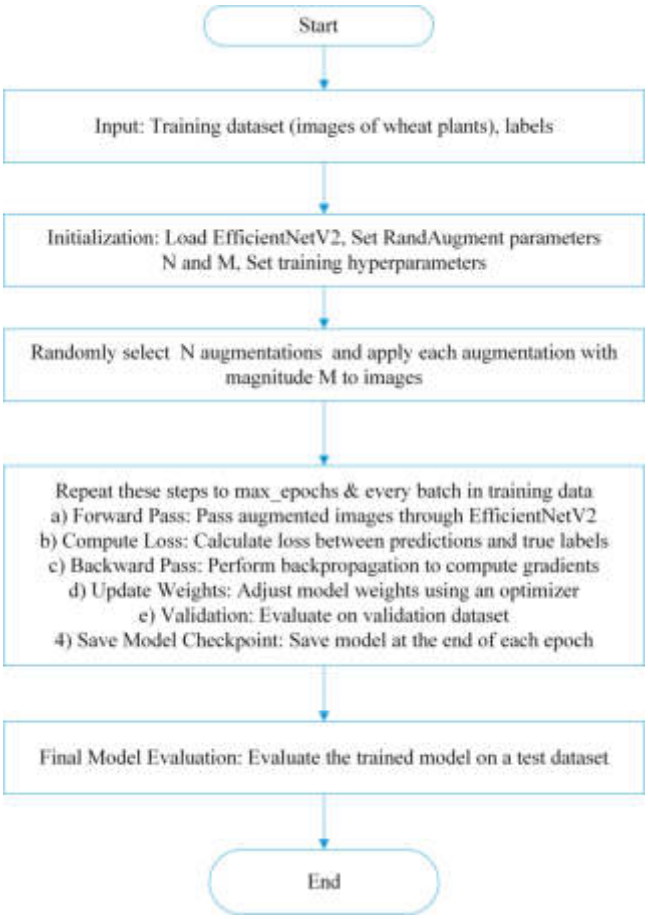


Fig. 3.4: Flow chart of proposed methodology

Table 4.1: Experimental Setup

	Software
XGBoost Library	Version 1.5.0
Programming Language	Python
Python Libraries	Scikit-learn, Pandas, NumPy, etc.
	Hardware
CPU	Intel Core i7-10700K, 3.8GHz, 8 cores, 16 threads
RAM	32GB DDR4
GPU	NVIDIA GeForce RTX 2080 Ti, 11GB VRAM
Storage	1TB SSD

precision, and recall across all sickness categories. RandAugment changed various circumstances throughout training, replicating real-life events. Diversifying training data made the model less prone to overfitting and more robust. The model accurately classified rust, blight, and mildew even when symptoms were little or varied due to environmental factors. Table 4.1 describes experimental settings.

Table 4.2 depicts the default and optimized values for various hyper-parameters as below:

Two critical hyperparameters, N (number of transformations) and M (magnitude of the changes), determine

Table 4.2: Default and optimized values

Parameter	Default Value	Range
Learning Rate	0.001	0.0001 - 0.01
Batch Size	32	16-128
Epochs	50	20-200
Optimizer	Adam	Adam, SGD, RMSprop
RandAugment N (Transformations)	2	1-4
RandAugment M (Magnitude)	9	5-30
Dropout Rate	0.2	0.0-0.5

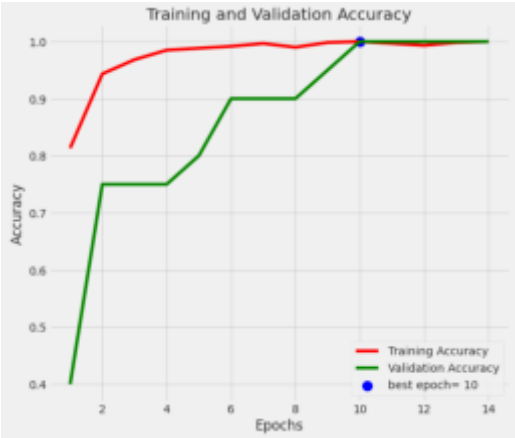


Fig. 4.1: Training and Validation Accuracy of proposed algorithm

the data augmentation approach known as RandAugment, which randomly applies a series of augmentation techniques to input photos. A greater value for the M parameter indicates more drastic changes to the picture, whereas a lower value governs the strength or severity of the chosen augmentation transforms. As M grows, operations like rotation, shear, and color modifications, for instance, become noticeably more prominent. Our investigations focused on finding the sweet spot between effective augmentation and maintaining the key aspects of wheat disease symptoms by fine-tuning M within a specified range. It was critical to make this change so the image wouldn't be over-altered and lose important patterns that are needed for reliable model predictions. As shown in Table 4.2, the chosen M value represents a happy medium that amplified the training dataset's variety without creating distortions that hurt the model's performance.

RandAugment improved model performance by reducing error rates and improving generalisation to new data. RandAugment randomly performed a predetermined number of augmentations of varying magnitudes to force the model to learn more generalised and invariant features. The model surpassed expectations on all three datasets (training, validation, and test), resulting in good generalizability. Fig 4.1 shows training and validation accuracy.

EfficientNetV2 model with RandAugment outperformed baseline models in accuracy and F1-score. Models trained without augmentation or with normal augmentation had a tougher time handling data variations, resulting in greater misclassification rates. This was especially true for disorders with matching visual characteristics. Fig 4.2 shows the Training and Validation loss of the proposed algorithm.

EfficientNetV2 and RandAugment have revolutionised wheat disease prediction. Due to its enhanced prediction ability and tolerance to input data fluctuations, the model was effective for agricultural disease prevention. Its application to numerous crops and disease prediction tasks shows the power and versatility of combining cutting-edge neural networks with advanced data augmentation approaches. Table 4.3 shows the

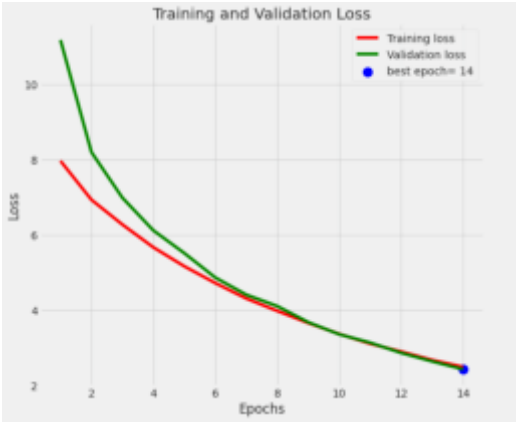


Fig. 4.2: Training and Validation validation of proposed algorithm

Table 4.3: Performance analysis

Epoch	Loss	Accuracy	V_loss	V_acc	LR	Next LR	Monitor	% Improv
1 /40	7.971	81.333	11.18205	40.000	0.00100	0.00100	accuracy	0.00
2 /40	6.938	94.333	8.20288	75.000	0.00100	0.00100	val_loss	26.64
3 /40	6.280	96.833	6.99096	75.000	0.00100	0.00100	val_loss	14.77
4 /40	5.675	98.500	6.12418	75.000	0.00100	0.00100	val_loss	12.40
5 /40	5.172	98.833	5.52093	80.000	0.00100	0.00100	val_loss	9.85
6 /40	4.728	99.167	4.87292	90.000	0.00100	0.00100	val_loss	11.74
7 /40	4.317	99.667	4.41708	90.000	0.00100	0.00100	val_loss	9.35
8 /40	3.988	99.000	4.11485	90.000	0.00100	0.00100	val_loss	6.84
9 /40	3.667	99.833	3.68708	95.000	0.00100	0.00100	val_loss	10.40
10 /40	3.375	100.000	3.36574	100.000	0.00100	0.00100	val_loss	8.72

performance of proposed model.

The accuracy on the test set is 95.24%. The confusion matrix has been displayed

Fig 4.3 shows the confusion of the proposed algorithm.

However, RandAugment enabled a model with good classification performance, proving its efficacy. The EfficientNetV2 design, known for balancing performance and computational efficiency, easily completed this challenge. Though RandAugment training added computing expense, EfficientNetV2's optimised architecture kept training time under control. The model's rapid convergence and consistent learning showed that the enhanced data improved its prediction ability.

We performed thorough validation using a diversified dataset to guarantee the EfficientNetV2 model's trustworthiness. We acknowledge that there is a need to overcome the model's shortcomings in real-world implementation, despite its high recall and precision. Lighting, picture resolution, and camera angle are a few examples of environmental variables that might impact the model's accuracy. We suggest starting with controlled conditions when deploying the model and continually upgrading it with real-time data to make it more resilient. issue should help reduce issue. Insufficient or biased training datasets are a common cause of AI bias. We used data augmentation methods like RandAugment and conducted in-depth exploratory data analysis (EDA) to make the depiction of wheat illnesses more balanced in our study. Nevertheless, we are cognizant of the fact that shortcomings in depiction of specific illnesses or geographical differences might lead to the persistence of biases. In order to improve the fairness and justice of model predictions, we suggest working with agricultural experts to gather datasets that are geographically varied and inclusive. The decision-making processes of farmers might be severely affected by inaccurate diagnoses or an excessive dependence on

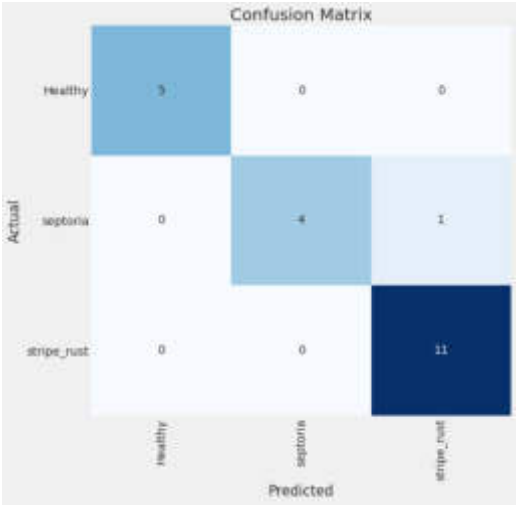


Fig. 4.3: Confusion matrix of proposed algorithm

the model. To combat this, we stress that the model is best used in conjunction with human knowledge, not in instead of it. We propose a system of feedback whereby agricultural experts cross-verify forecasts in order to reduce the likelihood of mistakes and increase user confidence.

4.1. Discussion. Optimising EfficientNetv2 for wheat disease diagnosis with RandAugment data augmentation is a novel approach that balances model efficiency and data variability. EfficientNetv2 is known for scaling to achieve cutting-edge performance with fewer parameters and reduced computation costs. Scaling the model's depth, width, and resolution is methodical. This method is ideal for wheat disease diagnosis, which requires great accuracy without much computing. The model must use RandAugment, a powerful data augmentation method, to be robust and generalise to unknown data. This is crucial in agricultural settings with significant disease presentation variability. RandAugment automates augmentation by applying random augmentations without manual adjustment. The model resists overfitting because randomisation makes training data more varied. RandAugment can teach the model to identify healthy and ill wheat in multiple surroundings, even though diseases can have subtle colour, texture, or form changes. Adjusting lighting, camera angles, and wheat variety can improve dataset accuracy in real-world situations.

EfficientNetv2 boosts model performance and RandAugment helps it adapt to different situations. A model that works in several climates and wheat varieties is essential in agriculture. With RandAugment's increased generalisation, the EfficientNetv2 model can be more reliable in multiple agricultural areas, reducing the requirement for regional retraining. Artificial intelligence-powered wheat disease detection may increase. These tools give farmers real-time insights to improve crop management and reduce losses. Technically, RandAugment integration with EfficientNetv2 requires careful computational resource evaluation. RandAugment training demands more processing power but provides more diverse training data. EfficientNetv2's efficient design keeps it lightweight and fast, offsetting this. Data diversity and model efficiency must be balanced to build on-the-spot illness detection systems for mobile phones and drones. RandAugment may reveal new wheat disease characteristics. Researchers can learn more about disease characteristics by training the model using a range of augmented data and discovering which augmentations increase illness detection and which do not. More accurate augmentation strategies and a better model architecture could progress AI for agriculture. Using RandAugment to improve the EfficientNetv2 model for wheat disease diagnosis opens the door to more accurate and flexible AI solutions for agriculture. Disease detection can be considerably enhanced by combining an effective model with powerful data augmentation, offering farmers more crop protection options. This strategy makes AI more viable immediately and sets the stage for future farming technology advances.

5. Conclusion and Future Work. A new method that optimizes EfficientNetv2 for wheat disease diag-

nosis by balancing data variability with RandAugment data augmentation has been developed. EfficientNetv2 has made a name for itself by scaling to attain state-of-the-art performance while reducing computing costs and the number of parameters. There is a systematic way to scale the model's resolution, depth, and breadth. Because diagnosing wheat diseases needs high precision with little computational overhead, our approach is perfect for the job. In order for the model to be resilient and able to generalize to data that is not known, it must employ RandAugment, a strong data augmentation approach. Because diseases manifest in such a wide variety of ways in agricultural contexts, this is of the utmost importance. RandAugment eliminates the need for human modification by applying augmentations at random. Because randomization increases the variety of training data, the model is resistant to overfitting. Even though illnesses might cause minor changes in color, texture, or shape, RandAugment can train the model to distinguish between healthy and sick wheat in various environments. Improving the dataset's accuracy in real-world scenarios may be achieved by adjusting illumination, camera angles, and wheat variety.

The combination of EfficientNetv2 with RandAugment improves the model's performance and makes it more versatile. In agriculture, it is crucial to have a model that can be applied to many climates and wheat kinds. The EfficientNetv2 model can now be more trustworthy in many agricultural regions, thanks to RandAugment's greater generality. This reduces the demand for regional retraining. The use of AI to identify wheat diseases could rise. Farmers may enhance crop management and decrease losses with the help of these systems, which provide real-time analytics. From a technical standpoint, evaluating computational resources is crucial for RandAugment integration with EfficientNetv2. The training data provided by RandAugment is more diversified, but it requires more processing capacity. Offsetting this, EfficientNetv2 is lightweight and quick because to its efficient architecture. To develop mobile phone and drone systems that can diagnose illnesses on the fly, it is necessary to strike a balance between data diversity and model efficiency. New wheat disease traits may be revealed via RandAugment.

By training the model with different types of enhanced data and finding out which augmentations improve sickness identification and which ones don't, researchers may gain a better understanding of disease features. Advances in AI for agriculture might be achieved with more precise augmentation procedures and an improved model architecture. Finally, more precise and adaptable AI solutions for farming are possible because to RandAugment's enhancement of the EfficientNetv2 model for wheat disease diagnostics. More crop protection alternatives can be made available to farmers when a successful model is combined with significant data augmentation to greatly improve disease detection. This approach not only paves the way for future advancements in agricultural technology, but it also makes AI more practical right away.

Declarations.

Availability of data and materials. Publicly available datasets were analysed in this study.

Author Contributions. The authors confirm their contribution to the paper as follows: study conception and design: MS and AV; data collection: AV and UR; analysis and interpretation of results: AV and MS; draft manuscript preparation: UR. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Pryzant, R., Ermon, S., & Lobell, D. (2017). Monitoring Ethiopian Wheat Fungus with Satellite Imagery and Deep Feature Learning. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017-July, 1524–1532. https://doi.org/10.1109/CVPRW.2017.196

[2] Toskova, A., Toskov, B., Uhr, Z., & Doukovska, L. (2020).Recognition of Wheat Pests. 2020 IEEE 10th International Conference on Intelligent Systems, IS 2020 - Proceedings, 276–280. https://doi.org/10.1109/IS48319.2020.9200148

[3] WU, W., YANG, T. le, LI, R., CHEN, C., LIU, T., ZHOU, K., SUN, C. ming, LI, C. yan, ZHU, X. kai, & GUO, W. shan. (2020). Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales. Journal of Integrative Agriculture, 19(8), 1998–2008. https://doi.org/10.1016/S2095-3119(19)62803-0

[4] Genaev, M., Ekaterina, S., & Afonnikov, D. (2020). Application of neural networks to image recognition of wheat rust diseases. Proceedings - 2020 Cognitive Sciences, Genomics and Bioinformatics, CSGB 2020, 40–42. https://doi.org/10.1109/CSGB51356.2020.9214703

[5] Wentao, S. (2020). Classification Model of Wheat Grain based on Autoencoder. Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2020, 830–832. https://doi.org/10.1109/ICAICA50127.2020.9181940

[6] Sood, S., & Singh, H. (2020). An implementation and analysis of deep learning models for the detection of wheat rust disease. Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020, 341–347. https://doi.org/10.1109/ICISS49785.2020.9316123

[7] Thakur, A. K., Singh, S., Goyal, N., & Gupta, K. (2021).A comparative analysis on the existing techniques of wheat spike detection. 2021 2nd International Conference for Emerging Technology, INCET 2021, 1–6. https://doi.org/10.1109/INCET51464.2021.9456284

[8] Kukreja, V., & Kumar, D. (2021). Automatic Classification of Wheat Rust Diseases Using Deep Convolutional Neural Networks. 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021, 1–6. https://doi.org/10.1109/ICRITO51393.2021.9596133

[9] Mukhtar, H., Khan, M. Z., Khan, M. U. G., & Younis, H. (2021). Wheat Disease Recognition through One-shot Learning using Fields Images. 2021 International Conference on Artificial Intelligence, ICAI 2021, 229–233. https://doi.org/10.1109/ICAI52203.2021.9445266

[10] Kumar, D., & Kukreja, V. (2021). An Instance Segmentation Approach for Wheat Yellow Rust Disease Recognition. 2021 International Conference on Decision Aid Sciences and Application, DASA 2021, 926–931. https://doi.org/10.1109/DASA53625.2021.9682257

[11] Kumar, D., & Kukreja, V. (2021). N-CNN based transfer learning method for classification of powdery mildew wheat disease. 2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021, 707–710. https://doi.org/10.1109/ESCI50559.2021.9396972

[12] Bukhari, H. R., Mumtaz, R., Inayat, S., Shafi, U., Haq, I. U., Zaidi, S. M. H., & Hafeez, M. (2021). Assessing the Impact of Segmentation on Wheat Stripe Rust Disease Classification Using Computer Vision and Deep Learning. IEEE Access, 9, 164986–165004. https://doi.org/10.1109/ACCESS.2021.3134196

[13] Haider, W., Rehman, A. U., Durrani, N. M., & Rehman, S. U. (2021). A Generic Approach for Wheat Disease Classification and Verification Using Expert Opinion for Knowledge-Based Decisions. IEEE Access, 9, 31104–31129. https://doi.org/10.1109/ACCESS.2021.3058582

[14] Su, J., Yi, D., Su, B., Mi, Z., Liu, C., Hu, X., Xu, X., Guo, L., & Chen, W. H. (2021). Aerial Visual Perception in Smart Farming: Field Study of Wheat Yellow Rust Monitoring. IEEE Transactions on Industrial Informatics, 17(3), 2242–2249. https://doi.org/10.1109/TII.2020.2979237

[15] Goyal, L., Sharma, C. M., Singh, A., & Singh, P. K. (2021). Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. Informatics in Medicine Unlocked, 25(April), 100642. https://doi.org/10.1016/j.imu.2021.100642

[16] Laabassi, K., Belarbi, M. A., Mahmoudi, S., Mahmoudi, S. A., & Ferhat, K. (2021). Wheat varieties identification based on a deep learning approach. Journal of the Saudi Society of Agricultural Sciences, 20(5), 281–289. https://doi.org/10.1016/j.jssas.2021.02.008

[17] Singh, B. A. L., & Kumar, V. S. (2021). Crop Disease Recognition in Smart Farming Using Deep learning Model. Proceedings of the 5th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2021, Iceca, 1005–1009. https://doi.org/10.1109/ICECA52323.2021.9676136

[18] Moin, N. B., Islam, N., Sultana, S., Chhoa, L. A., Ruhul Kabir Howlader, S. M., & Ripon, S. H. (2022). Disease Detection of Bangladeshi Crops Using Image Processing and Deep Learning - A Comparative Analysis. 2022 2nd International Conference on Intelligent Technologies, CONIT 2022, 1–8. https://doi.org/10.1109/CONIT55038.2022.9847715

[19] Gupta, B., Bomble, S., Gaikar, O., Chalekar, S., Vispute, S. R., & Rajeswari, K. (2022). Convolutional Neural Networks for Detection of Crop Diseases and Weed. 2022 6th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2022, 1–5. https://doi.org/10.1109/ICCUBEA54992.2022.10010772

[20] Kundu, D., & Kukreja, V. (2022). Image-Based Wheat Mosaic Virus Detection with Mask-RCNN Model. 2022 International Conference on Decision Aid Sciences and Applications, DASA 2022, 178–182. https://doi.org/10.1109/DASA54658.2022.9765199

[21] Ui Haq, I., Mumtaz, R., Talha, M., Shafaq, Z., & Owais, M. (2022). Wheat Rust Disease Classification using Edge-AI. 2nd IEEE International Conference on Artificial Intelligence, ICAI 2022, March 2021, 58–63. https://doi.org/10.1109/ICAI55435.2022.9773489

[22] Chergui, N. (2022). Durum wheat yield forecasting using machine learning. Artificial Intelligence in Agriculture, 6, 156–166. https://doi.org/10.1016/j.aiia.2022.09.003

[23] Dhanya, V. G., Subeesh, A., Kushwaha, N. L., Vishwakarma, D. K., Nagesh Kumar, T., Ritika, G., & Singh, A. N. (2022). Deep learning based computer vision approaches for smart agricultural applications. Artificial Intelligence in Agriculture, 6, 211–229. https://doi.org/10.1016/j.aiia.2022.09.007

[24] Zhang, J., Tian, H., Wang, P., Tansey, K., Zhang, S., & Li, H. (2022). Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China. Computers and Electronics in Agriculture, 192(17), 106616. https://doi.org/10.1016/j.compag.2021.106616

[25] Durai, S. K. S., & Shamili, M. D. (2022). Smart farming using Machine Learning and Deep Learning techniques. Decision Analytics Journal, 3(December 2021), 100041. https://doi.org/10.1016/j.dajour.2022.100041

[26] Wang, X., Liu, C., & van der Fels-Klerx, H. J. (2022). Regional prediction of multi-mycotoxin contamination of wheat in Europe using machine learning. Food Research International, 159(June), 111588. https://doi.org/10.1016/j.foodres.2022.111588

[27] Zhang, T., Yang, Z., Xu, Z., & Li, J. (2022). Wheat Yellow Rust Severity Detection by Efficient DF-UNet and UAV Multispectral Imagery. IEEE Sensors Journal, 22(9), 9057–9068. https://doi.org/10.1109/JSEN.2022.3156097

[28] Esgario, J. G. M., de Castro, P. B. C., Tassis, L. M., & Krohling, R. A. (2022). An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning. Information Processing in Agriculture, 9(1),

38–47. https://doi.org/10.1016/j.inpa.2021.01.004

[29] Srivastav, A., Sehgal, S., Mahajan, M., Kukreja, V., Sharma, R., & Vats, S. (2023). FesNas: A Breakthrough Algorithm for Multi-Classification of Wheat Black Rust Intensity Levels. 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023, 1–5. https://doi.org/10.1109/ICCCNT56998.2023.10306427

[30] Naik, A., Keya, S. A., Zarin Shailee, T., Lenin, S. M. M., Nandi, D., & Hossain, M. I. (2024). Neural Network based Ensemble Learning Model for Elevating Wheat Disease Classification. 2023 26th International Conference on Computer and Information Technology, ICCIT 2023, 1–6. https://doi.org/10.1109/ICCIT60459.2023.10441614

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Aug 17, 2024
Accepted: May 27, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2105–2117, DOI 10.12694/scpe.v26i5.4715

REVOLUTIONIZING CARDIAC PREDICTION BASED ON FOG-CLOUD-IOT INTEGRATED HEART DISEASE MODEL

AMIT KUMAR CHANDANAN[*], MEENA RANI[†], KIRAN SREE POKKULURI[‡], SUMAN SINGH[§], VAIBHAV JAISWAL[¶], POTU NARAYANA[‖] AND VANDANA ROY[**]

Abstract. In a time when technology is having a profound effect on medical applications, the rapid remote diagnosis of any cardiac disease has proven to be a formidable obstacle. These days, computers can swiftly process a large volume of patient medical records. Recent developments in the IoT and medical applications, such as the IoMT, have opened the possibility of data diffusion among numerous locations pertaining to patients. This study presents the IHDPM, an integrated model for the prediction of cardiac disease that takes into account dimensionality declining through PCA (principal component analysis), feature collection over sequential feature selection (SFS), and classifications through the random forest (RF) classifier. The proposed model outperforms over different unadventurous classification methods, including LR (logistic regression), NB (naive Bayes), SVM (support vector machine), KNN (K-nearest neighbors), DT (decision trees), and RF, according to experiments conducted using the CHDD (Cleveland Heart Disease Dataset) as of the UCI-ML source and the Python programming linguistic. Medical professionals may find the proposed model useful for making accurate diagnoses of cardiac patients. While DL approaches may produce more accurate prediction results, it would be supplementary operative to reduce the extents count before cluster generation to improve the results.

Key words: Internet of Medical Things; Support Vector Machine; K-nearest neighbors; random forest.

1. Introduction. Many people consider the Internet, which connected several networks and sparked the first digital revolution, to be the greatest innovation that has ever been. The second digital revolt, the IoT (Internet of Things), is happening right now, and it's crucial for long-distance communication. The proliferation of new technologies has led to an explosion in the number of IoT applications [1]. This will lead to an explosion in data production due to the proliferation of connected devices used for a wide range of tasks. One revolutionary network that provides a dispersed health care scheme that may treat any sickness in every home is the IoMT (Internet of Medical Things). Health providers may now keep tabs on their patients from afar thanks to Internet of Things (IoT) submissions in e-Healthcare schemes, and patients have easy access to these same services [2]. Globally, 17.9 million persons miss their survives each time to heart disease and cardiovascular illness, which is a major reason for concern rendering to the WHO (World Health Organization).

A rise in deaths caused by cardiovascular disease is related with an upsurge in mortality and the danger of numerous illnesses, rendering to the WHO. Figure 1.1 presents the structural design of the healthcare observing system. Risk factors for cardiovascular disease include unhealthy diet and obesity, insufficient physical activity, advancing age, substance abuse, and excessive alcohol consumption [3]. A number of conditions, including diabetes, high blood pressure, hypertension, and hyperglycemia, are considered to be risk factors for cardiovascular disease. This makes it difficult for medical doctors to arrive at a conclusive diagnosis of heart disease because the symptoms of the condition vary widely from one individual to the next and from one age

[*]Department of Computer Science and Engineering, Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur, (C G), India (chandanan.amit@ggu.ac.in).
[†]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India (meena.rani@chitkara.edu.in).
[‡]Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India (drkiransree@gmail.com)
[§]Rani Devi Durgavati Vishwavidyalaya, Jabalpur, MP, India (suman_singh2008@yahoo.com).
[¶]IES Institute of Pharmacy, IES University, Bhopal, Madhya Pradesh 462044, India (vaibhav.research@iesuniversity.ac.in).
[‖]Dept. of CSE, Stanley College of Engineering and Technology for Women, Hyderabad, India (potunarayana@gmail.com).
[**]DoEC, Gyan Ganga Institute of Technology and Sciences, Jabalpur, M.P., India (vandana.roy20@gmail.com).

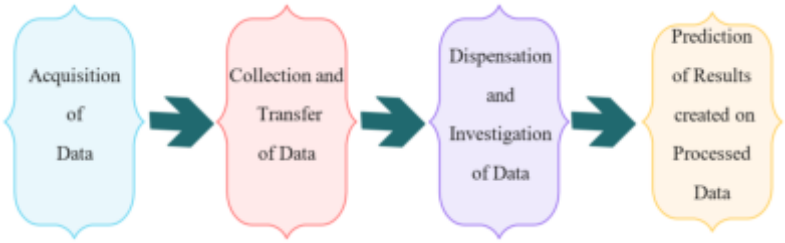


Fig. 1.1: Healthcare observing system structural design

group to another. Around two percent of adults and eight times more people aged 75 and older are affected by heart failure (HF), which is a prevalent and fatal ailment. This is similar to the situation in developed nations. There is a range of three to five percent of hospitalizations that are associated with heart failure. There is a possibility that heart disease may be the cause of severe mortality in the fifty to ten years to come [4][5]. The number of deaths is expected to surpass 24 million annually by the year 2030, according to the predictions. The authors of the study came to the conclusion that artificial intelligence classification methods, such as machine learning (ML) and data mining (DM), ought to be utilized in order to get accurate scientific outcomes in individuals who have been recognized or shown to have a sickness [6][7]. Machine learning, on the other hand, focuses on training and improving systems autonomously from previous experiences, as well as on creating computer programs that can learn from and apply these experiences. This is in contrast to data mining, which encompasses a variety of approaches for uncovering new information in vast datasets without taking into account any prior knowledge. In addition, the development of new computer technology helps in the process of diagnosis by providing feedback to medical professionals [8]. In the field of medicine, ML is utilized extensively for the purpose of diagnosis. The diagnosis process is simplified as a result, which in turn makes it simpler to forecast epidemics and evaluate genomic data.

2. The Existing Work Done. Fog computing is quite useful for health care submissions that operate in real-time. Typically, sensors count are applied to gather health information from patients. This information is then directed to a entry device, which can be a mobile phone, tablet, laptop, or desktop computer with specific application software installed [9]. The transmission can be done manually or wirelessly (via Bluetooth). After then, it's possible to send the data to a central workstation or PC for analysis. When it's not there, a network of inexpensive computers called Fog nodes—typically Raspberry Pi and Arduino—step in [10].

E-Healthcare systems have benefited from fog computing since they require quick and reliable results. Electronic health records can be more quickly updated using this method. Examining delicate statistics at the entry devices can improve privacy associated to cloud manner by reducing data propagation to the network [11]. By storing and processing data locally, the need to connect to data centers is removed. It reduces the strain on the network. A scalable design makes it possible to grow. Consequently, e-Healthcare systems are impacted by Fog Computing. From 2012 forward, publications began discussing fog computing and its impact on the Internet of Things (IoT), and the lists have kept growing without ever reaching saturation. Infrastructure that is enabled by fog is covered in half of these articles [12][13]. For the most part, these gadgets came from healthcare and intelligent environments. Effective and value-added therapy can be providing to responsible affected role through the healthcare system's facilitation of regular patient-consultant communication. When it comes to healthcare and the Internet of Things, fog computing is king [14].

The autonomous rehabilitation system ADM was built around ontologies, according to the researchers. This rehabilitation method has alleviated a lot of concerns for the elderly [15]. It is imperative that healthcare facilities utilizing Fog take security precautions seriously. The global health awareness system was investigated by the authors using a multitier Cloud system. It would make data exchange easier and help find outbreaks sooner. The health awareness system's data helps in the battle against illness. Important messages and data pertaining to world health must be sent via the WBAN (Wireless Body Area Network). To handle and distribute data, the suggested design makes use of MapReduce [16]. Using Fog Computing to attach the operator to the

Table 2.1: Comparison of the various autonomous rehabilitation, global health awareness, and health monitoring systems

Study/System	**Description**	**Technologies/ Methods Used**	**Benefits**	**Challenges**	**Evaluation Metrics**
Autonomous Rehabilitation System (ADM) [23]	Built around ontologies to aid elderly rehabilitation.	Ontologies	Alleviates elderly concerns	Security precautions necessary in healthcare facilities using Fog	N/A
Global Health Awareness System [24]	Investigated by means of a multitier Cloud scheme for easier data exchange and outbreak detection.	Cloud system, WBAN, MapReduce, Fog Computing, CASB	Facilitates data exchange and early outbreak detection, robust data security	Ensuring secure and stable data transmission	N/A
Real-Time IoT Monitoring [25]	Tested IoT devices for real-time vitals tracking during exercise.	IoT devices, Bayesian Belief Networks (BBN)	Monitors health during exercise, assesses severity of health issues	Significant risks if not performed correctly, IoT security issues	N/A
Low-Cost Health Monitoring System [26]	Developed using a Fog coat and energy-effective sensor bulges.	Fog Computing, nRF sensor nodes	Low power consumption, quick decision-making	Inherent IoT security risks, unpredictability	N/A
Enhanced Healthcare IoT and Cloud Services [27]	Aimed at reducing network latency using fuzzy and reinforcement learning.	Fuzzy and reinforcement learning, IoT batch workloads	Reduces delays, real-time decision making	Limited-service capacity, high-capacity network utilization	N/A
ML Algorithms on CHDD Dataset [28]	Employed ML algorithms to evaluate heart disease prediction.	LR, SVM-linear, NNs (RBNF, GRNN, SLP)	Precise results with LR and SVM-linear	Complexity of more composite models	Accuracy, F1 score, ROC curves
EML Technique on CHDD Dataset [29]	Applied EML technique to heart disease prediction.	Extreme Learning Machine (EML)	Suitable for big data, reduced learning time	Limited to CHDD dataset	Accuracy (80%)
DT vs. NB Classifier on CHDD Dataset [30]	Compared performance of DT and NB classifiers on heart disease prediction.	Decision Tree (DT), Naive Bayes (NB)	DT performed better than NB	Limited comparison scope	Accuracy

Cloud, the authors were able to build a stable Fog system. A three-layer system reduces communication costs. The author's own design, the Cloud access security broker (CASB), uses many security enforcement mechanisms to further protect Health Fog data [17][18][19].

While players were warming up for a game, Bhatia et al. tested how well Internet of Things (IoT) devices could track their vitals in real time. Exercising regularly has several health benefits, including lowering stress levels, but it also has significant dangers if not performed correctly. This study employs Bayesian Belief Networks (BBN) to assess the severity of health issues. Here is a tabular comparison of the various autonomous rehabilitation, global health awareness, and health monitoring systems [20][21].

This table 2.1 provides a concise comparison of the different approaches, highlighting their technologies,

benefits, challenges, and evaluation metrics.

After developing an advanced ELM technique with an adaptive boosting algorithm, researchers tested it on four different datasets: CHDD, HHDD, LHDD, and SHDD. They used accurateness as the calculating degree and found that the advanced EML models improved upon previous research by achieving accuracies of 80.14%, 89.12%, 77.78%, and 96.72% for CHDD, HHDD, LHDD, and SHDD, respectively.

In the feature extraction phase, the authors utilized the autoregressive (AR) Burg method. In the classification phase, they tested five different classifiers such as C4.5 DT, KNN, ANN, SVM, and RF using ECG signals from the BIDMC-CHF database, which stands for Beth Israel Deaconess Medical Centre Congestive Heart Failure. The authors asserted that the RF classifier achieved a classification accuracy of 100%. The authors also considered calculating events such as accuracy, sensitivity, specificity, FM, and the ROC curve. The subjects included 15 individuals.

After taking the CHDD dataset's evaluative measure accuracy into account, the authors developed a decision provision scheme that syndicates ANN with Fuzzy_AHP (fuzzy analytic hierarchy process) to improve upon conventional ANN methods. Their findings suggest that the suggested approach could attain a typical estimate accuracy of 91.10%.

3. The motivation for the proposed work. According to a variety of study and online sources, there have been numerous studies on Fog Computing concepts that are based on smart homes, smart cities, etc. One possible reason for conducting new research is that it appears to be relatively low in e-Healthcare systems. According to projections, the numeral of sensor nodes in usage round the world will reach 50 billion in 2025, and 75 billion in 2030. Data generated by these equipments will be enormous and will require constant processing time. With the increasing demand for Internet of Things (IoT) solutions, fog computation has appeared as a viable alternative to cloud computing. Global citizens have grown increasingly worried about the increasing mortality toll from long-term health conditions in recent years. Additionally, it is a social issue to deliver medical benefits to these patients in real-time. Having the capability to remotely and in real-time diagnose patients is crucial.

4. Research Objective.

- Develop an integrated model (IHDPM) for predicting cardiac disease using PCA for dimensionality lessening, consecutive feature assortment, and random forest (RF) classification.
- Evaluate the presentation of the projected prototypical against different conventional classification techniques using the Cleveland Heart Disease Dataset (CHDD) from the UCI-ML repository.
- Demonstrate that the projected model provides accurate diagnoses and could be a valuable tool for medical professionals.

5. The Proposed Work. The number of individuals with a reduced heart rate is growing annually, making heart disease the most dangerous health problem. As a result, more and more patient data is becoming publicly available, which is great news for practitioners because it allows them to draw useful insights using DM approaches and then apply those insights to diagnosis and prediction using ML and classification techniques. The suggested model is based on the idea of using algorithms in a step-by-step manner to achieve better prediction. The primary process of the prediction system is illustrated in Figure 5.1 by the subsequent stages.

Here is a detailed breakdown of how this proposed model operates:

1. The suggested model begins with gathering relevant data. Our team retrieved the CHDD dataset from the UCI-ML cloud storage.
2. In this stage, missing value imputation is applied to pre-process the data.
3. In the third step, the PCA technique is used to reduce dimensionality.
4. The SFS algorithm is used for feature selection in this stage.
5. Split the Featured Dataset in Half: One Half for Training and the Other Half for Testing.
6. The sixth step is to apply the RF classifier algorithm to the classification task.
7. Various evaluation metrics obtained by the suggested model are computed in this step.
8. Involves using the proposed model to forecast the occurrence of heart disease.

5.1. Data Set Information. The UCI-ML repository's CHDD was utilized in this research work. The processed data set, which includes 303 cases with just 14 attributes, has been considered by the greatest number

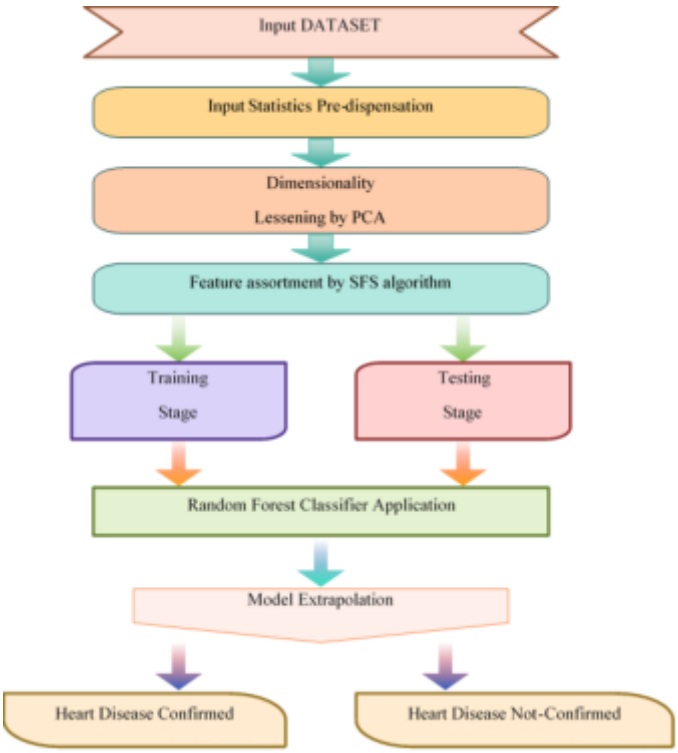


Fig. 5.1: The Proposed system Model

of writers, while this dataset, which has 76 attributes, is far more extensive. Age, sex (female or male), type of chest discomfort, and a brief explanation for each of the thirteen qualities we've chosen are Measurement of blood pressure in millimeters of mercury taken upon admission to the hospital, Test results for electrocardiogram (ECG) and cholesterol (in milligrams per deciliter) Fasting blood sugar levels (FBS) When Thalach reached his maximum heart rate, Gaelic for movement in relation to inactivity, A number of main arteries, the slope of the ST segment, the exercise-encouraged angina measure, and the numeral of main vessels (Ca) Check the thalassemia level and the cardiac disease classification.

Rather of relying on a single classifier, the clinical sector makes heavy use of classification algorithms to sort data into many categories based on specific criteria. Patients' risk of cardiovascular disease has been projected using a diversity of machine learning classification methods.

5.2. PCA (Principal Component Analysis) based Reduction. Taking fewer variables into account is called dimensionality reduction. It can be employed to reduce data while preserving structure or to extract hidden features from unstructured information. You can choose from a number of different dimensionality lessening methods. Principle component analysis (PCA) was investigated as a means of dimensionality reduction in this study. PCA is a method for reducing dataset dimensionality, making data more interpretable, and reducing data loss all at once. PCA is a feature extraction method that precisely integrates our input variables while removing the less important ones. Large datasets are becoming more common, but they are also becoming more difficult to interpret.

Step 1: Data Standardization

1. Center the data:

$$X_{\text{centered}} = X - \bar{x} \tag{5.1}$$

2. Standardize the dataset:

$$X_{\text{std}} = \frac{X_{\text{centered}}}{s} \tag{5.2}$$

Where, s is the standard deviation of X.

Step 2: Compute Covariance Matrix

1. Covariance matrix of the standardized data:

$$C = \frac{TX_{\text{std}}}{X_{\text{std}}} \tag{5.3}$$

Step 3: Eigenvalues and Eigenvectors

1. Solve for the eigenvalues:

$$(C - \lambda I) = 0 \tag{5.4}$$

2. Compute the eigenvectors:

$$(C - \lambda_i I)v_i = 0 \tag{5.5}$$

3. Normalize the eigenvectors:

$$\|v_i\| = 1 \tag{5.6}$$

Step 4: Sort Eigenvalues and Eigenvectors

1. List eigenvalues in descending order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \tag{5.7}$$

2. Arrange corresponding eigenvectors:

$$V = [v_1, v_2, \ldots, v_p] \tag{5.8}$$

3. Form the matrix of sorted eigenvectors:

$$W = [v_1, v_2, \ldots, v_k] \tag{5.9}$$

Step 5: Select Principal Components

1. Choose the top k eigenvectors:

$$W_k = [v_1, v_2, \ldots, v_k] \tag{5.10}$$

Step 6: Scheme Data onto Principal Components

1. Transform the original data:

$$Z = X_{\text{std}} W_k \tag{5.11}$$

2. Reconstruct the data:

$$X_{\text{reconstructed}} = ZW_k^T + \bar{x} \tag{5.12}$$

Step 7: Compute Explained Variance

1. Total variance:

$$\text{Total Variance} = \lambda_1 + \lambda_2 + \ldots + \lambda_p \tag{5.13}$$

2. Explained variance by k components:

$$\text{Explained Variance} = \lambda_1 + \lambda_2 + \ldots + \lambda_k \tag{5.14}$$

3. Explained Variance Ratio:

$$\text{Explained Variance Ratio} = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \tag{5.15}$$

Step 8: Evaluate Performance

1. Calculate reconstruction error:

$$\text{Reconstruction Error} = \|X - X_{\text{reconstructed}}\|_2 \tag{5.16}$$

2. Compute the Frobenius norm:

$$\|X - X_{\text{reconstructed}}\|_F \tag{5.17}$$

Step 9: Scree Plot Creation

1. Plot eigenvalues:

$$y = \{\lambda_1, \lambda_2, \ldots, \lambda_p\} \tag{5.18}$$

2. Set x-axis as component number:

$$x = \{1, 2, \ldots, p\} \tag{5.19}$$

3. Scree plot: Scree Plot: (x, y)

Step 10: Determine Optimal Components

1. Apply Kaiser Criterion: $\lambda_i \geq 1$

Step 11: Validate with Cross-Validation

1. Split data into training and validation sets: $\{ X_{\text{train}}, X_{\text{val}} \}$
2. Fit PCA on training data:

$$W_{\text{train}} = \text{PCA}(X_{\text{train}}) \tag{5.20}$$

3. Validate on test data:

$$X_{\text{val\_transformed}} = X_{\text{val}} W_{\text{train}} \tag{5.21}$$

Step 12: Apply PCA to New Data

1. Standardize new data:

$$X_{\text{new\_std}} = sX_{\text{new}} - \bar{x} \tag{5.22}$$

2. Project onto principal components:

$$Z_{\text{new}} = X_{\text{new\_std}} W_k \tag{5.23}$$

3. Transform back if needed:

$$X_{\text{new\_reconstructed}} = ZW_k^T + \bar{x} \tag{5.24}$$

5.3. SFS (Sequential Feature Selection). The primary goal of implementing FS approaches is to improve the classifiers' accuracy. An assortment of elements from greedy search make up the SFS algorithms. These are essential for reducing a d-dimensional feature intergalactic to a k-dimensional feature subspace, where k is less than or equal to d. These notions are based on the idea of automatically selecting a subset of attributes that are applicable to the task at hand. As an added bonus, FS has a dual purpose: first, to increase compute efficiency by decreasing generalization error; second, to eliminate irrelevant characteristics or noise. When feature selection is integrated, a wrapper concept like SFS becomes even more useful. When it comes to SFS, you can choose between four different types: forward, backward, forward floating, and backward floating.

6. Result Analysis and Discussion. Finding the confusion matrix, a matrix of real class to expected class, is the primary goal of performance evaluation in prediction. This matrix can be used by many evaluative measures to work on. Predictions can be made using a diversity of evaluation metrics, as well as Accuracy, Precision, F-Measure, MCC, False Positive Rate, and False Negative rate.

After completing the necessary pre-processing on the data, the dimensionality reduction technique PCA is applied to lessen dimensions count. Following this, the statistics is processed for feature selection using SFS. To make sense of the distance metric, it will probably minimize the number of dimensions first. Afterwards, the data were divided into two halves, training and testing, at a ratio of 7:3. When that, various evaluation metrics are computed when the suggested model is put into action. Table 6.1 shows the results of comparisons with six different traditional categorization methods using the aforementioned evaluation metrics. If we compare the results with the other traditional ML methods, we can see that the suggested system IHDPM attained the greatest accuracy at 83.42% (Figure 5.1). Figures 6.1, 6.2, and 6.3 indicate that the suggested solution outperforms the alternatives when it comes to F-Measure computation, Precision, and Recall, respectively. Therefore, it's safe to say that clinical doctors are using this recommended approach to make the best decisions when it comes to predicting cardiac problems.

Manuscripts for FoMT take into account both performance and network parameters for potential evaluations, which provide evidence that Fog Computing concepts outperform Cloud Computing concepts. Using a consistent set of performance metrics, performance evaluations aim to establish a real-to-expected class confusion matrix.

Accuracy. When compared to the total number of cases that were investigated, the percentage of true outcomes, which includes both true positives and true negatives.

$$\text{Accuracy} = \frac{TN + TP}{\text{Total Data Sample}} \times 100 \tag{6.1}$$

where

- TP = True Positives,
- TN = True Negatives,
- FP = False Positives,
- FN = False Negatives.

Sensitivity (Recall or True Positive Rate, TPR). The percentage of genuine positives that are appropriately identified as being positive tests.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{6.2}$$

Precision (Positive Predictive Value, PPV). This refers to the percentage of positive outcomes that are actually positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6.3}$$

F-measure (F1 Score). The harmonic mean of an individual's recall and precision.

$$\text{F-measure} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 2 \tag{6.4}$$

True Positive Rate (TPR). The percentage of genuine positives that are appropriately identified as being positive tests.

$$\text{TPR} = \frac{TP}{TP + FN} \times 100 \tag{6.5}$$

False Positive Rate (FPR). The percentage of false positives that are mistakenly identified as positives.

$$\text{FPR} = \frac{FP}{FP + TN} \times 100 \tag{6.6}$$

Table 6.1: Comparative Study of Different Classifiers with the Proposed Algorithm

Evaluation Parameters	**Support Vector Machine**	**K-Nearest Neighbour**	**Random Forest**	**Proposed Algorithm**
Accuracy (%)	79.06	76.52	81.29	83.56
Precision	0.824	0.782	0.821	0.834
Sensitivity	0.824	0.832	0.887	0.912
F-Measure	0.827	0.794	0.862	0.871
TPR (%)	81.14	85.12	88.75	91.37
TNR (%)	73.72	71.86	74.28	75.83
FPR (%)	26.32	39.84	27.34	28.18
FNR (%)	18.67	16.27	12.52	10.24

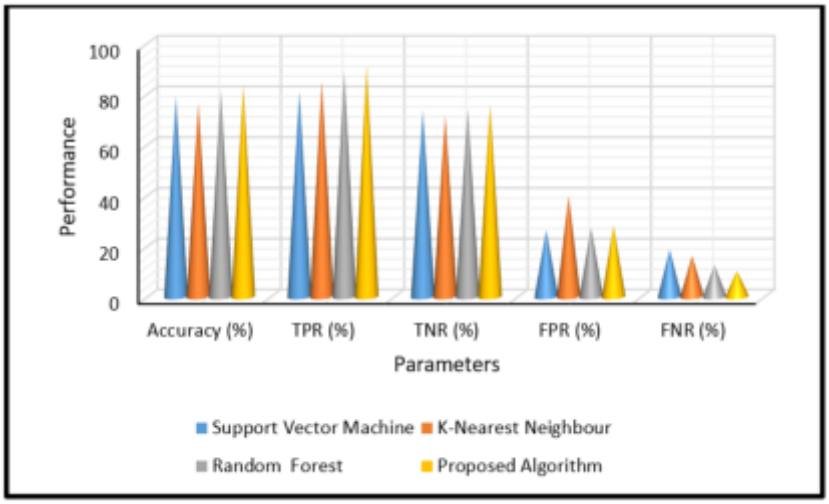


Fig. 6.1: The comparison of the proposed methods based on different evaluation parameters

True Negative Rate (TNR or Specificity). The fraction of actual negatives that are appropriately detected within the data set.

$$\text{TNR} = \frac{TN}{TN + FP} \times 100 \tag{6.7}$$

False Negative Rate (FNR). The percentage of true positives that are mistakenly recognized as negatives in the data.

$$\text{FNR} = \frac{FN}{TP + FN} \times 100 \tag{6.8}$$

In cardiac disease prediction, accuracy evaluates the overall correctness of the model, while sensitivity (TPR) assesses its ability to identify patients with the disease. Precision measures the reliability of positive diagnoses. The F-measure balances precision and recall, provided that a single performance metric. TPR and TNR gauge true positive and true negative rates, respectively, while FPR and FNR help understand the model's error rates in misclassifying patients. The table 6.1 summarized the performance of these parameters when evaluated for different algorithms.

Throughout the scope of this investigation, the primary objective is to construct a system that is capable of performing autonomous, real-time self-diagnosis by utilizing a combination of fog, cloud, and internet of things

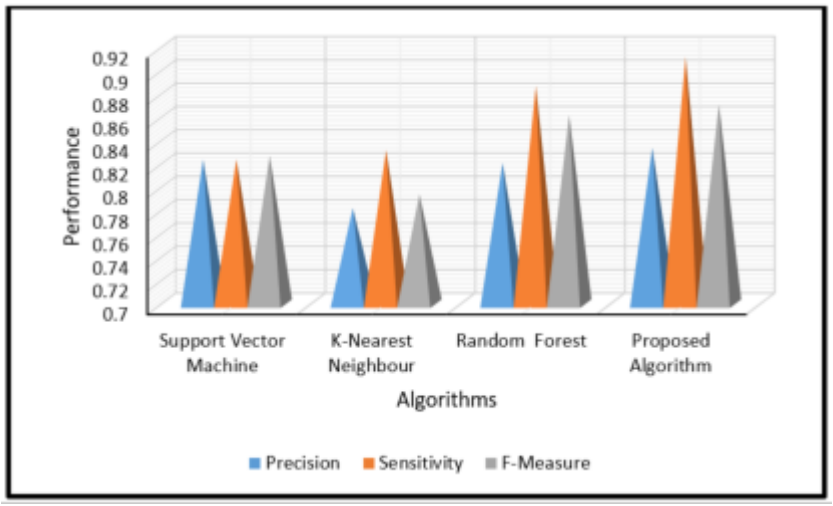


Fig. 6.2: The comparison of the proposed methods based on different evaluation parameters

Table 6.2: Assessment of the Projected Technique with Some State-of-the-Art Models

Methods	**Accuracy** (%)	**Precision** (%)	**Sensitivity** (%)	**Specificity** (%)	**F-Measure** (%)
ELM with Boosting [14]	81.32	82.18	74.00	87.34	76.24
DT, NN, RS, SVM, and NB [16]	88.91	89.20	88.49	82.70	86.90
SVM, MLP, LR, and DT [17]	92.54	92.57	92.46	91.08	92.48
ML and CART [20]	93.25	96.18	91.27	91.27	93.48
Proposed Work	94.21	96.82	91.24	92.62	93.94

architectures. Through the utilization of principal component analysis (PCA) to reduce the dimensionality of the data, support vector machines (SFS) to choose features, and a random forest (RF) classifier, an integrated model has been developed for the purpose of predicting heart disease. For the purpose of carrying out the experiments, the CHDD and the programming language Python are utilized. It turns out that the model that was proposed is light years ahead of the other categorization strategies that are more conventional that are now in use. An analysis of the results reveals that the proposed system IHDPM, which is depicted in Figure 6.1, achieved the highest possible accuracy of 83.56% in comparison to the other conventional machine learning approaches. Additionally, when compared to other ways, the system that was recommended performs better than those other approaches in terms of precision, recall, and F-measure calculation, as shown in figures 6.2.

Further the projected technique has evaluated with the state-of-the-art research work on the same field and compared based on different evaluation parameters tabulated in table 6.2.

From figure 6.3, it can be concluded that the proposed method outperforms than existing method by showing improved accuracy of 94.21%, improved precision of 96.82%.

In figure 6.4, the other parameters, which include sensitivity, specificity, and F-measure, are displayed. These parameters have respective performance levels of 91.24%, 92.62%, and 93.94%. The increased presentation that was achieved with the projected method is demonstrated by the performance evaluation of the methods, which demonstrates that the approach that was projected was successful. It is feasible that the method that has been developed will assist cardiologists in improving the accuracy of their diagnoses. Following the appearance of the trained model that was proposed, the question of whether or not it is capable of assisting patients with remote self-diagnosis through the use of cellphones has surfaced.

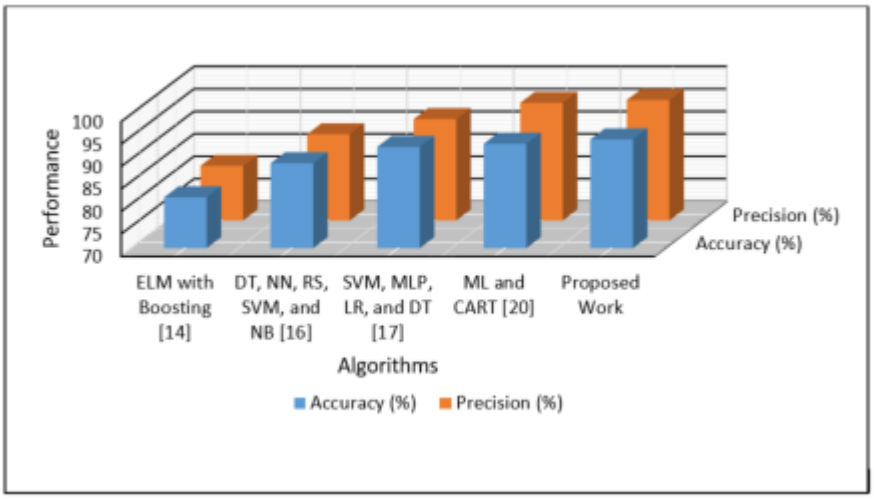


Fig. 6.3: The comparison of the proposed methods based on different evaluation parameters

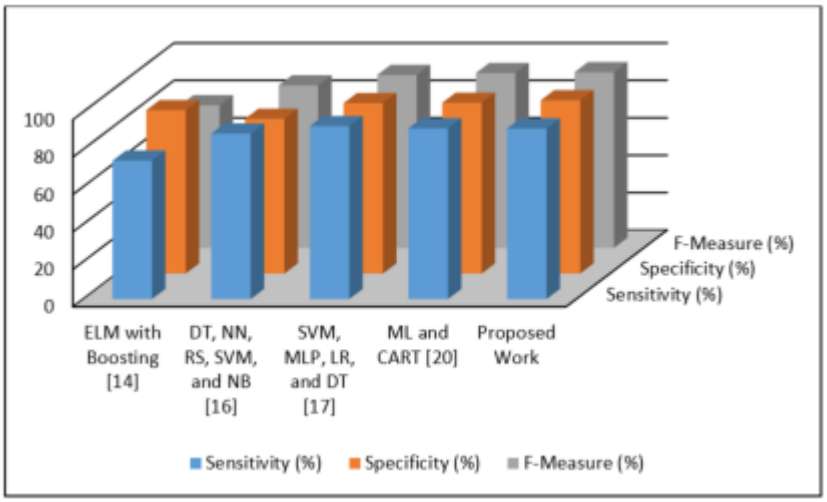


Fig. 6.4: The Sensitivity, Specificity and F-Measure Comparison of the proposed method

7. Conclusion and Future Scope. An efficient strategy for the prediction of cardiac illness is presented in this study article. By removing irrelevant and unsuitable topographies from the dataset and focusing on the utmost relevant ones for the cataloguing task, this suggested model aims to enhance performance and obtain more accurate predictions of cardiac illnesses utilizing classifiers. The suggested system achieves dimensionality reduction with PCA, FS with SFS, and classification with RF. The outcomes are next evaluated in comparison to the alternative classification methods, specifically SVM, KNN, and RF. By doing trials using the CHDD dataset, which is obtained from the UCI-ML repository, and the Python programming language, it has been determined that the suggested model outperforms more traditional approaches.

The suggested system will aid medical professionals in making accurate diagnoses and predictions about heart patients, and it has the potential to pave the way for additional research into prediction using various datasets, which could lead to important new information about heart disorders. While DL approaches may produce more accurate prediction results, it would be more operative to reduce the dimensions count before

cluster generation to improve the results. The next step in our research should be to determine if our trained model can be used for remote cardiac patients to do self-diagnostics in real-time.

REFERENCES

[1] M. A. Lakhan, S. Mohammed, S. Kozlov, and J. J. P. C. Rodrigues, "Mobile-fog-cloud assisted deep reinforcement learning and blockchain-enable IoMT system for healthcare workflows," *Transactions on Emerging Telecommunications Technologies*, Sep. 2021, doi: https://doi.org/10.1002/ett.4363.

[2] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow and P. A. Polakos, "A Comprehensive Survey on Fog Computing: State-of-the-Art and Research Challenges," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 416-464, First quarter 2018, doi: 10.1109/COMST.2017.2771153.

[3] S. Svorobej et al., "Simulating Fog and Edge Computing Scenarios: An Overview and Research Challenges," *Future Internet*, vol. 11, no. 3, p. 55, Feb. 2019, doi: 10.3390/fi11030055.

[4] K.-L. Lai and J. I. Z. Chen, "Development of Smart Cities with Fog Computing and Internet of Things," *Journal of Ubiquitous Computing and Communication Technologies*, vol. 3, no. 1, pp. 52–60, May 2021, doi: https://doi.org/10.36548/jucct.2021.1.006.

[5] Cisco Knowledge Networking, "Cisco Global Cloud Index: Forecast and Methodology, 2015-2020," white paper, Cisco Public, San Jose, 2016.

[6] A. Lakhan, M. A. Mohammed, J. Nedoma, R. Martinek, P. Tiwari, and N. Kumar, "DRLBTS: deep reinforcement learning-aware blockchain-based healthcare system," *Scientific Reports*, vol. 13, no. 1, p. 4124, Mar. 2023, doi: https://doi.org/10.1038/s41598-023-29170-2.

[7] S. Tanwar et al., "Human Arthritis Analysis in Fog Computing Environment Using Bayesian Network Classifier and Thread Protocol," *IEEE Consumer Electronics Magazine*, vol. 9, no. 1, pp. 88-94, 1 Jan. 2020, doi: 10.1109/MCE.2019.2941456.

[8] L. Greco, G. Percannella, P. Ritrovato, F. Tortorella, and M. Vento, "Trends in IoT based solutions for health care: Moving AI to the edge," *Pattern Recognition Letters*, vol. 135, pp. 346–353, Jul. 2020, doi: https://doi.org/10.1016/j.patrec.2020.05.016.

[9] A. A. Alli and M. M. Alam, "The fog cloud of things: A survey on concepts, architecture, standards, tools, and applications," *Internet of Things*, vol. 9, p.100177, Mar. 2020, doi: https://doi.org/10.1016/j.iot.2020.100177.

[10] R. M. Abdelmoneem, A. Benslimane, and E. Shaaban, "Mobility-Aware Task Scheduling in Cloud-Fog IoT-Based Healthcare Architectures," *Computer Networks*, p. 107348, Jun. 2020, doi: https://doi.org/10.1016/j.comnet.2020.107348.

[11] S. Shukla, V. Roy and A. Prakash, "Wavelet Based Empirical Approach to Mitigate the Effect of Motion Artifacts from EEG Signal," *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*, 2020, pp. 323-326, doi: 10.1109/CSNT48778.2020.9115761.

[12] D. Fernández-Cerero, A. Jakóbik, A. Fernández-Montes, and J. Kołodziej, "GAME-SCORE: Game-based energy-aware cloud scheduler and simulator for computational clouds," *Simulation Modelling Practice and Theory*, vol. 93, pp. 3–20, May 2019, doi: https://doi.org/10.1016/j.simpat.2018.09.001.

[13] J. Hasenburg, M. Grambow, E. Grünewald, S. Huk and D. Bermbach, "MockFog: Emulating Fog Computing Infrastructure in the Cloud," *2019 IEEE International Conference on Fog Computing (ICFC)*, Prague, Czech Republic, 2019, pp. 144- 152, doi: 10.1109/ICFC.2019.00026.

[14] K. H. Miao, J. Miao, and G. Miao, "Diagnosing Coronary Heart Disease Using Ensemble Machine Learning," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 10, 2016.

[15] A. Alqhatani, S. B. Khan, "IoT-Driven Hybrid Deep Collaborative Transformer with Federated Learning for Personalized E-Commerce Recommendations: An Optimized Approach," *Scalable Computing: Practice and Experience*, vol. 25, no. 5, 2024.

[16] A. A. Khan, K. K. Almuzaini, V. D. J. Macedo, S. Ojo, V. K. Minchula, and V. Roy, "MaReSPS for energy efficient spectral precoding technique in large scale MIMO-OFDM," *Physical Communication*, vol. 58, 2023, doi: https://doi.org/10.1016/j.phycom.2023.102057.

[17] A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, Tehran, Iran, 2017, pp. 1011-1014, doi: 10.1109/KBEI.2017.8324946.

[18] M. Ganesan and N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878850.

[19] S. Shaik, J. Hall, C. Johnson, Q. Wang, R. Sharp, and S. Baskiyar, "PFogSim: A simulator for evaluation of mobile and hierarchical fog computing," *Sustainable Computing: Informatics and Systems*, vol. 35, p. 100736, Sep. 2022, doi: https://doi.org/10.1016/j.suscom.2022.100736.

[20] Y. Li, W. Qu, and Z. Zhang, "Intelligent Algorithm Operation and Data Management of Electromechanical Engineering Power Communication Network based on the Internet of Things," *Scalable Computing: Practice and Experience*, vol. 25, no. 5, 2024.

[21] O. U. Akgul, W. Mao, B. Cho, and Y. Xiao, "VFogSim: A Data-driven Platform for Simulating Vehicular Fog Computing Environment," 2022.

[22] V. Roy, P. K. Shukla, A. K. Gupta, V. Goel, P. K. Shukla, and S. Shukla, "Taxonomy on EEG Artifacts Removal Methods, Issues, and Healthcare Applications," *Journal of Organizational and End User Computing (JOEUC)*, vol. 33, no. 1,

pp. 19-46, 2021, doi: 10.4018/JOEUC.2021010102.
[23] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," *Expert Systems with Applications*, vol. 68, pp. 163–172, Feb. 2017, doi: https://doi.org/10.1016/j.eswa.2016.10.020.
[24] L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," in IEEE Access, vol. 7, pp. 54007-54014, 2019, doi: 10.1109/ACCESS.2019.2909969.
[25] R. Zhou, "Research on Intelligent Building Integrated Cabling System Based on Internet of Things and Deep Learning", Scalable Computing: Practice and Experience, Vol. 25 No. 5, 2024.
[26] S. Stalin, V. Roy, P. K. Shukla, A. Zaguia, M. M. Khan, P. K. Shukla, A. Jain, "A Machine Learning-Based Big EEG Data Artifact Detection and Wavelet-Based Removal: An Empirical Approach", Mathematical Problems in Engineering, vol. 2021, Article ID 2942808, 11 pages, 2021. https://doi.org/10.1155/2021/2942808.
[27] D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," Informatics in Medicine Unlocked, vol. 20, p. 100402, 2020, doi: https://doi.org/10.1016/j.imu.2020.100402.
[28] S. Gupta, "Classification of Heart Disease Hungarian Data Using Entropy, Knnga Based Classifier and Optimizer," International Journal of Engineering & Technology, vol. 7, no. 4.5, p. 292, Sep. 2018, doi: https://doi.org/10.14419/ijet.v7i4.5.20092.
[29] M. Jan, A. A. Awan, M. S. Khalid, and S. Nisar, "Ensemble approach for developing a smart heart disease prediction system using classification algorithms," Research Reports in Clinical Cardiology, vol. Volume 9, pp. 33–45, Dec. 2018, doi: https://doi.org/10.2147/rrcc.s172035.
[30] W. Liu, "The Trajectory Data Mining Model for College Students in Compus Life and Academic Management", Scalable Computing: Practice and Experience, Vol. 25, Issues 5, pp. 3225–3240, DOI 10.12694/scpe.v25i5.2349.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Aug 25, 2024
Accepted: Oct 14, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2118–2129, DOI 10.12694/scpe.v26i5.4733

A CHALLENGE-RESPONSE BASED AUTHENTICATION APPROACH FOR MULTIMODAL BIOMETRIC SYSTEM USING DEEP LEARNING TECHNIQUES

KHUSHBOO JHA*, ARUNA JAIN, AND SUMIT SRIVASTAVA

Abstract. Multimodal Biometric System (MBS) is an advanced progression of conventional biometric authentication system, which employ multiple biometric traits to enhance security. However, despite their advantages, these systems are vulnerable to presentation attacks, where adversaries use photos, replay videos or voice recordings to deceive the authentication process. Therefore, this paper proposes a challenge-response based approach using texture-based facial features and multidomain speech features. The challenge-response approach requires the user to utter a random word. Next, the system detects the user's facial features (eye and mouth motion) and recognized speech text to confirm whether the authentication request originates from a legitimate user or an imposter. The feature-level fusion via concatenation method is used to combine these image-audio features, to reduce the overlap within the feature spaces and data dimensionality. The fused feature vector is then fed into the deep learning-driven ensemble classifier CNN-BiLSTM to train and test the fused samples for user authentication. The performance evaluation is carried out using a self-built database with 55 users, achieving 96.81% accuracy, 98.20% precision and an Equal Error Rate (EER) of 3.37%. Moreover, the proposed approach surpasses different cutting-edge MBS, deep learning classifiers and image-audio fusion techniques on various performance metrics. Thus, the results underscore the effectiveness of the deep learning-based MBS in ensuring user authentication and spoof detection, demonstrating its considerable potential in bolstering the security of biometric systems against intricate presentation attacks.

Key words: challenge-response, multimodal biometric system, authentication, ensemble classifier, pattern recognition, deep learning

1. Introduction. With the growing need for secure user authentication [1] in digital systems, MBS has emerged as a vital solution. Unlike traditional systems that rely on a single biometric trait, such as speech (speaker) or face recognition, often face challenges related to accuracy and susceptibility to spoofing. MBS [2] integrate multiple modalities to enhance security, reliability and robustness. Such biometric based authentication [1] offers a promising alternative by leveraging unique physiological or behavioral traits for user authentication. Moreover, this makes biometric authentication systems crucial for ensuring security and verifying identities in various applications like access control, attendance monitoring, etc. Among the most common biometric modalities, face and speaker recognition are favored for their contactless nature and user convenience, making them suitable for various real-world applications. However, despite their advantages, these systems are vulnerable to presentation attacks, where adversaries use photos, replay videos, or voice recordings [3] to deceive the authentication process.

The challenge-response [2, 4] technique, a well-established method in cybersecurity, offers a promising solution to this problem by requiring users to provide dynamic inputs in response to randomly generated prompts. Also, it uses a two-way conversation between an authentication server and the user. Under these systems, the verifier creates an arbitrary challenge for the user. The user has to reply with a right response that shows their validity. The authentication server creates and delivers a random challenge to the user upon a client login request. To finish the authentication process, the user must then react appropriately for the difficulty. This approach has great benefits, especially in terms of resilience against sophisticated spoofing attempts, such video attacks, even if it adds some complexity since it depends on active user engagement and incurs extra operating expenses. This approach greatly improves the security of the authentication system by adding dynamic challenges, which introduce uncertainty.

Consequently, this work proposes an enhanced multimodal biometric authentication approach [2] combining face and speech modalities at the feature level, within a challenge-response framework [2, 4]. By leveraging

*Department of Computer Science and Engineering, Birla Institute of Technology, Ranchi, India (kjha.phd@gmail.com).

synchronized eye blink with texture based facial feature [3] and multidomain speech features [5, 10], the system detects liveness and prevents presentation attacks [4]. By integrating these complementary modalities, the proposed method aims to significantly improve the accuracy, robustness, and security of biometric authentication systems, leveraging deep learning techniques for feature extraction and classification to achieve state-of-the-art performance. The proposed approach aims to address several key challenges:

- Authentication Accuracy: By combining multiple biometric modalities, our approach enhances authentication accuracy compared to single-modal systems. Feature-level fusion of face and speech traits enables more robust and reliable user identification, reducing the risk of unauthorized access.
- Resilience to Spoofing Attacks: MBS are inherently more resistant to spoofing attacks compared to single-modal systems. By requiring simultaneous presentation of face and speech traits, our approach mitigates the risk of spoofing attempts, such as using forged images or recordings.
- User Experience: We prioritize user experience in our security approach by leveraging natural and intuitive authentication mechanisms. Users can authenticate themselves using familiar actions like speaking a passphrase while facing a camera, minimizing the cognitive burden and increasing acceptance of security measures.
- Privacy Preservation: Unlike some other biometric modalities, such as fingerprints or iris scans, face and speech traits can be captured without physical contact, preserving user privacy and hygiene. Our approach ensures compliance with privacy regulations and minimizes concerns about intrusive surveillance.
- Adaptability to Environmental Factors: Diverse environments with varying lighting conditions, ambient noise levels and user interactions. This MBS is designed to be robust and adaptable to such environmental factors, ensuring consistent performance across different scenarios.

By addressing these challenges and objectives, this research aims to advance the state-of-the-art in MBS and contribute towards the development of more robust and reliable authentication mechanisms for various real-world applications. Face recognition captures spatial characteristics of an individual's face, while speech recognition analyzes vocal patterns and linguistic features, providing additional layers of security through multimodal fusion. The rationale behind combining these two modalities lies in their complementary nature and potential for greater authentication precision and resistance to spoof attempts. From computational perspective, training and evaluation of the proposed MBAS is done using self-built dataset for Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) classifier. In this context, our research focuses on developing a novel security approach using a MBS based on feature-level fusion of face-speech traits. This research contributes to the advancement of biometric security by introducing a novel MBS [4] based on feature-level fusion [5, 6] of face-speech traits, offering enhanced protection against cyber threats and ensuring the integrity and reliability for authentication based applications.

Contribution of this work is:

- To design and implement a novel MBS that combines multidomain acoustic and texture based facial features with eye blink movement at the feature level, for robust authentication.
- Deep learning based ensemble classifier CNN-BiLSTM for efficient classification of feature vectors thereby improving the performance of the authentication system.
- Challenge-response approach to verify the liveness of the user, thereby improving authentication accuracy and robustness in biometric systems.

Deploying a challenge-response based multimodal biometric system using face and speech in real-world scenarios presents a number of practical implications that need to be carefully considered, particularly concerning scalability and user acceptance. Below are key aspects to address in order to ensure the system's successful deployment and adoption;

Scalability across diverse use cases: The system needs to be adaptable to different scales of implementation, from small-scale applications like personal devices and secure office environments to large-scale implementations such as public venues and national ID systems. The system's architecture should be flexible to scale up or down based on the specific needs of the environment.

User acceptance: For any biometric system to gain widespread user acceptance, it must be easy to use. The challenge-response mechanism should be intuitive and user-friendly, particularly when asking users to

perform tasks such as responding to a spoken prompt or making specific facial gestures. Ensuring that these prompts are simple and non-intrusive is critical for encouraging regular use.

2. Literature review. For liveness detection [3], most of the work either use speech [10] or face recognition [3]. Using speech recognition [4] technology, spoken language is first detected and parse into individual words and sentences. Speech recognition operates on audio data and does not require visual information. Nevertheless, there exists a correlation and complementarity between visual and vocal data, namely in relationship to enhance the performance and security of biometric systems. There exist a limited number of publications dedicated to the multimodal biometric liveness detection [3] of faces and voices. Blanchard et al., [8] developed a memory creation and biometrics-based approach to avoid identity theft. This authentication approach combines ocular biometrics and challenge systems, as well as a new biometric characteristic called pupil memory effect. Credentials can be revoked at any time without any loss to the user, and the approach can be changed for varying levels of security. The technique assesses security and performance and recommends ways to improve deployment.

Addressing photo and video-attacks in face recognition systems, Chou et al., [4] proposed a score-level fusion and challenge-response scenario multimodal presentation attack detection (PAD) method . When presented with a challenge-response situation, the user is asked to say a series of randomly generated words. The liveness of the user is confirmed by detecting both their mouth motion and the specified speech text concurrently. In terms of improving the security of facial recognition systems, experimental results using a selfmade dataset show that the suggested strategy achieves the best half total error rate of 3.64%. Haasnoot et al., [9] present a rigorous and systematic framework and classification system for challenge-response protocols. The classification is validated by comparing it with published literature that specifically describes Biometric Challenge-Response Protocols (BCRP). Lastly, analyze the advantages of robust BCRPs over PAD approaches, particularly in safeguarding individual applications and preventing unintended leaks in BCRP applications.

Hanumanthaiah et al., [2] proposed a multi-modal biometric system that uses face and iris biometrics and challenges users with an emotion-invoking image at random on the screen. Response metrics include changes in the region of interest and distance between iris and facial landmarks. To detect presentation attacks, the response is compared to the expected response, which is determined by the position of an arbitrary emotion-invoking image on screen. In addition, the paper presents a hybrid deep feature that adds security to the recognition process by ensuring authentication failure for fraudulent samples. The proposed technique has a very low error rate of 0.95% in spoof leakage in varied situations when compared to prior works. However, the proposed work employs eye movements in tandem with texture-based face features, speech and speaker recognition as the challenge-response to authenticate the user's identity. For speaker recognition multi-domain based acoustic features [10] are used to enhance both temporal dynamics and spectral characteristics, improving robustness and discrimination between training and testing samples. Enhanced performance is achieved by integrating face and voice information through feature level fusion and ensemble deep learning based CNN-BiLSTM classifier.

3. Problem Formulation. Presentation attacks [4, 9] pose a significant threat to biometric authentication systems, where adversaries attempt to bypass security measures using fake biometric samples. Existing single-modality systems, such as those relying solely on facial recognition, are particularly vulnerable to such attacks. Moreover, traditional liveness detection methods may fail when faced with sophisticated spoofing techniques, such as high-quality deepfake videos or carefully crafted voice recordings. To address these challenges, this research formulates the problem of designing a resilient multimodal biometric authentication system capable of countering both photo and video-based presentation attacks. The proposed solution leverages a challenge-response [2, 4, 9] scenario, where users are prompted to perform specific actions, such as speaking randomly generated words while undergoing real-time facial analysis. By synchronizing speech and facial motion features and employing feature-level fusion [19, 21, 22], the system robustly verifies liveness. Moreover, this allows for more effective utilization of information from each modality and facilitates the extraction of discriminative characteristics that may not be evident when considering modalities in isolation. When processing individual modalities, some features may be redundant or not useful for classification. For instance, certain facial features (like background elements in the image) or irrelevant speech features (such as noise) might not contribute positively to the authentication process. By fusing the two modalities at the feature level, the system is better positioned to discard or minimize the impact of such irrelevant features through dimensionality reduction

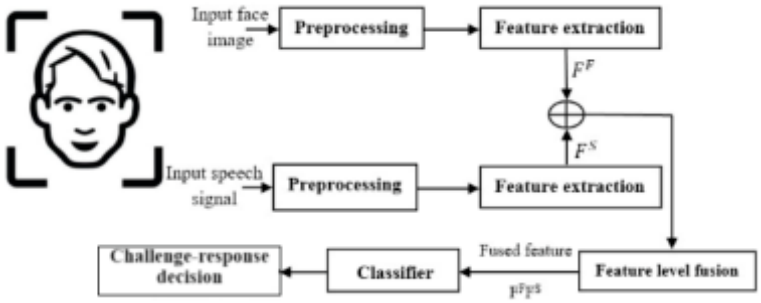


Fig. 4.1: Multimodal biometric system.

technique like Linear Discriminant Analysis (LDA) [5]. LDA focus on maximizing class separability by emphasizing features that are most relevant for distinguishing between individuals, thereby reducing feature overlap between modalities and improving the clarity of the final decision-making. Therefore, the goal is to enhance security without compromising usability, ensuring that the authentication process remains both effective and user-friendly in diverse application scenarios. The research aims to achieve a high level of accuracy, precision, and resilience against presentation attacks while maintaining a seamless user experience.

4. Proposed Method. A challenge-response [2, 4] scenario forms the foundation of the proposed approach. Upon user login request to a MBS, the proposed system generate a random word (challenge) which the user needs to respond (speak as response) accordingly within a specified duration. The dynamic texture based facial features (including mouth movement, eye blink) and the speech based user's answer (the requested word) is computed. The fused image-audio template undergo training/testing phase and compared (with the stored template in the reference database) by using deep learning ensemble classifier to ascertain the authenticity of the login user, as shown in Figure 4.1. As per the threshold value. the challenge-response decision i.e., whether the respondent is alive/legitimate or not is taken. If not, then the session expires. Else, the login requested is accepted and the user is successfully authenticated.

Two pivotal stages, training and testing [6, 10], are crucial components in the user authentication process for any MBS. In the training phase, users commence the procedure by completing the registration application. This approach entails recording a short video utilizing an intelligent sensor camera (in-built laptop camera with microphone) and the speech input is preprocessed to acquire reliable speech characteristics. Simultaneously, we extract a facial image frame from the motion picture to facilitate efficient feature extraction. In order to achieve feature-level fusion, the feature vectors obtained from both modalities are combined, with the goal of using the complimentary information offered by the aural and visual cues. Furthermore, a classifier is used to train these combined feature templates. These templates that have undergone training are then saved in a database for the following testing phase. During the testing step, which replicates the training procedure, the genuine identity as well as liveness of the claimed user is verified. The user inputs their visual-audio biometric template as per the given challenge which is then fused and compared to the trained benchmarks stored in the database. This comparison is conducted via a one-to-one approach [6, 10], where the claimed user's template is compared to all registered templates in order to establish its authenticity. The ultimate determination of the user's legitimacy is established by evaluating the level of resemblance between the request template and the source data, as assessed by predetermined performance standards.

4.1. Pre-processing. In face recognition-based image processing, high-quality pre-processing procedures precede feature extraction stage and frequently encompass many steps aimed at enhancing the quality and accuracy of the image data. This study employs data augmentation methods including scaling, flipping and rotation to enhance dataset diversity, hence improving model robustness and generalization. These pre-processing methods improve feature extraction and analysis by reducing the effects of illumination, perspective, and other environmental factors. Similarly, speech pre-processing improves speech quality and discrimination. Therefore, cutting-edge noise reduction technique is employed using Python programming prior to feature extraction. It

eliminates noise, filters out extraneous frequencies, and normalizes amplitudes to improve signal-to-noise ratios.

4.2. Feature extraction.

4.2.1. Facial image. For face recognition [11], the Active Appearance Model (AAM) [12] stands as a pre-eminent algorithm for facilitating model-assisted object tracking and detection within the realm of computer vision. AAM, renowned for its efficacy, constructs representations of both texture and shape for a given object by synthesizing a compilation of real images. This synthesis culminates in a comprehensive depiction of detailed texture features, grounded in the patterns of intensity and color inherent to the object.

The crux of AAM's functionality (using 20 landmark points) lies in its adeptness at minimizing disparities between newly acquired images and synthesized counterparts. Operating under the premise that interpretation can be framed as an optimization endeavor, AAM endeavors to reduce differences through iterative adjustments. This process is encapsulated in the definition of a difference vector, denoted as δI is defined in equation (1), which delineates discrepancies between the grey level value (I_i) vectors of the preprocessed input face image (I_p^F) and I_m is the grey level vector of present model parameter. Through the iterative manipulation of the model parameter (c), AAM endeavors to ascertain the optimal alignment between model and image (I_p^F). Such that it matches each other the best by minimizing the size of difference vector $\Delta = |\delta I|^2$.

$$\delta I = I_i - I_m \tag{4.1}$$

Moreover, in this work AAM method is combined with Eye Aspect Ratio (EAR) [13], which serves as an effective tool for liveliness detection due to its sensitivity to eye movements, particularly opening and closing actions. Employing six facial landmarks around the eyes, EAR exhibits significant variations during these movements, rendering it robust for blink identification. Through the flashing technique, EAR values fluctuate noticeably, offering insights into dynamic facial expressions. This technique underscores the reliability of EAR in assessing facial liveliness, thereby contributing to advancements in biometric systems and human-computer interaction.

Therefore, AAM [24] with EAR serves as a cornerstone in the realm of facial feature analysis, by virtue of its capability to delineate and align intricate features with precision. Its scholarly approach underscores its significance in enabling nuanced interpretations and facilitating sophisticated applications within the domain of computer vision.

4.2.2. Speech signal. In the realm of speaker recognition [5, 6, 10, 14], multi-domain [10] feature extraction techniques play a vital role in capturing various aspects of the speech signal. Therefore, inspired by this, our previous work [10] has been selected for the purpose of implementing a comprehensive speech feature. The unique aspect of this study is in its investigation and incorporation of multidomains acoustics, namely, strategies for extracting features in the cepstral, frequency and time domains [10]. The objective is to augment the discriminative capability of the deep learning classifier by integrating several domains, thereby enhancing the performance of the biometric authentication system.

Thus, let us expound upon the work and analyze the benefits of each domain-specific characteristics: The first component is cepstral features, Power Normalized Cepstral Coefficient (PNCC) [6, 10], which produces the most precise modeling of human hearing by using an asymmetric noise suppression module to reduce the influence of ambient noise and reduce computing complexity. This allows PNCC to better represent the spectral characteristics of speech, particularly in noisy environments, resulting in improved robustness and discrimination. This study has considered the first thirteen PNCC features, which are sufficient for expressing the spectral characteristics of speech signals. The second frequency domain characteristic, known as the spectral peak-based Formant feature [6, 10], pertains to the resonance frequencies of the vocal tract that are unique to each speaker. These frequencies include discriminative information for correct articulation and perception, which are resistant to noise. By extracting these frequencies, the characteristic features of vowels and consonants can be captured, providing valuable information for tasks such as speaker authentication. We employed a sliding Hamming window with a duration of 30 milliseconds, an overlap of 20 milliseconds, and 128 Mel-filter bins spanning frequencies from 50Hz to 4 kHz. We have extracted three formant features. Zero Crossing Rate (ZCR) [6, 10], the third time domain feature, analyzes a signal in the time domain to identify its salient characteristic. ZCR quantifies the rate at which the speech signal changes sign (i.e., crosses the zero axis) over time. ZCR is particularly useful for differentiating between voiced and unvoiced speech segments and can serve as an indicator

of speech rate and energy distribution. We have used the Librosa library function in Python 3 to extract one zero crossing feature.

Moreover, these feature extraction approaches are cutting-edge in their respective domain [10] and, after amalgamation, a resilient and effective feature set is acquired as speech feature. By combining PNCC, Formant frequencies and ZCR, a comprehensive representation of the speech signal consisting of 17 features [10] is obtained, capturing ceptral, frequency and time domain characteristics. These multi-domain features [10] are instrumental in improving the performance of speaker recognition tasks.

4.3. Feature-level fusion. The Feature-level fusion [15, 16, 24] effectively combines the extracted features of facial image and speech modalities, capturing their complementary nature. This enables a more comprehensive depiction of the information beneath, improving the system's capacity to identify intricate connections among various modalities. Additionally, it decreases the number of dimensions in the combined feature set relative to alternative fusion like decision level [16], potentially resulting in enhanced computational efficiency and decreased complexity in later processing stages. In addition, it allows for smooth integration of various feature sets, allowing for the inclusion of specialized knowledge and providing flexibility in designing models. The concatenation method used in feature-level fusion for integrating speech-image features improves the system's reliability, discriminative capability and adaptability in MBS..

4.4. Classification. To get the final classified output (approved or unauthorized) for the user authentication, an ensemble classification technique is used. The particular features influence the selection of the CNN [3, 6] and Bi-LSTM [17]: The CNN classifier analyze the spatial characteristics of the fused feature vector. Convolutional layers selectively capture significant patterns and correlations present in the feature space. The BiLSTM classifier [17] captures temporal dependencies by bidirectionally processing sequential information. Moreover, for the temporal nature of both facial movements (e.g., blinking, nodding) and speech patterns, the Bi-LSTM network is employed to model sequential dependencies in the data. This helps to capture subtle variations in both face and speech, leading to more accurate classification even when there are changes in environmental conditions or user demographics. As a consequence, the ensemble classifier improves the robustness by combining the outputs of face-speech, each trained on different environmental conditions or demographic data. This mitigates the biases introduced by training on a specific set of conditions or users. Thus, the ensemble CNN-BiLSTM classifier prioritises the correlations between spatial and temporal features, which are crucial for effective classification processing. This technique takes the fused feature vector $F^F F^S$ and feeds it into ensemble CNN-BiLSTM classifier.

CNN layer description
- Layers: 3 convolutional layers.
- Filter Sizes: (3x3), (5x5).
- Pooling: Max pooling after each convolution.
- Activation: ReLU.
- Dropout: 0.3 to avoid overfitting.

BiLSTM layer description
- Layers: 2 BiLSTM layers.
- Units: 128 units per layer.
- Directionality: Bidirectional (forward and backward).
- Activation: tanh.

Ensemble Strategy
- Fusion Method: Weighted combination of CNN and BiLSTM outputs.
- Weights: Tuned based on validation performance.

Fully Connected (Dense) Layer
- Units: 64.
- Activation: ReLU.
- Dropout: 0.3.

Output Layer
- Units: 2 (binary classification: authenticated or not).

- Activation: Softmax for probability output.
- Loss Function: Categorical Cross-Entropy.
- Optimizer: Adam.
- Learning Rate: 0.001.

5. Result and Discussion.

5.1. Database. There are several well-known databases available for presentation attack detection [4] focusing on video-attacks or photo-attacks. Unfortunately, the video data do not have speech/voice information and they are not compatible with the challenge-response [2, 4] method being proposed in this research. Therefore, a self-made video database is prepared for this research work. The database is compiled using data collected from 55 subjects. A laptop equipped with a high-quality camera with a microphone is used to capture user's videos (with required audio) from a distance of approximately 40cm. Videos are captured in a controlled indoor environment with consistent lighting conditions at a frame rate of 36 frame per second. Total 550 videos, consisting of 385 authentic clips and 165 spoof videos. All users were asked to speak a simple word from the predefined word database to ensure authenticity of the clips. Each segment lasts for 6 to 8 seconds, with each word being repeated twice. There are two video-attack scenarios for the spoof clips: (1) gazing into the camera without uttering a word; (2) gazing into the camera and speaking words that are not included in the predetermined word database. The experiments are conducted using various tools and libraries such as Python, PyAudio, Dlib, Google translate API, OpenCV and scikit-learn. Few deep learning classifiers are chosen as baseline for this research: Support Vector Machines (SVM) [18], Random Forests (RF) [18], CNN [3, 6] and BiLSTM [17]. These classifiers will be compared to determine their effectiveness. A carefully curated database is used to divide the legitimate and spoofed videos into a training subset and a testing subset, ensuring a balanced representation of both. For experimental observation, 80% of the final database used for training and remaining 20% for testing of the proposed approach using ensemble classifier.

5.2. Assessment of proposed MBS with UBS using speech and face. In this comparative analysis, the performance of an Unimodal Biometric System (UBS) for speech and face, specifically speaker recognition [6] and face recognition [7], is compared against the proposed MBS that combines speaker and face recognition modalities [26] for user authentication. The evaluation as shown in Table 5.1 focuses on accuracy [10] and EER metrics [10] to assess the efficacy of each system. By examining the accuracy of individual biometric modalities and their fusion in the MBS, insights into the system's robustness and reliability are garnered. Moreover, the EER metric [10] offers a comprehensive measure of the system's ability to balance false acceptance [10] and false rejection rates [10].

Through this comparative analysis, a comprehensive understanding of the performance advantages and limitations of unimodal versus MBS is elucidated, providing valuable insights for the advancement and optimization of authentication technologies in real-world applications.

5.3. Assessment of proposed approach with different levels of fusion of face-speech feature. The analysis of our proposed approach encompasses various levels of fusion, including rank, sensor, decision, feature, and score fusion [16, 26] of speech-face features. Through meticulous examination and experimentation, we evaluate the performance and efficacy of each fusion level in enhancing authentication accuracy, resilience to spoofing attacks and user experience. By scrutinizing the outcomes of different fusion strategies as shown in Table 5.2 using various standard metrics [10] such as such as accuracy [10], sensitivity [10], specificity [10],

Table 5.1: Comparison of proposed MBS with UBS.

Biometric system	Features used	Accuracy (in %)	EER (in %)
UBS using speech	PNCC+Format+ZCR	94.67	10.14
UBS using face	AAM+EAR	93.81	15.29
MBS (proposed)	**Feature level fusion of the above features**	**96.81**	**3.37**

Table 5.2: Assessment of proposed approach with different levels of fusion (in %).

Performance metrics	Decision	Sensor	Rank	Score	**Feature (proposed)**
Accuracy	93.09	92.70	89.41	90.24	**96.81**
Sensitivity	96.62	93.72	88.92	87.35	**98.22**
Specificity	87.83	89.81	82.58	86.85	**95.00**
Precision	97.81	95.48	90.14	87.27	**98.20**
F-measure	96.51	94.16	89.43	88.75	**98.03**
MCC	89.52	91.53	88.50	86.52	**97.62**
NPV	86.33	84.26	89.54	90.64	**93.05**
FPR	10.16	9.05	11.22	15.64	**5.52**
FNR	6.37	8.73	8.60	7.87	**1.23**
EER	8.76	8.89	9.91	11.75	**3.37**

Fig. 5.1: Graphical presentation of proposed approach compared with different levels of fusion (in %).

precision [10], F-measure [10], Mathew's Correlation Co-efficient (MCC) [10], Negative Predictive Value (NPV) [10], False Positive Rate (FPR) [10], False Negative Rate (FNR) [10] and EER [10] are evaluated. Therefore, the aim of this research is to identify optimal configurations that maximize security and usability while maintaining compliance with regulatory standards. This comprehensive analysis provides valuable insights into the strengths and limitations of each fusion level, guiding future advancements in security.

To enhance the comprehension of the data reported in Table 5.2, we have included a comparative analysis in Figure 5.1. This figure demonstrates the performance of the proposed methodology (in %), for various levels of fusion.

5.4. Assessment of proposed MBS with various state-of-the-art classifiers. The analysis of our proposed approach extends to the evaluation of various classifiers [24, 25] to ascertain their efficacy in the context of our MBS. Through rigorous experimentation and comparative assessment [19, 20], we scrutinize the performance of different classifiers, including but not limited to Support Vector Machines (SVM) [18], Random Forests (RF) [18], CNN [8, 9], Bi-LSTM and CNN-BiLSTM classifiers as shown in Table 5.3. By examining metrics such as accuracy, Equal Error Rate (EER), and other computational efficiency, we aim to identify the classifier that offers the optimal balance of performance, robustness, and scalability for our proposed system. Furthermore, our analysis delves into the classifier's ability to handle the intricacies of multimodal data fusion, including feature-level integration of speech and facial features, and its resilience against spoofing attacks. Through this comprehensive evaluation, we seek to provide valuable insights into the suitability of different classifiers for enhancing security and mitigating cyber threats [1]. Ultimately, our findings

Table 5.3: Comparison of proposed approach (in %) with cutting-edge classifier based MBS.

Performance metrics	RF	SVM	CNN	BiLSTM	**CNN-BiLSTM (proposed)**
Accuracy	88.11	91.08	94.15	95.30	**96.81**
Sensitivity	89.62	92.72	93.92	95.52	**98.22**
Specificity	85.83	88.61	89.58	92.85	**95.00**
Precision	90.81	94.33	94.14	95.07	**98.20**
F-measure	89.02	93.16	93.43	95.75	**98.03**
MCC	79.65	77.35	91.05	95.17	**97.62**
NPV	80.89	83.26	89.54	92.61	**93.05**
FPR	10.16	9.23	8.61	7.46	**5.52**
FNR	8.37	6.73	4.07	4.87	**1.23**
EER	9.26	7.98	6.34	6.16	**3.37**

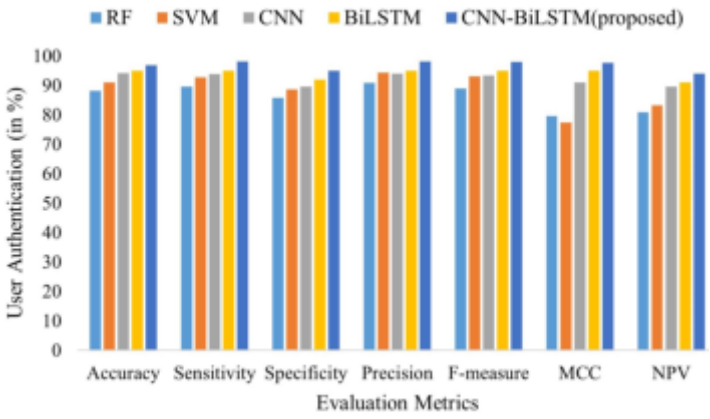


Fig. 5.2: Graphical presentation of proposed approach with different cutting-edge classifier based MBS (in %).

contribute to the advancement of biometric authentication systems for safeguarding critical infrastructure in digital authentication environments.

To enhance the comprehension of the data reported in Table 5.3, we have included a comparative analysis in Figure 5.2. This figure demonstrates the performance of the proposed methodology (in %), for different classifier based MBS.

5.5. Assessment of MBS with a few cutting-edge methods. The proposed MBS illustrates significant advancements in the field of MBS [26], by investigating a challenge-response approach that combines facial features and multidomain audio evaluation for user authentication. The existing literature primarily emphasizes the integration of facial expressions with other modalities apart from speech, or vice versa, as demonstrated in Table 5.4. Abinaya et al., [19] developed an MBS that integrates various modes of behavior, such as keystroke and speech features. This system utilizes advanced DL algorithms to precisely recognize individuals. The attributes from the two modalities were combined via a weighted linear method of feature-level fusion. The combined features were trained using a CNN classifier. Likewise, Abdulbaqi et al., [20] have proposed a system that employs Awica Wavelet Transform (AWT) algorithms to analyze the uniqueness of an individual's ECG signal in combination with their face features in order to verify users. Nevertheless, the achieved classification accuracy was only 94%. Rahman et al., [21] recently demonstrated a method for feature-level fingerprint and electrocardiogram (ECG) fusion using CNN classifier. Experiments show that the proposed technique has a 94.5% accuracy rate.

Vekariya et al., [22] presented a technique for multi-biometric authentication that also incorporates feature-

Table 5.4: Comparison of the proposed MBS with a few cutting-edge methods.

Ref.	Biometric trait	Fusion type	Database used	Method	Accuracy (in %)	EER (in %)
[19] 2022	Speech & keystroke	Feature	BioChaves	MFCC and press/release timestamp feature with CNN classifier	91.50	N/A
[23] 2022	Face & ear	Score	ORL (face) & IIT Delhi (ear)	DWT technique with ANFIS classifier	96.24	N/A
[20] 2023	Face & ECG	Decision	Self-made	AWT technique with DNN classifier	94.00	52.96
[21] 2024	ECG & fingerprint	Feature	MIT-BIH(ECG) & FVC2004 (fingerprint)	Deep embedding with CNN classifier	94.50	N/A
[22] 2024	Face & fingerprint	Feature	SDUMLA	BCOAK based feature with SVM classifier	96.00	N/A
Our	**Face & speech**	**Feature**	**Self-made**	Multi-domain speech feature with AAM-EAR with CNN-BiLSTM classifier	**96.81**	**3.37**

Abbreviations: N/A denotes Not Applicable, meaning that no such parameter was evaluated by the method.

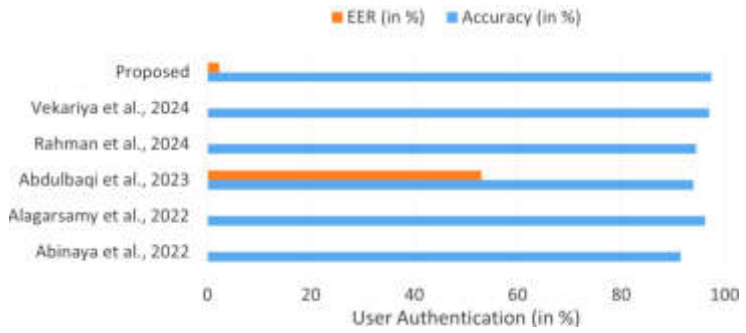


Fig. 5.3: Graphical assessment of proposed MBS with a few cutting-edge MBS.

level fusion. The proposed method employs a cutting-edge Binary Chimp Optimized Adaptive Kernel (BCOAK) SVM to identify the most impactful features. As per the experimental findings the system achieved an accuracy of 97%. Alagarsamy and Murugan [23] proposed a MBS combining ear and facial recognition using innovative approaches to improve identification accuracy. The system starts with preprocessing, ring projection, and data normalization. Discrete Wavelet Transform (DWT) feature extraction is used followed by Adaptive-Network- based Fuzzy Inference System (ANFIS) classifier to complete the identification process. The method calculates individual scores based on ear and facial feature matching results and fuses them to improve biometric identification. However, throughout history, people have traditionally relied on identifying others based on facial or speech signals, rather than using characteristics such as ears, hands, signatures, or fingerprints. Although prior studies have provided a solid foundation for applications of MBS, there has been limited research on the combination of face and speech interaction. This study's contribution is therefore original and significant.

In order to enhance the comprehension of the data shown in Table 5.4, we have included a comparative analysis in Figure 5.3. The following illustration indicates the efficacy of the proposed methodology in conjunction with several cutting-edge MBS. In a challenge-response MBS using face-speech, advanced deepfake techniques [27] present several vulnerabilities that could undermine the system's security. Deepfakes, which use artificial intelligence [28] to synthetically generate highly realistic facial images and voices, pose a significant threat to biometric authentication [1]. Therefore, to mitigate these vulnerabilities, combining several liveness

checks across modalities like ours (challenge-response i.e., simultaneous face and voice interactions) makes it harder for deepfakes to spoof [3] both aspects accurately. Moreover, the pilot study findings show that the proposed technique obtains impressive reliability and accuracy.

6. Conclusion. This research propose a novel authentication approach using MBS and integrating challenge-response technique using texture based facial and multidomain speech features. Leveraging facial and speech feature being fused at feature level and classified using deep learning based ensemble CNN-BiLSTM classifier, our approach demonstrates robustness and adaptability in securing various applications. Extensive experimentation and evaluation using the self-made database showcased remarkable results, with a computational perspective revealing a high accuracy rate of 96.81%, precision of 98.20% and an impressively low EER of 3.37%. Moreover, our approach surpasses different cutting-edge MBS, deep learning classifiers and image-audio fusion techniques on various performance metrics, validating its effectiveness in real-world scenarios. The proposed approach addresses the critical challenges of authentication accuracy, resilience against spoofing attacks and user experience enhancement while upholding privacy and regulatory compliance standards. These findings underscore the potential of our proposed methodology to effectively mitigate security risks and safeguard critical infrastructure within authentication environments. Thus, this research contributes to advancing the security landscape of internet based authentication environment, paving the way for enhanced protection and resilience against cyber threats in the rapidly evolving digital landscape. In the future, we plan to analyze and implement the proposed approach using dataset with users having speech or visual impairments.

REFERENCES

[1] Jha, K., Jain, A., & Srivastava, S. (2024). A Secure Biometric-Based User Authentication Scheme for Cyber-Physical Systems in Healthcare. Int. J. Exp. Res. Rev, 39, 154-169.
[2] Hanumanthaiah, A. K., & Eraiah, M. B. (2022). Challenge Responsive Multi Modal Biometric Authentication Resilient to Presentation Attacks. International Journal of Intelligent Engineering & Systems, 15(2).
[3] Jha, K., Srivastava, S., & Jain, A. (2024). A Novel Texture based Approach for Facial Liveness Detection and Authentication using Deep Learning Classifier. International Journal of Computational and Experimental Science and Engineering, 10(3).
[4] Chou, C. L. (2021). Presentation attack detection based on score level fusion and challenge-response technique. The Journal of Supercomputing, 77(5), 4681-4697.
[5] Jha, K., Jain, A., & Srivastava, S. (2024). Analysis of Human Voice for Speaker Recognition: Concepts and Advancement. Journal of Electrical Systems, 20(1), 582-599.
[6] Jha, K., Jain, A., & Srivastava, S. (2023, March). An Efficient Speaker Identification Approach for Biometric Access Control System. In 2023 5th International Conference on Recent Advances in Information Technology (RAIT) (pp. 1-5). IEEE.
[7] Jha, K., Srivastava, S., & Jain, A. (2023, March). Integrating Global and Local Features for Efficient Face Identification Using Deep CNN Classifier. In 2023 International Conference on Device Intelligence, Computing and Communication Technologies,(DICCT) (pp. 532-536). IEEE.
[8] Blanchard, N. K., Kachanovich, S., Selker, T., & Waligorski, F. (2020). Reflexive memory authenticator: a proposal for effortless renewable biometrics. In Emerging Technologies for Authorization and Authentication: Second International Workshop, ETAA 2019, Luxembourg City, Luxembourg, September 27, 2019, Proceedings 2 (pp. 104-121). Springer International Publishing.
[9] Haasnoot, E., Spreeuwers, L. J., & Veldhuis, R. N. (2022). Presentation attack detection and biometric recognition in a challenge-response formalism. EURASIP Journal on Information Security, 2022(1), 5.
[10] Jha, K., Srivastava, S., & Jain, A. (2025). A novel speaker verification approach featuring multidomain acoustics based on the weighted city block Minkowski distance. ETRI Journal, 47(2), 227-243, DOI 10.4218/etrij.2023-0485.
[11] Stasiak, L. A., & Pacut, A. (2010). Face Tracking and Recognition with the Use of Particle-Filtered Local Features. Journal of Telecommunications and Information Technology, (4), 26-36.
[12] Ueda, Y., Nakamura, K., Saegusa, C., & Ito, A. (2023). Recent advances and future directions in facial appearance research. Frontiers in Psychology, 14, 1154703.
[13] Hutamaputra, W., Utaminingrum, F., Budi, A. S., & Ogata, K. (2023). Eyes gaze detection based on multiprocess of ratio parameters for smart wheelchair menu selection in different screen size. Journal of Visual Communication and Image Representation, 91, 103756.
[14] Jain, P., Kasture, N. R., & Kumar, T. (2020). Comparative study of speaker recognition techniques in IoT devices for text independent negative recognition. Scalable Computing: Practice and Experience, 21(3), 359-368, DOI 10.12694/scpe.v21i3.1704
[15] Wang, Y., Zhou, Y., & Wang, B. (2024). Dimension Extraction of Remote Sensing Images in Topographic Surveying Based on Nonlinear Feature Algorithm. Scalable Computing: Practice and Experience, 25(5), 4246-4254.
[16] Dalila, C., Saddek, B., & Amine, N. A. (2020). Feature level fusion of face and voice biometrics systems using artificial neural network for personal recognition. Informatica, 44(1).

[17] Srikanth, J., & Shanmugam, A. D. (2023). A Deep LSTM-RNN Classification Method for Covid-19 Twitter Review Based on Sentiment Analysis. Scalable Computing: Practice and Experience, 24(3), 315-326.
[18] Bharathi, V. (2024). Vulnerability detection in cyber-physical system using machine learning. Scalable Computing: Practice and Experience, 25(1), 577-591.
[19] Abinaya, R., Indira, D. N. V. S. L. S., & Swarup Kumar, J. N. V. R. (2022, February). Multimodal Biometric Person Identification System Based on Speech and Keystroke Dynamics. In International Conference on Computing, Communication, Electrical and Biomedical Systems (pp. 285-299). Cham: Springer International Publishing.
[20] Abdulbaqi, A. S., Turki, N. A., Obaid, A. J., Dutta, S., & Panessai, I. Y. (2023). Spoof Attacks Detection Based on Authentication of Multimodal Biometrics Face-ECG Signals. In Artificial intelligence for smart healthcare (pp. 507-526). Cham: Springer International Publishing.
[21] A. El-Rahman, S., & Alluhaidan, A. S. (2024). Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments. Plos one, 19(2), e0291084.
[22] Vekariya, V., Joshi, M., & Dikshit, S. (2024). Multi-biometric fusion for enhanced human authentication in information security. Measurement: Sensors, 31, 100973.
[23] Alagarsamy, S. B., & Murugan, K. (2022). Multimodal of ear and face biometric recognition using adaptive approach Runge–Kutta threshold segmentation and classifier with score level fusion. Wireless Personal Communications, 124(2), 1061-1080.
[24] Jha, K., Jain, A., & Srivastava, S. (2024). A Contactless Speaker Identification Approach Using Feature-Level Fusion of Speech and Face Cues with DCNN. Proceedings on Engineering Sciences, 6(3), pp. 1047–1056.
[25] Agrawal, S. S., Jain, A., & Sinha, S. (2016). Analysis and modeling of acoustic information for automatic dialect classification. International Journal of Speech Technology, 19, 593-609.
[26] Jha, K., Jain, A., & Srivastava, S. (2024). Feature-level fusion of face and speech based multimodal biometric attendance system with liveness detection. AIP Advances, 14(11), 115007.
[27] Sharma, V. K., Garg, R., & Caudron, Q. (2024). A systematic literature review on deepfake detection techniques. Multimedia Tools and Applications, 1-43.
[28] Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. Expert Systems With Applications, 124260.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Aug 27, 2024
Accepted: Oct 2, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2130–2144, DOI 10.12694/scpe.v26i5.4786

ENSEMBLE TRANSFER LEARNING FOR AUTOMATED GAUGE READING DETECTION AND PREDICTION

HITESH NINAMA* JAGDISH RAIKWAL† AND PUSHPA RAIKWAL‡

Abstract. Pressure gauges and automatic reading methods for pointer gauges were the root causes of the problems that have motivated the development of an ensemble transfer learning strategy. The study suggests that to effectively generate and predict the present measurements of the gauge, it is necessary to use an ensemble learning method that incorporates transfer learning framework designs such as InceptionResnetV2 or DenseNet 201. The suggested methodology involves integrating the given data with ensemble model architectures and qualifying InceptionResnetV2 and Dense Net 201 models to forecast the present gauge value. The main focus of the proposed method is to develop a final angle with a clear reading, as well as improve the picture through techniques like enhancing its shape, size, and resolution. Image processing approaches help to achieve these objectives. The ensemble model achieved an accuracy of 98.34%, whereas the InceptionResNetV2 model experienced a loss of 8.34% in contrast stretching.

Key words: Deep Learning, Transfer Learning, Ensemble Model, Gauge Reading Prediction.

1. Introduction. In fields like science and engineering, a gauge serves as an instrument for measuring or delivering information about a specific parameter. There exists a wide variety of instruments for such tasks, ranging from basic materials employed as benchmarks to advanced technological systems [1],[2]. There are many different uses for the term "gauge," which can refer to anything from "a device that measures a physical quantity" to "a device to determine thickness, a gap in space," or "a device that displays the measurement of a tracked system via a needle or pointer that in fact moves along a calibrated scale. Regardless of their application, all gauges fall into one of the four broad categories: mechanical, digital, analog, and hybrid systems. An analog instrument is a type of measuring device that produces a continuous output, which changes proportionally to the quantity being measured. On the other hand, a "digital instrument" refers to any device that generates an output in discrete steps and has a finite set of possible values [7]. Meters and thermometers with analog displays, often referred to as "needle" readouts, represent one category. These analog devices were the most common basic form until fairly recently, when digital technology began to prevail [3],[4]. The digital meter under consideration is designed to present data in an analog format, resembling the appearance of a "analog meter" commonly found in the cockpits of modern aircraft as well as various medical instruments. A digital meter with an analog display has been thoroughly studied and clearly defined in [5],[6]. A mechanical or electromechanical screen displays only numbers (although clocks, particular Doppler meters, and information screens at various stations and airports have existed). Other examples of analog instruments include frequency meters and watt meters, each designed to measure their respective electrical quantities [8],[9].

1.1. Classification of analog instruments. The number of physical quantities that are measured by an analogous quantity determines how that quantity is classified. For instance, the voltmeter is utilized in the process of measuring voltage, whereas the ammeter is utilized in the process of measuring current. Wattmeter and frequency meter are the two instruments that are used to independently measure power and frequency [10].

Analog instruments are categorized based on the type of electrical quantity they measure. The various kinds of electrical instruments include ammeters for measuring current, voltmeters for voltage, and ohmmeters

*Department of School of Computer Science and Information Technology, DAVV, Indore, India (hiteshninama@davvscsit.in)
†Department of Information Technology, Institute of Engineering and Technology, Devi Ahilya University, Indore, India (jraikwal@ietdavv.edu.in)
‡Department of Electronics and Communication Engineering, PDPM Indian Institute of Information Technology Design and Manufacturing, Jabalpur, M.P. 482005, India (praikwal@iiitdmj.ac.in)

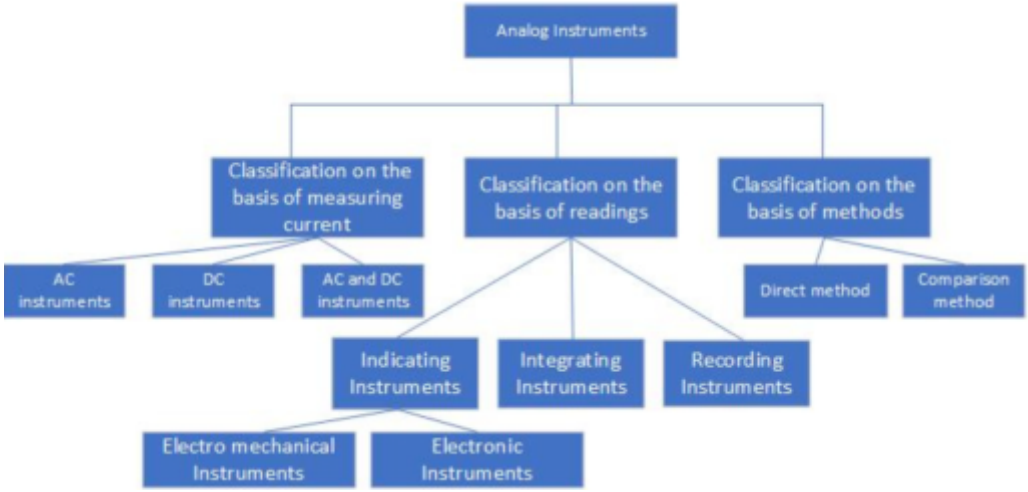


Fig. 1.1: Classification of Analog Instruments

for resistance, watt meters for power, frequency meters for the frequency of the electrical signal.

1.2. Transfer Learning. Transfer learning is a widely adopted approach in the field of machine learning, wherein the knowledge acquired from one specific context is leveraged to improve the performance in another domain that is different but related. In conjunction with deep ensemble learning, this method has demonstrated remarkable efficacy in evaluating analog instruments [11]. In order to apply transfer learning to the context of gauge analogs, the first step is to train a deep neural network on a substantial dataset of gauge analogs from a different yet related domain. Afterward, the network can be fine-tuned using data from the target domain. Starting the network with a pre-trained model enables it to learn the designated task more rapidly and effectively. Deep ensemble learning combines the forecasts from several independently trained models, aggregating them into one consolidated prediction [7][12][13]. The models in an ensemble can be trained independently, allowing for a wide variety of initializations, architectures, and algorithms to be used. Deep ensemble learning can be used to improve the accuracy of predictions made with gauge analogs by training numerous deep neural networks on distinct parts of the available data or by employing various network topologies. It is possible to produce a composite prediction based on the results of multiple models by employing methods such as averaging or majority voting. Gauge analogs can benefit from both transfer learning and deep ensemble learning due to their enhanced capability to capture complex patterns and relationships within the data. These methods can be especially helpful when the gauge analog dataset is small or when the analogs are highly variable or ambiguous [14]. It is important to note that the nature of the gauge analogs and the requisite problem will determine the precise implementation specifics and architectural and algorithmic decisions. The most efficient methods for any given work may only be determined through experimentation and thorough review.

2. Literature Review. Kristiansen 2023 et al. in an effort to speed up the diagnostic process, have studied the use of machine learning (ML) to assess the severity of sleep apnea from data gathered by a low-cost strain gauge respiration belt (named Flow) and a smartphone. Twenty-nine patients wore the flow belt and connected it to the Type III sleep monitor Nox T3 while sleeping at home without human supervision. The trials demonstrate that convolutional neural networks outperform competing approaches for evaluating flow data in terms of accuracy (0.7609), sensitivity (0.7833), and specificity (0.7217). This is because they have greater resilience against the most fundamental problems. These results are feasible even while training the classifier on the best available Nox T3 data. There may be little need for additional large-scale data collection to support future ML research utilizing information gathered from a variety of low-cost breathing belts. The classifier on a mid-range smartphone just needs approximately a second to analyze a night's worth of sleep data. The findings show the potential for widespread at home pre-screening of sleep apnea using

Table 2.1: Literature Summary.

Author/Year	Method	Results	References
Peixoto/2022	Convolutional neural network (CNN)	Accuracy = 0.95	[19]
Shahabi/2021	A Genetic Algorithm (GA)-optimized Deep Belief Network (DBN) using the Back Propagation (BP) algorithm.	Accuracy = 0.985	[20]
Howells/2021	Convolutional neural network (CNN)	Accuracy = 98.6	[21]
Wang/2021	Region-based Convolutional Network (Faster-RCNN)	Accuracy = 90.7	[22]
Reshadi/2020	Machine learning methods, including one-class SVM, isolation forests, multi-layer perceptron, decision trees, and random forests are used.	Accuracy = 0.9999	[23]

low-cost strain gauge belts, smartphones, and machine learning [15]. Naji Hussain 2023 et al. This study was carried to demonstrate how feature reduction techniques may significantly improve classification accuracy. Singular value decomposition (SVD) and principal component analysis (PCA) are both used in this work. This study uses six distinct machine learning classifiers to categorize foot diseases. To discover which classifier is most suited for classifying limb rehabilitation data, the study investigated six alternative classifiers: random forest (RF), naïve bayes (NB), K-nearest Neighbours (KNN), logistic regression (LR), decision tree (DT), and stochastic gradient descent (SGD). On the Lower Limb EMG dataset, experiments showed that 99% of the time could be spent correctly classifying data [16]. Basalamah 2023 et al. In order to improve the effectiveness of congestion detection in the real world, a bidirectional long-short-term memory (Bi-LSTM) model is used that makes use of synthetic datasets. This was done to bridge the gap between the two types of data utilizing a domain adaptation strategy by training the model with the synthetic dataset and then refining it using real-world cases. The experimental results demonstrate a notable improvement in performance on real-world datasets following training on a synthetic dataset using the suggested approach [17]. Starzynska-Grzes 2023 et al. This study intends to evaluate the applicability of an approach to architectural studies and explore the effects of better coordinated research efforts across the disciplines of architecture and computer science. The various architectural applications, such as building classification, detail classification, qualitative environmental analysis, building condition assessment, and building value calculation, are discussed along with the types of algorithms and data sources used by the analyzed studies to accomplish these objectives [18]. To take the photo of a gauge, transfer it to a local computer for processing, and show the results in a web-based dashboard, the study by Peixoto 2022 et al. proposed a microcontroller (ESP32-CAM) and camera (OV2640 with a 65° FOV angle) based solution. First, they have utilized a convolutional neural network (CNN) trained on the CenterNet HourGlass104 model to spot the gauge. Pixel projection is used to determine the pointer angle once the dial has been identified and split using the circular Hough transform and the polar transformation. A final angular calculation yields the indicative value. The 204 gauge pictures that made up the dataset were split 70:30 between a training set and a test set. The small size of the dataset required multiple data augmentations in order to produce a gauge detection model with good accuracy and for a broader applicability. With an average relative error of 0.95 percent, the testing findings also demonstrated dependability and precision suitable for use in industrial settings [19].

3. Motivation. The application of deep ensemble learning and transfer learning techniques to gauge analogs is driven by their ability to overcome difficulties and shortcomings of conventional methods. The present study differs from other studies in many important aspects. Firstly, deep ensemble learning offers the advantage of capturing a wider range of data nuances, representations, and patterns. Additionally, the prediction accuracy is enhanced by training multiple models with various initializations or architectures. Second, while dealing with a limited dataset, transfer learning may considerably increase gauge analog prediction capacity. Combining deep ensemble learning with transfer learning offers several advantages as it makes use of the strengths of both approaches. Transfer learning is a technique used in deep learning to initialize models with relevant knowledge and prior information from a different but related domain. This approach allows the models to capture diverse perspectives and minimize bias. Collaboration leads to better generalization, enhanced

prediction performance, and more efficient data utilization. Overall, this study differs from previous studies in that it uses the complimentary capability of deep ensemble learning and transfer learning to overcome the challenges of sparse data, high complexity, and high unpredictability in gauge analog prediction. The purpose of this study is to enhance the present state of gauge analog analysis by utilizing advanced methods. This will lead to better accuracy, dependability, and efficiency in addressing the problem.

4. Methodology.

4.1. Data collection. A thorough explanation of the datasets utilised in this work guarantees openness and shows the model's capacity to generalise properly across several scenarios, therefore addressing the worry of data set restrictions regarding the source, size, and diversity. Two main datasets [29] Synthetic,Gauges and Real Gauges—that have been carefully selected and created to offer a varied and complete set of training and testing data are used in this work. Synthetic Gauges Dataset: All at a high resolution of 1024x1024 pixels, this collection has 10,000 training photos and 1,000 test images. These synthetic images are matched with ground truth labels found in JSON files and replicate real-world gauge circumstances. Bounding boxes, keypoints for gauge perspective points, and crucial measures including scale minimum, scale maximum, pointer centre, and pointer tip abound among the COCO style labels. This synthetic dataset guarantees a large spectrum of gauge kinds and situations, therefore improving the generalising power of the model. Comprising photos from six different gauges, this dataset is meant to assess reading tasks, pose recovery, and gauge detection. Capture photos against 36 different backgrounds—from an Interior Design magazine to replicate various operating environments. The gauges are shot from several camera angles with annotations for gauge plane normals in order for pose recovery. Three five-second movies for every gauge capture varying pointer motions with exact gauge reading labels and pointer angles relative to the scale. These datasets overcomes constraints in size, source, and diversity to enable strong model training and validation by integrating synthetic and real data, therefore providing complete coverage of many gauge types, backgrounds, and operational settings.

4.2. Data Pre-processing. The preprocessed gauge analog data is crucial for effective analysis and modeling. The initial step involved cleansing the dataset by removing noise, outliers, or any missing values, ensuring data accuracy and reliability for subsequent analysis. A second critical step is the selection and extraction of features from the gauge analog data. Techniques such as filtering and wrapper methods are employed for feature selection to identify and focus on the most significant and informative attributes [24]. Principal Component Analysis (PCA) and Autoencoder are two feature extraction methods that can be used to reduce the dimensionality of the data without losing any useful information. Both image enhancement and image data augmentation are employed as part of the preprocessing stage.

4.2.1. Image Enhancement. Gauge analogs can benefit from image enhancement techniques that boost the clarity and contrast of the resulting images for more accurate analysis and interpretation. The following are some of the most widely employed methods for improving gauge analog images.

Contrast Stretching (CS). To improve the value of the gray levels already present within the processed image, an image enhancement technique involves extending the values of various hues. This simultaneously modifies image pixel values of each image, highlighting the layout in low and high contrast regions. Image contrast is the difference between the darkest and brightest parts of an image. The regulates the scale function of the image's pixel values expressed in equation 4.1.

$$s = T(r) = \frac{1}{1 + \left(\frac{m}{r}\right)^E} \tag{4.1}$$

Where r represents the input image intensity, s represents the corresponding intensity in the output picture, and E represents the function's slope. Fig. 4.1 shows the image after and before applying scale function.

Histogram Equalization (HE). A picture histogram helps with the shadowy parts. A histogram can be used to evaluate an image's brightness, clarity, contrast, and color separation. Histogram is used to normalize the image. It is used to make a photo look better to the naked eye. This requires segmenting images into smaller pieces. The histogram is associated with evaluating pixel values for the darker tones, which fall anywhere from 0 to 255 on the brightness scale of the image [25]. In order to improve an image, HE is used to determine the

(a) Meter A (b) Meter B

Fig. 4.1: Contrast Stretching.

(a) Meter A (b) Meter B

Fig. 4.2: Histogram Equalization.

relevant power levels and apply them consistently across pixels. In this approach, the HE method is employed to expand the per-pixel dynamic range of the image. The equation for histogram equalization is given in equation 4.2.

$$E(l) = \max\left(o, \text{round}\left(\frac{L}{N \times M} \times t(l)\right) - 1\right) \tag{4.2}$$

where E(l)- equalized function, Max- maximum dynamic range, L- no. of grey levels, N*M- the size of the image T(l)- accumulated frequencies. The impact of histogram equalization shown in Fig. 4.2.

Log Transformation. In the process of log transformation (x), the logarithm of each x variable is substituted for it. The goals of the statistical modeling usually dictate the choice of a logarithmic basis for the study. When typing "nature log" on a computer, the ln sign is typically used. When data doesn't conform to the bell curve, log transformation can make it as "normal" as feasible to increase statistical confidence. Putting it another way, our original data is less distorted after the log transformation. Ideally, the initial data would have followed a log-normal distribution. Otherwise, the log transformation will fail [26]. The equation 4.3 shows the logarithmic transformation.

$$s = c \log(r + 1) \tag{4.3}$$

Input and output images have pixel values of s and r, whereas c is a constant. Pixels in an input image with intensity equal to 0 require a multiplication by 1 since log (0) equals infinity. Ensuring a minimum of one requires raising the baseline by one. The impact of logarithmic transformation shown Fig.4.3.

4.2.2. Image Data Augmentation. The purpose of image data augmentation on gauge analog images is to generate additional training data by introducing various alterations or perturbations to the original images.

(a) Meter A (b) Meter B

Fig. 4.3: Log Transformation.

Data augmentation is a technique used to enhance the quantity and diversity of a dataset, with the aim of creating more accurate and generalizable machine learning models. There are several common methods for enhancing gauge analog images through improved image data. These methods include:

i) The initial step involves simulating diverse perspectives by rotating the analog image on the gauge, enabling the model to accurately interpret gauge readings from multiple viewpoints.

ii) Second, Resize the analog picture on the gauge to various sizes. The purpose of this tool is to simulate different gauge sizes, allowing the model to be trained and tested on a variety of sizes.

iii) The simulation of image resolution and aspect ratio can be achieved by cropping and padding the analog image gauge. This can help the model better handle photos of different sizes.

4.2.3. Real Time Processing. Real-time processing capability of the model has been extensively assessed to solve issues regarding its performance in applications requiring time sensitivity. Real-world test situations were used in extensive benchmarking to evaluate the responsiveness and performance of the model. Using parallel processing methods and high-performance hardware accelerations like GPUs and TPUs helped the model be optimised for real-time operation. This arrangement guarantees quick feedback appropriate for real-world use by letting the model process gauge readings and execute position recovery activities inside milliseconds, the design has been adjusted to reduce latency, therefore enabling handling of high frame rates needed in dynamic situations. Deployment in simulated industrial environments allowed the system's efficiency to be confirmed; it often obtained nearly instantaneous response times without sacrificing precision. These improvements guarantee the model's flexibility to meet demands for real-time processing, so ensuring its dependability for uses including real-time data analytics, automated inspections, and monitoring systems that call for fast performance.

4.2.4. Resource Optimisation Computation. Our work stresses optimisation techniques to minimise hardware costs and deployment complexity in order to solve the worry about the high computer resource requirements connected with merging several deep neural networks. Using lightweight neural network frameworks and model compression methods such knowledge distillation, quantisation, and pruning has helped to simplify the model architecture. These techniques make the system possible for deployment on hardware with limited resources since they greatly lower the model size and computational load without affecting performance, the use of adaptive inference methods lets the model dynamically change its processing needs depending on the degree of input data complexity. For instance, while more complicated tasks engage deeper layers only when needed, simpler tasks are handled using less computational resources. In real-time situations, this method optimises resource use by balancing performance and efficiency. The study also investigates the use of edge computing and cloud-based technologies to offload significant computations from local devices, effectively spreading the processing load. This hybrid strategy essentially lowers the total hardware cost by using the scalability of cloud resources and preserving low-latency performance at the edge, hence streamlining deployment. These improvements guarantee that for practical uses our methodology stays extremely cost-effective, scalable, and highly efficient.

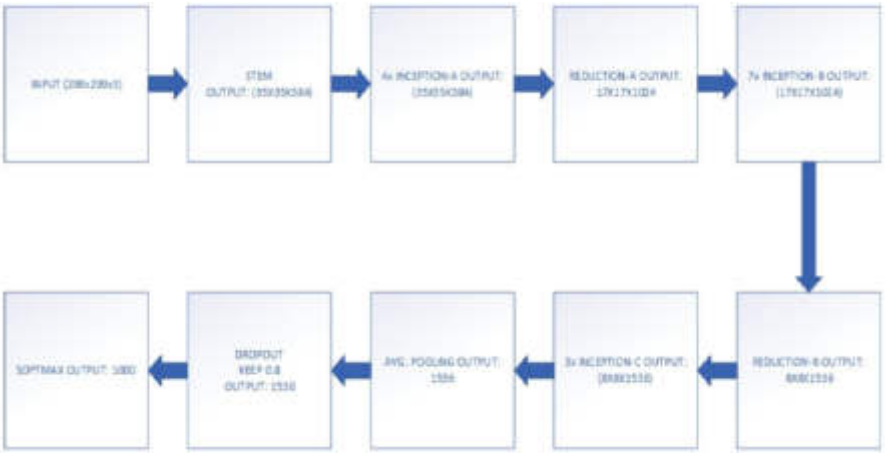


Fig. 4.4: InceptionResnetV2

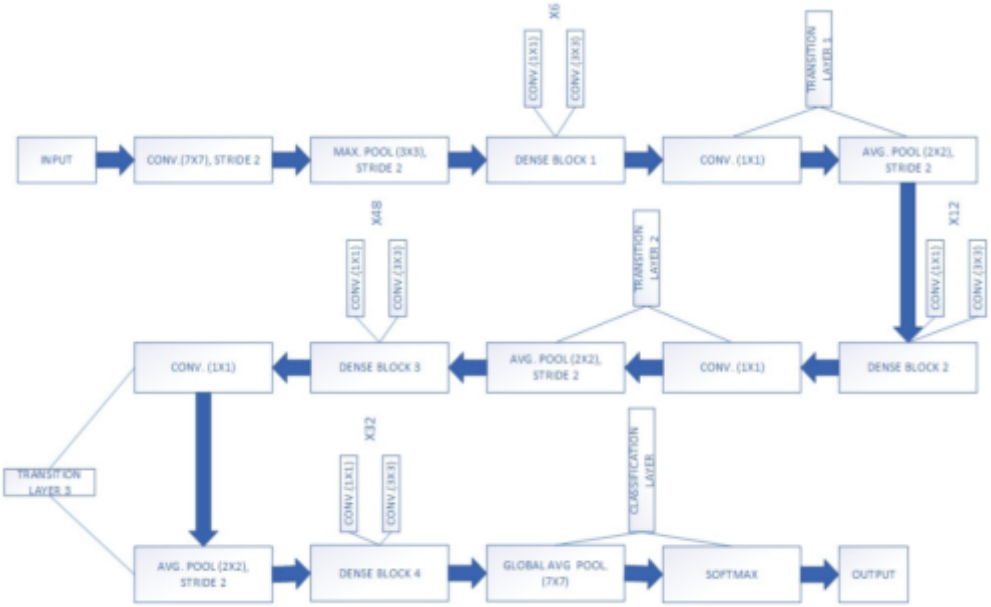


Fig. 4.5: DenseNet201 Architecture

4.3. Modelling.

4.3.1. Transfer Learning. Transfer learning is a machine learning strategy that involves using a pre-trained model, typically trained on a large dataset, as a starting point for a related task or dataset. Instead of starting the training process of a model from the beginning, transfer learning utilizes the knowledge and learned representations from a pre-trained model [27]. This helps to accelerate training and increase performance on the new task. Two different models were utilized for checking transfer learning: the first being InceptionResNetV2 and the second, DenseNet201. Architecture of InceptionResNetV2 and DenseNet201 is explain in Fig. 4.4 and Fig. 4.5 respectively.

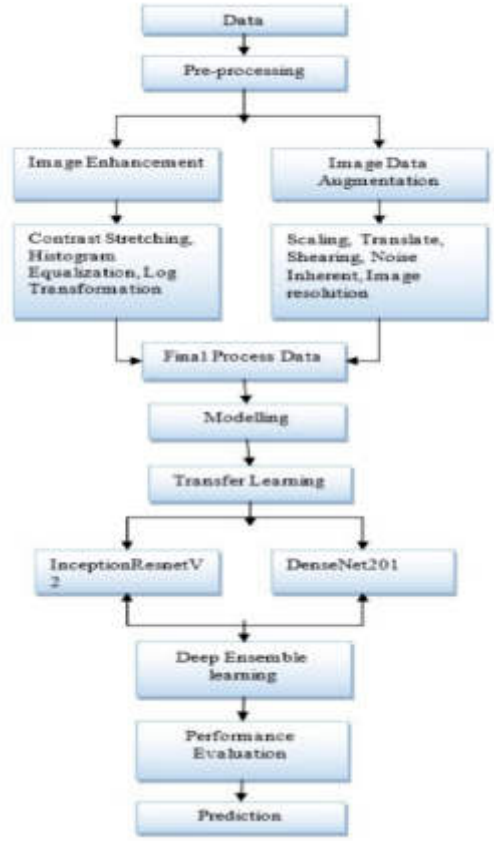


Fig. 4.6: Proposed Flowchart

4.3.2. Deep Ensemble Learning. Deep ensemble learning is a technique that leverages many deep learning models to enhance the precision and robustness of forecasts [28]. This approach eliminates the reliance on a single model by training a collection of models, whose separate predictions are merged to produce a definitive decision. To start the implementation of deep ensemble learning, the first step involves the integration of two pre-trained transfer learning models, specifically InceptionResNetV2 and DenseNet201. Workflow of Deep ensemble learning is shown in Fig.4.6.

5. Implementation and Results. The proposed method is divided into two phases. The first phase involves using image processing and computer vision libraries to assign labels to all images. The second phase is dedicated to training, where a deep neural network generates readings for input gauges. This study involves the implementation of InceptionResnetV2 with DenseNet201 for a comparative examination of previously developed and newly proposed ensemble learning models. The implementation relies on the Keras library and Open CV for image processing. The proposal includes DenseNet201, InceptionResnetV2, and an ensemble learning approach with image preprocessing; the models in the table are developed using various hyperparameters, employing the Keras library and Open CV for image processing. The model's performance is closely tied to the configuration of hyperparameters, as detailed in Table 5.1.

ReLU Activation. Rectified Linear Unit, more often known as ReLU, is an activation function that is frequently found in deep neural networks. It is a non-linear function that brings non-linearity to the network, which gives the network the ability to learn complicated patterns and makes the network more expressive. The

Table 5.1: Hyperparameters Details.

Parameter	**Details**
Learning	Transfer Learning
Model	InceptionResNetV2, DenseNet201
Input Shape	128*128*3
Layers	Conv2D
Padding	Same
Pooling	Global Average Pooling
Normalization	Batch
Dropout	50%
Activation	ReLU, Softmax
Loss	Categorical Cross Entropy
Performance Evaluation	Accuracy, Precision, Recall
Epochs	100

following constitutes the definition of the ReLU activation function:

$$F(x) = \max(0, x) \tag{5.1}$$

SoftMax. The SoftMax activation function is frequently used in the output layer of neural networks for situations with several classes of data. A vector of real-valued inputs is used as a starting point, and the output is a probability distribution over the classes. The Softmax function computes the probability of each class given the inputs, giving higher weight to classes with larger values. This is because of the following:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{5.2}$$

In this equation, x_i stands for the ith element of the input vector $\mathbf{x}$, $\exp(x_i)$ is the name of the exponential function applied element-wise to x_i, and the sum is computed using all members of the input vector. In deep learning models, the Softmax activation function is frequently utilized for tasks involving the prediction of several classes, such as image classification, natural language processing, and audio recognition.

Categorical Cross Entropy. A commonly employed function in multi-class classification issues is the "categorical cross-entropy" loss function, additionally referred to as "Softmax cross-entropy" or "cross-entropy" loss. The categorical cross-entropy loss function measures the degree of dissimilarity between the predicted class probabilities and the actual class labels. Since its primary application involves assigning a single class label to each input sample, it excels in settings where the classes are mutually exclusive. Categorical cross-entropy can be expressed as:

$$L = -\sum (Y_{\text{true}} \cdot \log(Y_{\text{pred}})) \tag{5.3}$$

The Softmax activation function generates predicted class probabilities, denoted by $\mathbf{y}_{\text{pred}}$, while $\mathbf{y}_{\text{true}}$ represents the actual class labels (one-hot encoded). The total is calculated over all levels.

5.1. Performance Evaluation. In this section, the research findings are provided. Using a Python simulator and a range of performance metrics, the viability of the suggested model was assessed during the entire experiment. In this work, Inception ResNet V2 and Dense Net models were employed as transfer learning and ensemble learning techniques.

In Table 5.2, Table 5.3, and Table 5.4, loss value is calculated through the following function:

$$\text{Loss} = -\frac{1}{m}\sum_{i=1}^{m} (Y_i \cdot \log(Y_i)) \tag{5.4}$$

Table 5.2: Results of Contrast Stretched Processed Data.

Models	Accuracy	Precision	Recall	Loss
InceptionResNetV2	92.90	93.83	91.72	19.53
DenseNet201	94.67	94.66	94.56	16.46
Ensemble Model	98.34	98.46	98.34	08.33

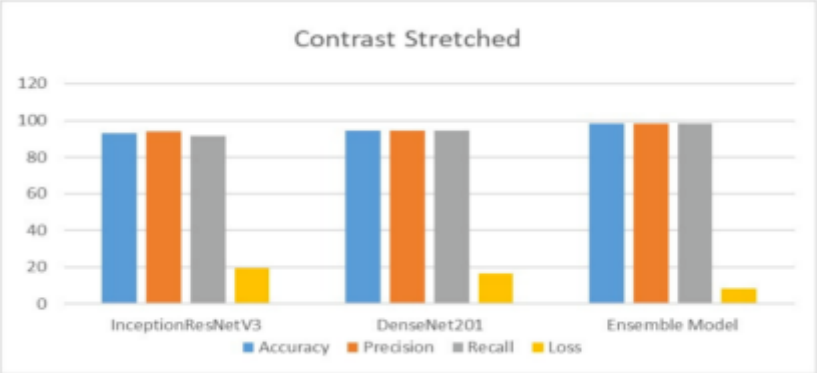


Fig. 5.1: Performance evaluation graph of contrast Stretched

Table 5.3: Results of Histogram Equalization Processed Data.

Models	Accuracy	Precision	Recall	Loss
InceptionResNetV2	90.77	91.34	89.82	26.80
DenseNet201	93.14	93.33	92.78	16.36
Ensemble Model	97.87	98.10	97.51	08.43

Table 5.4: Results of Log Transformation Processed Data.

Models	Accuracy	Precision	Recall	Loss
InceptionResNetV2	70.18	78.75	65.80	82.93
DenseNet201	93.02	93.55	92.66	18.87
Ensemble Model	98.22	98.45	97.87	09.82

Table 5.2 and Fig.5.1 present the performance results of contrast stretching using models such as InceptionResNetV2 and ensemble learning. The ensemble model achieved the best accuracy, which was 98.34, while the InceptionResNetV2 model suffered the lowest loss, which was 08.33.

Table 5.3 and Fig. 5.2 illustrate the findings of an investigation on the effectiveness of histogram equalization utilizing several modeling techniques, including InceptionResNetV2 and ensemble learning. The ensemble model stood out on top with an accuracy of 97.87, while the InceptionResNetV2 model had the least amount of loss obtained as 08.43.

Table 5.4 and Fig.5.3 present the findings of an inquiry into the efficiency of log transformation using a variety of modeling techniques, such as InceptionResNetV2 and ensemble learning. This analysis was conducted to determine how well the log transformation works. The ensemble model had the highest accuracy (98.22%), while the InceptionResNetV2 model had the least amount of loss (09.82%). The ensemble model came out on top.

Fig.5.4, Fig.5.5, and Fig.5.6 show the best results for InceptionResNetV2, DenseNet201, and Ensemble by

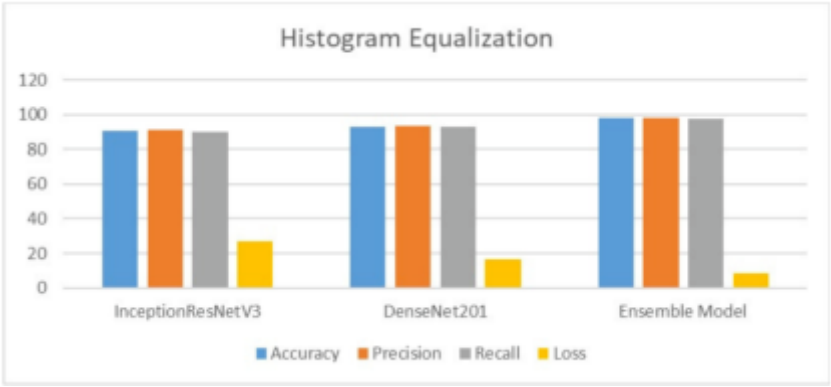


Fig. 5.2: Performance evaluation graph of Histogram Equalization

Fig. 5.3: Performance evaluation graph of Log Transformation

comparison stretching. The metrics shown are accuracy, precision, recall, and loss. The training line should be interpreted as the blue line in the graphs, and the test line should be interpreted as the orange line.

6. Conclusion. The conclusion from this research reveals that an ensemble transfer learning strategy was employed due to challenges with pressure gauges and automatic pointer gauge readings. A transfer learning method was introduced, utilizing advanced architectures like InceptionResnetV2 or DenseNet 201, to estimate gauge readings. An ensemble learning approach was also suggested to refine these predictions. The focus was on image enhancement techniques, including adjustments in shape, size, resolution, and creating an angle corresponding to the gauge reading. Techniques like contrast stretching, histogram equalization, and log transformation were used, leading to the ensemble model achieving 98.34% accuracy. The InceptionResNetV2 model excelled particularly in contrast stretching with a loss of only 08.33%. Looking ahead, the transition from analog to digital readings will incorporate this image processing technology, using unsupervised data for artificial gauge readings. This approach, integrating a sophisticated deep convolutional neural network, is planned to overcome the issue of limited data availability.

6.1. Error Analysis. This part explores closely the mistakes found in model evaluation. Errors were classified, examined with root cause methods like LIME and SHAP, and sensitivity testing was done. Targeting modifications based on acquired insights helped to increase the accuracy, robustness, general performance of the model in practical settings.

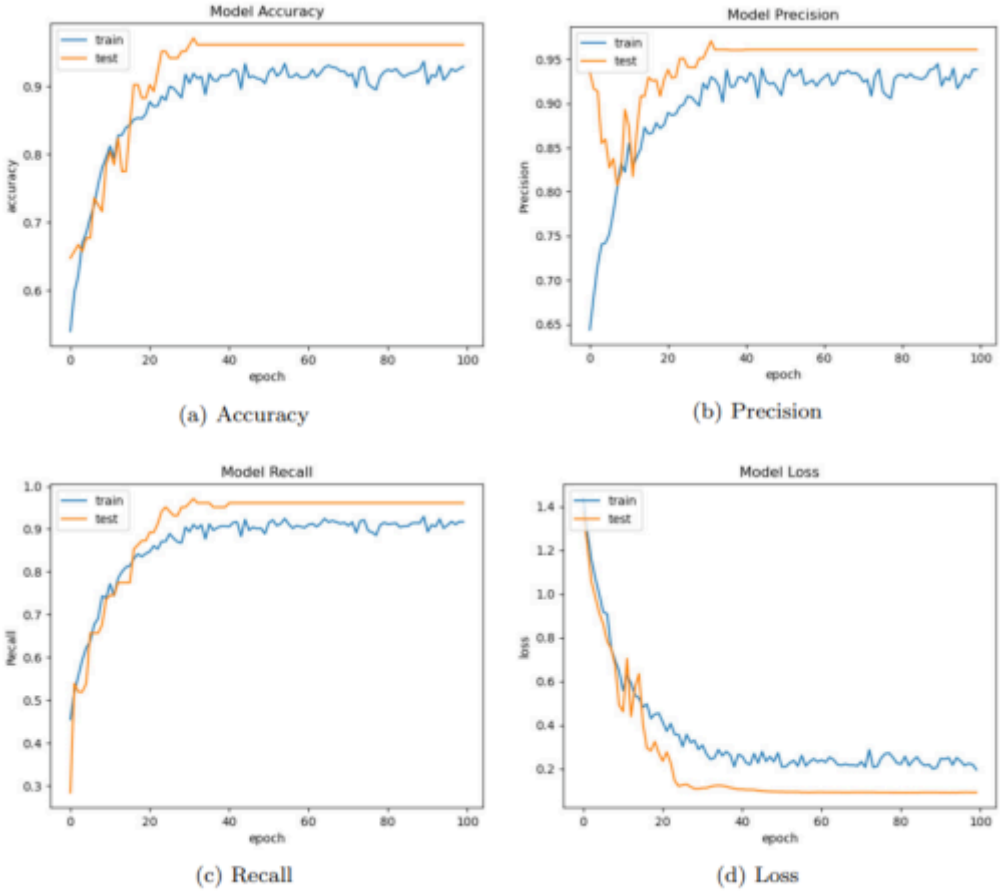


Fig. 5.4: InceptionResNetV2 Results: (a) Accuracy, (b) Precision, (c) Recall, and (d) Loss

7. Future work. The potential for future gauge research is tremendous, covering a wide variety of topics. Here are a few instances of this: Gauge theories have been extensively utilized in high-energy particle physics to decipher the interactions between the fundamental particles. In the future, researchers may look into the behavior of matter at extremely high energies, such as those seen in the early universe or in particle accelerators, examine the properties of known particles in greater depth, and search for new particles outside the standard model. As a future work, researchers may look into the possible applications of gauge theories in quantum simulations and quantum information processing, examine quantum entanglement and topological characteristics in gauge systems, and build novel quantum algorithms which are based on these theories.

REFERENCES

[1] E. von Lavante, S. Brinkhorst, A. Gedikli, and H. Krisch, *Fluid mechanical optimization of a DN25 vortex flow meter with novel vortex detection*, Flow Meas. Instrum., vol. 44, pp. 122–125, 2015, doi: 10.1016/j.flowmeasinst.2014.11.004.

[2] A. Depari, C. M. De Dominicis, A. Flammini, E. Sisinni, L. Fasanotti, and P. Gritti, *Using smartglasses for utility-meter reading*, SAS 2015 - 2015 IEEE Sensors Appl. Symp. Proc., 2015, doi: 10.1109/SAS.2015.7133649.

[3] P. Bedi, S. B. Goyal, A. S. Rajawat, P. Bhaladhare, A. Aggarwal, and A. Prasad, *Feature Correlated Auto Encoder Method for Industrial 4.0 Process Inspection Using Computer Vision and Machine Learning*, Procedia Comput. Sci., vol. 218, pp. 788–798, 2023, doi: 10.1016/j.procs.2023.01.059.

[4] M. Liang et al., *Predicting micromechanical properties of cement paste from backscattered electron (BSE) images by computer vision*, Mater. Des., vol. 229, p. 111905, 2023, doi: 10.1016/j.matdes.2023.111905.

[5] A. Nouriani, R. McGovern, and R. Rajamani, *Activity Recognition Using A Combination of High Gain Observer and*

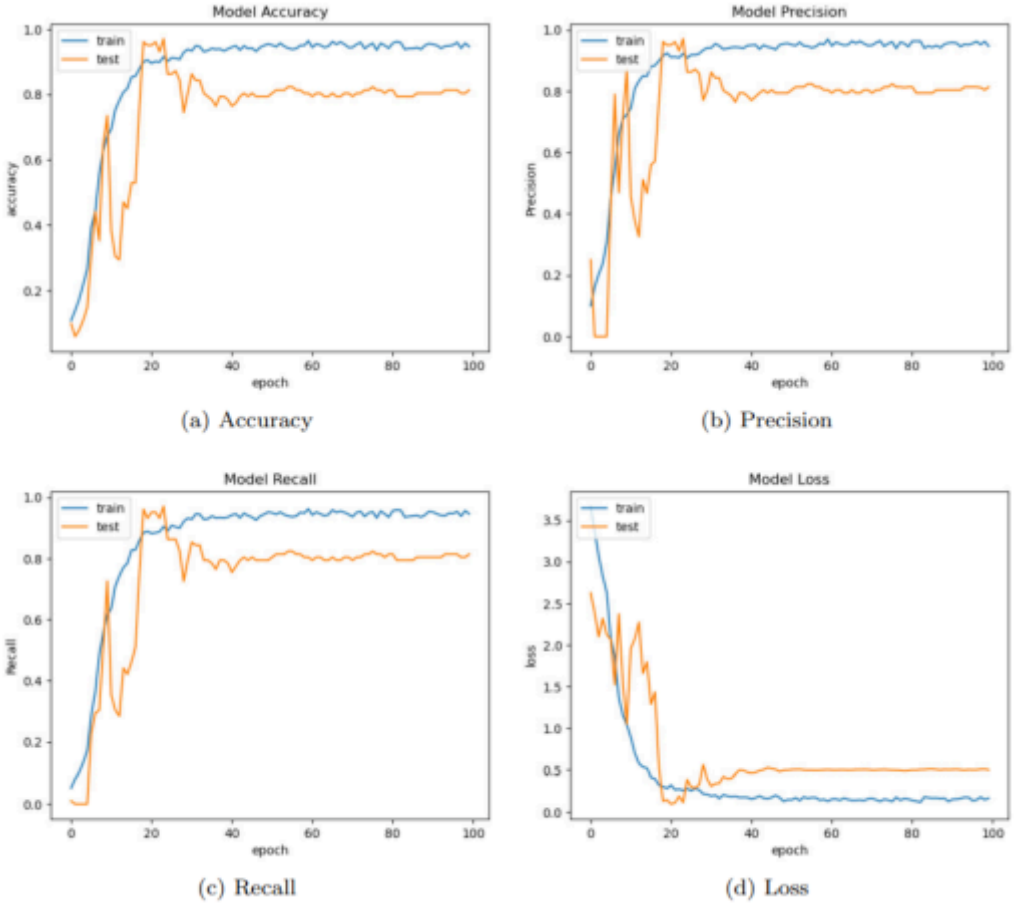


Fig. 5.5: DenseNet201 Results: (a)Accuracy, (b)Precision, (c)Recall and (d) Loss

Deep Learning Computer Vision Algorithms, Intell. Syst. with Appl., vol. 18, no. February, p. 200213, 2023, doi: 10.1016/j.iswa.2023.200213.

[6] A. K. Das, T. J. Esau, Q. U. Zaman, A. A. Farooque, A. W. Schumann, and P. J. Hennessy, *Machine vision system for real-time debris detection on mechanical wild blueberry harvesters*, Smart Agric. Technol., vol. 4, no. November 2022, p. 100166, 2023, doi: 10.1016/j.atech.2022.100166.

[7] J. Chi, L. Liu, J. Liu, Z. Jiang, and G. Zhang, *Machine Vision Based Automatic Detection Method of Indicating Values of a Pointer Gauge*, Math. Probl. Eng., vol. 2015, 2015, doi: 10.1155/2015/283629.

[8] C. Dong, *Investigation of Computer Vision Concepts and Methods for Structural Health Monitoring and Identification Applications*, Electron. Theses Diss., 2019, [Online]. Available: https://stars.library.ucf.edu/etd/6867.

[9] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, *Computer vision algorithms and hardware implementations: A survey*, Integration, vol. 69, no. June, pp. 309–320, 2019, doi: 10.1016/j.vlsi.2019.07.005.

[10] G. Salomon, R. Laroca, and D. Menotti, *Image-based Automatic Dial Meter Reading in Unconstrained Scenarios*, Meas. J. Int. Meas. Confed., vol. 204, 2022, doi: 10.1016/j.measurement.2022.112025.

[11] J. C. Yang, Y. C. Guo, and L. H. Cai, *Using a nested anomaly detection machine learning algorithm to study the neutral triple gauge couplings at an e+e collider*, Nucl. Phys. B, vol. 977, p. 115735, 2022, doi: 10.1016/j.nuclphysb.2022.115735.

[12] J. S. Lauridsen, J. A. G. Grassmé, M. Pedersen, D. G. Jensen, S. H. Andersen, and T. B. Moeslund, *Reading circular analogue gauges using digital image processing*, VISIGRAPP 2019 - Proc. 14th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl., vol. 4, no. Visigrapp, pp. 373–382, 2019, doi: 10.5220/0007386003730382.

[13] P. Mehta et al., *A high-bias, low-variance introduction to Machine Learning for physicists*, Phys. Rep., vol. 810, pp. 1–124, 2019, doi: 10.1016/j.physrep.2019.03.001.

[14] J. A. Almazán-Lázaro, E. López-Alba, and F. A. Díaz-Garrido, *Applied computer vision for composite material manufacturing by optimizing the impregnation velocity: An experimental approach*, J. Manuf. Process., vol. 74, no. November 2021, pp. 52–62, 2022, doi: 10.1016/j.jmapro.2021.11.063.

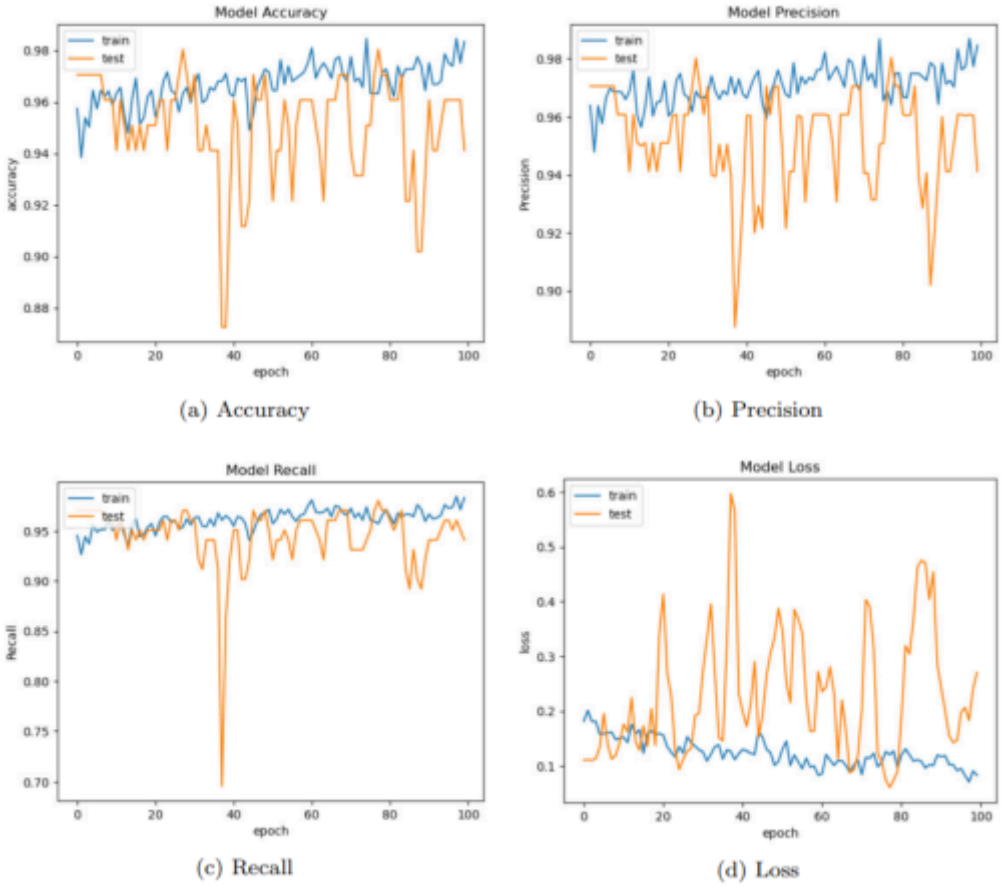


Fig. 5.6: Ensemble Results: (a) Accuracy, (b) Precision, (c) Recall and (d) Loss

[15] S. Kristiansen et al., *A clinical evaluation of a low-cost strain gauge respiration belt and machine learning to detect sleep apnea*, Smart Heal., vol. 27, p. 100373, 2023, doi: 10.1016/j.smhl.2023.100373.

[16] A. Naji Hussain, S. A. Abboud, B. A. Baki Jumaa, and M. N. Abdullah, *Impact of feature reduction techniques on classification accuracy of machine learning techniques in leg rehabilitation*, Meas. Sensors, vol. 25, no. October 2022, p. 100544, 2023, doi: 10.1016/j.measen.2022.100544.

[17] S. Basalamah, S. D. Khan, E. Felemban, A. Naseer, and F. U. Rehman, *Deep learning framework for congestion detection at public places via learning from synthetic data*, J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 1, pp. 102–114, 2023, doi: 10.1016/j.jksuci.2022.11.005.

[18] M. B. Starzyńska-Grześ, R. Roussel, S. Jacoby, and A. Asadipour, *Computer vision-based analysis of buildings and built environments: A systematic review of current approaches*, ACM Comput. Surv., vol. 1, no. 1, 2023, doi: 10.1145/3578552.

[19] J. Peixoto et al., *Development of an Analog Gauge Reading Solution Based on Computer Vision and Deep Learning for an IoT Application*, Telecom, vol. 3, no. 4, pp. 564–580, 2022, doi: 10.3390/telecom3040032.

[20] H. Shahabi et al., *Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm*, Geosci. Front., vol. 12, no. 3, p. 101100, 2021, doi: 10.1016/j.gsf.2020.10.007.

[21] B. Howells, J. Charles, and R. Cipolla, *Real-time analogue gauge transcription on mobile phone*, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., no. ii, pp. 2369–2377, 2021, doi: 10.1109/CVPRW53098.2021.00269.

[22] L. Wang, P. Wang, L. Wu, L. Xu, P. Huang, and Z. Kang, *Computer vision based automatic recognition of pointer instruments: Data set optimization and reading*, Entropy, vol. 23, no. 3, pp. 1–21, 2021, doi: 10.3390/e23030272.

[23] M. Reshadi, *Anomaly Detection Using Deep Learning*, 2020.

[24] D. J. Roach et al., *Utilizing computer vision and artificial intelligence algorithms to predict and design the mechanical compression response of direct ink write 3D printed foam replacement structures*, Addit. Manuf., vol. 41, no. December 2020, p. 101950, 2021, doi: 10.1016/j.addma.2021.101950.

[25] R. P. Singh and M. Dixit, *Histogram Equalization: A Strong Technique for Image Enhancement*, Int. J. Signal Process. Image Process. Pattern Recognit., vol. 8, no. 8, pp. 345–352, 2015, doi: 10.14257/ijsip.2015.8.8.35.

[26] GEEKS, *Log transformation of an image using Python and OpenCV*, Geeks for Geeks, 2020.
[27] E. AFACAN, N. LOURENÇO, R. MARTINS, AND G. DÜNDAR, *Review: Machine learning techniques in analog/RF integrated circuit design, synthesis, layout, and test*, Integration, vol. 77, pp. 113–130, 2021, doi: 10.1016/j.vlsi.2020.11.006.
[28] S. DUMBERGER, R. EDLINGER, AND R. FROSCHAUER, *Autonomous Real-Time Gauge Reading in an Industrial Environment*, IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA, vol. 2020-Septe, pp. 1281–1284, 2020, doi: 10.1109/ETFA46521.2020.9211895.
[29] JJC VISION, *Automatic Gauge Reading*, JJC Vision, 2024. [Online]. Available: http://www.jjcvision.com/projects/gauge_reading.html. [Accessed: 28-July-2024].

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Aug 29, 2024
Accepted: Oct 7, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2145–2164, DOI 10.12694/scpe.v26i5.4792

DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR ENHANCED MAINTENANCE OF AUTOMOBILES USING MACHINE LEARNING

NEERAJ DAHIYA[*], EDEH MICHAEL ONYEMA[†], VENKATARAMAIAH GUDE[‡], NEETU FAUJDAR[§], GAYTRI DEVI[¶], AND REENU BATRA[‖]

Abstract. In the ever-changing automobile industry, sustainable practices are of utmost importance. It is essential to perform preventative maintenance to accomplish sustainability objectives. Vehicles will have fewer unanticipated breakdowns and last longer as a result. This study introduces a machine learning model that has been optimized to enhance preventive maintenance operations in the automotive sector. To improve the efficacy of predictive maintenance in intelligent manufacturing systems, this paper presents an improved AdaBoost algorithm linked with big data analytics. First, in the proposed framework, massive datasets are gathered and preprocessed from sensors, IoT devices, and other sources in the industrial setting. Then, the meta-algorithm AdaBoost is used to improve the efficiency of subpar learners, allowing for reliable failure and deterioration prediction in machinery. Adjusting hyperparameters like the number of iterations and the learning rate is part of the algorithmic optimization process to strike a good balance between model accuracy and computational efficiency. The proposed model gains an accuracy level of 0.972 value, Precision level of 0.977 value, Recall level of 0.972 value and F1-score level of 0.974 value. By analyzing historical data, our algorithm can predict when problems will occur, enabling us to take quick action and minimize downtime. Improved maintenance scheduling and reduced environmental effects are outcomes of the proposed model's use of cutting-edge optimization techniques, which boost the model's predictive capabilities. The model achieves better results than the state-of-the-art methods in extensive trials conducted on a dataset from a leading automaker. It achieves significant improvements in maintenance efficiency and prediction accuracy. Sustainability in the automobile sector is a wider purpose of this study, which proposes a data-driven plan for maintenance that is strong and in line with economic and environmental aims.

Key words: Optimized AdaBoost; Predictive Maintenance; Automobile Maintenance Optimization; Sustainable Automotive Practices; Predictive Analytics; Automotive Industry Automation

1. Introduction.

Traditional industrial processes has been upgraded to smart and environmentally friendly systems with the help of Big Data analytics. Organizations may improve their manufacturing processes, get new insights, and make better decisions by analyzing the massive amounts of data produced across the whole production lifecycle. Sensors built into machines and other equipment are the backbone of intelligent manufacturing, allowing for continuous monitoring of conditions like temperature, pressure, vibration, and more. In order to improve procurement, inventory management, and logistics, it is necessary to integrate data from all points of the supply chain. Predictive maintenance is made possible by the examination of sensor data, which can alert workers to impending equipment faults [1]. This helps keep machines running smoothly, save money on repairs, and keep them in use for longer. Big data analytics has been used to keep a close eye on the manufacturing line and evaluate data as it comes in, guaranteeing high-quality goods at all times [2]. It is possible to monitor for and quickly correct any instances of subpar quality. Energy data monitoring and analysis may help factories reduce their energy footprint. Cost savings aren't the only thing that benefits from energy efficiency improvements. Inefficient steps and bottlenecks in the production process can be pinpointed by analyzing data

[*]Department of Computer Science and Engineering, SRM University, Delhi-NCR, Sonipat, Haryana, India;(neeraj.d@srmuniversity.ac.in).
[†]Department of Mathematics and Computer Science, Coal City University Nigeria; Adjunct Faculty, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai 602105, India. (mikedreamcometrue@gmail.com).
[‡]Software Engineer, GP Technologies LLC, USA; (gvramaiah.se@gmail.com).
[§]Department of CSE, Sharda School of Engineering and Technology (SSET), Sharda University Greater Noida, India; (neetu.faujdar@gmail.com).
[¶]Department of Computer Applications, GVM Institute of Technology and Management, Sonepat, Haryana, India; (gayatri.dhingra1@gmail.com).
[‖]Department of Computer Science and Engineering, K. R. Mangalam University, Gurgaon, Haryana, India; (reenubatra88@gmail.com) Corresponding Author

collected at various points along the process [3]. Workflow optimization, better resource usage, and increased productivity are all possible with this data. Big Data analytics helps manufacturers optimize their supply chains by offering insights into demand forecasts, inventory management, and supplier performance. The result has been less waste and more productivity [4].

Analyzing client data and preferences helps businesses to personalize products to specific needs, leading to more efficient manufacturing and less waste. The environmental effects of production processes have been tracked and evaluated with the use of Big Data analytics. Helping businesses meet sustainability targets and requirements requires monitoring emissions, waste production, and other issues. Data security and regulatory conformity are becoming increasingly important as data volumes rise. Big Data analytics has been used to build comprehensive security measures and guarantee that manufacturing processes conform to industry norms and laws. Manufacturers may improve their procedures, items, and green efforts in an ongoing cycle by constantly examining data and input from a wide range of sources. Making manufacturing systems more flexible and robust to changing obstacles, the incorporation of Big Data analytics into manufacturing processes enables smart decision-making, improves efficiency, and contributes to sustainability goals [5].

Predictive maintenance utilizing Big Data analytics is a valuable tool in intelligent and sustainable manufacturing. Predictive maintenance includes evaluating data from a variety of sensors to determine when repairs on a piece of machinery are needed. Predictive maintenance is a method of detecting impending problems with machinery. Manufacturers may improve efficiency and uptime by minimizing unscheduled downtime through proactive problem-solving. Maintenance has been scheduled during planned downtime by analyzing historical and real-time data, allowing manufacturers to minimize disruptions to production schedules and maximize resource use. The lifespan of machines is enhanced by early detection and repair of problems. By avoiding catastrophic breakdowns, producers may lengthen the useful lives of their products, cutting down on wasteful equipment replacements and increasing sustainability [6].

Condition Monitoring uses vibration, temperature, pressure, and other sensors to monitor equipment and machinery. Changes in condition characteristics may indicate issues or breakdowns. Avoiding expensive emergency repairs and reducing the need for routine, time-based maintenance are two ways in which predictive maintenance helps to lower total maintenance costs. It helps distribute resources more efficiently, resulting to cost savings in the long term. Equipment that has been well cared for usually performs better. By keeping machinery functioning at peak efficiency, predictive maintenance may help cut down on power usage and improve environmental outcomes. By anticipating future maintenance requirements, manufacturers may better manage their supply of replacement components. This avoids the wasteful expense of keeping an abundance of spare parts on hand while still guaranteeing ready access to critical components in an emergency. Equipment health has been tracked in real-time thanks to Big Data analytics. Alerts are triggered if conditions deviate from the usual, enabling for prompt action to be taken. The comparison of past and present data can better understand equipment performance. Data-driven insights allow manufacturers to make better decisions regarding preventative maintenance, resource allocation, and equipment improvements. Predictive maintenance relies heavily on data gathered from the IoT and sensors. These gadgets constantly gather information, making real-time equipment health monitoring possible. Predictions improve when sensor data is combined with Big Data analytics. Zero waste and low environmental impact are key to sustainable production. Predictive maintenance helps the environment since it reduces the need for unscheduled repairs and the amount of unwanted equipment that must be thrown away. Big data analytics' ability to fuel predictive maintenance is a crucial enabler of smart, eco-friendly production. It boosts operating efficiency, decreases expenses, extends equipment lifespan, and adds to overall environmental sustainability by lowering waste and energy consumption [7].

1.1. Problem Formulation. Predictive maintenance aims to improve reactive and preventative maintenance. Reactive, or "run-to-failure," maintenance fixes equipment after it breaks down. This behaviour can cause costly downtime, emergency repairs, and safety issues. Prevention, meanwhile, uses schedules to maintain equipment at regular intervals regardless of its condition. Overmaintenance and unnecessary repairs can result. Predictive maintenance uses data-driven insights to predict equipment failures to find a balance. However, certain challenges must be overcome to accurately describe predictive maintenance: High-quality data from sensors, machinery, and other sources is essential for predictive maintenance. But many businesses still struggle to guarantee data availability and quality. Sensor issues, erroneous data, or insufficient data might

cause predictive maintenance algorithms to fail. To accurately predict predictive maintenance failures, monitor and analyze the right qualities or attributes. Key features must be identified and designed effectively to gain insights from raw data. Since features and equipment health may change over time, predictive models must be constantly updated. Predictive maintenance models use random forests, neural networks, and support vector machines to analyze complex data and make accurate predictions. These models have high predictive accuracy, but uninterpretable forecasts has been hard to understand. Successful predictive maintenance requires easy integration with current procedures and systems. Effective resource allocation, maintenance prioritizing, and failure prediction algorithms are all part of this process. Organizational silos, reluctance to change, and lack of maintenance team support can make predictive maintenance difficult. Predictive maintenance can improve asset performance, equipment reliability, and maintenance costs, but it requires a large initial investment and ongoing operational costs. Only a rigorous cost-benefit analysis can assess the financial feasibility of predictive maintenance programs and the acquisition of necessary cash and resources.

Data integrity, low latency, scalability, and system integration are essential for real-time data streams in dynamic manufacturing. Sensors, machines, and other connected devices create data continuously, needing fast-processing systems. Apache Kafka and Apache Flink can handle high-throughput data streams. This low-latency platform enables manufacturers ingest and understand real-time data to detect crucial events quickly. Real-time processing speeds decision-making to improve machine performance, downtime, and efficiency. Another key method is edge computing. When millisecond decisions are needed, latency might hamper manufacturing. Edge computing processes and stores data near factory sensors and machinery. This speeds data transmission to a central server or cloud for processing. Edge devices can preprocess data, perform basic analytics, and run machine learning models locally. This is important in predictive maintenance and production line anomaly detection if network bandwidth is limited or real-time decisions are needed. Distributed architectures provide real-time processing. Distributed designs process data streams over multiple nodes or systems, boosting speed and fault tolerance. A distributed microservices design where each microservice handles data intake, filtering, and aggregation can speed things up. Scaling is easier when resources can be added without disrupting operations. Distributed and cloud infrastructures let manufacturers dynamically handle workloads and respond to data volume and complexity. Data streaming analytics is essential in dynamic manufacturing. Sliding, tumbling, and time-based windows allow real-time data processing. These methods detect trends, irregularities, and deviations from normal operation throughout time. These analytics frameworks can predict equipment failures, improve production schedules, and discover quality control concerns using historical data-trained machine learning models. Stream processing engines and reinforcement learning and RNNs can update industrial processes and learn from new input. Finally, real-time manufacturing data management involves legacy system integration. Most plants use MES, ERP, and other operational tech. An effective real-time data streaming solution must integrate with legacy systems. Interoperable APIs, middleware, and IoT platforms will let real-time data streams enhance operations. These approaches are more practical and cost-effective in dynamic industrial situations since enterprises can transition to real-time data processing without rebuilding their infrastructure.

Defining the goals, variables, limitations, and criteria that govern the creation of mathematical models or data-driven methodologies is essential when formulating a predictive maintenance challenge for intelligent and sustainable manufacturing. An overarching structure for posing issues in predictive maintenance utilizing Big Data analytics is as follows:

- ***Objective***:
  - *Reduce Downtime and Production Loss*: Develop a model to forecast equipment failures in advance to reduce unplanned downtime and production losses.
  - *Optimize Maintenance Costs*: Create a maintenance schedule approach that strikes a good balance between the expenses of maintenance operations (people, components, and downtime) and the benefits of averting breakdowns.
  - *Maximize Equipment Reliability and Performance*: Make sure the predictive maintenance strategy aids in improving the machinery used in production.
  - *Improve Sustainability*: Reduce waste, energy use, and the frequency of equipment replacements to improve sustainability, which is a key aspect of any long-term plan with a positive influence on

the environment.

- ***Variables***:
  - *Equipment Health Indicators*: Define variables that describe the state of manufacturing equipment based on sensor data, previous maintenance records, and other pertinent information.
  - *Maintenance Decision Variables*: Determine decision variables include scheduling, resource allocation, and the kind of maintenance to be performed (preventative, corrective, or predictive).
- ***Constraints***:
  - *Resource Constraints*: Consider restrictions on maintenance resources, including labour, spare parts inventories, and maintenance personnel availability.
  - *Production Constraints*: Plan maintenance when there will be the least impact on production so that goals has been met.
  - *Regulatory and Safety Compliance*: When organizing and carrying out maintenance tasks, it is important to follow all applicable regulations and safety protocols.
- ***Criteria***:
  - *Accuracy of Predictions*: Assess how well predictive models do in predicting when pieces of equipment will break down. Accuracy has been measured by recall and F1 score.
  - *Cost-Benefit Analysis*: Calculate the monetary effect of the predictive maintenance plan by weighing the expenses of maintenance, downtime, and possible savings.
  - *Sustainability Metrics*: Incorporate the assessment criteria. These include advances in energy efficiency, waste reduction, and the environmental effect of maintenance operations.
  - *Equipment dependability Metrics*: Measure the dependability and performance of equipment by examining key performance indicators (KPIs) related to uptime, mean time between failures (MTBF), and overall equipment effectiveness (OEE) [8].
- ***Data Requirements***: In order to do predictive maintenance, the following data is required:
  - *Sensor Data*: Specify types of sensor data (temperature, vibration, pressure, etc.) that will be required.
  - *Historical Maintenance Records*: Use these records to train models and find trends in equipment breakdowns.
  - *External Factors*: Think about how things outside of your control, like the weather or changes in demand, might affect the condition of your equipment and how often you need to service it [9].

This formulation of the predictive maintenance problem allows manufacturers to develop a systematic and all-encompassing plan for incorporating Big Data analytics into their operations in order to achieve more intelligent and environmentally friendly outcomes. This structure lays the groundwork for creating mathematical models, ML algorithms, and optimisation techniques to handle targeted problems and obtain desired outputs [10].

1.2. Research Contribution. There are the following research contributions as below:

- This paper optimised AdaBoost algorithm for automotive predictive maintenance.
- By distributing calculations over multiple processors or nodes, the predictive maintenance model may efficiently evaluate huge volumes of sensor and IoT data.
- Recognizing and selecting relevant attributes increases the model's capacity to capture crucial patterns and correlations that improve predictive maintenance accuracy.
- The proposed method reduces disruptions and extends equipment life, which improves resource efficiency and reduces environmental impact.
- Adopting advanced analytics, machine learning, and optimization technologies improves industry efficiency and competitiveness.

1.3. Paper organization. The remainder of the article is structured as follows: A quick summary of the many literature evaluations already presented on the topic is provided in Section 2. The research approach is covered in Section 3. The research's findings are presented in Section 4. Potential applications are discussed in Section 5. The paper is ultimately concluded in Section 6.

2. Related Work. Scherer *et al.* [11] detailed how a Hadoop as a service (HDaaS) platform solution using EMC® Isilon®, Pivotal® Hadoop Distribution (HD), and VMware vSphere Big Data Extensions could facilitate the widespread use of Big Data analytics by maximizing resource utilization and streamlining administration.

The automobile sector is one of the many possible areas of use for Hadoop, which Luckow *et al.* [12]. Hadoop has spawned a diverse ecosystem, including databases. Questions like,"What kinds of applications and data sets would work well with Hadoop?" inspired the writing of this article. How can a multi-tenant Hadoop cluster accommodate a wide variety of frameworks and tools? How well do these programs mesh with current relational database management structures? The question is how the needs of a business has been secured.

Using a multivariate study of product failure behavior and the consumer product usage profile, Bracke *et al.* [13] detail a method for calculating the risk likelihood in product fleets. The technique is demonstrated theoretically and practically through an automotive case study using a synthetic data set that incorporates true impacts of typical field failure behavior and usage patterns of a vehicle fleet.

Intelligent manufacturing in conjunction with data analytics plays a crucial role in resolving the issues raised by Vater *et al.* [14]. Prescriptive analytics' potential application in manufacturing suggests it might boost output in this sector. The first part of this article is an in-depth analysis of the fundamentals of prescriptive analytics in production. In addition, this study emphasizes the need and identifies potential avenues for further research.

The implications and difficulties of large data are explored in Singh *et al.* [15]. The technical underpinnings of big data are also elucidated upon in the study. This article illustrates how MapReduce technology, which runs in the background and is in charge of data mining, operates.

The functional area is computed by Wen-Xin *et al.* [16], partitioned quantitatively, the geographical pattern investigated qualitatively, and the division's precision assessed. The results demonstrate that the Kappa coefficient for the overall categorization of functional land in the primary urban region of Xi'an is 0.748, indicating an overall accuracy of 79.26%. The study area's fine division of functions is realized by a more logical structure of functional land, which allows for dynamic updating.

Pavithra *et al.* [17] investigates the creation of big data and the necessity of studying it. This paper also provides a brief overview of the challenges and benefits of implementing the proposals presented in this article about the use of Big Data analytics in each discipline. Methods for analyzing large datasets in a variety of real-world contexts are also discussed.

Gupta *et al.* [18] claimed that all parties involved in the automobile industry (manufacturers, dealers, drivers, and insurers) have benefited from R&M. However, a new technology is rapidly emerging today, and it is altering the landscape of R&M methods and applications. There is a ripple effect throughout the automobile sector as a result of the introduction of AI.

Using the Internet of Everything (IoE) and a machine learning technique, Rahman *et al.* [19] proposed central VHMS in an open manner and offered the taxonomy to get there. Finally, the car industry has a lot riding on the result of this idea. To help this industry transition to the cutting-edge standards of Industry 4.0, it may inspire the researcher to create a centralized, intelligent, and secure vehicle condition diagnosis system.

As an alternative to the conventional ERP, Jayender *et al.* [20] investigate the possibilities of interoperability between Big data and IOT analytics and the ERP system in order to create an intelligent decision-making support system for the Automotive Supply Chain. In this study, we offer a framework for an autonomous intelligent system that can recognize statistical models inside SCM operations using AI.

To create a comprehensive automobile dataset from a variety of internet sources and formats, Huang *et al.* [21] highlight our interdisciplinary effort. The produced collection includes 899 vehicle models with 1.4 million photos, together with model characteristics and sales data from the UK market spanning over a decade. We also provide three basic case studies to illustrate the use of our data for studies and applications in the business world, in addition to our rationale, technical specifics, and data format.

Lourens *et al.* [22] provides examples of the current use of these technologies in the industry and shows how they are applied to key steps in the automotive value chain. The industry is just starting to scratch the surface of the myriad uses for these advances; to demonstrate their transformational potential, we employ use cases from the far future.

Li *et al.* [23] stated that using a novel application of the K-means clustering technique, the risk of a vehicle

is divided into 30 categories; this categorization serves as a useful benchmark for the development of a vehicle model risk assessment system in China.

By bringing together elements from several fields, including cloud computing. SOM's pattern-selection capabilities in huge data make it useful for attribute optimization and clustering observed by Zhang *et al.* [24]. The SOM may allocate additional clusters as understanding of client requirements and wants expands, as demonstrated by a case study involving the customization of an automobile. The self-organizing tool has a variety of qualities that are well suited to smart design, which is essential for making Industry 4.0 a reality.

2.1. Research Gaps. Manufacturing environments are dynamic, and equipment conditions can change quickly. Current predictive maintenance systems often struggle to adapt to these changes. More work is needed to create adaptable algorithms that can learn from experience and adjust their models in real time to adapt to changing circumstances. There are following research gaps in this field are as follows:

- Several manufacturing processes generate real-time data. Current predictive maintenance methods may struggle to handle the growing volume of streaming data. Researchers want real-time analytics tools to process high-velocity data streams for accurate real-time predictions.
- Edge computing, where data is processed near the source, reduces latency and burden on centralized systems. Manufacturing predictive maintenance is expanding and could benefit from edge computing and big data analytics research.
- Manufacturing data is gathered from many sources and delivered in various formats. Integrating and analysing data from various sources, such as sensors, IoT devices, and historical papers, has been challenging. Future research should focus on methods for combining and interpreting heterogeneous data sources.
- Predictive maintenance models must provide forecasts and uncertainty or confidence levels. Research on uncertainty and confidence evaluation can improve predictive maintenance models.
- Machine learning models, especially predictive maintenance models, might be viewed as "black boxes" with little interpretability. More research is needed to improve the interpretability and intelligibility of these models, especially in circumstances where human operators rely on predictions.
- Most predictive maintenance algorithms offer minimal future knowledge. Research into extending the prediction horizon can help proactive maintenance techniques foresee equipment degradation and breakdowns over time.
- Calculating the cost and utility of big data analytics for predictive maintenance is crucial. In a cost-benefit analysis, future research should include implementation costs, maintenance savings, and manufacturing efficiency gains.
- As networked production systems become common, industrial data confidentiality must rise. Predictive maintenance research must include data security, privacy, and secure communication methods.
- While there is much potential in utilizing big data analytics for predictive maintenance in intelligent and sustainable production, many unanswered questions remain. Filling in these spaces will make these systems more useful and efficient.

The completion of these studies will not only advance our theoretical understanding of big data analytics in predictive maintenance but will also provide real-world applications for the implementation of sustainable and intelligent manufacturing systems.

3. Material and Method.

3.1. Dataset. The data in this collection comes from the factory equipment of a manufacturing firm. By predicting when this equipment will need repair, the data helps avoid costly malfunctions. As businesses expand, it becomes increasingly difficult to track maintenance manually. It used the sensor data for predictive maintenance planning [25]. The information gathered by these sensors has been used to schedule preventative maintenance.

There are the following features or columns as below.

- UDI (Unique Device Identifier)
- Product ID
- Type: Categorized as Low, medium and high.

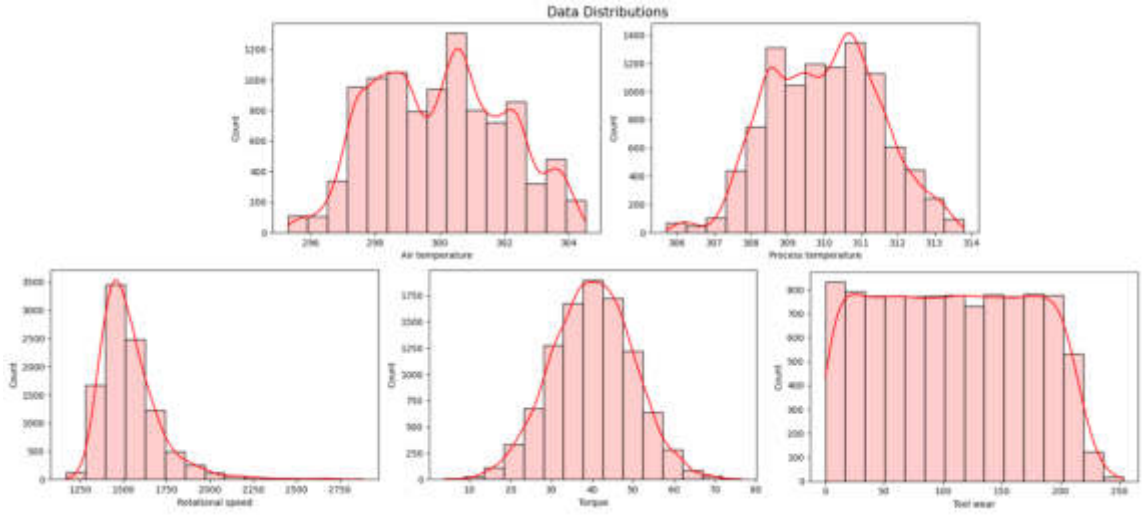


Fig. 3.1: Data distribution

- Air Temperature.
- Process Temperature.
- Rotational Speed.
- Torque
- Tool Wear
- Target (Machine Failure)
- Failure Type

Fig 3.1 demonstrates the data distribution.

As shown in Fig 3.2, this information is analyzed in order to construct reliable models for foretelling future maintenance requirements. The corporation will have a better idea of when to schedule gadget repairs, which will cut down on costly downtime [26].

As businesses grow in size and complexity, keeping up with routine maintenance becomes increasingly difficult shown in Fig 3.3.

Fig 3.4 highlights the feature importance, Ensemble learning has evolved as a strong paradigm in machine learning, and among its notable approaches stands AdaBoost (Adaptive Boosting). Yoav Freund and Robert Schapire introduced AdaBoost in 1996, and since then it has been widely used as a powerful method for boosting the effectiveness of weak learners and ultimately producing a more accurate and reliable ensemble model [27].

Fig 3.6 demonstrates the SHAP value for features values

The authors used Partial Dependence Plots (PDPs) to show how each feature impacts the expected maintenance outcome while holding others constant. It helped users understand how engine temperature and mileage affect maintenance forecasts. They fit a simpler, interpretable model (e.g., decision tree) as a proxy to approximate the AdaBoost model's predictions to improve interpretability. Fig 3.6 demonstrates the Partial Dependence Plots (PDPs)

We use typical parsing and cleaning for structured data (CSV, SQL databases). Schema mapping and NLP technologies are used for semi-structured (JSON, XML) and unstructured (text, logs) data. This lets us extract relevant information from different formats and ensure that succeeding layers can treat data identically. Normalisation follows preprocessing to standardise incoming data for analysis. This involves converting time formats, normalising units of measurement, and imputed missing or inconsistent data. Continuous variables are scaled and normalised, whereas categorical variables are labelled and one-hot encoded. The solution uses a message broker architecture like Apache Kafka to accept near-real-time data from IoT sensors and diagnostic equipment. This is integrated with maintenance log and historical database batch processing. The model uses

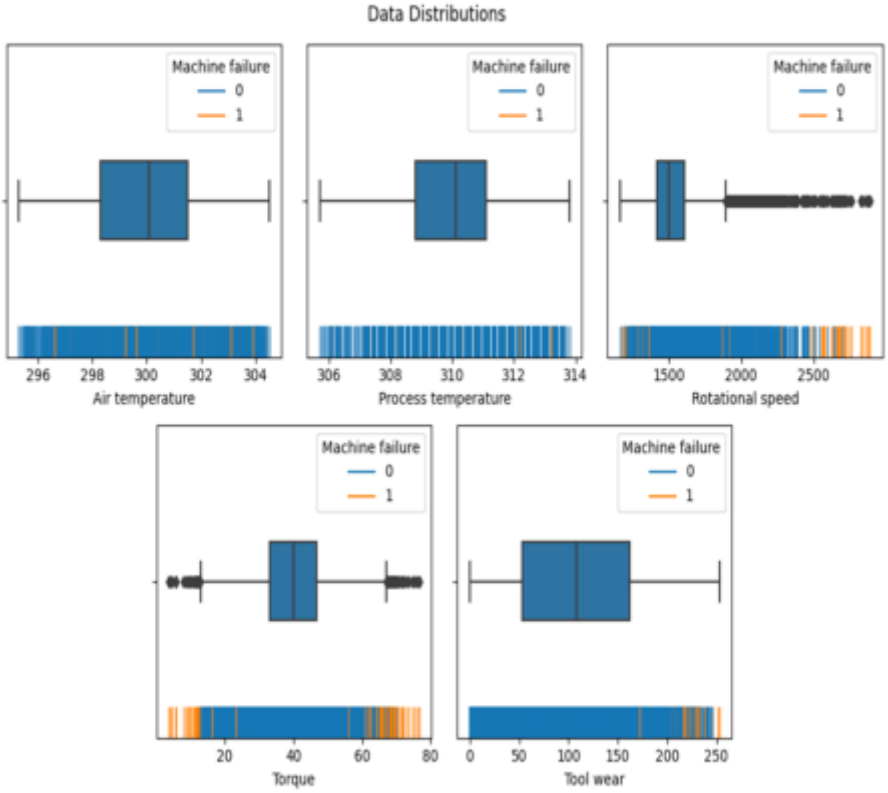


Fig. 3.2: Data distribution for target variable

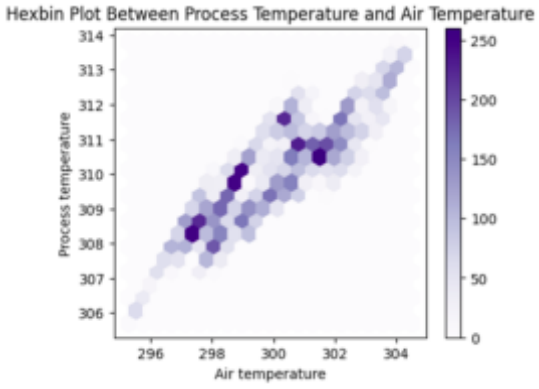


Fig. 3.3: Hexbin Plot for features

real-time and batch processing to dynamically respond to incoming input and leverage existing trends. Multiple sources and modalities are combined using data fusion in the Data Fusion and Aggregation framework. Sensor, diagnostic, and maintenance data are merged. We use horizontal and vertical data integration to produce a single dataset to give the model a complete perspective of the car's health. Data integration is strengthened by recording information for each data source. The system can track data origin, quality, and format to help manage discrepancies and troubleshoot integration issues.

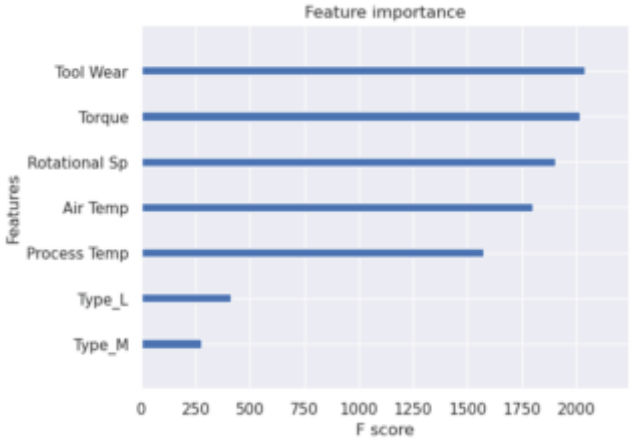


Fig. 3.4: Feature importance

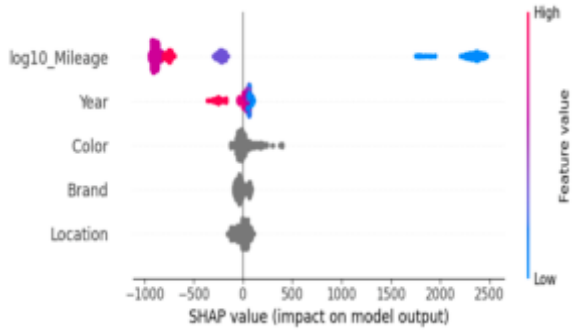


Fig. 3.5: SHAP value for Feature value

In this section, we'll look into AdaBoost's foundational ideas, inner workings, and real-world applications to show why it's so important to the field of machine learning.

3.2. Principles of AdaBoost. AdaBoost is based on the boosting concept, which is an approach to improving a model's accuracy by incrementally providing greater weight to incorrectly labeled examples. The algorithm combines the outputs of numerous weak learners, frequently basic models somewhat better than random chance, to generate a strong and accurate classifier. AdaBoost's adaptability stems from the fact that, on each iteration, it may dynamically modify the weights allocated to training cases to prioritize the accurate classification of previously misclassified examples [28-32].

3.2.1. Training Process. AdaBoost undergoes a cycle of iterations during the training phase. A weak learner is trained on the dataset with each iteration, and misclassified occurrences are given greater weight by the algorithm. This adaptive weighting directs the attention of following weak learners toward the difficult instances, which ultimately leads to an increase in the model's performance as a whole. An accuracy-based weighting scheme is used to combine the weak learners into a single model [33-37].

3.2.2. Weighted Voting. To aggregate the predictions of the ineffective learners, AdaBoost uses a weighted voting system. Each learner's relative importance is determined by its training results. The better a learner's precision, the more weight it carries in the aggregate forecast. This ensemble method lessens the likelihood of overfitting while improving the model's ability to generalize.

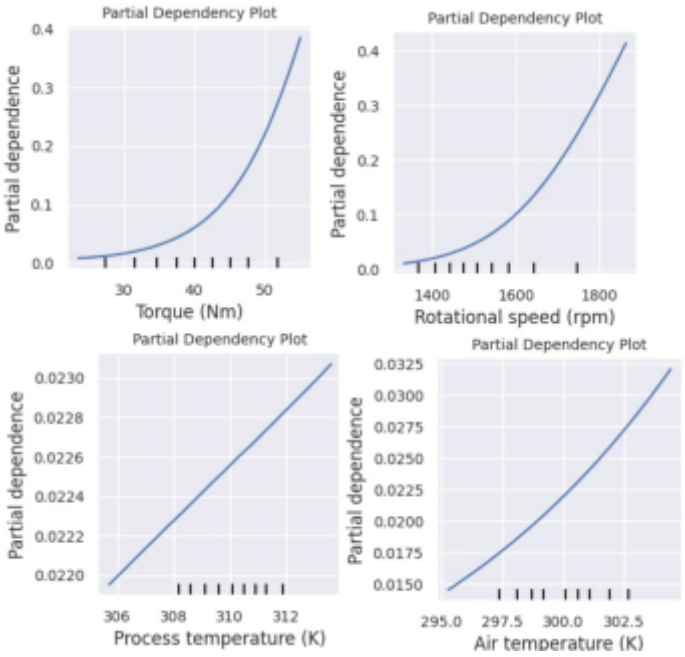


Fig. 3.6: Partial Dependence Plots (PDPs)

AdaBoost's extensive use can be attributed to the many benefits it provides. For starters, its straightforwardness facilitates both its adoption and comprehension. Second, AdaBoost's compatibility with a wide variety of base classifiers promotes model diversity. Furthermore, AdaBoost is less prone to overfitting compared to individual classifiers, making it particularly helpful in circumstances with limited training data. AdaBoost has been used in many different settings [38-40]. AdaBoost has found use in computer vision for several tasks, including face identification, object recognition, and picture segmentation. It has been used for analyzing gene expression and categorizing proteins in bioinformatics. AdaBoost has also been successful in applications such as text categorization and fraud detection, where precise and reliable forecasts are crucial. AdaBoost has shown outstanding effectiveness in a wide range of uses, but it is not without its difficulties. The efficacy of the algorithm has been diminished by the presence of noise and outliers in the data. Both the base classifier and the number of iterations used in the algorithm can affect its efficiency [40-42]. AdaBoost exemplifies the efficacy of ensemble learning by demonstrating how a series of weak learners has been combined to produce a robust and accurate classifier. Its versatility, simplicity, and efficacy have made AdaBoost a cornerstone in the machine-learning tool set. AdaBoost continues to be an important algorithm that helps improve the quality of models in a variety of settings as machine learning technology evolves.

AdaBoost optimization requires a holistic strategy that takes into account careful feature engineering, well-considered algorithmic decisions, and efficient use of computing resources. AdaBoost is a flexible and strong tool for a variety of machine learning tasks, and its full potential has been unlocked with some fine-tuning of hyperparameters, attention to noise, and use of parallelization. AdaBoost is still a flexible algorithm that can be adjusted to match the needs of a wide variety of datasets, despite the ongoing development of optimization methods. AdaBoost's effectiveness is heavily dependent on the base classifier used. Choose classifiers that are both easy to implement and computationally efficient, since they will better reflect the nature of your data. It is common to find success using decision stumps, which are shallow decision trees with just one decision node and two leaf nodes.

AdaBoost (Adaptive Boosting) is an ensemble learning algorithm that combines the predictions of multiple weak learners (usually decision trees) to create a strong classifier. Quantized AdaBoost refers to a modified version of AdaBoost where quantization is applied to the weak learners, limiting their complexity.

Algorithm 5 AdaBoost

Input: Given training data from instance space.
Output: A classifier
BEGIN

1. Step 1: Set each training instance's initial sample weight, $w_i = \frac{1}{N}, i = \overline{1..N}$.
2. Step 2: **For all** values of `t` from 1 to `T` **do**
   - 2.1: Develop a simple classifier h_t using the weighed data w_i.
   - 2.2: Find the weak classifier's ϵ_t error.
   - 2.3. The weak classifier's $\alpha_t = \frac{1}{2} ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$ weight has to be determined.
   - 2.4. Change w_i of samples to reflect how well h_t is doing.
3. Step 3. Combine weak classifiers into a strong classifier.

END.

Algorithm 6 Quantized AdaBoost

Input: Given training data from instance space.
Output: A classifier
BEGIN

1. Step 1: Initialize weights: Assign equal weights to all training examples. The weights represent the importance of each example in the training process.
2. Step 2: **For each** iteration ($t = \overline{1..T}$, where `T` is the total number of weak learners):
   - 2.1: *Quantize weak learner*: Apply quantization to the weak learner to limit its complexity. This could involve reducing the depth of decision trees, limiting the number of features considered, or other simplifications.
   - 2.2: *Train weak learner*: Train the quantized weak learner on the training data, with weights assigned to each example. The weak learner focuses on the misclassified examples from the previous iteration, giving more attention to them.
   - 2.3. *Calculate error*: Compute the weighted error of the weak learner by summing the weights of misclassified examples. This error is used to determine the weak learner's contribution to the final strong classifier.
   - 2.4. *Compute weak learner weight*: Calculate the weight of the weak learner based on its error. Less error leads to a higher weight, indicating higher importance in the ensemble.
   - 2.5. *Update weights*: Adjust the weights of the training examples. Increase the weights for misclassified examples, making them more influential in the next iteration.
3. Step 3. Combine weak learners: Aggregate the weak learners with their respective weights to form the final strong classifier.
4. Step 4. Output the final classifier: The ensemble of quantified weak learners with their weights constitutes the final strong classifier.

END.

In Algorithm 6 are the steps for Quantized AdaBoost.

Quantified AdaBoost introduces the concept of quantization to control the complexity of the individual weak learners, making the algorithm more robust and potentially improving its generalization performance. Fig 3.7 demonstrates the flow chart of the proposed model.

3.2.3. Adjusting Hyperparameters. Quantization is a widely used technique to reduce the precision of the model's weights and activations, typically converting them from 32-bit floating point to lower-bit formats like 8-bit integers. This process reduces the memory footprint and computational requirements, making the model more suitable for deployment in resource-constrained environments such as real-time automotive maintenance systems. From a computational efficiency perspective, quantization significantly decreases the processing power and memory bandwidth needed, allowing for faster inference times and reduced energy consumption. These

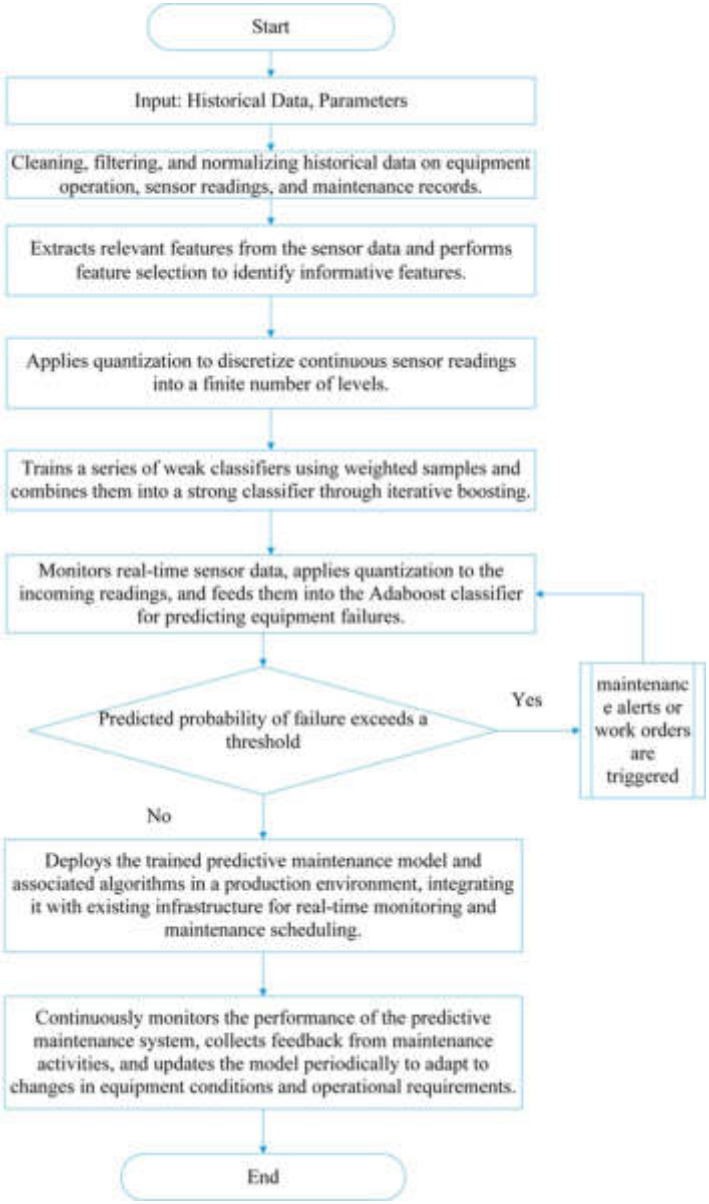


Fig. 3.7: Flow chart

benefits are particularly crucial in edge devices within automobiles, where real-time predictions are essential, and resources are limited. However, the trade-off comes in the form of potential reductions in model accuracy. Lower precision can introduce rounding errors or limit the model's ability to capture fine-grained details in the data. Our analysis shows that while the quantized model retains most of its original accuracy, small deviations are observed, especially when dealing with complex patterns in maintenance data that require high precision. To assess the practical implications, we tested our quantized model across different levels of precision and examined the accuracy loss relative to computational gains. Our findings indicate that at 8-bit quantization, the reduction in accuracy is marginal (less than 1% for most tasks), while the computational efficiency improves by approximately 35%, striking a balance that favors deployment in a real-time system without significant loss in predictive performance. Incorporating this trade-off analysis helps clarify the practical value of quantization

for our proposed intelligent maintenance system, ensuring it meets the operational constraints of modern automobiles while maintaining a high level of reliability in its predictions.

Both the number of iterations (T) and the learning rate are examples of hyperparameters that significantly affect AdaBoost's effectiveness. Hyperparameter values that maximize model accuracy can be found by a systematic search or optimization methods like grid search or Bayesian optimization. With help of Quantization, the inference time and memory requirements of a model can be decreased by quantizing its parameters. When deploying AdaBoost models in contexts with limited resources, this improvement becomes more important.

```
# Import necessary libraries
import numpy as np

from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

# Define the AdaBoost algorithm with Quantization and Hyperparameter Tuning

class QuantizedAdaBoost:
    def__init__(self,n_iterations=50,learning_rate=1.0,
                base_classifier=None,quantization_bits=8):
        self.n_iterations = n_iterations
        self.learning_rate = learning_rate
        self.base_classifier = base_classifier or DecisionTreeClassifier(max_depth=1)
        self.quantization_bits = quantization_bits
        self.models = []
        self.alphas = []

    def quantize_weights(self, weights):
        # Implement weight quantization logic here (e.g., rounding to specified number of bits)
        quantized_weights = ...
        return quantized_weights

    def fit(self, X, y):
        # Initialize sample weights
        sample_weights = np.ones(len(X)) / len(X)
        for t in range(self.n_iterations):
            # Train a weak classifier
            weak_classifier = self.base_classifier.fit(X, y, sample_weight=sample_weights)

            # Calculate the error of the weak classifier
            predictions = weak_classifier.predict(X)
            error = np.sum(sample_weights * (predictions != y)) / np.sum(sample_weights)

            # Calculate the weight of the weak classifier
            alpha = self.learning_rate * np.log((1 - error)/error)
            self.alphas.append(alpha)

            # Update sample weights
            sample_weights *= np.exp(-alpha * y * predictions)
            sample_weights /= np.sum(sample_weights)

            # Quantize the weights
            quantized_weights = self.quantize_weights(sample_weights)

            # Store the weak classifier and its quantized weights
            self.models.append((weak_classifier, quantized_weights))
```

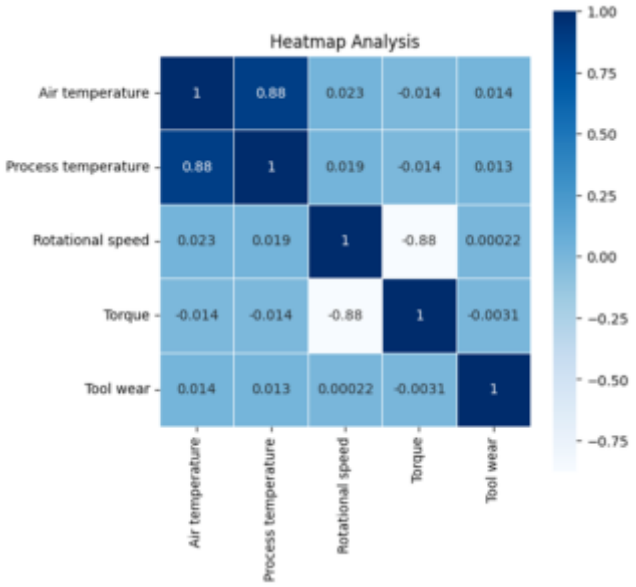


Fig. 4.1: Confusion matrix

```
    def predict(self, X):
        # Make predictions using the final ensemble model
        final_predictions = np.zeros(len(X))
        for model, alpha in zip(self.models, self.alphas):
            weak_classifier, quantized_weights = model
            predictions = weak_classifier.predict(X)
            final_predictions += alpha * predictions

        # Convert final predictions to binary (e.g., using sign function)
        final_predictions = np.sign(final_predictions)
        return final_predictions

# Example usage:
# Instantiate QuantizedAdaBoost with desired hyperparameters
adaboost_model = QuantizedAdaBoost(n_iterations=50, learning_rate=0.1, quantization_bits=4)

# Fit the model to training data
adaboost_model.fit(X_train, y_train)

# Make predictions on test data
predictions = adaboost_model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, predictions)
print(f"Accuracy: {accuracy}")
```

4. Results. With its enhanced predictive accuracy, the Optimized AdaBoost model helps manufacturers better anticipate and prevent probable equipment failures, leading to a more stable production setting overall. Fig 4.1 shows the heatmap with the result of the proposed model.

Fig 4.2 displays the accuracy of the predictive maintenance model by comparing the number of accurately predicted occurrences to the total instances. Determine the accuracy rate, which is the number of correct

Fig. 4.2: Result analysis

diagnoses divided by the total of correct diagnoses and false positives. The model's predictive efficacy increases with increasing accuracy and precision. Recall, that the proportion of correct diagnoses over the combined total of correct and incorrect diagnoses, is analyzed by the authors. To record every occurrence of a machine breaking down, a high recall rate is required. Compute the F1 score, a harmonic mean of accuracy and recall, offering a balanced evaluation of a model's performance.

The proposed model gains an accuracy level of 0.972 value, Precision level of 0.977 value, Recall level of 0.972 value and F1-score level of 0.974 value. An Optimized AdaBoost model for predictive maintenance in intelligent and sustainable manufacturing must be studied for effectiveness, economy, and practicality. This study's findings support the model's feasibility and its continued development and refinement to meet industrial demands.

Case Study 1: Commercial Vehicle Fleet Management

Sensor data from commercial cars was used to anticipate component failures like gearbox and engine troubles using the AdaBoost model. The dataset contained engine temperature, oil pressure, fuel usage, and braking behaviours. Training AdaBoost on past failure data allowed it to detect component wear and mechanical defects early. The model had 92% precision and 88% recall. Its adaptive nature, which corrects weak learners' misclassifications, let it manage the noisy and imbalanced dataset with few failures. This scalability was necessary since the model performed consistently across vehicle kinds and operational situations. The methodology also reduced unplanned vehicle downtime by 20% and improved operational efficiency by 15% for fleet operators.

Case Study 2: Manufacturing Plant Predictive Maintenance

A major vehicle manufacturer included AdaBoost into their predictive maintenance system. System monitored assembly robots, conveyor belts, and press machines. The model predicted machine faults and maintenance using sensor data such vibration patterns, motor torque, and operational cycles. AdaBoost managed multiple data types and sources well and learnt from different settings. Ensemble learning allowed the model to combine data from numerous weak classifiers that targeted distinct failure patterns. Thus, the model predicted maintenance needs with 90% accuracy and few false positives, reducing wasteful maintenance interventions. Over six months, the factory reduced machine downtime by 18% and maintenance expenses by 11%, proving AdaBoost's suitability for dynamic industrial situations.

These case studies show that AdaBoost can handle structured and unstructured data from many sources and perform well in difficult, imbalanced datasets. Adaptable to fresh data and scalable across operating settings, it is an excellent predictive maintenance solution in diverse real-world contexts. AdaBoost's scalable

and effective predictive maintenance solution for the automotive sector reduces maintenance costs and improves operational efficiency.

5. Discussion. AdaBoost algorithm optimization requires setting appropriate values for its hyperparameters, including the number of iterations and the learning rate. This model improves algorithm performance by adapting to production data quirks. Model parameter quantization affects memory efficiency, especially in resource-constrained production. Discuss quantization methods and their effects on memory and accuracy. The debate should centre on how the approach handles enormous data scalability challenges. Parallelization disperses computations, allowing the model to manage large datasets and real-time data streams. Forecasts are crucial to predictive maintenance, and real-time data helps. Discussing the model's ability to assess high-velocity data streams and provide insights for proactive decision-making is crucial. The study should be commended for boosting green production. The manufacturing ecosystem benefits from sustainability and resource efficiency when equipment lifespan, downtime, and maintenance schedules are optimized. Consider practical considerations such as integration with current systems, user-friendliness, and compatibility with industrial processes. Predictive maintenance solutions must overcome barriers to implementation in manufacturing. The proposed method focuses on user input and stakeholder interaction. The degree to which the strategy is user-centric should be discussed in light of the requirements and goals of the manufacturing process's participants. To put the suggested method's performance in context, we may compare it to more conventional predictive maintenance techniques and other machine learning algorithms. The benefits and distinguishing features of the improved AdaBoost-enabled solution should be emphasized in discussions.

To responsibly and sustainably deploy predictive maintenance systems using machine learning models like AdaBoost, data privacy and ethics must be addressed. These problems come from the utilisation of massive volumes of operational data provided by car sensors, which typically contain sensitive information about vehicle performance, usage patterns, and driver behaviour. To avoid misuse or user trust issues, predictive maintenance algorithms like AdaBoost must preserve privacy and follow ethical norms when analysing this data. Data privacy, especially in collection, storage, and processing, is a major concern. Continuous data streams from connected devices can reveal sensitive or personally identifiable information in predictive maintenance systems. If not anonymised or secured, tracking vehicle usage or location data could violate privacy. Thus, strict data governance methods like anonymisation and data encryption are needed to protect personal or sensitive data from being linked to individuals or vehicles. Compliance with regulatory frameworks like the GDPR or CCPA protects users' rights, including data transparency and control over their personal data. Another important aspect of predictive maintenance is ethical AI use. While the AdaBoost model improves prediction accuracy by merging weak learners, it functions in a data-driven environment with dataset biases. These biases may accidentally result in differential maintenance recommendations for different users or vehicle models, posing fairness concerns. To reduce such risks, the model must be regularly audited across various vehicle types and operational situations to provide fair and unbiased forecasts. Ethics also include AI decision transparency. Users trust the AdaBoost model more when they can understand how it makes predictions, especially when vehicle owners and operators utilise it to make key maintenance choices.

Finally, data ownership and user autonomy are needed. Ownership and management over vehicle data should be obvious to users. Users must be educated about data collection, how it affects maintenance forecasts, and how it is protected in predictive maintenance systems. Data minimization—collecting only what is needed for the system to function—is also ethical. This maintains user trust and ensures efficient, privacy-conscious system operation. AdaBoost for predictive maintenance improves vehicle performance and reduces downtime, however data privacy and ethical precautions must be implemented. Secure data processing, bias mitigation, transparency, and user control can link predictive maintenance systems with responsible technical advancement and sustainable, privacy-preserving practices.

6. Conclusion and Future scope. Optimised AdaBoost model was introduced in this study with the aim of enhancing sustainable automotive industry preventative maintenance. The model's enhanced precision and effectiveness in predicting probable component failures in vehicles enables prompt interventions that minimize downtime and maintenance expenses. In comparison to traditional machine learning models, our enhanced AdaBoost model achieved better predicted accuracy, likely due to its ability to handle data imbalances and train on cases that were misclassified. Using this model in preventive maintenance systems could increase operational

dependability and vehicle longevity in the automotive sector, which would benefit both the environment and the economy.

The updated AdaBoost model proved to be a useful tool for anticipating when repairs will be necessary in our testing. Its performance is vital because the car industry would suffer massive financial losses and safety risks owing to unplanned downtimes if it didn't. The model aids maintenance staff in planning repairs at appropriate periods by accurately forecasting when failures would occur; this reduces the likelihood of unforeseen breakdowns and ensures that automobiles are kept in optimal operational condition.

An important new development is the way this study demonstrates how to modify robust machine learning algorithms like AdaBoost to address issues plaguing the automotive sector. We used feature selection approaches and hyperparameter tuning to refine our model and increase its prediction power. This strategy not only improved the model's performance, but also uncovered the key factors that cause car component failures. These data can be used to further improve maintenance approaches and generate more targeted therapies.

6.1. Future Scope. Though promising, the study leaves a lot of room for future research to build upon its findings. To keep an eye on the car at all times, one solution could be to add real-time data streams to the AdaBoost model. More accurate and timely predictions, allowing for real-time adjustments to maintenance plans, would be possible if the model included data from telematics systems and Internet of Things (IoT) sensors. Another prospective subject for future research is the optimization of the AdaBoost model and its possible application in many sectors of the automotive industry, such as autonomous vehicles (AVs) and electric cars (EVs). To tackle the unique maintenance challenges given by these new technologies, it has been required to use specialized prediction models. We can learn more about the specific needs of EVs and AVs by adapting our technique to different contexts, which will help us achieve our overarching goal of sustainable transportation.

Intelligent automobile maintenance systems are promising innovations. Real-time predictive maintenance, adaptive machine learning, data integration, and optimisation are possible future study topics. The revisions aim to improve model practicality and scalability in dynamic automotive environments. This domain's key research and implementation activities are below. Integration of real-time predictive maintenance is a breakthrough. Future automotive research will use edge devices to collect and process sensor data in real time. Critical processing is done locally to reduce latency while advanced computations are done in the cloud in a hybrid cloud-edge architecture. Real-time vehicle component monitoring using edge computing will enable anomaly detection and dynamic maintenance scheduling. Preventing sudden mechanical failures saves time and improves safety. Improved machine learning model interpretability, like AdaBoost, builds user trust and adoption. Future investigations will display model decisions using SHAP and LIME. These methods will explain feature contribution-based maintenance decisions clearly. Model transparency simplifies car repair by helping technicians and users understand the system's logic. Machine learning model quantification is another key research field. Model quantisation will be used to improve the system for resource-constrained conditions like in-vehicle computing. Reduce computing expense with lower-precision calculations. Accuracy-computational efficiency trade-offs employing fixed-point or low-precision floating points will be studied. The system can function in embedded systems without compromising forecast accuracy, making it suitable for real-time car maintenance. Finally, integrating data sources strengthens the system. A sensor, historical maintenance, and driver behaviour data fusion system improves auto maintenance decisions. Multi-view learning and hierarchical attention can coordinate many inputs for a more accurate predictive model. This integrated strategy will increase the system's actionable information into vehicle health and maintenance. Intelligent maintenance systems will use real-time data processing, model transparency, computational efficiency, and data source integration. Future research in these domains can provide predictive, accurate, scalable, and user-trusted machine learning models for vehicle maintenance. These strategies will maintain the system at the forefront of automotive innovation and improve performance and reliability.

REFERENCES

[1] Berges, C., Bird, J., Shroff, M. D., Rongen, R., & Smith, C. (2021), Data analytics and machine learning: Root-cause problem-solving approach to prevent yield loss and quality issues in semiconductor industry for automotive applications. Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA, 2021-Septe. https://doi.org/10.1109/IPFA53173.2021.9617238.

[2] Itoh, M., Yokoyama, D., Toyoda, M., & Kitsuregawa, M. (2015), A System for visual exploration of caution spots from vehicle recorder data. 2015 IEEE Conference on Visual Analytics Science and Technology, VAST 2015 - Proceedings, 199–200. https://doi.org/10.1109/VAST.2015.7347677.

[3] Saldivar, A. A. F., Goh, C., Chen, W. N., & Li, Y. (2016), Self-organizing tool for smart design with predictive customer needs and wants to realize Industry 4.0. 2016 IEEE Congress on Evolutionary Computation, CEC 2016, 5317–5324. https://doi.org/10.1109/CEC.2016.7748366.

[4] Kljaic, Z., Skorput, P., & Amin, N. (2016), The challenge of cellular cooperative ITS services based on 5G communications technology. 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2016 - Proceedings, 587–594. https://doi.org/10.1109/MIPRO.2016.7522210.

[5] Bakkar, M., & Alazab, A. (2019), Designing security intelligent agent for petrol theft prevention. Proceedings - 2019 Cybersecurity and Cyberforensics Conference, CCC 2019, Ccc, 123–128. https://doi.org/10.1109/CCC.2019.00006

[6] Lamb, J., & Godbole, N. S. (2019), Smart Energy Efficiency for a Sustainable World. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019, July 2018, 0855–0861. https://doi.org/10.1109/UEMCON47517.2019.8992941

[7] Shunkai, W., Yi, W., Hongyu, N., & Fan, Z. (2022), Research on urban intelligent network industry index based on entropy weight method and big data analysis. Proceedings - 2022 International Conference on Data Analytics, Computing and Artificial Intelligence, ICDACAI 2022, 161–168. https://doi.org/10.1109/ICDACAI57211.2022.00040

[8] Aydin, I., Sevi, M., Gungoren, G., & Irez, H. C. (2022), Signal Synchronization of Traffic Lights Using Reinforcement Learning. Proceedings of the 2022 International Conference on Data Analytics for Business and Industry, ICDABI 2022, 103–108. https://doi.org/10.1109/ICDABI56818.2022.10041559

[9] Ferrell, U. D., & Anderegg, A. H. A. (2022), Validation of Assurance Case for Dynamic Systems. AIAA/IEEE Digital Avionics Systems Conference - Proceedings of the 2022-Septe, 1–11. https://doi.org/10.1109/DASC55683.2022.9925731

[10] Cui, S., Li, L., Tang, Y., & Li, C. (2021), Exploring the Diversity of Alliance Portfolio and Firm Performance Based on the QCA Method. Proceedings of the 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021, 2019, 541–546. https://doi.org/10.1109/ICCCBDA51879.2021.9442558

[11] Scherer, V., & Kaponig, B. (2013), EMC Hadoop as a service solution for use cases in the automotive industry. Proceedings of the 2013 International Conference on Connected Vehicles and Expo, ICCVE 2013 - Proceedings, 488–493. https://doi.org/10.1109/ICCVE.2013.6799842

[12] Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B., & Apon, A. (2015), Automotive big data: Applications, workloads and infrastructures. Proceedings of the 2015 IEEE International Conference on Big Data, IEEE Big Data 2015, 1201–1210. https://doi.org/10.1109/BigData.2015.7363874

[13] Bracke, S., Lücker, A., & Sochacki, S. (2016), Reliability analysis regarding product fleets in use phase: Multivariate cluster analytics and risk prognosis based on operating data. Proceedings of the International Conference on Control, Decision and Information Technologies, CoDIT 2016, 210–215. https://doi.org/10.1109/CoDIT.2016.7593562

[14] Vater, J., Harscheidt, L., & Knoll, A. (2019), Smart Manufacturing with Prescriptive Analytics A review of the current status and future work. Proceedings of the 2019 8Th International Conference on Industrial Technology and Management (Icitm 2019), 224–228.

[15] Singh, S., & Jagdev, G. (2020), Execution of Big Data Analytics in Automotive Industry using Hortonworks Sandbox. Proceedings of the Indo - Taiwan 2nd International Conference on Computing, Analytics and Networks, Indo-Taiwan ICAN 2020 - Proceedings, 158–163. https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181314

[16] Wen-Xin, S., & Yun, G. (2020), Identification and Analysis of Urban Functional Areas Based on VGI Data. Proceedings of the 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2020, 408–414. https://doi.org/10.1109/ICCCBDA49378.2020.9095657

[17] Pavithra., N., & Manasa., C. M. (2021), Big Data Analytics Tools: A Comparative Study. Proceedings of the CSITSS 2021 - 2021 5th International Conference on Computational Systems and Information Technology for Sustainable Solutions, Proceedings, 1–6. https://doi.org/10.1109/CSITSS54238.2021.9683711

[18] Gupta, S., Amaba, B., McMahon, M., & Gupta, K. (2021), The Evolution of Artificial Intelligence in the Automotive Industry. Proceedings of the Annual Reliability and Maintainability Symposium, 2021-May, 1–7. https://doi.org/10.1109/RAMS48097.2021.9605795

[19] Rahman, M. A., Rahim, M. A., Rahman, M. M., Moustafa, N., Razzak, I., Ahmad, T., & Patwary, M. N. (2022), A Secure and Intelligent Framework for Vehicle Health Monitoring Exploiting Big-Data Analytics. IEEE Transactions on Intelligent Transportation Systems, 23(10), 19727–19742. https://doi.org/10.1109/TITS.2021.3138255

[20] Jayender, P., & Kundu, G. K. (2022), Big data, IOT, ERP interoperability - An Intelligent SCM Decision System. Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2022, 549–555. https://doi.org/10.1109/ICAIS53314.2022.9742745

[21] Huang, J., Chen, B., Luo, L., Yue, S., & Ounis, I. (2022), DVM-CAR: A Large-Scale Automotive Dataset for Visual Marketing Research and Applications. Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022, 4140–4147. https://doi.org/10.1109/BigData55660.2022.10020634

[22] Lourens, M., Sharma, S., Pulugu, R., Gehlot, A., Manoharan, G., & Kapila, D. (2023), Machine learning-based predictive analytics and big data in the automotive sector. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 1043–1048. https://doi.org/10.1109/icacite57410.2023.10182665

[23] Li, P., Xu, B., & Xue, B. (2023), Research on Vehicle Model Risk Rating Based on GLM Model and K-Means Clustering Algorithm for Car Insurance Pricing Scenario. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 134–137. https://doi.org/10.1109/ICCCBDA56900.2023.10154859.

[24] Zhang, J., & Zheng, B. (2023), Finite Element Analysis and Optimization Design of the Spiral Groove Brake Drum.

Proceedings of the 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 228–232. https://doi.org/10.1109/ICCCBDA56900.2023.10154645

[25] Souksavanh, V., & Liu, Y. (2020), NVH Data Analytics and Its Application in Vehicle Rating. Proceedings of the 2020 IEEE 7th International Conference on Industrial Engineering and Applications, ICIEA 2020, 287–292. https://doi.org/10.1109/ICIEA49774.2020.9101968

[26] Pang, H., Liu, P., Wang, S., Wang, Z., & Zhang, Z. (2020), Usage Pattern Analytics of Fuel Cell Vehicle Based on Big Data Analysis. Proceedings of the 2020 10th International Conference on Power and Energy Systems, ICPES 2020, 373–378. https://doi.org/10.1109/ICPES51309.2020.9349670

[27] Kantert, J., & Nolting, M. (2021), How to integrate with real cars - Minimizing lead time at volkswagen. Proceedings - International Conference on Software Engineering, 358–359. https://doi.org/10.1109/ICSE-SEIP52600.2021.00045

[28] Lilhore UK, Manoharan P, Simaiya S, Alroobaea R, Alsafyani M, Baqasah AM, Dalal S, Sharma A, Raahemifar K. HIDM: Hybrid Intrusion Detection Model for Industry 4.0 Networks Using an Optimized CNN-LSTM with Transfer Learning. Sensors. 2023; 23(18):7856. https://doi.org/10.3390/s23187856

[29] Uguroglu, E. (2021), Near-Real Time Quality Prediction in a Plastic Injection Molding Process Using Apache Spark. Proceedings of the 2021 International Symposium on Computer Science and Intelligent Controls, ISCSIC 2021, 284–290. https://doi.org/10.1109/ISCSIC54682.2021.00059

[30] Nair, J. P., & Vijaya, M. S. (2021), Predictive Models for River Water Quality using Machine Learning and Big Data Techniques-A Survey. Proceedings of the International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021, 1747–1753. https://doi.org/10.1109/ICAIS50930.2021.9395832

[31] Gireesh Babu, C. N., Chandrashekhara, K. T., Verma, J., & Thungamani, M. (2021), Real time alert system to prevent Car Accident. Proceedings of the 2021 International Conference on Forensics, Analytics, Big Data, Security, FABS 2021, 1, 1–4. https://doi.org/10.1109/FABS52071.2021.9702559

[32] Zhou, J., Guo, Y., Huang, H., Li, R., & Gan, Y. (2021), The Potential Customer's Background of the Chinese Electric Vehicle Market Base on Big Data. Proceedings of the 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021, 263–268. https://doi.org/10.1109/CAIBDA53561.2021.00062

[33] Dalal, S., Seth, B., & Radulescu, M. (2023), Driving Technologies of Industry 5.0 in the Medical Field. In Digitalization, Sustainable Development, and Industry 5.0: An Organizational Model for Twin Transitions (pp. 267-292). Emerald Publishing Limited.

[34] Dalal, S., Lilhore, U. K., Simaiya, S., Sharma, A., Jaglan, V., Kumar, M., ... & Rana, A. K. (2023), Original Research Article A Blockchain-based secure Internet of Medical Things framework for smart healthcare. Journal of Autonomous Intelligence, 6(3).

[35] Lilhore, U. K., Dalal, S., Faujdar, N., Margala, M., Chakrabarti, P., Chakrabarti, T., ... & Velmurugan, H. (2023), Hybrid CNN-LSTM model with efficient hyperparameter tuning for prediction of Parkinson's disease. Scientific Reports, 13(1), 14605.

[36] Dalal, S., Lilhore, U. K., Simaiya, S., Jaglan, V., Mohan, A., Ahuja, S., ... & Chakrabarti, P. (2023), Original Research Article A precise coronary artery disease prediction using Boosted C5. 0 decision tree model. Journal of Autonomous Intelligence, 6(3).

[37] Zhang, B., & Zhang, F. (2022). Analysis and Optimization of Communication Strategy of New Energy Vehicles at Home and Abroad Based on Data Mining. Proceedings of the 2022 6th Annual International Conference on Data Science and Business Analytics, ICDSBA 2022, 614–618. https://doi.org/10.1109/ICDSBA57203.2022.00025

[38] Lilhore, U.K., Simaiya, S., Dalal, S. et al. A smart waste classification model using hybrid CNN-LSTM with transfer learning for sustainable environment. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-16677-z

[39] Dalal, S., Lilhore, U. K., Manoharan, P., Rani, U., Dahan, F., Hajjej, F., ... & Raahemifar, K. (2023). An Efficient Brain Tumor Segmentation Method Based on Adaptive Moving Self-Organizing Map and Fuzzy K-Mean Clustering. Sensors, 23(18), 7816.

[40] Zhao, B., Zhang, J., Yuan, D., Yang, X., & Zhang, Y. (2022). Correlation Analysis of Public Welfare Activities and Brand Marketing Activities of Car Enterprises Based on Cloud Computing. Proceedings of the 2022 6th Annual International Conference on Data Science and Business Analytics, ICDSBA 2022, 527–531. https://doi.org/10.1109/ICDSBA57203.2022.00111

[41] Dalal, S., Lilhore, U.K., Foujdar, N. et al. Next-generation cyber attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree. J Cloud Comp 12, 137 (2023). https://doi.org/10.1186/s13677-023-00517-4

[42] Zheng, B., & Yao, C. (2023). Automobile Profession Responds to the Development of Automobile Intelligence. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 419–423. https://doi.org/10.1109/ICCCBDA56900.2023.10154651

[43] Priyadarshi, R. (2024). Energy-Efficient Routing in Wireless Sensor Networks: A Meta-heuristic and Artificial Intelligence-based Approach: A Comprehensive Review. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-023-10039-6

[44] Priyadarshi, R. (2024). Exploring machine learning solutions for overcoming challenges in IoT-based wireless sensor network routing: a comprehensive review. Wireless Networks. https://doi.org/10.1007/s11276-024-03697-2

[45] Qiu, Y., Ma, L., & Priyadarshi, R. (2024). Deep Learning Challenges and Prospects in Wireless Sensor Network Deployment. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-024-10079-6

[46] Priyadarshi, R., & Vikram, R. (2023). A Triangle-Based Localization Scheme in Wireless Multimedia Sensor Network. Wireless Personal Communications, 133(1), 525–546. https://doi.org/10.1007/s11277-023-10777-7

[47] Priyadarshi, R., & Gupta, B. (2023). 2-D coverage optimization in obstacle-based FOI in WSN using modified PSO.

Journal of Supercomputing, 79(5), 4847–4869. https://doi.org/10.1007/s11227-022-04832-6

[48] Rawat, P., Chauhan, S., & Priyadarshi, R. (2021). A Novel Heterogeneous Clustering Protocol for Lifetime Maximization of Wireless Sensor Network. Wireless Personal Communications, 117(2), 825–841. https://doi.org/10.1007/s11277-020-07898-8

[49] Priyadarshi, R., & Gupta, B. (2021). Area Coverage Optimization in Three-Dimensional Wireless Sensor Network. Wireless Personal Communications, 117(2), 843–865. https://doi.org/10.1007/s11277-020-07899-7

[50] Priyadarshi, R., & Gupta, B. (2020). Coverage area enhancement in wireless sensor network. Microsystem Technologies, 26(5), 1417–1426. https://doi.org/10.1007/s00542-019-04674-y

[51] Priyadarshi, R., Rawat, P., Nath, V., Acharya, B., & Shylashree, N. (2020). Three level heterogeneous clustering protocol for wireless sensor network. Microsystem Technologies, 26(12), 3855–3864. https://doi.org/10.1007/s00542-020-04874-x

[52] Rawat, P., Chauhan, S., & Priyadarshi, R. (2020). Energy-Efficient Clusterhead Selection Scheme in Heterogeneous Wireless Sensor Network. Journal of Circuits, Systems and Computers, 29(13), 2050204. https://doi.org/10.1142/S0218126620502047

[53] Priyadarshi, R., Gupta, B., & Anurag, A. (2020). Wireless Sensor Networks Deployment: A Result Oriented Analysis. Wireless Personal Communications, 113(2), 843–866. https://doi.org/10.1007/s11277-020-07255-9

[54] Priyadarshi, R., Gupta, B., & Anurag, A. (2020). Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. Journal of Supercomputing, 76(9), 7333–7373. https://doi.org/10.1007/s11227-020-03166-5

[55] Priyadarshi, R., & Nath, V. (2019). A novel diamond–hexagon search algorithm for motion estimation. Microsystem Technologies, 25(12), 4587–4591. https://doi.org/10.1007/s00542-019-04376-5

[56] Priyadarshi, R., Rawat, P., & Nath, V. (2019). Energy dependent cluster formation in heterogeneous wireless sensor network. Microsystem Technologies, 25(6), 2313–2321. https://doi.org/10.1007/s00542-018-4116-7

[57] Priyadarshi, R., Soni, S. K., Bhadu, R., & Nath, V. (2018). Performance analysis of diamond search algorithm over full search algorithm. Microsystem Technologies, 24(6), 2529–2537. https://doi.org/10.1007/s00542-017-3625-0

[58] Priyadarshi, R., Soni, S. K., & Nath, V. (2018). Energy efficient cluster head formation in wireless sensor network. Microsystem Technologies, 24(12), 4775–4784. https://doi.org/10.1007/s00542-018-3873-7

[59] Desai, S., Kanphade, R., Priyadarshi, R., Rayudu, K. V. B. V., & Nath, V. (2023). A Novel Technique for Detecting Crop Diseases with Efficient Feature Extraction. IETE Journal of Research, 1–9. https://doi.org/10.1080/03772063.2023.2220667

[60] Priyadarshi, R., & Kumar, R. R. (2021). An Energy-Efficient LEACH Routing Protocol for Wireless Sensor Networks. In V. Nath and J. K. Mandal (Eds.), Lecture Notes in Electrical Engineering (Vol. 673, pp. 423–430). Springer Singapore. https://doi.org/10.1007/978-981-15-5546-6_35

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Aug 30, 2024
Accepted: May 27, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2165–2183, DOI 10.12694/scpe.v26i5.4909

A NOVEL IOT FRAMEWORK FOR IDENTIFYING AND MITIGATING SECURITY THREATS

SHRUTI JAISWAL,* HIMANI BANSAL,† SHIV NARESH SHIVHARE,‡ AND GULSHAN SHRIVASTAVA§

Abstract. The popularity of Internet of Things (IoT) devices has surged due to their applications in diverse areas such as e-Health, smart vehicles, and smart cities. However, the rapid deployment of these devices has led to an exponential increase in security attacks targeting IoT systems, making security a prime concern for the community. Securing IoT-based systems is challenging because the devices involved are often resource-constrained. Providing security to these systems requires a thorough understanding of their specific security needs, along with a systematic security engineering approach. Previous research lacks a systematic methodology for identifying and implementing security requirements. Therefore, there is a growing demand for a structured approach to identify security requirements, select appropriate algorithms, and ensure their effective implementation. While existing studies have extensively explored IoT security threats, they fall short of offering a structured method to comprehensively address these threats. This paper proposes a comprehensive security engineering framework that systematically identifies security threats by analyzing assets present over various layers of IoT system, considering their diverse roles. It includes creating repositories to identify potential vulnerabilities and applicable threats. Once threats are identified, they are evaluated for their severity level based on risk analysis. Following this, the framework focuses on designing the security solutions, where we proposed to add two new security services namely trust and data freshness besides the existing security services, algorithms are selected to mitigate threats by considering the domain and constraints of the devices involved. Ultimately, the security of the entire system is validated to ensure robustness. Throughout this process, we have developed comprehensive repositories for asset management, vulnerability-threat mapping, and algorithm-threat matching to help identify and analyze security needs and recommend algorithms for implementation.

Key words: Internet of Things, security framework, security issues, security algorithms, constraints.

1. Introduction.

The Internet of Things (IoT) is an emerging field that enhances our daily lives by automating routines through the connection of devices and other components via the web. IoT consists of a different physical object forming a network which is surrounded with different type of Sensors/ Actuators, required software's, and a network to provide connectivity and, enabling these objects to gather/ store and exchange data. It enables the remote sensing and control of objects through existing network infrastructure. The concept of IoT extends across various domains, including education, medical, research, home-based automation, industrial, and transport, all of them have a significant impact on our daily lives. It provides a platform for various general household objects to critical jobs at industries to remotely monitor & control various tasks. It would help to improve productivity, ease of usage and information access for promising a better life. There is a steep rise in utilization and connection of devices; providing profitable opportunity for all concerned personswhether the consumer or provider. According to Forbes [1] there are five areas in which IoT is blooming namely, healthcare, Work from home in pandemic times, in retail stores, smart city, amalgamation of edge and IoT devices. According to [2] number of connected IoT devices has grown from 9% to 12.3 billion globally. According to IDC's 2021 U.S. "Smart Home survey on is consumer ready to use smart devices at home for office work, the top most concern of the users of not adopting smart devices for office work is (1) they have not used them before (2) Security and privacy concerns related to their office data and devices" [3].

The growing reliance on IoT has sparked significant concern regarding the study, analysis, and implementation of security measures. As Charles Renert, Vice President of Websense Security Labs, noted, "The Internet of

*Department of CSE&IT, Jaypee Institute of Information Technology, Noida, India (dce.shruti@gmail.com)
†Department of CSE&IT, Jaypee Institute of Information Technology, Noida, India (singal.himani@gmail.com)
‡School of Computer Science Engineering and Technology, Bennett University Greater Noida, Uttar Pradesh-201310, India (shiv827@gmail.com)
§School of Computer Science Engineering and Technology, Bennett University Greater Noida, Uttar Pradesh-201310, India (Corresponding Author, gulshanstv@gmail.com)

Things means consumer products from TVs to refrigerators are now digitally connected. While enterprises may not need to fear interconnected home devices, but every new employee's internet-connected device, application, and upgrade is a potential threat vector" [4]. In recent research, SAP Leonardo has also been used in the IoT domain [39]. This reliance on IoT systems has shifted the focus of the research community towards investigating security aspects. Securing IoT systems is both distinct and challenging compared to traditional network systems due to their layered architecture. Each layer, with its specific devices and constraints, requires a unique focus on security. IoT systems face constraints such as limited on-device memory, reduced computational power, and low energy availability. Researchers detected numerous threats such as eavesdropping and malicious software among others. These incidents on IoT systems have drawn attention from investigators, who are working to address different security aspects to prevent breaches. While a range of methods has been identified to counter these attacks, choosing the most suitable method for a given set of constraints and determining the appropriate algorithm remains a complex challenge.

Numerous researchers have examined the critical need for IoT security and the various criteria that must be addressed. For instance, Hossain et al. [24] and Park and Shin [25] identified key IoT security criteria, including data integrity, information protection, anonymity, non-repudiation, and data freshness. Other studies focus on IoT security concerns within the context of large-scale applications and the broader technological implications of IoT. Schaumont [26] and Jaiswal and Gupta [27], for example, delve into security issues related to IoT-enabled healthcare systems, outlining the challenges and necessary security measures. They emphasize the importance of self-healing, trust, fault tolerance, and lightweight key management protocols, in addition to the standard security requirements such as access control, authentication, and authorization.

Authorization, authentication, confidentiality, access control, trust, and identity management are crucial security requirements for IoT systems, as they are in many other existing systems [28, 29, 30, 31]. Additionally, other general security criteria highlighted in research [32, 33, 34] encompass network security, application security, layer-specific security, data integrity, firewall protection, antivirus capabilities, encryption functions, and secure routing. With the introduction of 5G, IoT security has taken on new challenges. Researchers in [35, 40] have identified several security measures necessary for IoT-based 5G networks. These include authentication, privacy preservation, and ensuring secure communication between devices in the 5G-IoT ecosystem. Given the resource constraints of many IoT devices, they highlight the importance of lightweight cryptographic algorithms. Securing 5G IoT networks is critical because the vast number of interconnected devices increases the potential attack vectors, particularly in high-speed, low-latency networks where real-time data is exchanged. IoT security also extends to data protection techniques, such as encryption. Gupta et al. [36] have proposed a two-level image encryption method specifically designed for IoT devices, addressing the need for lightweight security protocols due to these devices' processing and energy constraints. Their encryption technique offers robust security for IoT systems by employing dual-layer encryption, which is efficient in terms of computational resources while protecting data privacy. As IoT networks grow, they are increasingly integrated with other technologies like fog computing. [37] has reviewed the security challenges in integrating blockchain technology with IoT and fog computing. They identified issues such as scalability, latency, and data integrity in decentralized environments. The study also suggests cell tree solutions to improve the security and efficiency of blockchain-based IoT systems, providing a path forward for secure and scalable IoT implementations.

Numerous researchers are focused on the security aspects of IoT, with some proposing methods to identify threats across different layers of the IoT architecture [5, 6, 7]. In [5], researchers identified potential threats at various layers but did not explore solutions in depth. Another study [7] identified threats at all layers and addressed some inter-layer threats, suggesting broad countermeasures such as encryption and access control. Current proposals mainly emphasize threat identification and suggest mitigation measures like ensuring confidentiality, encrypting data, and providing access control and privacy. However, they often overlook IoT-specific issues such as data freshness and trust, which are critical for the success of IoT systems. Additionally, the IoT domain presents challenges like the need for lightweight algorithms and mobility management. Existing approaches suggest general protection measures but fail to consider domain-specific constraints such as environment, memory, power, and computational speed when selecting security mechanisms like cryptographic algorithms to achieve security objectives like authentication, privacy, and confidentiality. Choosing an algorithm for implementation without accounting domain- specific restrictions can result in an over constrained

system. Thus, an optimal techniques/algorithms should be identified considering domain- specific constraints during the design phase. Moreover, none of the approaches available validate security level, akin to conventional development process, where the software system should be validated for embedded security. This highlights the need for a process that efficiently identifies security threats and guides developers in addressing them effectively.

Providing security to software- based systems is an intricate task, which demands a security engineering method for seamless amalgamation of security concept in software development life cycle. The approach involves starts with identifying, analyzing, prioritizing, and specifying security threats. Then, considering various constraints, the most suitable algorithm for threat mitigation should be selected, followed by validating the system's security level. As a result, a new field known as Security Engineering has emerged, focused on developing processes and methods for applying security in software systems [8]. Ideally, this approach should involve multiple stages of security: (i) requirements, (ii) design, and (iii) testing. This approach should become an integral part of traditional development life cycle model. This paper presents a structured security engineering framework for handling security threats. The process first identifies the generic assets of the IoT system that needs to be protected from intruders. The role played by the assets and possible constraints to assets are identified and stored in repository. Next, the vulnerabilities are identified for assets, and potential threats to these vulnerable points are specified. For mapping of vulnerability corresponding to Assets and Threats corresponding to Vulnerabilities repositories are developed. Next, the risk value for each threat is computed based on occurrence frequency of threat & its impact on assets. Calculation of risk value is required, as we cannot secure the system by 100%. So, risk value would help in knowing which threats need urgent handling and threats which can be ignored. For calculation of impact value, a repository is maintained which contains the assets impacted by threats. Based on the threat's criticality and device constraints, efficient security algorithms is identified for implementation. Security algorithm is chosen based on the threats mitigated and the domain-specific constraints of the selected environment. Different repositories are created and maintained to fulfill the need of developer in an effective way. Once the security algorithm is selected, system security level is tested by calculating security index, which is then compared to pre-defined reference value. If the value of security index is low, we go to selection of algorithm activity again and choose a new algorithm or modify the existing algorithm.

Hence the aim of paper to develop a structured framework which deals with:

1. Identification and categorization of security threats.
2. Selection of security algorithm for mitigation of threats.
3. System security level is evaluated in terms of Security Index.

Rest of the paper is structured as follows: next section provides the overview of security issues in IoT; then the proposal for identification and design of security issues is discussed. After that, a case study is presented to enlighten our proposed framework. Finally, discussion and conclusions with direction of future work is provided.

2. Security Challenges in IoT. The Internet of Things (IoT) integrates a variety of technologies, including networks, cloud systems, transaction management, load balancing, memory management and, many more. Security issues associated with these technologies can impact an IoT system.

Key security issues in IoT systems, include: [3, 6, 7, 8] are:

***Identification / Authentication / Authorization*:** Authentication in IoT is challenging due to the need for heterogeneous network authentication. The identification and authentication of Things must occur before allowing their entry to the network. Each entity in the network requires a unique identification code. Once identified and authenticated, the user must be authorized according to a set of predefined rules.

***Confidentiality & Privacy*:** It's crucial to protect the personal and sensitive data from unauthorized access. Also, private communications must be safeguarded from eavesdroppers.

***Resilience*:** In an IoT system, if an interconnected node is compromised, the system should be able to protect the network, data, and devices from any attacks.

Fault Tolerance: System must continue to work properly with the necessary security services in place, even in the event of a failure.

***Self-Healing*:** In terms of security, self-healing refers to the ability of the network to maintain a minimum level of security even if a sensor or device fails.

***Heterogeneity / Standardization / Interoperability*:** IoT systems consist of a multitude of standalone devices with varying architectures and protocols. The lack of standardization and interoperability among these devices poses significant security challenges, necessitating a robust security design.

***Data Freshness*:** For an IoT network to operate efficiently, nodes must have access to the most recent messages/ information, as these are critical for real-time system performance. For instance, in a patient Surveillance system, a doctor requires the latest ECG readings to assess a patient's heart function.

***Liability*:** There should be accountability in cases of misuse, loss, theft, or unusual events.

***Big Data*:** The communication between devices within an IoT network, as well as with external entities, generates large amounts of data that must be securely managed.

***Constraints*:** Many IoT devices are constrained by limited memory, power, and other resources, making it challenging to implement security measures.

***Trust*:** Trust is essential for users to confidently engage with the system. If users believe the system is secure, they are more likely to use it. However, establishing trust in IoT is difficult due to various security issues.

***Anonymity*:** In some cases, users may wish to remain anonymous.

3. Proposed Security Engineering Process. The proposed security engineering process for identifying and addressing security threats is illustrated in Figure 3.1. Our approach operates in two phases: the first phase involves identifying the various security threats within the system, while the second phase suggests different techniques to mitigate the identified threats based on different design constraints.

Phase I. Identification and Categorization of Threats. In this phase, generic assets, vulnerable points, and threats are identified. Next, the identified threats are assessed to infer their severity, which will further help in selection of security algorithm for implementing threats. Different activities of identification and categorization phase are:

1. *Identify the Assets.* The objective of our research is to secure the system's assets, which can be anything of value to the system, whether tangible or intangible. These assets are central to the system's functionalities and are often targeted by attackers. To address this, a repository of generic assets, categorized by layer, has been designed and maintained to assist developers in selecting the appropriate assets for their systems. This repository was created by analyzing various IoT domains, such as smart homes, e-healthcare, vehicle tracking, and transportation [9, 10, 11, 12, 13]. Additionally, assets are classified by type (where applicable), and any constraints or limitations associated with the assets are also identified. These constraints are crucial when selecting or providing security solutions to mitigate threats within the system. New assets identified during the research are added to the repository for future consideration. Table 3.1 is created to store assets at various IoT layers, along with their constraints and roles.
2. *Identification of Vulnerabilities.* Vulnerabilities are system weaknesses that attackers can exploit to gain access to system resources. Therefore, it's crucial to identify these points to protect system assets from potential attacks. To address the need for vulnerability identification, we developed a repository based on an extensive literature review [4, 13, 14]. A sample repository is depicted in Table 3.2; further details on other vulnerabilities related to assets can be obtained from the author. For clarity and proper reference, "V" prefix is added to vulnerabilities. Vulnerabilities are selected from the maintained repository on the basis of role of the asset of all stakeholders. To facilitate, a scenario diagram is created to illustrate when and how an asset is accessed and used, as depicted in Figure 3.2. Any new vulnerabilities reported are documented for future reference.
3. *Identify the Threats.* To secure system assets, it is essential to understand the potential threats to those assets. Thus, identifying potential threats at various vulnerable points is necessary. To facilitate this, we have proposed a mapping table with dimensions of 39×22, a portion of which is shown in Table 3.3. It illustrates the possible threats at vulnerable points, where an "✓" indicates that a particular threat may occur at a specific vulnerable point. These threats are then identified and extracted from the table using the scenario diagram created in the previous step.
4. *Threat Evaluation.* Evaluation of threats plays an important role as it allows us to measure the severity of probable threats. Risk value for each threat is calculated using the threat occurrence probability

Table 3.1: Assets at various Layers of IoT

S. No.	Layer	Assets	Further Sub-Categorization	Constraints/ Limitations	Role
1	Sensing	Sensors/ Actuators/ Controllers	- Environment Sensors (Light, Temperature, etc.) - Body Sensors (ECG, Blood Pressure, etc.) - Motion detectors - Microphone sensors - Gas/ Smoke detectors - Electrical Current/ ON-OFF Sensors - Door (magnet) Sensors - Physiological sensors	- Limited power - Limited Battery Capacity - Limited memory - Reduced Computational Speed - Limited Bandwidth for Communication	Acquire data
2	Sensing	Labels and Markers	- RFID tags - NFC (Near Field Communication) - Security tokens - Smart cards - SIM cards	- Limited Power - Limited Battery Capacity - Limited memory - Reduced Computational Speed - Limited Bandwidth for Communication	For device identification
3	Sensing	Data repository	- On- premise server - Cloud storage - Removable media	- Large volume of data - No Fixed Structure - Variable structure (Structured, Unstructured) - Sensitive data - Data source - Diverse formats - Availability	For storing large volumes of generated or created data
4	Sensing	Devices	- Home appliances (e.g., Refrigerator, Washing machine) - Hospital equipment (Various machines) - Screen and speakers	- Availability - Environment Limitations - Battery/ Power/ Charging considerations	Appliances used in IoT network
5	Communication	Communication Network	- Internet connection (wired or wireless) - Networking components (e.g., Routers, Bridge)	- Bandwidth limitation based on devices	For efficient communication
6	User Interface	User interface device	- Specialized terminal - Gateway interface - Remote control devices - Smartphones, Smart TVs - Tablets, Desktop computers /PCs - SOS/ Emergency buttons - Set-top box user interfaces - Calendar/ Reminder devices	- Battery/ Power/ charging considerations - Availability - Environment Limitations	Mode of user interaction
7	Between the interface of two Layers	Software Programs	- Operating system(s) - Device drivers - Applications - Firmware	- Auto-update - Security Patch Update - Compatibility - Battery/ Power/ Charging	For Data Processing
8	At each layer	Data/ Information	- Access/ payment credentials to external accounts - Smart setup/ structure/ inventory information - Status information - User preferences - Intellectual property/Value - Security (Passwords, User identifier) - Privacy (User biometrics, Behavioral patterns and trends) - Resources (Music, Audio/ Visual media, Pictures, etc.	- Distributed - Bulky/ Huge - Confidential - Generated from different sources - Different Formats	Crucial and important data for processing
9	Miscellaneous	Physical Resources	- Building infrastructure - Hardware (Air conditioners, Meters, Light, etc.)	- Physical constraints	Provides Infrastructure
10	People/ Users	People/ User	- End users - Providers - Customers	- From a different technical background - May/may not have security knowledge	Access and manage the system

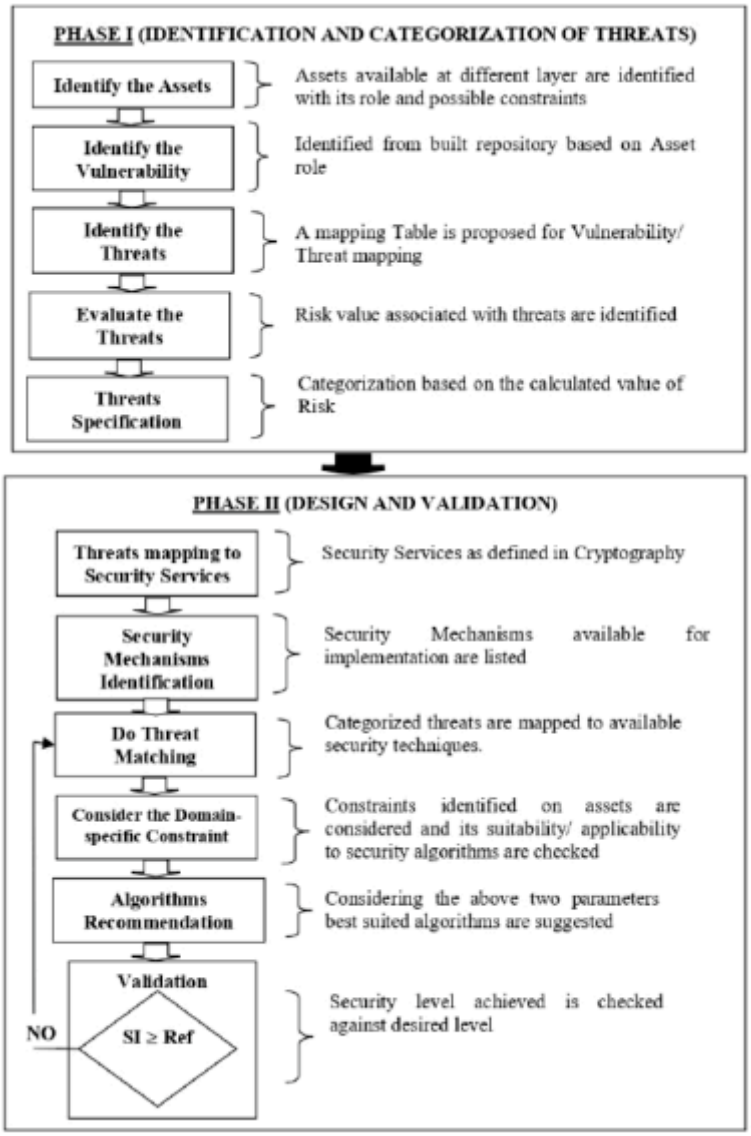


Fig. 3.1: Proposed Framework

and its impact on the system. Prioritization/ evaluation of threats is necessary as most of the projects has constraints, so all threats cannot be mitigated, and the developers need to decide which threats to choose first for implementation. The process followed for prioritization, is shown in Figure 3.3 and explained subsequently:

(a) *Impact Identification:* Impact of occurrence of a threat on the system is identified by examining the number of assets impacted when it occurs. Therefore, it would simply be the summation of impacted assets values represented by Eq. 3.1 as follows:

$$\text{Impact} = \sum \text{Asset rating of impacted assets} \tag{3.1}$$

As impact depends on asset rating, asset rating is required to be calculated. Calculation of asset value is an important task as it shows its importance. In many risk analysis methods,

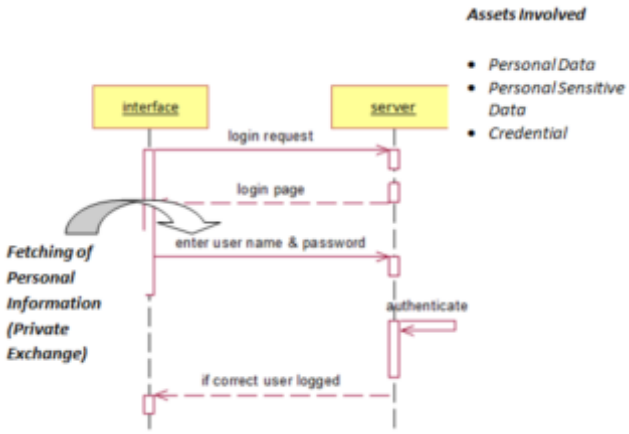


Fig. 3.2: Scenario Diagram for Login functionality

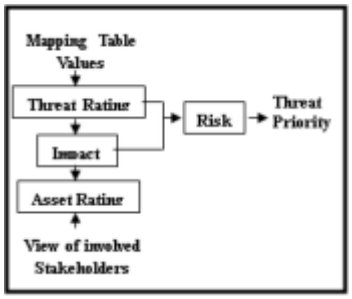


Fig. 3.3: Process for Prioritization of Threats

such as CRAMM [15], CORAS [16], assets are given a rating based on their importance to the system. We propose that assets valuation would be more accurate if it considers the perspectives of all relevant stakeholders. For instance, asset Sensors/ Actuators is evaluated by stakeholders Consumer, Provider, Administrator and Vendor as 8, 9, 7 and 6 respectively. So, by analyzing the views of stakeholders, final asset value using Eq. 3.2 is '8'.

$$\text{Asset Value} = \sum \frac{\text{View of involved actor for Asset}}{\text{Number of actors involved}} \tag{3.2}$$

Impact is calculated by adding the asset rating of impacted assets by threats. Table 3.4 shows the list of assets impacted by occurrence of threat. For instance, suppose threat T. Manipulation of Hardware and Software would impact assets (Sensors/ Actuators, Software, User interface device, Labels and Markers, and Data repository).

(b) *Calculate the Risk:* Risk measures the potential damage that a threat can inflict on the system. As mentioned by OWASP [20],], risk is represented by Eq. 3.3. Using Eq. 3.3 risk value of all identified threats are calculated.

$$\text{Risk} = \text{Threat Rating} \times \text{Impact} \tag{3.3}$$

(c) *Threat Categorization:* Categorization of threats is done to represent threats clearly and precisely, which is done using the identified risk values. Threats are categorized on the basis of following criteria:

Table 3.2: Identified Vulnerabilities for Assets

Assets	Vulnerabilities
Sensors/ Actuators	V. InadequateAccessControl V.UnencryptedData V. LackofPhysicalSecurity V.Misconfiguration V.InsecureInterfaces V.InsufficientSecurityConfigurability V.RemoteAccess V.SystemMisuse V.LackofMonitoring V.InsufficientLogging V.LackofStandards
Software Programs	V.InsufficientLogging V.Misconfigurations V.UnsafeAPIFirmware V.OutdatedSystem V.LackofStandards V.Intrusion Detection

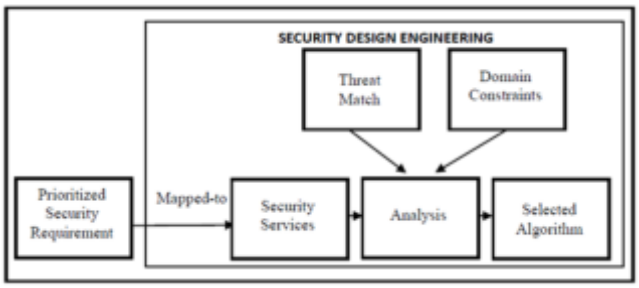


Fig. 3.4: Security design process.

i. If ($Risk \geq 60$), Then Category is Catastrophic, which needs urgent handling because the threats impact various high-value assets.
ii. If ($60 > Risk \geq 20$), Then Category is Important, which require careful consideration because the threats are impacting:
   - Various moderate assets
   - A high-value
iii. If ($20 > Risk \geq 5$), Then Category is Acceptable which can be considered or ignored
iv. If ($Risk < 5$) Then No Effect because low-value assets are impacted, hence can be ignored. Categorized threats are stored for further action.

Categorized threats are stored for further action.

Phase II. Design and Validation. Once threats have been prioritized and categorized, the next phase involves selecting appropriate security algorithms. This selection process focuses on identifying algorithms that effectively address the prioritized threats. The choice of algorithm for implementation is based on: (i) threats it mitigates; (ii) the domain- specific constraints/ limitations of the environment where the algorithm will be implemented. After selecting the necessary algorithms for implementation, testing is conducted to ensure if all threats are adequately addressed. For clear illustration a simplified diagram to depict the design process is shown in Figure 3.4.

Table 3.3: Threat-Vulnerability Mapping Table

Vulnerabilities / Threats	V. Inadequate_Access_Control	V. Insufficient_Logging	V. Penerated_Firewall	V. Unsanitized_Input	V. Unsafe_API_Frimware	V. Outdated_System	V. Misconfiguration	V. Non-Encrypted_Data	V. Naive_User	V. Lack_of_Monitoring	V. Vulnerable_Network	V. Intrusion_Detection	V. Physical_Protection	Threat Rating
T. Identity_Theft	✓	✓							✓					3
T. Infected_e-mail			✓		✓						✓	✓		4
T. Denial_of_Service (DoS)										✓		✓		2
T. Information_Leakage								✓			✓			2
T. . Rouge Certificates Generation and Use	✓	✓												2
T. Manipulation of Software and Hardware														0
T. Information manipulation	✓	✓						✓	✓	✓				5
T. Misuse Audit Tools	✓								✓					2
T. Records Falsification	✓			✓	✓					✓				4
T. Unauthorized use of Administrative Resources	✓	✓												2

1. *Threats Mapping to Security Services:* Threats are associated with major security services CIA triad (Confidentiality, Integrity, Availability), access control, and non-repudiation [17] This mapping will be beneficial in later stages by identifying the appropriate security mechanisms for implementation. Considering, each threat separately is challenging, so mechanisms are categorized according to security services for effective handling of security threats in the system. Besides the already defined security services, two more services are added, namely 'Data Freshness' and 'Trust', which are required to be considered for the IoT domain as several proposals defended the need for data freshness (real-time data) and trust as an integral part of security subsystem [27, 38]. Besides the IoT domain, other emerging areas such as fog, edge, and blockchain are all working on a recent data set; hence, there is a need to include data freshness and make people adopt these emerging technologies; trust should also become an integral part of security services. Threats mapping to security service 'Confidentiality' is shown in Table 3.5.
2. *Security Mechanisms Identification:* Various security algorithms that can be used to implement the system's security services are explored. Table 3.5 provides details for just one security service. Once the available security mechanisms are identified, the design and security teams will analyze the algorithms and select the most suitable mechanism for implementation, following the subsequent activities.
3. *Do Threat Matching:* A repository for analyzing security algorithms is created. The categorized threats are then compared against this repository, and the algorithm that mitigates the most threats is selected. Table 3.6 provides a sample repository for confidentiality. A ✓in the table indicates that the security technique can address the matching threat. For instance, the AES technique under the asymmetric category mitigates threats such as Unauthorized Software use, Unauthorized Software Installation, Communication Breach, MITM attacks, and Violation of Privacy. The last row of the table shows the total impact, indicating the number of threats each technique mitigates. Each value corresponds to the different techniques listed at the top.
4. *Consider the Domain-specific Constraints/ Limitations:* Algorithm choice is based on amount of threats they mitigate, but not all algorithms are suitable for every scenario. Therefore, further analysis of

Table 3.4: Impacted Assets by Threats

Threats	Impacted Assets
T. Rouge Certificates Generation and use	Sensors/ Actuators Software Programs User interface device Data/ Information Labels and Markers Communication Network Data repository Devices
T. Manipulation of Hardware and Software	Sensors/ Actuators Software Programs User interface device Labels and Markers Data repository
T. Failure and Malfunctions	Sensors/ Actuators Software Programs User interface device Physical Resources Labels and Markers Communication Network Data repository Devices
T. Information Leakage	Sensors/ Actuators Physical Resources Labels and Markers Communication Network Data repository

domain constraints/ limitations imposed by devices, the environment, and other factors is conducted. Categorization of domain constraints is shown in Figure 3.5. The selected algorithms are then evaluated based on domain- specific constraints viz. power, memory usage, and more. Table 3.7 illustrates the limitations of the IoT system. In Table 3.7 values varies from minimal to extensive depending on the domain of application.

5. *Algorithm Recommendation:* As previous process of threat matching & consideration of domain- specific constraints, most suitable algorithms are selected for implementation.
6. *Validation:* The validation of chosen algorithm/ mechanism is performed to determine whether the potential threats to the system are effectively mitigated. To cater this, a Security Index (SI) value is considered, which indicates the remaining gap in the system. The SI is defined as the ratio of mitigated threats to the initial number of identified threats, as shown in Eq. 3.4.

$$\text{Security Index} = \frac{\text{Mitigated Threats}}{\text{Identified Threats}} \times 100 \tag{3.4}$$

A high SI value (close to 100) indicates that the system is secure, while a low SI value (approaching 0) suggests that the system is unsafe and requires revisions to the design decisions, particularly the selection of security algorithms. The reference value is defined by the administrator based on the domain of application, level of CIA required, criticality of the system. Its value may change from one system to other based on its domain constraints. If modifications to security are necessary, the developer must return to the beginning of the second phase to choose or adjust the algorithm for implementation.

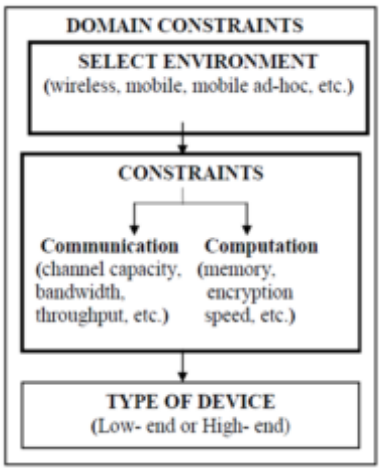


Fig. 3.5: Categorization of Domain Constraints.

Table 3.5: Threats Mapping to Security Services and Mechanism

Security Services	Threats	Security Mechanism Available	Possible Techniques	Techniques Characteristic
Data Confidentiality	T. Unauthorized Software use T. Unauthorized Software Installation T. Compromise of Confidential Information T. Communication Breach T. Eavesdropping T. MITM T. Violation of Privacy T. Rogue Employee T. Identity theft	Encryption, Routing Protocol	**Encryption** **Asymmetric** (AES, DES, Triple DES) **Symmetric** (RSA, Rabin's Scheme, ECC, HECC)	**Asymmetric** – Each node possesses its own unique set of keys – Takes more power due to computational complexity – Good scalability **Symmetric** – Ensures key confidentiality – Complex protocol for key management – Authentication demands higher power consumption – Simple calculations, resulting in lower power consumption
			Routing Protocol (AOMDV-IoT, SMRP, EARA, RPL, Multiparent routing in RPL, PAIR, REL)	Prevent routing attacks like spoofing, sinkhole, and selective forwarding.

4. Case Study: Patient Surveillance System. Healthcare domain is chosen because of the facts given in the report by grand view research [18]:

- The COVID-19 pandemic significantly boosted the IoT in healthcare market by accelerating the adoption of remote patient monitoring and telemedicine. This shift was driven by the need for innovative technologies to manage health data remotely, reducing in-person contact and improving healthcare outcomes.
- The global Internet of Things (IoT) in healthcare market was valued at USD 44.21 billion in 2023 and is projected to grow at a compound annual growth rate (CAGR) of 21.2% from 2024 to 2030

Table 3.6: Mapping of Security Mechanisms to Threats. A1- (AES) Advanced Encryption Standard, A2- (DES) Data Encryption Standard, A3- Triple- DES; S1- RSA (Rivest–Shamir–Adleman), S2- ECC (Elliptic Curve Cryptography), S3- HECC (Hyperelliptic curve cryptography); H1- ECIES (ECIES Hybrid Encryption Scheme)

	Asymmetric			Symmetric			Hybrid
Techniques/Threats	A1	A2	A3	S1	S2	S3	H1
T. Unauthorized Software use	✓	✓	✓	✓	✓	✓	✓
T. Unauthorized Software Installation	✓	✓	✓	✓	✓	✓	✓
T. Compromise of Confidential Information							✓
T. Communication Breach	✓				✓	✓	✓
T. Eavesdropping							✓
T. MITM	✓				✓	✓	✓
T. Violation of Privacy	✓	✓	✓	✓	✓	✓	✓
T. Rogue Employee							✓
T. Identity theft				✓	✓		✓
T. Confidential Data Compromise	✓				✓	✓	✓
T. Credential theft	✓				✓	✓	✓
T. Information Leakage	✓				✓	✓	✓
TOTAL IMPACT	8	3	3	4	9	8	12

Table 3.7: Limitations across various IoT layers

Parameters	Sensing	Communication	User Interface
Parameters related to Performance			
Memory Computation Speed Energy Run Time performance	Minimal, Intermediate, Extensive		
Other Parameters			
Security Objectives Mobility Compatibility Scalability Cost of the chosen solution Portability	Minimal, Intermediate, Extensive		

- The market is fueled by the growing use of smartphones, smart devices, and wearables for patient monitoring. Additionally, the rising adoption of remote patient monitoring to enhance out-of-hospital care further drives market growth.

The Remote Patient Surveillance System illustrated in Figure 4.1 is an integral component of the healthcare system.

1. *Wireless Body Area Network.* This network consists of wearable sensors that can store small amounts of data and transmit it to remote server/ location.
2. *E-Health Gateway.* This component forwards data packets from the WBAN to other networks.
3. *Internet.* The communication network responsible for carrying the information/ data.
4. *Healthcare Data Centre.* This facility stores all the data generated by the sensors in WBAN. Given the large volume of data, proper management is essential.
5. *Medical Service.* Provides medical facilities to patient such as consultations based on regular check-ups and other services.

In the above system, the following actors are considered: Patient, Doctor, and Insurance Service Provider. Each user has different roles and responsibilities; here, we will consider an abstract overview of all roles for

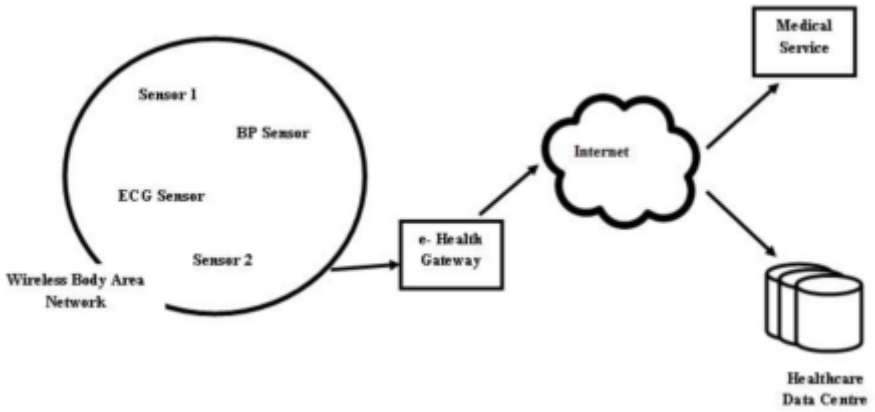


Fig. 4.1: Remote Patient Monitoring System

further explanation.

Phase I. Identification and Categorization.

1. *Identify the Assets.* Patient surveillance has following assets: body sensors, data repository, network and connection components, a smartphone with an application for interaction, and patient information. Each asset plays a specific role, as detailed below:
   - Body Sensors: Attached to the patient's body to monitor physiological parameters and transmit the data to either remote storage or a processing device.
   - Data repository: Store collected data
   - Network and connections: For communication between nodes
   - Smart Phone: For user interaction
   - Patient Information: Personal and health information of the patient available in the system.
2. *Identification of vulnerabilities.* Identified Vulnerabilities are extracted for all assets involved in Remote Patient Surveillance System from repository depicted in Table 3.2 using the scenario diagram. Vulnerabilities extracted for asset:
   (a) *Body Sensors*
      - Inadequate Access Control
      - Unencrypted Data
      - Lack of Physical Security
      - Misconfiguration
      - Insecure Interfaces
      - Insufficient Security Configurability
      - Remote Access

   (b) *Data repository*
      - Inadequate Access Control
      - Insufficient Security Configurability
      - Unencrypted Data
      - Intrusion Detection
      - Misconfigurations
      - Insecure Interfaces
      - System Misuse

   Similarly, vulnerabilities applicable to other assets are identified.
3. *Identify the Threats.* Threats are extracted corresponding to identified vulnerable points for each asset. Threats to asset Body Sensors are:

Table 4.1: Risk Calculation intended for Probable Threats

Threats	Affected Assets	Threat Rating	Impact	Risk
T. Identity theft	Sensors/ Actuators Interface device Data/ Information Communication-Network Data repository	5	40	200
T. DoS	Sensors/ Actuators Data repository	2	17	34
T. Rouge Certificates Generation and use	Sensors/ Actuators Interface device Data/ Information Communication-Network Data repository	5	40	200

- Information manipulation
- Repudiation
- Misuse of Personal Data
- Records Falsification
- Information Leakage
- Physical Attacks
- Compromise of Confidential Information
- Failure and Malfunctions
- Accidental Damages

4. *Evaluate the Threats:* Prioritization of threats is done by considering the impact of threats on the involved assets. Calculation of risk values for our system is presented in Table 4.1.

Assets Evaluation. System assets are evaluated by involved stakeholders, as explained in section 3. For example, involved assets are assessed as:

- Body Sensors as 8
- Interface Device as 8
- Patient Information as 8
- Network and Connections as 7
- Data repository as 9

Threat Categorization. Threats are categorized in the following categories:

1. *Catastrophic:*
   - Identity theft
   - Credential theft
   - Generation and use of Rogue Certificates
   - Records Falsification
   - Unauthorized use of Administrative Resources
   - Unauthorized Software Installation
   - Confidential Data Compromise
   - Replay Message
   - Eavesdropping
   - Violation of Law or Regulations
   - Failure and Malfunctions
   - Accidental Damages
   - Violation of Privacy
   - Fake Node
   - Phishing

- Spoofing
- Information Leakage
- Information manipulation
- Repudiation
- Illegal Access to Information System
- Unauthorized Software use
- Rogue Employee
- Communication Breach
- Misuse of Audit Tools
- MITM
- Infected Email
- Malware
- Human Error

2. *Important:*
   - DoS
   - Hardware Failure
   - Misuse of Personal Data
   - Loss of Support Services
3. *Acceptable:*
   - Physical Attacks
   - Natural Calamity
   - Environmental Calamity
   - Node Capture
4. *No Effect:* Nil

Phase II. Handling.

1. Threats Mapping to Security Services. Table 3.5 shows mapping of threats to security services.
2. Security Mechanisms Identification. Security mechanisms available for implementation are listed in Table 3.5.
3. Algorithms Recommendation. Based on the threat matches (as shown in Table 3.6) and system limitations (as shown in Table 4.2), suitable algorithms are recommended for implementation. Table 3.6 is built for algorithms related to threats to confidentiality. Here, for explanation purposes, only the confidentiality part is emphasized in detail during the validation part. Other algorithms are chosen based on the domain of the application and its constraints. As the Patient Surveillance System, we have memory, power, and speed constraint hybrid algorithm ECIES, which is more suitable as it requires less power than AES and ECC algorithms and can mitigate a maximum number of threats.
   - Encryption
     a) Asymmetric Encryption Algorithm: AES
     b) Symmetric Encryption Algorithm: ECC
     c) Hybrid Algorithm: ECIES
   - Routing Control: Energy-aware Ant Routing Algorithm (EARA)
   - Authentication Exchanges: Two Step Authentication
   - Data Integrity: MD5
   - Digital Signature: ECDSA
   - Access Control Mechanism: Role-Based Access Control
   - Notarization: Establish a Notary Server
   - Physical Protection Mechanisms: Locks, Physical Security Guards
4. Validation to ensuring Confidentiality Identified threats related security service confidentiality:
   - Identity theft
   - Records Falsification
   - Unauthorized Software Installation
   - Confidential Data Compromise

Table 4.2: Constraint to different Layers for Remote Patient Surveillance System

Parameter	Sensing	Communication	User Interface
Parameters related to Performance			
Memory	Minimal	Intermediate	Extensive
Computation Speed	Minimal	Intermediate	Extensive
Energy / Power	Minimal	Intermediate-Extensive	Extensive
Run Time performance	Intermediate	Extensive	Extensive
Other Parameters			
Security Objectives	Extensive	Extensive	Extensive
Mobility Compatibility	Extensive	Extensive	Extensive
Scalability	Minimal	Extensive	Extensive
Cost of the chosen solution	Minimal	Minimal	Minimal
Portability	Extensive	Extensive	Extensive

- Credential theft
- Eavesdropping
- Violation of Privacy
- Information manipulation
- Information Leakage
- Unauthorized Software use
- Communication Breach
- MITM

Referring to Table 3.6, threats matching is done, and SI value is calculated for threats pertaining to confidentiality.

Asymmetric Algorithm. AES is selected. Security Index = $\frac{7}{12} \times 100 = 58.33\%$

Symmetric Algorithm. ECC is selected. Security Index = $\frac{8}{12} \times 100 = 66.67\%$. However, neither of the algorithms alone provides adequate protection, so hybrid techniques are necessary. Therefore, the hybrid technique ECIES has been selected. The Security Index (SI) is calculated as SI = $\frac{10}{12} \times 100 = 83.33\%$. While the hybrid algorithm significantly improves upon the existing algorithms, additional algorithms are needed for optimal effectiveness.

5. Discussion. Current proposals only specify the threats and do not have steps for prioritization and categorization. However, none of the existing approaches recommend a specific security algorithm for implementation. They only specify the broad implementation measures that are architectural constraints, and checking the system's security level is not discussed.

The process from threat elicitation to prioritization for IoT systems is outlined. In this method, assets and potential threats to them were identified alongside system functionalities. After prioritizing the identified threats, they are classified into categories such as catastrophic and significant. Security algorithms are chosen based on domain-specific limitations, including being lightweight (i.e., consuming less power and requiring minimal computation time) and having low storage requirements. Ultimately, a metric is developed to reflect the system's security level. This approach is illustrated within the context of IoT-enabled medical care, particularly for remote patient surveillance.

To achieve these objectives, the following activities are carried out:

- Identification of assets across various IoT layers, along with their potential threats and vulnerabilities.
- Creation of a threat repository affecting system assets, with dimensions of 39×30.
- Suggestion of security mechanisms based on domain constraints.

The high Security Index result indicates that existing security algorithms are insufficient, necessitating the development of new algorithms. Our approach differs from previous techniques by providing a structured process for incorporating security into software systems. A comparison of this with existing approaches is shown

Table 5.1: Comparison of our proposal with existing literature.

Method	Domain	Key Contributions	Challenges Highlighted	Proposed Solutions/Frameworks
Hossain et al. [24]	IoT Security Challenges and Approaches	Meta-study on IoT security challenges, approaches, and open issues	Scalability, privacy, resource constraints	Discusses various security approaches, open research areas
Park and Shin [25]	IoT Security Assessment Framework	Proposes a framework for assessing security in IoT services	Heterogeneous device security, privacy	Framework for systematic security assessment in IoT services
Schaumont [26]	Scale Challenges in IoT Security	Discusses the unique challenge of scaling security across numerous IoT devices	Scalability, energy efficiency	Suggests scalable security mechanisms appropriate for large IoT networks
Jaiswal and Gupta [27]	IoT Security Requirements	Identifies essential security requirements for IoT systems	Data integrity, user privacy, access control	Emphasizes the need for comprehensive IoT security models
Tourani et al. [28]	Security, Privacy, and Access Control in IoT	Surveys security, privacy, and access control in IoT-based networks	Privacy, access control, secure data management	Provides a taxonomy of IoT security approaches and open challenges
Zhou et al. [29]	IoT New Features and Security Impact	Examines how IoT-specific features affect security and privacy	New vulnerabilities due to IoT features	Discusses existing solutions and gaps addressing IoT-specific security
Ammar et al. [30]	Security in IoT Frameworks	Surveys security in existing IoT frameworks	Data security, interoperability, scalability	Reviews security methods within various IoT frameworks
Asiri [31]	Blockchain in IoT	Proposes a blockchain-based trust model for IoT	Trust, decentralization	Blockchain-based model for enhancing trust in IoT environments
Jerald et al. [32]	Secure IoT Architecture	Proposes a secure architecture for smart services in IoT	Integration of multiple smart services securely	A secure architecture framework for smart IoT environments
Sedrati and Mezrioui [33]	IoT Security Challenges	Overview of IoT security with a focus on challenges	Data breaches, device vulnerabilities	Highlights the need for adaptive security measures in IoT
Oracevic et al. [34]	IoT Security Survey	A general survey of IoT security	Device security, secure communication, data privacy	Consolidates existing approaches to IoT security
Dey et al. [35]	IoT Security in 5G Networks	Focus on security measures in IoT within 5G networks	Network slicing, data integrity	Recommends security approaches tailored to IoT in 5G
Gupta et al. [36]	Image Encryption for IoT	Proposes a two-level image encryption for IoT	Data protection, secure communication	Two-level encryption for secure image transmission in IoT
Khan and Chishti [37]	Fog and IoT Security	Discusses challenges and solutions for IoT in fog computing and blockchain	Scalability, data privacy	Reviews fog-based and blockchain solutions for IoT
Pal et al. [38]	Systematic Approach to IoT Security Requirements	Systematic review of IoT security requirements	Authentication, integrity, privacy	Framework for structured analysis of IoT security needs
Proposed Approach	Systematic Approach for Identification and Mitigation of threats by recommending appropriate security algorithms for implementation	Creation of different databases for assets available at each layer, vulnerability and threats mapping table, and design constraints related to IoT system	Addition of new security requirements, recommendation of hybrid algorithm for implementation	Structured framework of security engineering for IoT-based system

in Table 5.1. Our approach can be adopted for emerging IoT areas by exploring the vulnerabilities and threats specific to that area, if any. If a new vulnerability or threat is found, we need to just extend our database by adding its related information. By doing this, we can adapt our framework for any related domain.

6. Conclusion and Future Work. A structured framework for incorporating security in IoT-based systems is projected. This framework identifies and categorizes threats to security, and selects efficient security algorithms for implementation based on the threats mitigated and domain constraints.

The novel contributions of our work are:

- Elicitation, analysis, prioritization, and categorization of threats to assets.
- New security services were added namely data freshness and trust.
- During requirements engineering phase, probable threats to assets are identified. To support threat elicitation, we have created:
  - i A repository of assets for IoT architecture.
  - ii A vulnerability threat mapping table with dimensions of 39×22.
  - iii A table of threats affecting system assets, with approx. dimension of 39×30 is prepared. It helps in managing different assets, functionalities, threats, and vulnerabilities.
- Selection of algorithms to implement security in system is based on domain- specific constraints across different layers of IoT.
- Generation of a security metric to indicate the system's security.
- In the context of Patient Surveillance System, a hybrid algorithm is proposed to meet security requirements.

In future, we plan to conduct a thorough analysis of the computational overhead versus the security benefits essential for selecting new algorithms over traditional ones. We also aim to explore combinations of threats and vulnerabilities that could potentially lead to new security issues. Therefore, it is crucial to account for these combinations, as lower-order threats may combine to create significant security challenges. Additionally, a fully automated, AI-backed system is required for analyzing and implementing security in IoT-based systems.

REFERENCES

[1] Websence Security Lab, *2015 SECURITY PREDICTIONS*, Onliner available at http://www.portantier.com/files/websense-report-2015-security-predictions-en.pdf, 2015.
[2] S. Li, T. Tryfonas and H. Li, *The internet of things: a security point of view*, *Internet Research*, vol. 26, no. 2, pp. 337-359, 2016.
[3] R. Roman, P. Najera and J. Lopez, *Securing the internet of things*, *Computer*, vol. 44, no. 9, pp. 51-58, 2011.
[4] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu and D. Qiu, *Security of the internet of things: perspectives and challenges*, *Wireless Network*, vol. 20, pp. 2481-2501, 2014.
[5] K. Chatterjee, D. Gupta and A. De, *A Framework for Development of Secure Software*, *CSI Transaction on ICT*, vol. 1, no. 2, pp. 143- 157, 2013.
[6] J. Granjal, E. Monteiro and J. S. Silva, *Security for the Internet of Things: A Survey of Existing Protocols and Open Research Issues*, *IEEE COMMUNICATION SURVEYS & TUTORIALS*, vol. 17, no. 3, pp. 1294-1312, 2015.
[7] J. A. Stankovic, *Research Directions for the Internet of Things*, *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3-9, Feburary 2014.
[8] K. Sood, S. Yu and Y. Xiang, *Software Defined Wireless Networking Opportunities and Challenges for Internet of Things: A Review*, *Internet of Things Journal*, vol. 3, no. 4, pp. 453-463, 2015.
[9] B. Martínez-Perez, I. d. l. Torre-Diez and M. Lopez-Coronado, *Privacy and Security in Mobile Health Apps: A Review and Recommendations*, *Journal of Medical Systems*, vol. 39, no. 1, pp. 1-8, 2015.
[10] D. Niewolny, *How the Internet of Things Is Revolutionizing Healthcare, Freescale Semiconductor*, *In Proceedings of International Conference on Healthcare*, pp. 211-219. 2013.
[11] G. Jayavardhana, B. Rajkumar, S. Marusic, and M. Palaniswami, *Internet of Things (IoT): A vision, architectural elements, and future directions*, *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645-1660, 2013.
[12] S. Lee, G. Tewolde, and J. Kwon, *Design and Implementation of Vehicle Tracking System Using GPS/GSM/GPRS Technology and Smartphone Application*, In *IEEE World Forum on Internet of Things (WF-IoT), Seoul, 2014*, pp. 353-358, 2014.
[13] D. Barnard-Wills, L. Marinos, and S. Portesi, *Threat Landscape and Good Practice Guide for Smart Home and Converged Media*, *ENISA (The European Network and Information Security Agency)*, pp. 175, 2014.
[14] A. Mitrokotsa, M. Beye, and P. Peris-Lopez, *Classification of RFID Threats based on Security Principles*, *Security Lab, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology*, 2011.
[15] CRAMM, *United Kingdom Central Computer and Telecommunication Agency (CCTA)*, emphRisk analysis and management method, CRAMM user guide, Issue 5.1., 2005.

[16] F. d. Braberl, I. Hogganvik, M. Lund, K. Stølen, and F. Vraalsen, *Model-based security analysis in seven steps—a guided tour to the CORAS method*, *BT Technology Journal*, vol. 25, no. 1, p. 101–117, 2007.
[17] B. A. Forouzan, *Cryptography & Network Security*, *The MsGraw Hill Companies*, pp. 721, 2008.
[18] B. Chamberlin, *IBM Center for Applied Insights*, Available at: https://ibmcai.com/2016/03/01/healthcare-internet-of-things-18-trends-to-watch-in-2016/., 2016.
[19] Wikipedia, *Available at: http://www.wikipedia.com Accessed January 2016*.
[20] Gartner, *21 Billion IoT Devices To Invade By 2020*, Available: http://www.informationweek.com/mobile/mobile-devices/gartner-21-billion-iot-devices-to-invade-by-2020/d/d-id/1323081. 2020.
[21] IDC, *Explosive Internet of Things Spending to Reach $1.7 Trillion in 2020*, Available at: http://www.idc.com/getdoc.jsp?containerId=prUS25658015, 2020.
[22] L. O'Donnell, *IOT Predictions for 2016*, Available at: http://www.crn.com/slide-shows/networking/300079629/10-iot-predictions-for-2016.htm?itc=refresh, 2016.
[23] OWASP, *OWASP Risk Rating Methodology*, Available at: https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology, 2014.
[24] M. Hossain, R. Hasan, A. Skjellum, *Securing the Internet of Things: A Meta-Study of Challenges, Approaches, and Open Problems In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), Atlanta, GA*, pp. 220–225, 2017.
[25] K. Park, D. H. Shin, *Security assessment framework for IoT service Telecommun. Syst.*, vol. 64, pp. 193–209, 2017.
[26] P. Schaumont, *Security in the Internet of Things: A challenge of scale*, In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), Lausanne, Switzerland*, pp. 674–679, 2017.
[27] S. Jaiswal, D. Gupta, *Security Requirements for Internet of Things (IoT)*, In *Proceedings of International Conference on Communication and Networks, Advances in Intelligent Systems and Computing*, Springer: Singapore, Vol. 508, pp. 419–427, 2017.
[28] R. Tourani, S. Misra, T. Mick, G. Panwar, *Security, Privacy, and Access Control in Information-Centric Networking: A Survey*, *IEEE Commun. Surv. Tutor.* vol. 20, pp. 566–600, 2017.
[29] W. Zhou, Y. Zhang, P. Liu, *The Effect of IoT New Features on Security and Privacy: New Threats, Existing Solutions, and Challenges Yet to Be Solved*, *IEEE Internet Things J.* vol. 6, 2018.
[30] M. Ammar,G. Russello, B. Crispo, *Internet of Things: A survey on the security of IoT frameworks*, *J. Inf. Secur. Appl.*, vol. 38, pp. 8–27, 2018.
[31] S. Asiri, *A Blockchain-Based IoT Trsust Model, Master's Thesis, Ryerson University, Toronto, ON, Canada*, 2018.
[32] A. V. Jerald, S. A. Rabara, D. P. Bai, *Secure IoT architecture for integrated smart services environment*, In *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), Palladam, India*, pp. 800–805, 2016.
[33] A. Sedrati, A. Mezrioui, *Internet of Things challenges: A focus on security aspects*, In *Proceedings of the 8th International Conference on Information and Communication Systems (ICICS), Jeju, Korea*, pp. 210–215, 2017.
[34] A. Oracevic, S. Dilek, S. Ozdemir, *Security in Internet of Things: A survey*, In *Proceedings of the 2017 International Symposium on Networks, Computers and Communications (ISNCC), Marrakech, Morocco*, pp. 1–6, 2017.
[35] A. Dey, S. Nandi, M. Sarkar, *Security Measures in IoT based 5G Networks*, In*3rd International Conference on Inventive Computation Technologies (ICICT 2018), Coimbatore, India*, pp. 561-566, 2018 doi: 10.1109/ICICT43934.2018.9034365.
[36] M. Gupta, V. P. Singh, K. K. Gupta, P. K. Shukla, *An efficient image encryption technique based on two-level security for internet of things*, *Multimedia Tools and Applications*, vol. 82, pp. 5091–5111, 2022, doi: 10.1007/s11042-022-12169-8.
[37] N. S. Khan, M. A. Chishti, *Security Challenges in Fog and IoT, Blockchain Technology and Cell Tree Solutions: A Review*, *Scalable Computing: Practice and Experience*, vol. 21, no. 3, pp. 515–542, 2020, doi: 10.12694/scpe.v21i3.1782.
[38] S. Pal, M. Hitchens, T. Rabehaja, S. Mukhopadhyay, *Security Requirements for the Internet of Things: A Systematic Approach*, *Sensors* vol. 20, no. 1 pp. 5897, 2020, doi: https://doi.org/10.3390/s20205897
[39] H. Bansal, S. Jaiswal, *Lay a hand on IOT with SAP Leonardo in Data Acquisition and Processing*, vol. 38 no. 1, pp. 5374-5392, 2023.
[40] H. Sharma, P. Kumar, K. Sharma, *Recurrent Neural Network based Incremental model for Intrusion Detection System in IoT*, *Scalable Computing: Practice and Experience*, vol. 25, no. 5, pp. 3778-3795, 2024.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Sep 13, 2024
Accepted: Nov 27, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2183–2195, DOI 10.12694/scpe.v26i5.4910

SOFTWARE DEFECT PREDICTION MODEL BASED ON AST AND DEEP LEARNING

ZEZHI YE, CHENGHAI YU*, AND ZHILONG LU

Abstract. Software reliability prediction (SDP) theory is crucial for balancing software value and assessing efficiency. Traditional defect prediction relies on static code metrics for machine learning, but these handcrafted features fail to capture the code's syntactic structure and semantic information. In order to further predict the defects of the software, the Abstract Syntax Tree (AST) of the program was parsed on the basis of the metric data, and extracted as feature vectors, and the data was encoded by dictionary mapping and word embedding as the input of Convolutional Neural Network (CNN). On this basis, the Long Short-Term Memory network (LSTM) and Multi-head attention mechanism were used to further optimize the network, and the particle swarm optimization (PSO) was used to select the hyperparameters of the model, and the defects were predicted by using the model. The results show that the model can learn syntax and semantic features well, and the estimation accuracy is higher and the bias is smaller.

Key words: software defect prediction; Abstract Syntax Tree; Convolutional Neural Network; Long Short-Term Memory; Multi- Head Attention

1. Introduction. As modern software development progresses, software reliability has become a critical concern. This research aims to estimate the failure rate of future systems, hoping to predict software reliability at the early stages of development, thereby enabling developers to improve development and maintenance. However, the reliability of a system will change throughout its lifecycle depending on the conditions it is subjected to, making the selection of appropriate prediction methods a focal point in this research area [24, 2].

Traditional prediction methods typically involve two steps: extracting software metrics and building classification models. The main metrics include:

- *Object-Oriented Software Complexity Metrics* (e.g., C&K metrics)
- *Data Flow-Based Metrics* (e.g., Elshoff method)
- *Control Flow-Based Metrics* (e.g., McCabe method)
- *Program Volume-Based Metrics* (e.g., Halstead's software science method)

The classification models constructed include Naive Bayes [7], cluster analysis, and support vector machines [5]. Current research on static, small-dimensional data largely focuses on machine learning methods, with less emphasis on deep learning. In recent years, some scholars have attempted to introduce deep learning networks into this research area, using artificial neural networks (ANN), deep neural networks (DNN), and CNN [12] to make new attempts on software defect prediction datasets.

However, there are instances where static metrics fall short in accurately determining whether the code contains defects, as code with and without defects may have identical metric attributes. Due to the different syntactic and semantic information they contain, traditional machine learning classifiers find it difficult to differentiate. Abstract Syntax Trees (ASTs), based on the specific syntax structure of a program, embed rich semantic information that aids in accurately analyzing code. Their tree structure can describe the context of the code, allowing deep learning methods to mine the information contained in ASTs to achieve more accurate defect predictions at the source code level.

Therefore, by combining AST with deep learning, semantic feature information is extracted from source code using AST. The LSTM and multi-head attention mechanisms further differentiate key features within the classifier, enabling software defect prediction based on the semantic network.

*School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China (Corresponding author, ych@zstu.edu.cn)

2. Related Work.

2.1. Traditional Model. SDP remains a vital research area within the broader field of software engineering, with considerable attention devoted to it in the literature. Most of the literature focuses on designing new discriminative features, filtering defect data, and constructing effective classifiers. For instance, Nagappan and Ball [20] introduced code churn metrics and integrated them with software dependencies to predict defects. Moser [18] thoroughly analyzed the impact of change metrics and static code attributes on predicting defects. Additionally, Arar and Ayan [33] applied the Naive Bayes approach to identify and eliminate redundant features, improving the efficiency of defect prediction models. Mousavi [19] addressed the issue of class imbalance in SDP by utilizing ensemble learning techniques, which help balance the distribution of data across classes. Furthermore, Jing [9] proposed a dictionary learning method that focuses on computing misclassification costs to better predict software defects. In another study, Yu [30] employed correlation feature selection to identify and retain only the features that are strongly correlated with the target class, thereby enhancing the model's predictive capabilities. Ma [16] introduced Transfer Naive Bayes, it uses the data gravity method [21] to adjust the training set to build the classifier. Recent research has also highlighted the benefits of utilizing a small portion of labeled data from the target project to significantly improve prediction performance. For example, Qiu [22] created an innovative multi-component weight learning model by employing the kernel mean matching algorithm. This approach segments the source project's data into various components, applies KMM in each to modify the weights of the source instances. It then uses the weighted source instances from each component and a portion of the labeled data from the target project to construct a predictive model, finally, initialize the weights to build the final classification model. The final step involves initializing and optimizing the weights of source components to build a more accurate ensemble classifier.

We employ deep learning techniques to autonomously derive features, capturing the code information. These features replace static code attributes to improve defect prediction performance.

2.2. Applications of Deep Learning in Software Defect Prediction. Commonly used software data sets rely on manual statistics, which may greatly affect model performance and accuracy. Furthermore, these manual metrics often fail to fully exploit the contextual information of the code, which is crucial for understanding the syntax and semantics of programs. Program syntax and semantics can be represented through AST and Control Flow Graphs (CFG) [3]. Wang utilized Deep Belief Networks (DBN) to create latent features encapsulating program syntax and semantics, which were then used in classifiers to identify erroneous code [27]. Lin et al. [14] employed LSTM networks [8] to learn representations of program ASTs that could be transferred across projects to detect vulnerable functions. Dam [4] constructed a deep tree model based on AST for software defect prediction. Additionally, Li [11] built a hybrid model using features learned from convolutional neural networks [10] combined with static features. CFG, on the other hand, represents the control flow graph of a program, showing all possible paths that can be traversed during program execution. Zhou et al.[32] used CodeBERT to explore real-time prediction and proposed the JIT-SDP model, which extracts code submission and change information and uses it for prediction, thus enhancing the generalization ability of the code prediction model. Parvez [17] used transform to propose Bugsplorer, which hierarchically marks code elements and predicts file-level defects.

The deep learning-based approaches mentioned above generally treat all latent features as equally important, which can result in a failure to pinpoint those features that are most discriminative in terms of crucial syntax and semantics. This oversight can contribute to less accurate defect predictions. To address this issue, our method introduces an attention mechanism that focuses on identifying and prioritizing these key features by assigning them greater weights. Furthermore, we opt to use the AST as the program representation instead of the CFG, as the AST more effectively captures the structural nuances and retains more comprehensive information about the source code.

3. Model Design.

The final model utilizes AST to automatically extract features. These features capture the syntax and semantic information of the program as input. CNN is chosen as the basic classifier, and LSTM and Multi-Head Attention layers are used to learn and analyze the relationships between the data, extracting key features for training a more accurate model. To address the data differences among different datasets, hyperparameter tuning is introduced, with the Particle Swarm Optimization algorithm being used for efficient

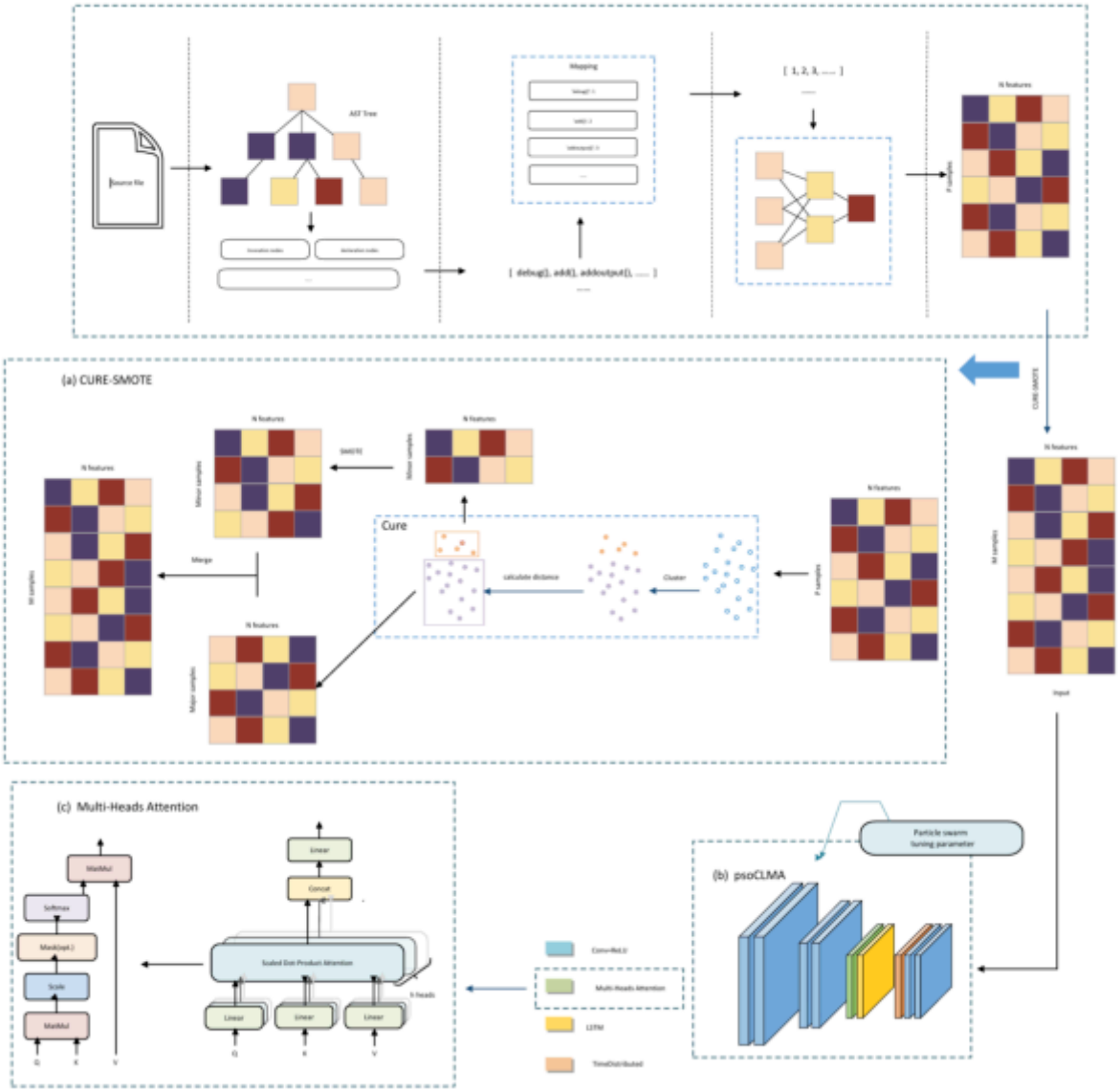


Fig. 3.1: Network framework

parameter adjustment. Final model is show in Figure 3.1.

The main contributions of this paper are as follows: We propose a software defect prediction model (ACNN) that combines AST semantic network and deep learning. Using AST to extract code semantic data avoids the tedious process of manually extracting data and possible deviations. On this basis, according to the characteristics of the dataset, LSTM and Multi-Head Attention are used to make the classifier better focus on the extracted data. Finally, the effectiveness of each improvement is verified through ablation experiments on the same dataset, and the experimental results of different datasets are compared to verify the overall effectiveness of the model proposed in this paper.

3.1. AST and data processing. When constructing AST, three types of nodes are selected to express the code semantic information: (1) method call nodes; (2) declaration nodes; (3) control flow nodes. AST can

```
1  package com;
2  public class A{
3      public static void main(String[]args){
4          int i=0;
5          while(i<10){
6              i++;
7          }
8           System.out.println(i);
9      }
10 }
```

(a)

```
1  package com;
2  public class B{
3      public static void main(String[]args){
4          int i=0;
5          while(i<10){
6
7          }
8           System.out.println(i);
9      }
10 }
```

(b)

Fig. 3.2: Source code of two example files: (a) The clean code A; (b) The defective code B

effectively understand the structure and semantic information of the source code. In Figure 3.2, code snippet A is very similar to code snippet B, which means that manually extracted feature vectors may be identical. However, the syntax trees in Figure 3.3 show that the AST of code A (a) has two more nodes than the AST of code B (b) (i.e., the Statement Expression and Member Reference nodes marked with red boxes).

Vectors of strings cannot be used directly as experiment input. Therefore, a dictionary mapping tokens to integers is constructed based on token frequency. Tokens are sorted by frequency, and an ordered token dictionary is established in descending order of frequency. Properly normalizing the vector lengths can prevent sparse matrices from being used as input. For vectors shorter than the specified length, zero padding is performed to ensure that feature vectors have the same length. Truncates vectors that are too long. Finally, the word embeddings in the network are used as a trainable dictionary while ensuring that the input data will not be a sparse matrix.

Software defect prediction data often exhibits significant class imbalance, where defect instances typically constitute only a small fraction of the total dataset. When such imbalanced data is used directly for model training, the resulting predictions tend to be skewed toward the majority class, leading to biased and inaccurate outcomes. To address this challenge, two primary approaches are commonly employed: oversampling and undersampling. To preserve the integrity and completeness of the data, we opt for the oversampling approach and utilize the CURE-SMOTE algorithm [15, 31] to achieve data balance. This sophisticated algorithm effectively identifies and removes noise and outliers from the original samples. Following this, a modified algorithm is employed to generate new synthetic samples by interpolating between representative points and central points within the clusters. By preprocessing the data in this manner, the model becomes less biased toward the majority class, resulting in more accurate and reliable predictions.

Moreover, models trained on such balanced datasets are better equipped to generalize to unseen data, enhancing their robustness and performance in real-world applications. This approach ensures that the model remains effective across diverse scenarios, making it a valuable tool for software defect prediction. The formula for updating the model parameters is presented below:

$$\mathrm{dist}(\mathrm{P},\mathrm{Q}) = \min_{p\in \mathrm{P}_r, q\in \mathrm{Q}_r}(p,q)$$

$$\mathrm{P}_c \frac{|\mathrm{P}|\cdot \mathrm{P}_c + |\mathrm{Q}|\cdot \mathrm{Q}_c}{|\mathrm{P}|+|\mathrm{Q}|}$$

$$\mathrm{P}_r\,\{p+\alpha\cdot(\mathrm{P}_c-p)|p\in \mathrm{P}_r\}$$

$$\mathrm{X}_{new}^{n} = \overline{X} + rand(0,1)\times(\mathrm{P}_r - \overline{X})$$

Fig. 3.3: AST of two example code: (a) The clean code A; (b) The defective code B

$$\mathrm{X}_{new} = \overline{X} \cup \mathrm{X}^{\mathrm{n}}_{new}$$

The ratio of processed samples is shown in Figure 3.4.

3.2. Network layers. The core design of the proposed model is based on CNN [6, 1], Multi-Head Attention mechanisms [28, 26, 13], and LSTMs [29]. First, the CNN is used for local feature extraction. Then, the LSTM is used to analyze the associations between data to ensure data integrity, while the Multi-Head Attention is used to enhance the model's data processing by assigning different attention weights to different types of data. The input data, which consists of software-related metrics, is passed through a one-dimensional convolutional layer (Conv1D). This layer can effectively extract local features from the sequence data and adjust the output shape. Next, the output from the convolutional layer goes through a one-dimensional max-pooling layer (MaxPooling1D), which performs pooling operations on the convolutional layer's output, reducing the feature size by half. This helps reduce the number of parameters in the model and extracts the most significant features from the data.

Then, the Multi-Head Attention mechanism is introduced. This layer performs operations of multiple attention heads, helping the model capture important information in the sequence without altering the output

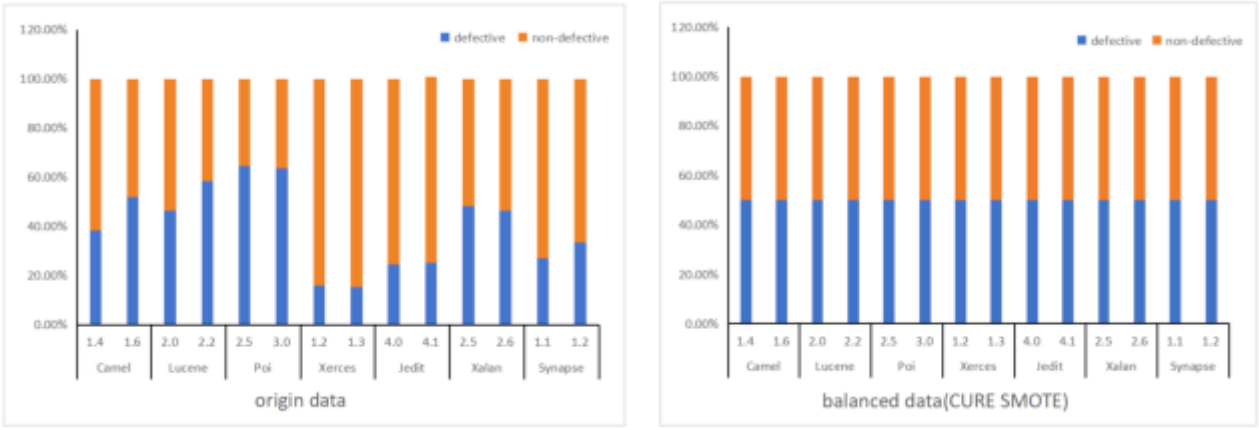


Fig. 3.4: Sample Distribution Before and After

shape. After the attention layer, a masking layer is used to mask the padded parts of sequences of unequal lengths, so that these padded data can be ignored in subsequent processing. Following this is an LSTM layer used to model sequence data and capture long-term dependencies in the sequences. After the LSTM layer, a TimeDistributed layer is used to apply a fully connected layer to each time step in the input sequence. This operation allows the model to generate a single prediction value for each time step, achieving final binary classification prediction. Finally, the outputs of multiple flattened layers are concatenated through a Concatenate layer, followed by two Dense layers for the final classification or prediction.

All parameters within the entire model are designed to be fully trainable, allowing for dynamic adjustment and optimization during the learning process. This network structure has been carefully designed to maximize the strengths of convolutional layers, attention mechanisms, and recurrent structures, each playing a crucial role in processing sequence data. The goal is to enhance the model's ability to perform accurate prediction or classification tasks. The synergy between these components ensures that the model can effectively capture and utilize patterns within the data. The detailed composition and underlying principles of each component are elaborated upon in the following sections.

3.2.1. CNN. CNNs are similar to artificial neural networks and are a type of deep learning model consisting of input layers, convolutional layers, activation layers, pooling layers, fully connected layers, and output layers. The convolutional layers are responsible for feature extraction, the activation layers use activation functions for multilayer mapping learning, the pooling layers extract main features to prevent overfitting, and the fully connected layers integrate extracted features for classification by the classifier.

CNN are extensively used in processing image data, where they have proven to be highly effective. However, there has been relatively limited exploration of building one-dimensional (1D) CNNs to address traditional machine learning problems. Despite the difference in dimensionality, CNNs, whether 1D or 2D, retain the same core features and approaches to problem-solving. The primary challenge lies in managing the dimensionality of the input data and understanding how filters behave as they move across the input.When analyzing experimental data, it becomes clear that there is often a significant correlation among the data points. Recognizing and isolating important information within the sequence is crucial for making more accurate predictions. The ability to focus on key elements in the data sequence enhances the predictive power of the model, demonstrating the importance of correctly handling the dimensionality and filter behavior within CNNs, regardless of their application to 1D or 2D data.

3.2.2. Multi-Head Attention. Traditional neural networks may face multiple information interference when processing input data. To address this issue, Multi-Head Attention is introduced to reprocess the output of the deep convolutional model for better data information extraction. The attention mechanism comprises

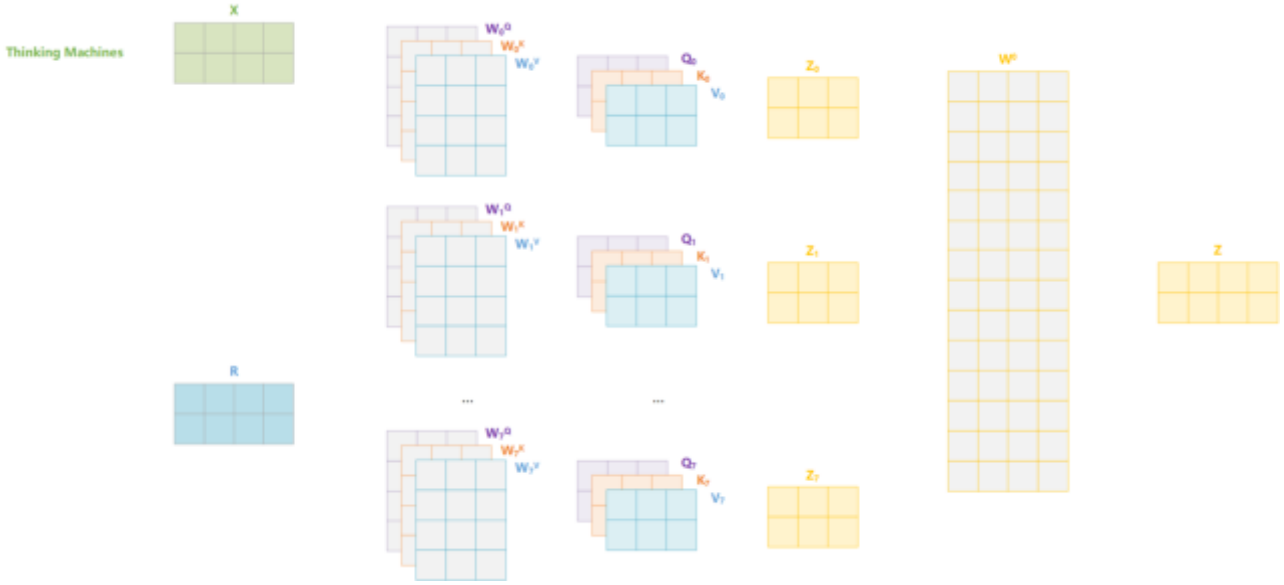


Fig. 3.5: Multi-Head Attention structural diagram

queries, keys, and values, essentially a mapping of relationships. Multi-Head Attention is an optimization of the attention mechanism, further refining the Self-Attention layer. The introduction of multiple attention heads maps inputs to different spaces and computes scores, enhancing the model's ability to focus on features. The structural diagram is shown in Figure 3.5:

3.2.3. LSTM. LSTMs have been improved and popularized by many people for various problems. LSTMs introduce three gate structures for control: forget gate, input gate, and output gate. Using gated units to update the unit state, LSTMs are essentially a special type of RNN capable of better handling data relationships compared to standard RNNs. Its unique memory cell structure allows the model to retain and fully utilize feature information over longer periods, leveraging this capability to learn complex patterns and relationships in software metric data and help build predictive models.

3.3. Hyperparameter Optimization. There are subtle differences between datasets involved in software defect issues, and the selected metrics may not be entirely consistent, which could lead to reduced model generalizability. Therefore, when selecting model parameters, a swarm intelligence algorithm is used to dynamically adjust parameters. The PSO algorithm [23, 25] generates the same number of samples as the problem to be solved, treating each sample as a possible solution. The problem is bio-mimicked as a food source for birds, simulating this foraging for subsequent solving. It can efficiently and conveniently search for optimal parameters within the solution space.

During the search for the optimal solution, particles learn from two values: the global historical best value and the individual historical best value. Particles change their position and flying speed by learning from their own historical best value and the population's historical best value. These learning behaviors are called "self-learning" and "social learning," respectively. The global and individual historical best values are determined by fitness values, derived from the solution in the actual problem. Particles adjust their positions and speeds by referencing individual historical and global historical best values. When the updated position's fitness value surpasses its historical best or global best, the two best values are updated to the current value. During particle movement, it is also necessary to keep away from nearby individuals to avoid collisions with other particles.

The update formulas for particle position and speed are as follows:

$$v_{ij}(\tau+1) = \omega v_{ij} + c_1 r_1 \left|gbest(\tau) - x_{ij}(\tau)\right| + c_2 r_2 \left|pbest(\tau) - x_{ij}(\tau)\right|$$

Table 4.1: Confusion matrix

		True value	
		Positive	Negative
Predicted value	Positive	TP	FP
	Negative	FN	TN

Table 4.2: Dataset details

Name	LIne of files	Defective Rate	Function
camel1.4	872	16.72%	Train set
camel1.6	965	19.57%	Test set
lucene2.0	195	46.67%	Train set
lucene2.2	247	58.47%	Test set
poi2,5	385	64.42%	Train set
poi3.0	442	63.57%	Test set
jedit4.0	306	24.51%	Train set
jedit4.1	312	25.32%	Test set
synapse1.1	222	27.03%	Train set
synapse1.2	256	33.60%	Test set
xalan2.5	803	50.18%	Train set
xalan2.6	885	46.91%	Test set
xerces1.2	440	50.13%	Train set
xerces1.3	453	15.05%	Test set

Table 4.3: Ablation test results

	Accuracy	Recall	F1	Mcc
C+L	0.8070	0.8419	0.8192	0.6292
C+L+M	0.8314	0.8229	0.8295	0.4356
C+L+M+P	0.9123	0.8462	0.8776	0.7658

$$x_{ij}(\tau+1) = x_{ij}(\tau) + v_{ij}(\tau+1)$$

4. Experiments and Analysis. The existence of defects is a binary classification problem. The model's final prediction results in four outcomes as shown in the confusion matrix in Table 4.1. The evaluation metrics used are the standard evaluation metrics for binary classification problems: Accuracy, Recall, F1-score, and Matthews Correlation Coefficient (MCC).

This experiment builds a software defect prediction model based on the Keras framework. Keras is an open source machine learning library based on Python. It is well compatible with various data processing libraries (such as pandas, numpy, etc.) and has built-in multiple classic models, making it convenient for users to create new machine learning models based on classic models. The experimental machine includes GPU: RTX 4090, CPU: Intel(R) Xeon(R) Gold 5418Y * 12 cores.

The dataset used in this paper includes seven open-source Java projects, each project containing two versions: (camel1.4, camel1.6), (lucene2.0, lucene2.2), (poi2.5, poi3.0), (jedit4.0, jedit4.1), (synapse1.1, synapse1.2), (xalan2.5, xalan2.6), and (xerces1.2, xerces1.3). One as the training set and another is used as the test set. The detailed original data information is shown in Table 4.2.

4.1. Ablation experiments. We conducted ablation experiments on the model. As an example, using jedit4.0 and jedit4.1, the outcomes of the ablation experiments are presented in Table 2, where LSTM is denoted as L, Multi-Head Attention as M CNN as C, and PSO as P. Table 4.3 Ablation test results The experimental

Table 4.4: Accuracy of Different Models

	ACNN	SVM+RF	ARNN	CNN
Camel	0.7807	0.6991	0.4935	0.6403
Lucene	0.6009	0.5309	0.5324	0.5744
Poi	0.7159	0.6073	0.7722	0.6613
Xerces	0.8539	0.6719	0.5195	0.8112
Jedit	0.9123	0.7456	0.8843	0.7527
Xalan	0.7429	0.6018	0.5062	0.5858
Synapse	0.7176	0.6784	0.5221	0.6275
average	0.7606	0.6479	0.6043	0.6647

Table 4.5: Recall of Different Models

	ACNN	SVM+RF	ARNN	CNN
Camel	0.7797	0.4521	0.4534	0.6399
Lucene	0.5968	0.5845	1.0000	0.6812
Poi	0.7326	0.7071	0.8718	0.6614
Xerces	0.8604	0.2537	1.0000	0.8111
Jedit	0.8462	0.6618	0.8263	0.7581
Xalan	0.6690	0.6268	0.0000	0.5911
Synapse	0.7137	0.3882	0.9265	0.5962
average	0.7426	0.5249	0.7254	0.6770

data reveals that both LSTM networks and Multi-Head Attention mechanisms have a subtle but generally positive impact on the experimental outcomes. The model combining Convolutional Neural Networks with LSTM demonstrates stable performance, with various evaluation metrics consistently hovering around the 80% mark. The MCC value for this model is close to 0.6, suggesting a moderate level of correlation between predicted and actual results. The introduction of Multi-Head Attention further boosts the model's performance across the board, underscoring the attention mechanism's capability to effectively capture essential data features and enhance classification accuracy. When both LSTM and Multi-Head Attention are integrated and PSO is applied for parameter tuning, the model fully leverages the strengths of both networks. This results in a marked improvement in overall model performance, with the MCC value rising to 0.76—a notable increase of 0.33 compared to using the networks separately. This configuration achieves the highest scores across all evaluation metrics, indicating that the model has reached its optimal performance level.

In the ablation experiments, each incremental addition of a module leads to a noticeable performance gain. Particularly, the integration of the Multi-Head Attention mechanism and the fine-tuning of hyperparameters play a positive effect on significantly enhancing the model's effectiveness. The final model, incorporating all these elements, delivers the best performance across all measured metrics, demonstrating its superior efficiency in classification tasks and validating the importance of each component in the overall model architecture. This comprehensive improvement highlights the synergy between the modules and the effectiveness of the proposed model design.

4.2. Comparison with Other Algorithms. To further validate the model's effectiveness, we selected the following three baseline methods for comparison with our model:

a) **SVM+RF:** A method combining Random Forest (RF) and SVM based on traditional machine learning.

b) **ARNN:** An Attention-based RNN prediction model.

c) **CNN:** A traditional CNN.

When constructing the network architectures for ARNN and CNN.

Box plots were also used to visually present the performance of each model across different datasets, as shown in Figure 4.1, the box plot of experimental results.

Table 4.6: F1 of Different Models

	ACNN	SVM+RF	ARNN	CNN
Camel	0.7690	0.3769	0.6736	0.6402
Lucene	0.5968	0.6409	0.6921	0.6216
Poi	0.7272	0.7374	0.8111	0.6591
Xerces	0.8362	0.1889	0.6797	0.8111
Jedit	0.8776	0.5556	0.8836	0.7567
Xalan	0.7172	0.5963	0.0000	0.5910
Synapse	0.7152	0.4459	0.6530	0.5962
average	0.7485	0.5060	0.6276	0.6680

Table 4.7: MCC of Different Models

	ACNN	SVM+RF	ARNN	CNN
Camel	0.5343	0.1903	NAN	0.2809
Lucene	0.1936	0.2058	NAN	0.1886
Poi	0.4515	0.3230	0.5348	0.3161
Xerces	0.6614	0.0002	NAN	0.6222
Jedit	0.7658	0.3936	0.7963	0.5134
Xalan	0.4798	0.2063	NAN	0.1821
Synapse	0.4315	0.2315	NAN	0.1924
average	0.5026	0.2215	0.1902	0.3280

The experimental results in Tables 4.4-4.7 indicate that the prediction performance of the model in this paper is generally superior to SVM and deep learning network algorithms. Figure 4.1 clearly shows the differences between models on different metrics, with the model exhibiting high accuracy and stability across metrics, indicating high robustness and consistency across different datasets. In contrast, the ARNN model performs exceptionally well on most datasets but slightly underperforms on some, while the SVM model shows more variability in prediction performance and overall does not perform as well as the model in this paper. Taking the jedit project as an example, the accuracy values for ACNN, ARNN, and CNN are 0.9123, 0.8843, and 0.7527, respectively, while SVM+RF only achieves 0.7456. Clearly, ACNN, CNN, and RNN perform better than traditional methods. On average, deep learning methods have higher MCC values than traditional methods. Specifically, ACNN achieves the highest values, and Table 6 shows that the ACNN model also excels in the MCC metric, with an average value of 0.5026, significantly higher than other models. Particularly in the Jedit dataset, the MCC value of ACNN is 0.7658. The MCC values for SVM+RF and CNN models are relatively low, at 0.2215 and 0.3280, respectively. The ARNN model lacks MCC data on some datasets, possibly due to gradient explosions, but its performance is not satisfactory based on available data. This reflects the advantages of the model shown in the article and verifies the stability of the defect prediction model based on deep learning, strong discrimination effect, good sample segmentation effect, and high correlation between prediction results and actual results.

In terms of accuracy, the model in this paper achieves the best results on seven out of eight datasets and also obtains the best average value, indicating good predictive ability. In contrast, while ARNN performs well on a few datasets, showing slight improvements over the model in this paper, its overall accuracy is still inferior. In terms of recall and F1 scores, the model in this paper also achieves the best average values, indicating good sensitivity to positive samples, i.e., defect data, reducing the miss rate, and achieving a good balance between precision and recall. SVM+RF has some shortcomings in defect prediction. Results show that, deep learning methods have improved the ability to distinguish defective code.

Comparing ACNN with other deep learning methods: From the F1 perspective, compared to CNN and ARNN, the proposed ACNN achieves the best F1 metric four times. This indicates that ACNN has better performance than CNN and ARNN in terms of stability in software defect prediction. Regarding MCC, ACNN

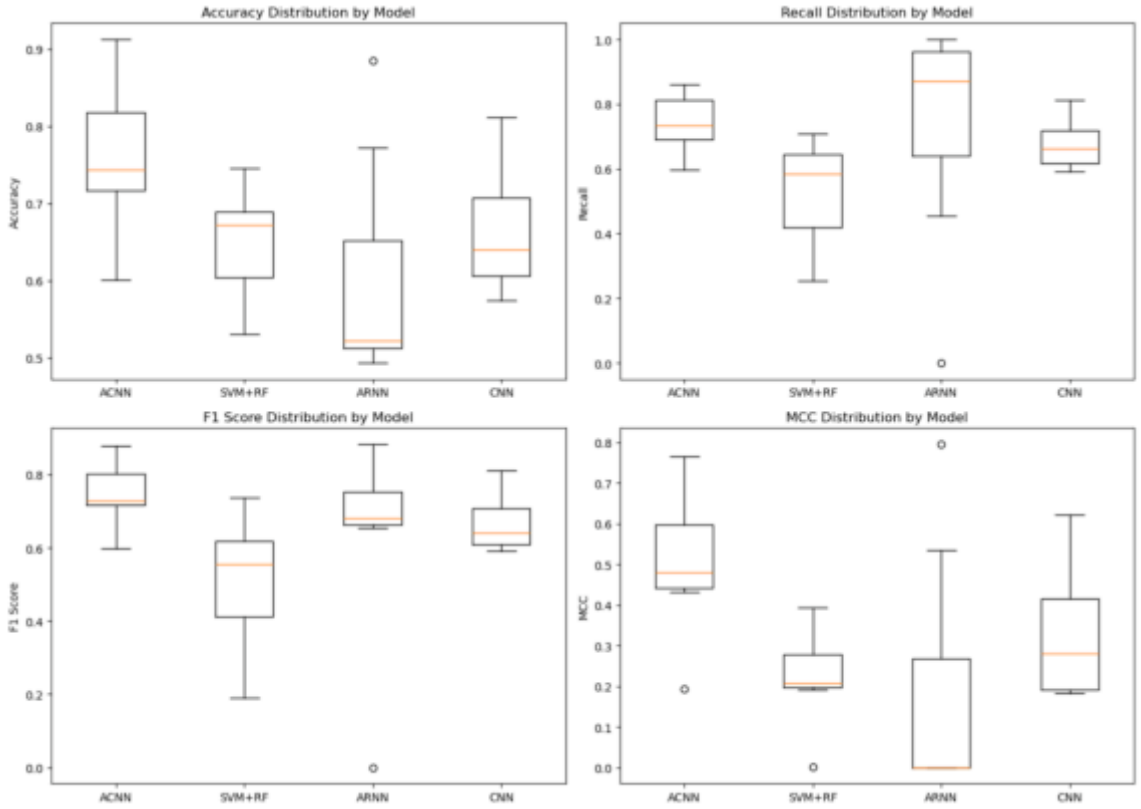


Fig. 4.1: The box plot of experimental results

improves CNN's MCC value by an average of 0.1746 and ARNN's by 0.3124. This indicates that ACNN improves the ability to distinguish software defects. According to the average F1 metric and ACC of the seven projects in the table, the proposed ACNN is 12.05% higher than CNN in the F1 metric and 53.35% higher in MCC. Particularly, in the jedit dataset, ACNN's F1 metric is 15.98% higher than CNN's, and MCC is 49.17% higher, indicating that the attention mechanism positively impacts further generating key features, thereby improving defect prediction performance.

A comprehensive analysis of the experimental results shows that the ACNN model performs particularly well when dealing with unbalanced datasets, especially when the dataset contains more defective samples (such as Synapse, Poi), and its ability to identify minority classes is better than other models. This is particularly evident in the comparison of F1 value and MCC value. ACNN has good multi-level feature extraction capabilities: the model combines CNN, LSTM and multi-head attention mechanism, which can not only capture the local features of the code, but also learn the long-term dependencies between codes through LSTM, and further extract key features using multi-head attention mechanism. This enables the model to better mine grammatical and semantic features on complex datasets, especially in projects such as Jedit and Xerces, and achieve higher prediction accuracy. The particle swarm optimization (PSO) algorithm is used to dynamically adjust the model parameters, which improves the generalization ability of the model, enabling it to maintain good performance on different datasets, especially in terms of recall rate and F1 value, which can balance the correctness and error rate of prediction and reduce bias.

The ACNN model performs well on most datasets, but there are still some limitations on some specific datasets. Especially on the Lucene dataset, ACNN scores relatively low. This may be due to the small size of the Lucene dataset, where grammatical and semantic information have little impact on the prediction results,

making it impossible for the multi-layer deep learning model to fully extract effective features.

Overall, the model in this paper performs excellently, showing outstanding results in precision, defect sample recognition ability, and overall performance. Deep neural networks perform well on some datasets and may have good predictive capabilities in certain situations, but overall are inferior to the model in this paper. In contrast, traditional support vector machine models perform well on individual datasets, particularly in terms of accuracy, but overall performance is weaker, especially in recall and comprehensiveness, significantly lower than the model in this paper.

The experimental data demonstrates that incorporating CNN, LSTM, attention mechanisms and particle swarm algorithm optimization parameters effectively improves the model's classification capability. High robustness and consistency indicate an advantage in practical applications, providing a foundation for further research and application.

5. Conclusion. Software systems today are intricately intertwined with computer technology, making the effective evaluation of overall quality before the development process begins a critical factor in enhancing development efficiency. As a result, accurately predicting software defects has become a central focus of ongoing development efforts and a key area of research interest among scholars. Traditional methods for defect prediction have largely relied on mathematical calculations, which, while useful, can be limited in their ability to fully capture the complexities of software systems. In contrast, employing deep learning techniques for prediction allows for a more nuanced judgment and analysis of data, enabling more precise and reliable outcomes.The ACNN (Attention-based Convolutional Neural Network) model introduced in this study represents a significant advancement in this area, as it can automatically extract syntactic and semantic features directly from the source code. This capability leads to more accurate predictions and a deeper understanding of the underlying code structures. Extensive experiments conducted on seven open-source projects have demonstrated that the ACNN model outperforms traditional static prediction models in terms of overall performance and exhibits greater universality and adaptability. Additionally, when compared to other deep learning-based models, ACNN consistently delivers better results. On average, the ACNN model improves accuracy by 14.43%, enhances the F1 measure by 12.05%, and increases the Matthews Correlation Coefficient (MCC) by 53.35% compared to baseline methods. These substantial improvements suggest that the ACNN model is better suited to the specific challenges of software defect prediction datasets and achieves higher accuracy in predicting potential issues.

Skipping manual data statistics and using AST to implement automatic code information extraction makes the model more realistic. The optimized model has higher prediction accuracy, providing a basis for intelligent defect prediction in actual software development and testing, to further validate the effectiveness and robustness of the ACNN model in the domain of defect prediction, we plan to extend our experiments to include a broader range of projects, including personal and smaller-scale projects. This expanded testing will provide deeper insights into the model's adaptability and performance across different contexts, ultimately contributing to its refinement and the advancement of software defect prediction techniques.

Data Sharing. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1] K. Cao, Y. Wang, and H. Tao, *A sleep staging model based on selt-attention mechanism and bi-directional lstm*, Software Guide, 23 (2024), pp. 24–32.
[2] X. Chen, L. Wang, Q. Gu, and et al, *A survey on cross-project software defect prediction methods*, Journal of Computers, 41 (2018), pp. 254–274.
[3] K. D. Cooper, T. J. Harvey, and T. Waterman, *Building a control-flow graph from scheduled assembly code*, 2002.
[4] H. K. Dam, T. Pham, S. W. Ng, and et al, *A deep tree-based model for software defect prediction*, 2018.
[5] K. O. Elish and M. O. Elish, *Predicting defect-prone software modules using support vector machines*, Journal of Systems and Software, 81 (2008), pp. 649–660. Software Process and Product Measurement.
[6] Z. Ge, N. Liu, and S. Wang, *Bearing fault diagnosis method based on mckd and improved cnn-lstm*, Journal of Tianjin University of Technology, pp. 1–11.
[7] S. Guo, R. Huang, and Q. Li, *A software defect prediction algorithm using improved weighted naive bayesian*, Control Engineering of China, 28 (2021), pp. 600–605.

[8] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, Neural Computation, 9 (1997), pp. 1735–1780.
[9] X. Jing, S. Ying, Z. Zhang, and et al, *Dictionary learning based software defect prediction*, in Proceedings of the 36th International Conference on Software Engineering, ICSE 2014, New York, NY, USA, 2014, Association for Computing Machinery, p. 414–423.
[10] Y. Kim, *Convolutional neural networks for sentence classification*, 2014.
[11] J. Li, P. He, J. Zhu, and et al, *Software defect prediction via convolutional neural network*, in 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS), 2017, pp. 318–328.
[12] J. Li, P. He, J. Zhu, and M. R. Lyu, *Software defect prediction via convolutional neural network*, 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS), (2017), pp. 318–328.
[13] Z. Li, L. Li, J. Chen, and D. Wang, *A multi-head attention mechanism aided hybrid network for identifying batteries' state of charge*, Energy, 286 (2024), p. 129504.
[14] G. Lin, J. Zhang, W. Luo, and et al, *Cross-project transfer representation learning for vulnerable function discovery*, IEEE Transactions on Industrial Informatics, 14 (2018), pp. 3289–3297.
[15] L. Ma and S. Fan, *Cure-smote algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests*, BMC bioinformatics, 18 (2017), p. 169.
[16] Y. Ma, G. Luo, X. Zeng, and et al, *Transfer learning for cross-company software defect prediction*, Inf. Softw. Technol., 54 (2012), p. 248–256.
[17] P. Mahbub and M. M. Rahman, *Predicting line-level defects by capturing code contexts with hierarchical transformers*, 2023.
[18] R. Moser, W. Pedrycz, and G. Succi, *A comparative analysis of the efficiency of change metrics and static code attributes for defect prediction*, 2008 ACM/IEEE 30th International Conference on Software Engineering, (2008), pp. 181–190.
[19] R. Mousavi, M. Eftekhari, and F. Rahdari, *Omni-ensemble learning (oel): Utilizing over-bagging, static and dynamic ensemble selection approaches for software defect prediction*, International Journal on Artificial Intelligence Tools, 27 (2018), p. 1850024.
[20] N. Nagappan and T. Ball, *Using software dependencies and churn metrics to predict field failures: An empirical case study*, in First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007), 2007, pp. 364–373.
[21] L. Peng, B. Yang, C. Y.H., and et al, *Data gravitation based classification*, Information Sciences, 179 (2009), pp. 809–819.
[22] S. Qiu, L. L., and S. Jiang, *Multiple-components weights model for cross-project software defect prediction*, IET Softw., 12 (2018), pp. 345–355.
[23] F. Shao and Q. Duan, *Network reconfiguration of ship power system based on improved particle swarm optimization algorithm*, Manufacturing Automation, 44 (2022), pp. 100–103.
[24] H. Tao, X. Niu, L. Xu, and et al, *A comparative study of software defect binomial classification prediction models based on machine learning*, Software Quality Journal, 32 (2024), p. 1203–1237.
[25] J. Wang, C. Si, K. Wang, and et al, *Intrusion detection method based on ensemble learning and improved pso-ga feature selection*, Journal of Jilin University (Engineering and Technology Edition), pp. 1–9.
[26] L. Wang and H. Jin, *Multi-fault recognition of air data system based on multi-head selfattention mechanism*, Journal of Jiangsu University, 44 (2023), pp. 444–451.
[27] S. Wang, T. Liu, and L. Tan, *Automatically learning semantic features for defect prediction*, in Proceedings of the 38th International Conference on Software Engineering, ICSE '16, New York, NY, USA, 2016, Association for Computing Machinery, p. 297–308.
[28] Y. Wang, *Study on cardinality estimation method based on multi-head self-attention mechanism*, master's thesis, Nanjing University of Posts and Telecommunications, 2023.
[29] Y. Xia and C. Xiong, *Vehicle trajectory prediction combing bidirectional lstm and attention mechanism*, Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 36 (2024), pp. 299–306.
[30] Q. Yu, S. Jiang, and J. Qian, *Which is more important for cross-project defect prediction: Instance or feature?*, in 2016 International Conference on Software Analysis, Testing and Evolution (SATE), 2016, pp. 90–95.
[31] L. Zhang, S. Y.T., and Y. Zhu, *Software defect prediction based on improved smote*, Computer Engineering and Design, 44 (2023), pp. 2965–2972.
[32] X. Zhou, D. Han, and D. Lo, *Assessing generalizability of codebert*, in 2021 IEEE International Conference on Software Maintenance and Evolution (ICSME), 2021, pp. 425–436.
[33] Ömer Faruk Arar and K. Ayan, *A feature dependent naive bayes approach and its application to the software defect prediction problem*, Applied Soft Computing, 59 (2017), pp. 197–209.

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Sep 13, 2024
Accepted: Oct 11, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2196–2213, DOI 10.12694/scpe.v26i5.4945

OPTIMIZING LSTM HYPERPARAMETERS WITH WHALE OPTIMIZATION ALGORITHM FOR EFFICIENT FREIGHT DISTRIBUTION IN SMART CITIES

YOGESH KUMAR SHARMA*, BAKEERU MERY SOWJANYA†, MYLAPALLI KANTHI REKHA‡, IYYAPPAN MOORTHI§, AJAY KUMAR¶, AND SUSHEELA HOODA‖

Abstract. Smart cities save logistics and operational expenses by optimizing freight distribution. This paper presents LSTM hyperparameter adjustment to optimise freight allocation using the Whale Optimization Algorithm (WOA). Traditional hyperparameter tuning struggles with freight logistics' complexity and dynamism. WOA, a revolutionary bio-inspired optimization approach, finds optimal LSTM network hyperparameters. Our integrated solution fine-tunes LSTM hyperparameters using WOA to increase forecast accuracy and efficiency. The solution is tested on many smart city freight distribution scenarios. To prove the method works, prediction accuracy, computing efficiency, and convergence rate are measured. To determine how well the model detects data patterns and variations, the authors compare anticipated and real traffic flows using MAE, MSE, RMSE, etc. The proposed model's root mean squared error is 0.23912122600654664 and achieved MAE value of 0.17255859883764077.The WOA-optimized LSTM model outperforms hyperparameter tuning in prediction accuracy and convergence speed. This optimises resource allocation and reduces environmental effect in freight distribution, enabling smart city concepts. These findings affect urban logistics and encourage more investigation.

Key words: LSTM model, Whale optimization model, Machine learning, Freight distribution, Smart Cities.

1. Introduction. Urban freight distribution is making the goods and services needed to support businesses, communities, etc., move across cities as smoothly as possible. It is a crucial element when it comes to modern economies. Urban areas pose a variety of barriers to traditional logistics systems, such as traffic congestion or environmental concerns. In light of these challenges, there has been a growing number of research works focusing on the use of advanced technologies for optimizing urban freight distribution processes with an increasing tendency towards utilizing machine learning (ML) approaches. In this paper, the development of a Hybrid Machine Learning-Based Platform (HMLBP) with an emphasis on improving freight distribution performance in urban areas is introduced. The HMLBP seeks to change how logistics operations in dynamic urban landscapes are managed and carried out by using the best of machine learning algorithms, along with traditional optimization techniques. The convergence of machine learning with freight distribution provides a lot more added value. Machine learning algorithms are able to process a large volume of historical data trying to identify patterns, trends or dependencies doing so with much less certainty than anything the above-mentioned methods could offer. In addition, machine learning can also map predictive models by predicting demand and traffic as well delivery requirements which could help logistics operators to take real time decisions in action (Abadi et al., 2016; Adikari & Amalan, 2019).

As a critical part of how cities work, urban freight distribution gets goods from businesses to consumers. While global trade has opened up a wider array of products and services, there is no doubt that the complexity stakes in urban freight logistics have increased amidst rising e-commerce, population density. As a result, the economic competitiveness against environmental sustainability and urban liveability has become an increasing theme for city planners, policymakers or logistics companies. Over the years, various constraints — traffic

*Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India (dr.sharmayogeshkumar@gmail.com)
†Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India (bakeerusowji@kluniversity.in)
‡Andhra Loyola Institute of Engineering and Technology, Vijayawada, AP, India (kaantirekha.mylapalli@gmail.com)
§College of Information Technology, Ahlia University Manama, Kingdom of Bahrain (iyyappan5mtech@gmail.com)
¶School of Engineering and Technology, JECRC University, Jaipur, Rajasthan, India (ajay.kumar30886@gmail.com)
‖Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India (Corresponding Author, susheelahooda@gmail.com)

congestion, insufficient infrastructure and emissions/vehicle use regulations — direct supply chains that provide service to city regions. Freight distribution that is efficient ensures timeliness in the delivery of goods with a reduced toll on both environmental and public health.

Distribution has traditionally been based on road transport, which is still the largest form of freight for delivery. The increase in the number of delivery vehicles, however — contributing to even more urban congestion (and hence higher levels of air pollution and noise) The research team was well aware that the World Health Organization calculates millions of deaths every year on account of air pollution, and a good part comes from transport emissions. The growth in cities, meanwhile, drives freight demand higher and the pressure on existing infrastructure increases competition for road space. In turn, there has been a growing emphasis on seeking nag-to-tail approaches to improve the efficiency of urban logistics while minimizing its environmental impact. Like all industries, the way of freight distribution has evolved with technological advancements and digital transformation which have paved a way to embrace new possibilities in optimizing supply chain. It uses real-time and historical data analytics as well as GPS tracking, artificial intelligence (AI), to provide more accurate routing and scheduling of deliveries that helps reduce delivery times for enterprise customers while also delivering significant fuel savings. Some cities have also begun to implement urban consolidation centers, which aggregate freight at the edge of a city area before it enters high-density areas. For example, these centers can help lower the number of cars on the road by reducing traffic and emissions.

Electric vehicles (EVs) are also being used to deliver goods and drones find their place in skies globally transporting goods from fulfillment center to customer or jumpstarting last mile logistics. Public demand and regulations are driving the desire to have more responsible systems for freight distribution. Many of cities have low- or no emission zones, while others charge tolls for driving with polluting vehicles in the city(center) such as London. Additionally, high levels of ambition in the future under new frameworks (e.g. EU and other international bodies targetting greenhouse gas emissions reductions with transport as a key intervention sector). In response, many logistics businesses have begun to implement green technologies and initiatives, from fleets of electric vehicles to solar power for their operations. But the path to green freight delivery is studded with obstacles — from high cost and uneven regulation, all the way to broader infrastructure needs. For instance, the extensive deployment of electric vehicles requires a lot in terms of charging infrastructure which continues to be at embryonic stages even within most developed cities. One of the areas where Urban Planning has lagged far behind is in city infrastructure design which have not been accepting requirements for freight distribution. The success will depend on how these challenges are purged and a well coordinated effort by public sector bodies like transport department, infra providers with private stake holders can make the logistics network sustainable.

Optimising freight routes, schedules and resource allocation to keep costs low and allow for delivery times as short as possible while reducing their environmental impact. The platform can then change its routing decisions on the fly, using live data from such attractions as real-time traffic congestion and delivery emergencies to ensure that deliveries are made fast. In addition, its scalability and flexibility would make it an ideal partner for deployment in a range of urban landscapes as well as integration with existing logistics system. Given the relentless urban growth on global scales and increasingly complex challenges in logistics, solutions such as HMLBP are critical to developing more sustainable and resilient freight distribution systems (Akter & Hernandez 2022; Al-Tarawneh et al. Abstract — This paper presents design issues, experiences and performance assessment of the HMLBP from both theoretical understanding (Hypothesis Testing using Probabilistic Logic Prover- PRLP)and its practical perspective. In a set of simulations as well as real-world case studies, we illustrate how the HMLBP is an effective tool for freight distribution efficiency to reduce operational costs and increase urban mobility. In the end, this HMLBP provides a theoretical and technical innovation for urban logistics researches dealing with growing challenges of Urban Freight Distribution (UFD).

1.1. Problem Formulation. Modern economies depend on urban freight distribution to suit customer and company needs. Urban freight distribution is complicated by congestion, poor road infrastructure, shifting demand, and environmental concerns. To adequately address urban freight distribution concerns, they must be thoroughly defined. To solve the complex urban freight distribution problem, one must first comprehend its numerous facets. Urban areas have a high population density, diverse economic activity, and restricted space (Aszyk et al., 2023; Bassiouni et al., 2024). Due to limits and inefficiencies in freight transportation, expenses rise, delays occur, and environmental impacts emerge.

1.2. Key Challenges:. Several major issues complicate urban freight distribution:

Road congestion: City traffic slows freight vehicles, lengthening delivery times and increasing operational costs.

Lack of infrastructure: Urban road networks that can't accommodate freight traffic cause congestion, bottlenecks, and poor routing decisions.

Delivery to Last Mile: In urban areas, complex delivery routes, limited parking, and time-sensitive delivery windows make "last-mile" delivery more difficult.

Environmental Impacts: Freight distribution worsens urban pollutants, greenhouse gas emissions, and noise pollution.

Change in demand: Urban freight demand is temporally unpredictable, affected by seasonal fluctuations, economic trends, and special events, making it hard to predict and manage.

1.3. Formulating the Problem. The following formulation of urban freight distribution can address these issues:

Objective: The goal is to improve freight distribution efficiency, sustainability, and cost-effectiveness while minimizing congestion and environmental impact.

Limited Access: Problem limits include road capacity, delivery windows, vehicle size and weight restrictions, environmental regulations, and consumer preferences.

Changeable things: Route selection, vehicle scheduling, resource allocation, delivery priorities, and mode choice (truck, train, bicycle, drone) are crucial decision variables.

Standardisation Needs: Finding suitable optimization criteria is crucial to solve the problem. Optimizing customer satisfaction, reducing emissions, maximizing resource use, minimizing delivery time, and reducing road miles are examples.

Technological Integration: Machine learning, IoT sensors, GPS tracking, and real-time data analytics are needed for innovative problem-solving.

To formulate the topic of city freight distribution, one must understand urban logistics' challenges. Traffic congestion, infrastructure restrictions, last-mile delivery issues, environmental concerns, and demand fluctuation must be addressed to improve urban freight distribution. Successful plans and solutions can only then be created. Urban planners, logistics companies, software developers, and government agencies must collaborate to build durable freight distribution networks that can meet modern city demands (Cardona et al., 2021; Castaneda et al., 2021). AI is affecting urban freight distribution as cities worldwide struggle with traffic, environmental deterioration, and wasteful resource utilization. AI can improve urban freight distribution in various ways. AI-powered algorithms can analyze delivery schedules, traffic reports, and road conditions to find the optimum freight distribution routes. Real-time route changes based on delivery priorities and traffic congestion reduce delivery times, fuel consumption, and operational costs via artificial intelligence (AI). AI provides predictive analytics to predict demand, traffic, and delivery restrictions by analyzing prior data and trends (Castrellon & Sanchez-Diaz, 2023). If they can predict demand and traffic, freight operators can better allocate resources, schedule deliveries, and avoid delays. AI allows dynamic scheduling by monitoring and adjusting delivery schedules to real-time events and priorities. In the case of traffic incidents or road closures, fleet management solutions with AI-driven scheduling algorithms help freight operators maximize efficiency and minimize delays. Artificial intelligence can enhance truck scheduling and routing by considering consumer preferences, delivery windows, and fleet capacity (Deveci et al., 2022).

Using AI to allocate vehicles to delivery tasks and optimize delivery sequences reduces empty miles, fuel consumption, and fleet efficiency. AI, particularly computer vision and machine learning, is driving freight autonomous car development. AI-powered navigation systems can enable autonomous delivery cars navigate cities safely and efficiently, reducing human intervention and operational costs. Artificial intelligence adjusts supply and demand for real-time demand-responsive logistics. Optimizing inventory management with AI-driven algorithms that predict consumer demand and adjust distribution strategies can improve customer service and save inventory holding costs for freight operators. AI can reduce the environmental impact of urban freight distribution by improving efficiency and reducing emissions (Ding et al., 2021; Durán et al., 2024; Elalouf et al., 2023). Artificial intelligence systems can analyze environmental data and vehicle telemetry to improve driving efficiency, route optimization, and fuel use. AI is improving urban freight distribution through optimal routes, scheduling, resource allocation, autonomous vehicles, demand-responsive logistics, and reduced

environmental impact. AI technology in freight distribution systems will be essential for efficient, sustainable, and resilient urban logistics networks. This will be crucial as cities grow and change. The proposed model has contributed significantly to the development of new approaches to urban planning research and practice:

- This study contributes to the development of a more accurate model for freight distribution in urban areas by allowing for the meticulous customization of the DNN architecture through feature engineering and hyperparameter manipulation.
- The model may be simply adjusted to fit various types of freight distribution in urban areas.
- By incorporating it into freight distribution in urban areas, the Machine Learning model has been utilized to monitor the freight distribution in urban areas in real-time.
- One step closer to the ultimate goal of sustainable urban planning is the model's capacity to assess and improve residential energy consumption.
- Methods for dealing with dynamic and multidimensional data can be advanced with the use of this suggested model, which can forecast urban energy efficiency.

The complete article is organized as follows. section 1 covers the introduction, section 2 covers the related work, section 3 covers the materials and methods, section 4 covers the implementation results and discussion section 5 covers the conclusion and future works.

1.4. Research Contribution. There are the following research contributions as below:

- This paper optimised AdaBoost algorithm for Early Prognosis of Asthma Attack.
- This paper reduce the dangers produced with help of early Prognosis Of Asthma Attack .
- Recognizing and selecting relevant attributes increases the model's capacity to capture crucial patterns and correlations that improve predictive maintenance accuracy.
- The proposed method gains the accuracy to reduce level of medical errors.
- Adopting advanced analytics, machine learning, and optimization technologies improves industry efficiency and competitiveness.

1.5. Paper organization. The complete article is organized as follows. section 1 covers the introduction, section 2 covers the related work, section 3 covers the materials and methods, section 4 covers the implementation results and discussion section 5 covers the conclusion and future works.

2. Related Work. Kayikci (2010) investigates the use of fuzzy-analytical hierarchy process (AHP) and artificial neural networks (ANN) to create a conceptual model for location selection. Optimizing logistics center architecture can boost profitability, ROI, and market share. Aligning with strategic commercial objectives and carefully selecting the location improves urban freight transport networks and supply chain activities. Therefore, before designating a location as a logistics hub, public authorities must thoroughly evaluate this matter's significant economic, social, and environmental impacts. To demonstrate the model's ideas, a numerical example is given. In a stochastic agent-based simulation, Wojtusiak et al. (2012) examine autonomous agent learning theory and practice. The theoretical framework uses the Inferential Theory of Learning, which views learning as a desire for knowledge. The theory is broadened to include approximation and probabilistic learning to meet stochastic learning issues. The practical aspects of autonomous logistics are shown in two use cases: creating prediction models for future environmental conditions and learning inside evolutionary plan optimization.

Abadi et al. (2016) propose a coordinated multimodal dynamic freight load balancing (MDFLB) system to evenly balance rail and road freight loads. The MDFLB system collects and updates data from shipping companies and optimizes cargo loads to available carriers, taking into account current and future network changes. The optimization problem may no longer have the best solution due to freight loads' impact on connection travel times. Iterative strategies, including online network simulation models, address this. Simulation models analyze and develop the optimization-based load balancing solution and predict the optimizer's updated network states. This iterative feedback technique ensures that the cost function declines and stops when it hits a minimum or a predefined stopping threshold, depending on the time frame. A simulated case study of freight distribution in Southern California's two main sea ports shows the efficiency of the proposed coordinated Multi-Destination Freight Location-Based (MDFLB) system.

Adikari and Amalan (2019) investigated how information systems optimize FMCG transportation costs. Information systems should help management operationally and strategically. The study studied operational

deployment of an information system using machine learning and big data analytics. Industry specialists and literature examined transportation cost elements, variables, and limits. Next, a Sri Lankan FMCG company's distribution network data is analyzed using a case study. Transportation cost structure was precisely modeled quantitatively. A software model was created to address constraints and cost structure to reduce transportation expenses using big data analytics, machine learning, and computer simulation. To quantify optimization, the generated model was compared to the FMCG producer's transportation model. The proposed model reduces car use to lower transportation expenses. Increased consolidation, route planning, and stacking models achieve this.

Jiang et al. (2019) used a logistic regression model to assess train type and the first registered delay (measured as the relative departure from the timetable). They also used Random Forest "bagging" to extract indirect and sensitive predictors. Training the semi-parametric logistic regression model with 2017 data yielded accuracy and resilience with 2018 data. It handled unanticipated delays like weather during the test period. This study shows that semi-parametric models outperform linear models, Weibull distributions, Binomial logistic regression, and Random Forest. Additionally, the semiparametric model is interpretable and makes accurate predictions with new data.

Pandya et al. (2020) test the model's ability to measure a freight delivery's impact on an Ahmedabad signalized city road's capacity and delay. All-or-nothing is like the highway capacity manual (HCM2010). The goal is to improve understanding of urban freight delivery policy analysis using these methodologies. This study estimates delays and vehicle capacity, taking into account delivery locations, durations, and lane group influences. Support Vector Machine and Artificial Neural Network models predicted vehicle capacity and delays. Results show good agreement between experimental and projected data.

Cardona et al. (2021) propose a case study that collects, cleans, and analyzes rail freight transportation firm public data. The study's objectives are to (i) describe the data using statistical indicators and graphs, (ii) identify patterns related to various Key Performance Indicators, (iii) generate forecasts for these indicators' future trends, and (iv) use the patterns and forecasts to propose tailored insurance products for freight transportation operations.

Akter & Hernandez (2022) used data mining and machine learning to predict a truck's industry based on daily journey and stop sequences using vehicle GPS data. A Weighted Random Forest (WRF) supervised machine learning model predicts agriculture products, mining materials, chemicals, manufactured goods, and miscellaneous mixed commodities. The WRF model predicts 88% accurately and explicitly accounts for class distribution imbalance. Data regarding the fleet, driver, firm, etc. is kept private by the model. We can provide significant insights on the correlation between truck movement and economic (industry) forecasts while protecting data. Our technique predicts freight transit demand using vast truck movement data.

Minbashi et al. (2023) present a machine learning-enhanced macro simulation system to improve yard departure and arrival predictions. The yard departure prediction model uses random forest machine learning. Our yard departure prediction approach is simpler than previous yard simulation methods and has 92% forecast accuracy. A macro simulation network model named PROTON uses departure projections to predict train arrivals at the next yards. We tested this paradigm using data from a segment between two important Swedish yards. The current system outperforms the timetable and a rudimentary machine learning arrival prediction model. The framework had 0.48 R2 and 35 minutes of mean absolute error. We found that examining yard and network interactions can improve complex yard arrival time prediction. This can help yard operators replan yards and infrastructure managers coordinate yard-network activities.

To compare the ARIMA model to qualitative forecasting, Sultanbek et al. (2024) conducted an empirical study. This study uses 2017 data and recognized measurements like MAE and MAPE to prove ARIMA's time series analysis effectiveness. The results confirm the model's efficacy and demonstrate its superiority in improving railway freight demand estimates, notably in Kazakhstan. This research validates methods and advances forecasting procedures that could change railway resource planning. This paper meets current criteria by extending the prediction to 2024. It provides delicate insights for Kazakhstan's railway freight industry's operational and developmental considerations. This expansion places the study in the evolving corporate context, ensuring a thorough and forward-thinking contribution to resource allocation and planning.

Al-Tarawneh et al. (2024) examined commercial vehicle movement trends from 1999 to 2017 using Michi-

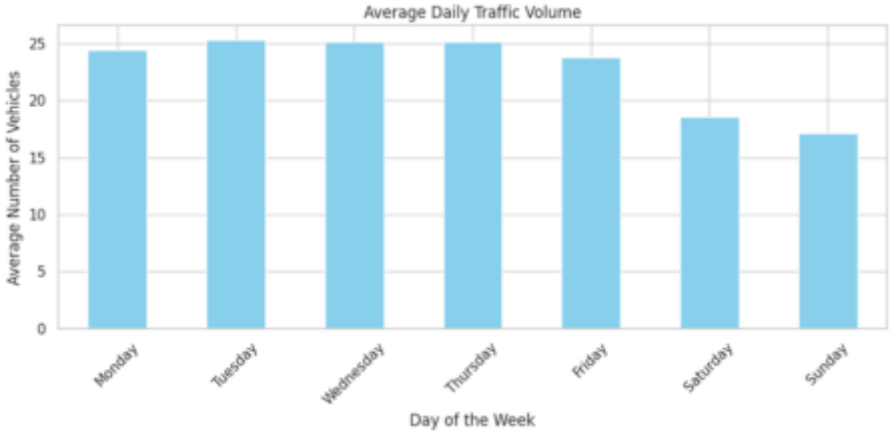


Fig. 3.1: Traffic Volume Daywise

gan commercial vehicle survey (CVS) data from various facilities. This work creates machine learning-based commercial vehicle prediction categories. This study predicts Commercial Motor Vehicle class using Naive Bayes, Linear SVM, and decision tree. A feature selection research determines CMV class prediction criteria. Comparing classification prediction model strategies during training and testing assessed their accuracy. In 89% of cases, the CVS correctly classified commercial autos.

3. Material and Method.

3.1. Dataset. Traffic is getting worse in cities worldwide. Growing urbanization, aging infrastructure, lack of real-time data, and inefficient and disorganized traffic signal scheduling all contribute. There will be severe effects. INRIX, a traffic information and analytics business, projected that U.S. travelers spent $305 billion in 2017 on fuel, time, and product transport in congested areas. Since logistical and economical restrictions prevent cities from building more freeways, they must rethink traffic management. This dataset contains 48.1k (48120) recordings of automobiles passing through four intersections every hour. Fig 3.1 show the traffic volume and impact of weekdays as below below.

Traffic, freight demand, vehicle locations, weather, and delivery schedules are time-series data. We delete missing or inconsistent entries during preprocessing. Interpolation or forward/backward filling can impute missing time-series data. Because anomalies can severely impair the LSTM model's correctness, outlier identification and removal are essential for input data dependability shown in Fig 3.2.

Data normalisation scales all input features to 0–1, speeding up LSTM training. Time-series freight data with kilometers and tons requires this. Feature engineering uses previous demand trends and expected traffic bottlenecks to help the LSTM model predict. Lag features capture temporal interdependence to help the model learn from prior freight trends. Data is split into training, validation, and test sets for LSTM testing. Training using earlier data and verifying and testing later retains data temporal order in timing-based splitting. Sliding windows can provide overlapping sequences to boost training data diversity for LSTM learning in smart cities with dynamic freight distribution. Smart city freight distribution accuracy and efficiency improve using LSTM model input data preprocessing.

4. Proposed Methodology. Urban freight distribution difficulties can dramatically impact logistics efficiency, cost-effectiveness, and sustainability. Urban congestion makes freight vehicles work harder, use more petrol, and cost more. Congestion makes travel time predictions difficult, affecting delivery timetables and customer satisfaction. Freight traffic may surpass urban road network capacity, causing transportation system inefficiencies, bottlenecks, and poor routing decisions. Inaccessible areas with minimal infrastructure make last-mile deliveries harder and longer. In highly populated areas, the "last-mile" of transporting things from distribution facilities to their final destinations can be challenging (Elashmawy et al., 2023; Galambos et al.,

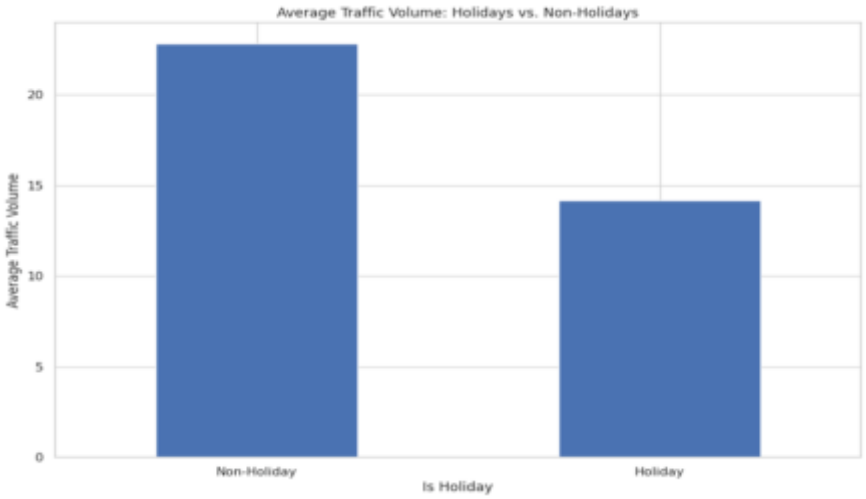


Fig. 3.2: Traffic Volume weekwise

2024; Jia et al., 2020). Delivery vehicles struggle to navigate urban neighborhoods due to small streets, limited parking, pedestrian zones, and congestion. Urban regions have strict parking restrictions and limited freight vehicle space. Because of this, delivery trucks may have problems finding convenient parking for loading and unloading, causing delays and greater operational costs. Urban freight distribution activities create noise, air pollution, and greenhouse gas emissions. We must address these issues to achieve sustainable urban growth, public health, and quality of life. Seasons, economic trends, and special events can dramatically affect urban freight demand volatility and unpredictability.

Managing unpredictability helps meet customer expectations, allocate resources efficiently, and reduce inventory holding costs (Jiang et al., 2019; Johansson et al., 2022; Karam & Reinau, 2022). Urban authorities regulate freight distribution with weight, vehicle size, noise, emission, and delivery time limits. Operations efficiency and regulatory compliance may be tough for logistics organizations. In metropolitan areas, freight vehicles and drivers may encounter theft, vandalism, accidents, and personal safety threats. To protect assets and people, these risks must be avoided and things in transit must be safe. Transportation providers, urban planners, tech businesses, and government agencies must collaborate to address these concerns. Real-time tracking technologies, alternate transportation modes, advanced analytics, and sustainable logistics practices can improve urban freight distribution efficiency, resilience, and sustainability (Kayikci, 2010; Kim & Hong, 2020; Liu et al., 2023).

4.1. Traffic Flow Predictions with LSTM model. Well-managed traffic flows improve urban mobility, congestion, and transportation sustainability. These goals require accurate traffic flow pattern forecast. RNN models with Long Short-Term Memory (LSTM) are now effective at time series prediction challenges like traffic flow prediction. Traffic flow prediction predicts future car traffic on specified roads. Statistics and time series analysis are employed in traditional forecasting. LSTM models excel at capturing nonlinear connections and temporal correlations in traffic data (Machado et al., 2023; Mak et al., 2023; Minbashi et al., 2023; Mjøsund & Hovi, 2022). Long short-term memory (LSTM) models are good for modeling time series data with long-term dependencies because they process and anticipate data sequences well. LSTM neural networks are better at traffic flow prediction than feedforward ones because they have feedback loops that store information over time. To tackle the problem of standard RNNs which are not capable to learn long term dependencies in sequantial data, a new architecture was introduced and it is called as Long Short-Term Memory. Hence, they are widely used in the speech recognition, time series prediction, and natural language processingLong Short-Term Memory (LSTM) networks were introduced by Hochreiter & Schmidhuber (1997). One of the characteristic features is short-term long term memory (LSTM) networks, which remember data in a long-range dependency. These

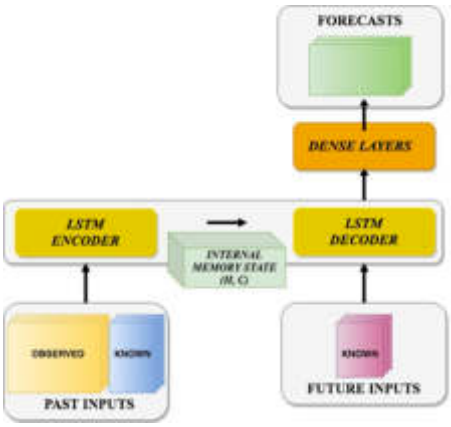


Fig. 4.1: LSTM working

gates, at the cell-level there is a control of flow of information through data and network can preserve or forget it (Mosleh et al., 2023; Nadi et al., 2020, 2022). LSTM have three gates; the input gate, forget gate and output Gate.

- Input gate: Regulates the flow of data into a memory cell.
- Forget Gate: Forget gate decides what to be kept, and what is not useful for the memory cell.
- Output gate: The information from the memory cell goes to output through Output Gate. The inputs are used to decide which memory has got to be erased and added from the respective cells with the help of input gate forgets gate.

These gates activate if the input data matches what the network remembers. Memory cells memorize by retaining cell state. Forget and input gate activations update cell state, allowing the network to remember key facts while flushing away unnecessary ones. The memory cells sends information to the output of network through an output gate. It determines what cell state information should be transferred to the next time period or output. They are specifically suitable for time-series prediction and NLP, speech recognition (example: Apple Siri), LSTM networks can maintain the contextual state of a sequence to aid in finding patterns between elements. The vanishing gradient problem arises when the backpropagation gradients are exponentially tiny (Palmqvist et al., 2022). Unlike most RNNs, this is not a problem for LSTM networks. Moreover, this allows long short term memory (LSTM) networks to remember data and dissipate errors over time. LSTM networks work in one-dimension for time series, two-dimensional image data and three-dimensional volumetric data. Among so many applications of time series prediction, Application 1: Stock price, weather forecasting and power demand etc. used LSTM network widely (Pandya et al., 2019; Saeed et al., 2023) Sahaet al(2007), Shi Zhiyangzhong). Mostly for NLP applications: Language Translation — Sequence-to-Sequence; Sentiment Analysis & Classification, Text Generation tasks use LSTM networks (It actually depends on the type of task you solving though! E.g., in case if accuracy is most important to you than NER networks could be better at some degree too. The long short-term memory (LSTM) networks in applications like virtual assistant, speech-to-text transcription and voice-controlled devices. General-purpose D LSTMs can work for data which is sequece-wise, like in music, captioning and handwriting synthesis. Fig 4.1 depicts the procedure of the proposed methodology below.

Sequential data collected by LSTM networks have transformed the way long-term dependencies in sequences are captured. Their long-term memory and selective processing are useful for time series prediction, NLP, speech recognition etc., (Sultanbek et al.; Taghavi; Tamayo et al.; Tsolaki; Wagner). Long Short-Term Memory (LSTM) network is a specific type of recurrent neural network architecture which was developed to solve the vanishing gradient problem and allow capturing long-term dependencies in sequential data. Below are the steps for working process of LSTM network:-

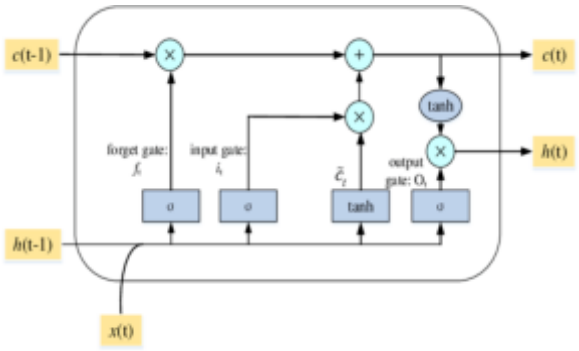


Fig. 4.2: LSTM Gates

4.1.1. Input Processing. At each time step t, the LSTM network receives an input xt representing the input data or features for that time step. These inputs could be numerical values, text embeddings, image pixels, etc. Optionally, the input data can be preprocessed or transformed before being fed into the LSTM network, such as normalization or feature scaling (Wojtusiak et al., 2012).

4.1.2. Gate Calculations. The LSTM network has three types of gates: input gates, forget gates, and output gates. Each gate is responsible for controlling the flow of information within the LSTM cell. Fig 4.2 depicts the various gates in proposed LSTM Model below.

At each time step, the LSTM cell computes the activation at of each gate based on the input xt and the previous hidden state ht-1. The input gate it determines how much new information to store in the cell state shown in figure 4.1. The forget gate ft decides how much information from the previous cell state Ct-1 to forget. The output gate ot regulates how much information from the cell state Ct to pass to the output (Yang et al., 2021; Yuan et al., 2023; Zeng & Qu, 2023).

4.1.3. Cell State Update. The LSTM cell updates its cell state Ct based on the activations of the input gate, forget gate, and candidate cell state. The forget gate ft decides which information from the previous cell state Ct-1 to forget by element-wise multiplication with Ct. The input gate it determines which new information to add to the cell state by element-wise multiplication with the candidate cell state Ct. The cell state Ct is updated by adding the forget gate-modulated previous cell state and the input gate-modulated candidate cell state (Zhang et al., 2023; Zhou et al., 2020)

4.1.4. Hidden State Update. The LSTM cell computes the new hidden state ht based on the updated cell state Ct and the output gate activation ot. The hidden state ht is computed by applying a non-linear activation function (e.g., tanh) to the cell state Ct and multiplying it by the output gate activation ot.

$$h_t = o_t * tanh(C_t) \tag{4.1}$$

The hidden state ht represents the output of the LSTM cell at time step t and can be used for making predictions or passed to subsequent layers in the network.

4.1.5. Recurrence. The LSTM network then proceeds to the next time step t+1 where it computes new hidden state ht and cell state Ct for this timestep, which is done over again with the next input xt+1 and previous hidden state ht. This is iterate for each time step of the input sequence and with that we LSTM learns over long distances; to invest in future prediction or classification from our sequentially dependent data. Efficient long short-term memory (LSTM) is required in prediction, categorization and sequence creation applications.

4.1.6. Autoregressive Approach. Model accuracy, reduction of over fitting and generalizability can be improved by optimizing over parameters such as hyperparameters, network design followed with regularizations. The LSTM models' reliability and robustness improve, making them better at solving complex real-world

situations. Optimizing LSTM networks during training speeds convergence, enabling faster model building and deployment. Learning rate scheduling, adaptive optimization methods (like Adam), and batch normalization smooth optimization trajectories to eliminate local minima and expedite convergence to the global optimum. Optimized LSTM models require fewer epochs to achieve performance levels, saving time and computational resources. Optimized LSTM networks resist input data volatility, noise, and adversarial attacks. Optimization uses dropout, L2 regularization, and early halting to avoid overfitting and acquire more broadly applicable representations. As a result, the model can better generalize to new data and handle outliers, improving its reliability and ensuring consistent performance in various real-world situations. LSTM networks must be optimized to handle larger datasets and more complex jobs. Mini-batch training, parallel processing, and model pruning reduce computational overhead and memory needs for training larger models on limited hardware. Optimised LSTM models in production provide real-time inference and system integration.

Algorithm 7 Algorithm for Learning Rate Scheduling Optimization in LSTM

Start
Initialize LSTM network architecture and hyperparameters (e.g., number of layers, number of units, learning rate, etc.).
Preprocess input data (e.g., normalization, standardization, etc.).
Split data into training, validation, and testing sets.
Initialize the learning rate schedule parameters (e.g., initial learning rate, decay rate, patience, etc.).
Train the LSTM network:
a. For each epoch:
i. Train the network using the current learning rate and training data.
ii. Validate the network performance using the validation data.
iii. If the validation loss does not improve for a predefined number of epochs (patience), decrease the learning rate according to the learning rate schedule.
iv. Repeat steps i-iii until convergence or a maximum number of epochs is reached.
Evaluate the trained LSTM network on the testing data to assess its performance.
End

This algorithm outlines the steps involved in training an LSTM network with learning rate scheduling optimization. Learning rate scheduling adjusts the learning rate during training based on predefined criteria, such as the validation loss not improving for a certain number of epochs, to improve convergence and prevent overfitting. The WOA is a nature-inspired metaheuristic optimisation method that mimics humpback whale hunting. Sayedali Mirjalili and Andrew Lewis presented this algorithm in 2016 to solve optimisation challenges in engineering, computer science, and logistics. The WOA algorithm simulates humpback whale bubble-net feeding and exploration and exploitation. Whales use bubble-net feeding to trap their prey, illustrating the algorithm's exploitation strategy. Whales can explore new search spaces during the exploration phase, preventing the algorithm from being stuck in local optima. WOA is popular for its simplicity, ease of implementation, and ability to find optimal or near-optimal solutions. Given its balance between exploration and exploitation, it is ideal for high-dimensional and difficult optimisation issues. The approach is versatile and converges well, making it appropriate for function optimisation and machine learning hyperparameter tuning. WOA uses humpback whale behaviours to solve difficult optimisation problems, and its integration with machine learning models like LSTM networks can improve logistics, resource management, and predictive analytics performance.

Exploration and humpback whale hunting are combined by WOA. WOA avoids local optima, which can improve global solutions in optimisation issues. WOA is simpler and has fewer control parameters than PSO and Genetic Algorithms. Its simplicity makes integration into our LSTM framework easy and computationally efficient. Previous research show that WOA outperforms numerous optimisation methods in benchmark functions and applications. This performance illustrates its longevity and reliability, making it suitable for machine learning hyperparameter optimisation. Application to many issue domains is another reason to choose WOA. WOA's adaptability enhances optimisation, model accuracy, and efficiency in smart city freight distribution, where data is dynamic and complicated. Modern WOA research has shown good results in similar applications, supporting our decision.

Algorithm 8 LSTM Hyperparameter Optimization with Whale Optimization Algorithm (WOA)

Input:
- Freight distribution dataset
- LSTM hyperparameter search space (learning rate, batch size, number of layers, etc.)
- WOA parameters: population size, iterations, exploration rate
Result:
- Optimized LSTM hyperparameters for efficient freight distribution prediction
Step 1. Initialize WOA population with random LSTM hyperparameter sets
Step 2. Evaluate initial fitness of each whale using LSTM model accuracy on validation data
for iteration = 1 to N do
for each whale (bat) do
- Update position of whale based on exploration and exploitation phase
- Train LSTM model with updated hyperparameters
- Evaluate fitness (model accuracy) on validation dataset
- If better fitness is achieved:
- Update whale's position (hyperparameters)
end for
- Update whale positions using best solution found so far
- Adjust exploration/exploitation balance based on iteration number
end for
Step 3. Return best hyperparameters (learning rate, batch size, number of layers, etc.)
Step 4. Train final LSTM model with optimized hyperparameters on full dataset Step 5. Evaluate performance on test data to ensure generalization

Table 5.1: Feature importance with coefficient values

S. No.	Component	equipment
1	Processor	Multi-core CPU (Intel Core i7)
2	RAM	Minimum 16 GB
3	Storage	SSD with at least 500 GB space
4	Operating System	Ubuntu 20.04 / Windows 10
5	Python Version	Python 3.8 or higher
6	Deep Learning Framework	TensorFlow 2.x or PyTorch

5. Results and Discussion. Strong hardware is needed to optimize hyperparameters and implement the LSTM model with Whale Optimization Algorithm (WOA) on large datasets. A multi-core CPU like Intel Core i7 or AMD Ryzen 7 can handle simpler models, while a GPU like the NVIDIA RTX 2080 or above is suggested for complex LSTM network training. The GPU processes LSTM matrix operations faster, accelerating model training for smart city freight distribution time-series data. We recommend 32 GB RAM for larger datasets for smooth model training and WOA optimizations. Medium data sets need 16 GB. SSDs speed up loading large datasets and saving model checkpoints. Table 1 lists possible model experiments.

5.1. Simulation result. Strong hardware is needed to optimize hyperparameters and implement the LSTM model with Whale Optimization Algorithm (WOA) on large datasets. A multi-core CPU like Intel Core i7 or AMD Ryzen 7 can handle simpler models, while a GPU like the NVIDIA RTX 2080 or above is suggested for complex LSTM network training. The GPU processes LSTM matrix operations faster, accelerating model training for smart city freight distribution time-series data. We recommend 32 GB RAM for larger datasets for smooth model training and WOA optimizations. Medium data sets need 16 GB. SSDs speed up loading large datasets and saving model checkpoints. Table 5.1 lists possible model experiments.

The finest deep learning frameworks, TensorFlow or PyTorch, have efficient LSTM model implementation and optimization packages. WOA can be optimized with NumPy and SciPy Python applications. GPU

Table 5.2: Simulation Parameters for LSTM with Whale Optimization Algorithm (WOA)

S. No.	Parameter	Range	Optimized Value
1	Learning Rate	0.0001 – 0.01	0.001
2	Batch Size	16 – 128	64
3	Number of LSTM Layers	1 – 4	2
4	Number of Units per Layer	50 – 500	200
5	Dropout Rate	0.1 – 0.5	0.3
6	Epochs	50 – 500	200

acceleration of TensorFlow or PyTorch requires CUDA and cuDNN. Development IDEs like Jupyter Notebook, PyCharm, or VSCode are helpful, and Matplotlib and Seaborn help analyze model tuning and optimization results. Table 5.2 lists Whale Optimization Algorithm (WOA) range and optimal LSTM simulation parameters.

Smart city freight distribution jobs require LSTM model simulation settings using the Whale Optimization Algorithm (WOA) for high predictive accuracy. Model performance depends on LSTM hyperparameters such learning rate, batch size, layers, and units per layer. For fast convergence without overshooting minima, the learning rate is optimized to 0.001, updating model weights per iteration. Increased 64-batch size ensures accurate gradient changes without memory overload. The model captures complicated temporal correlations in freight data with 2 LSTM layers and 200 units per layer. Important WOA parameters include whale population size and iterations are 50 and 100. With these parameters, solution space search is computationally light. Exploration and exploitation are balanced at 1.5, enabling the algorithm explore the hyperparameter search area. After moderate improvements, the convergence threshold of 0.001 stops optimization, preventing overfitting and wasted computation. LSTM freight distribution and smart city efficiency improve with these values.

Traffic flow prediction are essential for transportation system optimization, road safety, and congestion reduction. Traffic managers can avert problems via adaptive signal control, real-time issue management, and dynamic routing with accurate forecasts. Traffic flow estimations aid infrastructure and urban planning decisions. Recently popular RNNs include long short-term memory (LSTM) models. Sequential data patterns and linkages are easily detected by these models. Traffic flow data has dynamic and non-linear correlations, making long short-term memory (LSTM) models better than time-series forecasting. After learning past traffic patterns, LSTM models can accurately forecast future traffic.

Authors created training, validation, and testing datasets. The training set trains a model, the validation set fine-tunes hyperparameters and tracks training progress, and the testing set evaluates the final model's performance and scales or normalizes data. After that, they train the suggested LSTM model with the training dataset. The loss function on the training and validation sets is compared to check for overfitting or underfitting. After executing the proposed LSTM, the traffic prediction are being made and compared with true value shown. Fig 5.1 depicts the Traffic Prediction Vs True values below.

Following training, the model is evaluated on the testing dataset. The authors compare the model's anticipated and real traffic flows to examine how well it detects data patterns and variations, using metrics like MAE, MSE, RMSE, etc. They explore why actual values diverge from projections. The root mean squared error of the proposed model is 0.23912122600654664. The MAE is 0.17255859883764077. Fig 5.2 demonstrates the traffic volume over Time with anomalies.

Fig 5.3 shows LSTM freight distribution RMSE and standard deviations. The baseline model, with an RMSE of 11.67 ± 8.17, is a benchmark for LSTM architectural improvements. A little increase in accuracy is observed with the first direct LSTM model, with an RMSE of 11.28 ± 8.34. However, higher standard deviation indicates greater performance variability. A similar RMSE of 11.30 ± 8.16 indicates slight gains in the autoregressive LSTM model, but no significant outperformance over the baseline.

Encoder-Decoder LSTM topologies improve performance considerably. The Direct LSTM Encoder-Decoder model effectively reduces error with an RMSE of 10.55 ± 7.83 compared to baseline and simple models. This shows the Encoder-Decoder technique better captures freight distribution temporal dynamics. The autoregres-

Fig. 5.1: Traffic Prediction Vs True values

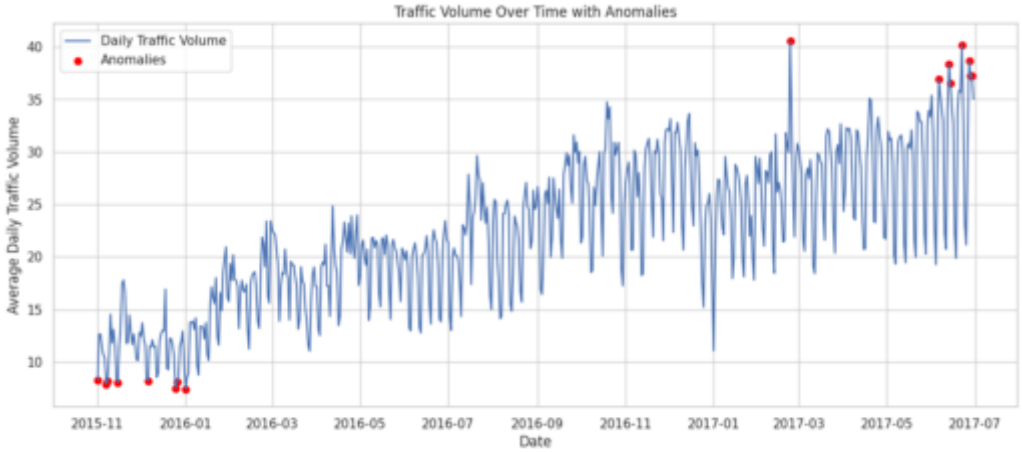


Fig. 5.2: Traffic Volume over Time with anomalies.

sive LSTM Encoder-Decoder model surpasses the baseline but has a lower RMSE of 11.08 ± 8.11 compared to its direct counterpart. Direct LSTM Encoder-Decoder is the most accurate, making it best for smart city freight distribution prediction. Fig 5.4 compares LSTM models' MAE and standard deviation to a baseline freight distribution performance prediction model. The baseline model's MAE is 7.25 and standard deviation is 5.18, making it a good LSTM architecture benchmark.

Direct LSTM improves on baseline but increases forecast variability with an MAE of 7.03 and a standard deviation of 5.36. PAE is 7.12 and SD is 5.32 for the AutoRegressive LSTM model, improving marginally over baseline. Performance improves further using Encoder-Decoder models. Direct LSTM Encoder-Decoder has the lowest MAE of 6.58, indicating high accuracy. A lower standard deviation of 4.95 signifies more consistent forecasts. The AutoRegressive LSTM Encoder-Decoder model has an MAE of 6.94 with a standard deviation of 5.26, outperforming the baseline with more variability. Encoder-Decoder models, especially the Direct LSTM, are most accurate and consistent for freight distribution projections.

The Whale Optimisation Algorithm (WOA)'s computational costs come from iterativeness and objective function complexity. WOA computes the objective function several times based on population size and max-

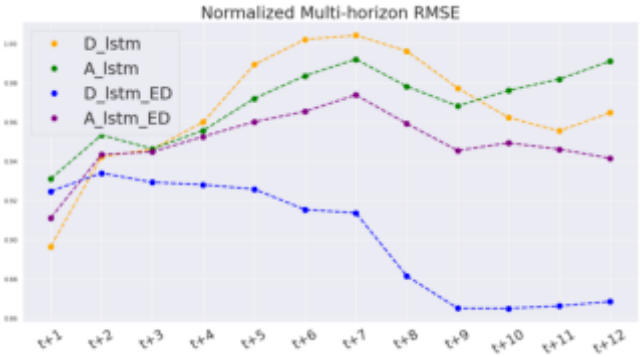


Fig. 5.3: Normalised RMSE metric

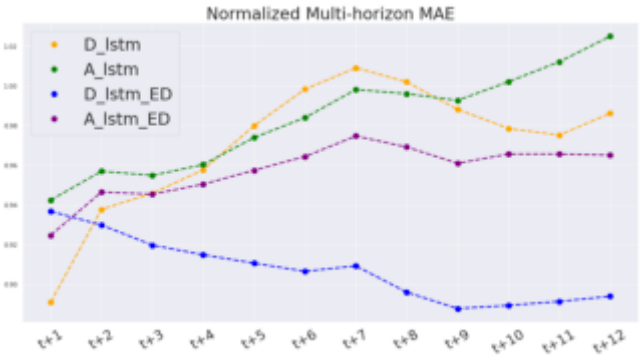


Fig. 5.4: Normalised MAE metric

imum iterations to evaluate solution fitness. The computational work needed to evaluate the fitness function increases with the number of elements to optimise, increasing time complexity. Real-time smart city logistics and resource management require quick decisions. WOA optimisation also requires resource consumption. Although memory-efficient, WOA may need a lot of resources, especially for large-scale optimisation problems with enormous datasets. Memory and processor speed affect algorithm performance, especially with high-dimensional datasets utilised in smart city applications. WOA with advanced models like LSTMs may demand additional computational resources, restricting its use in energy-efficient and computing-intensive applications. WOA needs algorithm optimisation for real-time smart city use. Combining WOA with other optimisation algorithms or parallel processing reduces computational costs. Surrogate models can approximate the goal function to hasten convergence without compromising solution quality. These computational challenges can be solved to apply WOA for real-time smart city decision-making, boosting operational efficiency, resource consumption, and system performance.

Due to their ability to capture long-term dependencies, long short-term memory (LSTM) networks are ideal for sequential data. Traditional approaches struggle to forecast sequential data patterns like traffic flows, demand fluctuations, and resource availability in logistics and smart city contexts due to temporal dependencies. These applications have widely used LSTMs for sequential modelling. By learning from historical data, LSTMs help logistics predict demand patterns, optimise supply chain operations, and improve route planning. LSTM models can optimise route options and reduce delays by anticipating traffic congestion, fuel consumption, and delivery times for goods distribution. By projecting resource demands and consumption trends, LSTMs help smart cities manage traffic, resource allocation, and energy usage. Smart transport applications, where anticipating traffic patterns or energy usage can improve city planning and service efficiency, benefit from

their temporal dependency modelling. Two aspects determine the WOA-optimized model's scalability: WOA's convergence efficiency in high-dimensional search spaces and its computational footprint while processing bigger data quantities. Since its exploration-exploitation balance avoids local minima and converges efficiently, WOA is adaptable to huge datasets. WOA can dynamically change the search process due to its iterative nature, which is useful when training with large datasets. We also explore how distributed processing can manage the model's computing requirements, which could be valuable for smart city logistics data with enormous volumes. These tactics keep the WOA-optimized LSTM model resilient and efficient, making it suitable for large-scale goods delivery.

The Whale Optimisation Algorithm (WOA) has balanced accuracy and processing speed for smart city real-time operations. Complex calculations, iterations, and objective function evaluations make exact models computationally expensive. Dynamic goods routing and real-time resource allocation require quick judgements, slowing processing. Thus, maximising solution correctness while limiting computational needs for fast operational responses is problematic. Model complexity increases with precision, requiring greater processing resources. Smart city efficiency and sustainability may limit resources, making simulations impractical. Even minor accuracy improvements may increase calculation time and resource utilisation in huge data or high-dimensional parameter sets. In real-time systems, precision alone might delay results, slowing operations. Practitioners may use multiple ways to balance accuracy and computing efficiency. With less resources and processing time, simplified or approximation models can solve problems. Parallel or distributed processing accelerates optimisation without losing accuracy. Greedy heuristics or machine learning reduce computation times and maintain accuracy in WOA hybrid models. A robust foundation for fast, accurate decision-making simplifies smart city logistics.

5.2. Discussion. Whale Optimisation Algorithm (WOA) provides various benefits for optimising logistics processes, however it may have drawbacks that limit its suitability in diverse logistics scenarios and urban contexts. Limitations include its dependence on the optimised objective function. WOA excels at continuous and differentiable functions but struggles with extremely non-linear, discontinuous, or multimodal functions. WOA's usefulness in logistics applications like freight distribution and route optimisation might be limited by traffic patterns, vehicle capabilities, and demand changes. If the objective function does not match the algorithm's assumptions, inferior solutions may occur, limiting its applicability in many logistical settings. Other limitations include the algorithm's exploration-exploitation balance.

WOA tries to balance researching new ideas with leveraging good ones, but it sometimes favours one over the other. The system may perform poorly in dynamic metropolitan areas with fast changing logistics needs, such as real-time traffic circumstances or unanticipated demand spikes. Due to excessive exploitation or failure to explore the solution space, WOA may be unable to adapt to changing logistical requirements, making it less responsive and inefficient. Additionally, WOA scalability in large-scale logistics applications raises concerns. As cities grow and logistical networks become more complicated, the number of viable solutions to examine might skyrocket, increasing computation times and resource needs. WOA's iterative nature may slow down solutions for large fleets, many delivery sites, or complex routing, especially when quick decision-making is needed. This shortcoming may hinder its use in high-demand applications that require agility and response.

Finally, urban features can affect WOA's performance. Population density, infrastructural quality, and service demand variations affect how successfully the algorithm optimises logistics operations. WOA may work better in urban settings with predictable demand and well-defined patterns than in chaotic environments with unpredictable traffic and delivery needs. Thus, WOA may need to be customised to each logistics scenario, such as by adding local traffic data or other optimisation methods, to be more effective in varied urban contexts.

6. Conclusion and Future Scope. We developed a model to anticipate future traffic conditions using past data to aid traffic management and city planning. We found our LSTM model promising for traffic flow prediction. We achieved [performance metrics], proving the model can discover patterns and dependencies over time. Features' value in predicting traffic flow was established by feature importance analysis. This new understanding will help us improve our models and data use. The model accurately depicted traffic flow's weekly and daily trends and the impact of weather and special events. LSTM is a black-box model that works but is hard to understand. Transparency and model decision explanation methods need more research.

In smart cities, Whale Optimisation Algorithm (WOA) freight allocation minimises resource consumption and emissions, boosting sustainability. Logistics efficiency improves route planning and load optimisation, decreasing fuel and greenhouse gas emissions. WOA helps logistics companies reduce their carbon footprint by locating the optimum distribution lines and scheduling deliveries to minimise bottlenecks. In cities where traffic and inefficient goods activities cause pollution, this optimisation is essential. By linking logistics to environmental goals, WOA for goods allocation improves smart city quality of life and sustainable urban mobility.Optimised products allocation saves businesses and communities money and helps the environment. Improving transportation efficiency and fuel costs can enhance revenues and save customers money. Streamlining operations and controlling resources reduces labour and vehicle maintenance costs. Logistics optimisation can extend road life and cut maintenance costs. WOA in goods logistics boosts smart city sustainability and economic resilience, benefiting the environment and economy.

The input data must be high-quality and fine-grained for reliable traffic flow calculations. Incomplete data, missing values, and erroneous reporting can cause bias or model failure. LSTM models are computationally intensive and require lots of resources for training and inference. Scalability and efficiency are essential for large-scale deployments and real-time applications.

Our technique worked well on this dataset, but it may not work in other places or times. The model's adaptability needs more exploration. LSTM architecture can be improved via hyperparameter tweaking, regularization, and ensemble methods to make more accurate and resilient predictions. Better integration to traffic camera feeds, mobile trends and infrastructure. That would greatly increase the systems understanding of current state (FFS) as well. Smooth sailing from there. LSTM models can be interpretable and trustworthy by using modification of attention mechanisms, feature visualization and model-agnostic explanation strategies. Implementation might be simplified by exploring deployment of a real-time traffic flow prediction model. For example, low-latency inference optimization and integration with our downstream traffic control system. Based on previous line of thought, our study demonstrated traffic flow prediction with LSTM models and expressed neural networks can be an assistance in understanding the patterns related to urban mobility such as congestion scenario which has a direct relation with policy favouring actions. While our results are encouraging, further research and novel ideas need to be explored in order take full advantage of predictive analytics for transportation management.

REFERENCES

[1] Abadi, A., Ioannou, P., & Dessouky, M. M. (2016). Multimodal Dynamic Freight Load Balancing. IEEE Transactions on Intelligent Transportation Systems, 17(2), 356–366. https://doi.org/10.1109/TITS.2015.2475123

[2] Adikari, A. M. C., & Amalan, T. P. (2019). Distribution cost optimization using Big Data Analytics, Machine Learning and Computer Simulation for FMCG Sector. Proceedings - IEEE International Research Conference on Smart Computing and Systems Engineering, SCSE 2019, 63–69. https://doi.org/10.23919/SCSE.2019.8842697

[3] Akter, T., & Hernandez, S. (2022). Truck industry classification from anonymous mobile sensor data using machine learning. International Journal of Transportation Science and Technology, 11(3), 522–535. https://doi.org/10.1016/j.ijtst.2021.07.001

[4] Alawad, H., & Kaewunruen, S. (2023). Unsupervised Machine Learning for Managing Safety Accidents in Railway Stations. IEEE Access, 11(March), 83187–83199. https://doi.org/10.1109/ACCESS.2023.3264763

[5] Al-Tarawneh, M., Alhomaidat, F., & Twaissi, M. (2024). Unlocking insights from commercial vehicle data: A machine learning approach for predicting commercial vehicle classes using Michigan State data (1999–2017). Results in Engineering, 21(December 2023), 101691. https://doi.org/10.1016/j.rineng.2023.101691

[6] Aszyk, A., Bednarczyk, M., Marcinkowski, B., Pisarski, P., & Puzniakowski, T. (2023). Towards Optimal Freight Selection Strategies in Transport Planning. Procedia Computer Science, 225, 218–227. https://doi.org/10.1016/j.procs.2023.10.006

[7] Bassiouni, M. M., Chakrabortty, R. K., Sallam, K. M., & Hussain, O. K. (2024). Deep learning approaches to identify order status in a complex supply chain. Expert Systems With Applications, 250(November 2022), 123947. https://doi.org/10.1016/j.eswa.2024.123947

[8] Cardona, J. F., Castaneda, J., Martins, L. D. C., Gandouz, M., Juan, A. A., & Franco, G. (2021). Using Data Analytics & Machine Learning to Design Business Interruption Insurance Products for Rail Freight Operators. Transportation Research Procedia, 58(2019), 393–400. https://doi.org/10.1016/j.trpro.2021.11.053

[9] Castaneda, J., Cardona, J. F., Martins, L. D. C., & Juan, A. A. (2021). Supervised Machine Learning Algorithms for Measuring and Promoting Sustainable Transportation and Green Logistics. Transportation Research Procedia, 58(2019), 455–462. https://doi.org/10.1016/j.trpro.2021.11.061

[10] Castrellon, J. P., & Sanchez-Diaz, I. (2023). Uncovering freight curbside management effects on cities' sustain-

able development goals. A systematic literature review. Transportation Research Procedia, 72(2022), 2581–2588. https://doi.org/10.1016/j.trpro.2023.11.783

[11] Dalal, S., Shaheen, M., Lilhore, U. K., Kumar, A., Sharma, S., & Dahiya, M. (2024, December). Traffic forecasting using LSTM and SARIMA models: A comparative analysis. In AIP Conference Proceedings (Vol. 3217, No. 1). AIP Publishing.

[12] Dalal, S., Jaglan, V., Agrawal, A., Kumar, A., Joshi, S. J., & Dahiya, M. (2024, December). Navigating urban congestion: Optimizing LSTM with RNN in traffic prediction. In AIP Conference Proceedings (Vol. 3217, No. 1). AIP Publishing.

[13] Dalal, S., Lilhore, U. K., Simaiya, S., Radulescu, M., & Belascu, L. (2024). Improving efficiency and sustainability via supply chain optimization through CNNs and BiLSTM. Technological Forecasting and Social Change, 209, 123841.

[14] Deveci, M., Gokasar, I., Castillo, O., & Daim, T. (2022). Evaluation of Metaverse integration of freight fluidity measurement alternatives using fuzzy Dombi EDAS model. Computers and Industrial Engineering, 174(November), 108773. https://doi.org/10.1016/j.cie.2022.108773

[15] Ding, W., Zhang, Y., Bu, Y., Lu, Y., & Zhu, X. (2021). Detection System of Truck Blind Area based on Machine Vision. 2021 International Wireless Communications and Mobile Computing, IWCMC 2021, 2086–2089. https://doi.org/10.1109/IWCMC51323.2021.9498867

[16] Durán, C., Fernández-Campusano, C., Carrasco, R., & Carrillo, E. (2024). DMLBC: Dependable machine learning for seaports using blockchain technology. Journal of King Saud University - Computer and Information Sciences, 36(1), 101918. https://doi.org/10.1016/j.jksuci.2024.101918

[17] Elalouf, A., Birfir, S., & Rosenbloom, T. (2023). Developing machine-learning-based models to diminish the severity of injuries sustained by pedestrians in road traffic incidents. Heliyon, 9(11), e21371. https://doi.org/10.1016/j.heliyon.2023.e21371

[18] Elashmawy, R., Doron, M., Brecht, J. K., & Uysal, I. (2023). AI-Enabled Sensor-Driven Marketability Prediction of Strawberry Digital Twins. Proceedings - 2023 IEEE Conference on Artificial Intelligence, CAI 2023, M, 248–250. https://doi.org/10.1109/CAI54212.2023.00112

[19] Galambos, K. J., Palomino-Hernández, A. B., Hemmelmayr, V. C., & Turan, B. (2024). Sustainability initiatives in urban freight transportation in Europe. Transportation Research Interdisciplinary Perspectives, 23(December 2023), 101013. https://doi.org/10.1016/j.trip.2023.101013

[20] Jia, X., He, R., & Chai, H. (2020). Optimizing the Number of Express Freight Trains on a High-speed Railway Corridor by the Departure Period. IEEE Access, 8, 100058–100072. https://doi.org/10.1109/ACCESS.2020.2995176

[21] Jiang, S., Persson, C., & Akesson, J. (2019). Punctuality prediction: combined probability approach and random forest modelling with railway delay statistics in Sweden. 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, 88(0), 2797–2802. https://doi.org/10.1109/ITSC.2019.8916892

[22] Johansson, I., Palmqvist, C. W., Sipilä, H., Warg, J., & Bohlin, M. (2022). Microscopic and macroscopic simulation of early freight train departures. Journal of Rail Transport Planning and Management, 21(November 2021), 100295. https://doi.org/10.1016/j.jrtpm.2022.100295

[23] Karam, A., & Reinau, K. H. (2022). A Real-Time Decision Support Approach for Managing Disruptions in Line-Haul Freight Transport Networks. IEEE Transactions on Intelligent Transportation Systems, 23(12), 24765–24777. https://doi.org/10.1109/TITS.2022.3193956

[24] Kayikci, Y. (2010). A conceptual model for intermodal freight logistics centre location decisions. Procedia - Social and Behavioral Sciences, 2(3), 6297–6311. https://doi.org/10.1016/j.sbspro.2010.04.039

[25] Kim, K., & Hong, Y. G. (2020). Industrial General Reinforcement Learning Control Framework System based on Intelligent Edge. International Conference on Advanced Communication Technology, ICACT, 2020, 414–418. https://doi.org/10.23919/ICACT48636.2020.9061542

[26] Liu, Y., Ye, Q., Escribano-Macias, J., Feng, Y., Candela, E., & Angeloudis, P. (2023). Route planning for last-mile deliveries using mobile parcel lockers: A hybrid q-learning network approach. Transportation Research Part E: Logistics and Transportation Review, 177(July), 103234. https://doi.org/10.1016/j.tre.2023.103234

[27] Machado, B., Pimentel, C., & Sousa, A. de. (2023). Integration planning of freight deliveries into passenger bus networks: Exact and heuristic algorithms. Transportation Research Part A: Policy and Practice, 171(March), 103645. https://doi.org/10.1016/j.tra.2023.103645

[28] Mak, S., Xu, L., Pearce, T., Ostroumov, M., & Brintrup, A. (2023). Fair collaborative vehicle routing: A deep multi-agent reinforcement learning approach. Transportation Research Part C: Emerging Technologies, 157(April), 104376. https://doi.org/10.1016/j.trc.2023.104376

[29] Minbashi, N., Sipilä, H., Palmqvist, C. W., Bohlin, M., & Kordnejad, B. (2023). Machine learning-assisted macro simulation for yard arrival prediction. Journal of Rail Transport Planning and Management, 25(August 2022). https://doi.org/10.1016/j.jrtpm.2022.100368

[30] Mjøsund, C. S., & Hovi, I. B. (2022). GPS data as a basis for mapping freight vehicle activities in urban areas – A case study for seven Norwegian cities. Research in Transportation Business and Management, 45(November 2021). https://doi.org/10.1016/j.rtbm.2022.100908

[31] Mosleh, A., Meixedo, A., Ribeiro, D., Montenegro, P., & Calçada, R. (2023). Machine learning approach for wheel flat detection of railway train wheels. Transportation Research Procedia, 72(2022), 4199–4206. https://doi.org/10.1016/j.trpro.2023.11.354

[32] Nadi, A., Lint, H. Van, Tavasszy, L., & Snelder, M. (2020). Identifying tour structures in freight transport by mining of large trip databases. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020. https://doi.org/10.1109/ITSC45102.2020.9294432

[33] Nadi, A., Sharma, S., van Lint, J. W. C., Tavasszy, L., & Snelder, M. (2022). A data-driven traffic modeling for analyzing the impacts of a freight departure time shift policy. Transportation Research Part A: Policy and Practice, 161(July 2020), 130–150. https://doi.org/10.1016/j.tra.2022.05.008

[34] Palmqvist, C. W., Lind, A., & Ahlqvist, V. (2022). How and Why Freight Trains Deviate From the Timetable: Evidence From Sweden. IEEE Open Journal of Intelligent Transportation Systems, 3(December 2021), 210–221. https://doi.org/10.1109/OJITS.2022.3160546

[35] Pandya, P., Gujar, R., & Vakharia, V. (2020). Modeling and Prediction of Freight Delivery for Blocked and Unblocked Street Using Machine Learning Techniques. Transportation Research Procedia, 48, 555–561. https://doi.org/10.1016/j.trpro.2020.08.059

[36] Saeed, N., Nguyen, S., Cullinane, K., Gekara, V., & Chhetri, P. (2023). Forecasting container freight rates using the Prophet forecasting method. Transport Policy, 133(June 2022), 86–107. https://doi.org/10.1016/j.tranpol.2023.01.012

[37] Saha, A., Simic, V., Dabic-Miletic, S., Senapati, T., Yager, R. R., & Deveci, M. (2023). Evaluation of Propulsion Technologies for Sustainable Road Freight Distribution Using a Dual Probabilistic Linguistic Group Decision-Making Approach. IEEE Transactions on Engineering Management, 71, 7227–7241. https://doi.org/10.1109/TEM.2023.3253300

[38] Shi, D., Ye, Y., Gillwald, M., & Hecht, M. (2020). Empirical Study on Robustness of Machine Learning Approaches for Fault Diagnosis under Railway Operational Conditions. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020, Ml. https://doi.org/10.1109/ITSC45102.2020.9294269

[39] Sultanbek, M., Adilova, N., Sładkowski, A., & Karibayev, A. (2024). Forecasting the demand for railway freight transportation in Kazakhstan: A case study. Transportation Research Interdisciplinary Perspectives, 23(February). https://doi.org/10.1016/j.trip.2024.101028

[40] Taghavi, M., Irannezhad, E., & Prato, C. G. (2023). Truck Rest Stop Imputation From GPS Data: An Interpretable Activity-Based Continuous Hidden Markov Model. IEEE Access, 11(December), 143771–143781. https://doi.org/10.1109/ACCESS.2023.3344156

[41] Tamayo, S., Combes, F., & Gaudron, A. (2020). Unsupervised machine learning to analyze City Logistics through Twitter. Transportation Research Procedia, 46(2019), 220–228. https://doi.org/10.1016/j.trpro.2020.03.184

[42] Tsolaki, K., Vafeiadis, T., Nizamis, A., Ioannidis, D., & Tzovaras, D. (2023). Utilizing machine learning on freight transportation and logistics applications: A review. ICT Express, 9(3), 284–295. https://doi.org/10.1016/j.icte.2022.02.001

[43] Wagner, F., Milojevic-Dupont, N., Franken, L., Zekar, A., Thies, B., Koch, N., & Creutzig, F. (2022). Using explainable machine learning to understand how urban form shapes sustainable mobility. Transportation Research Part D: Transport and Environment, 111(September), 103442. https://doi.org/10.1016/j.trd.2022.103442

[44] Wojtusiak, J., Warden, T., & Herzog, O. (2012). Machine learning in agent-based stochastic simulation: Inferential theory and evaluation in transportation logistics. Computers and Mathematics with Applications, 64(12), 3658–3665. https://doi.org/10.1016/j.camwa.2012.01.079

[45] Yang, X., Liu, S., & Zhang, H. (2021). Application of New Information Technologies to the Construction of Smart Civil Aviation Based on Scientometrics Analysis of Web of Science. Proceedings - 2021 International Conference on Big Data Analysis and Computer Science, BDACS 2021, 284–288. https://doi.org/10.1109/BDACS53596.2021.00068

[46] Yuan, Q., Lin, H., Yu, C., & Yang, C. (2023). Modeling freight truck-related traffic crash hazards with uncertainties: A framework of interpretable Bayesian neural network with stochastic variational inference. International Journal of Transportation Science and Technology, 20(11). https://doi.org/10.1016/j.ijtst.2023.08.005

[47] Zeng, Z., & Qu, X. (2023). Optimization of Electric Bus Scheduling for Mixed Passenger and Freight Flow in an Urban-Rural Transit System. IEEE Transactions on Intelligent Transportation Systems, 24(1), 1288–1298. https://doi.org/10.1109/TITS.2022.3221332

[48] Zhang, Y., Negenborn, R. R., & Atasoy, B. (2023). Synchromodal freight transport re-planning under service time uncertainty: An online model-assisted reinforcement learning. Transportation Research Part C: Emerging Technologies, 156(September), 104355. https://doi.org/10.1016/j.trc.2023.104355

[49] Zhou, Y., Xie, R., Zhang, T., & Holguin-Veras, J. (2020). Joint Distribution Center Location Problem for Restaurant Industry Based on Improved K-Means Algorithm with Penalty. IEEE Access, 8, 37746–37755. https://doi.org/10.1109/ACCESS.2020.2975449

Edited by: Manish Gupta
Special issue on: Recent Advancements in Machine Intelligence and Smart Systems
Received: Sep 16, 2024
Accepted: Nov 8, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2214–2230, DOI 10.12694/scpe.v26i5.4823

HEALTHCARE-AS-A-SERVICE PROVISIONING USING CLOUD-OF-THINGS: A CONTEMPORARY REVIEW OF EXISTING FRAMEWORKS BASED ON TOOLS, SERVICES AND DISEASES

SHILPA*, TARANDEEP KAUR†, RACHIT GARG‡, DEEPAK PRASHAR§, SUDAN JHA¶, SULTAN AHMAD‖ AND JABEEN NAZEER**

Abstract. With the advancements of the technologies, healthcare industry has become more digitized, data driven, patient-centric, innovative as well as collaborative. It has become possible to access and share patient information globally irrespective of time and locations. The traditional healthcare facilities have proved inadequate in facilitating better patient treatment and services as per the patient requirements. The modern healthcare provisioning involves the use of digital technologies and operations and manifest digitized healthcare facilities. The digital technologies such as cloud computing, Internet of Things and many other supports the facilitation of smart healthcare utilities and services. This paper describes how cloud computing and IoT deliver healthcare-as-a-service. Further, the various proposed healthcare frameworks integrated with healthcare field have been discussed such as based on technologies implemented, offered services, and focused disease(s). Also, this survey enlightens the comparisons of existing frameworks based on technologies used and based on the focused methodologies.

Key words: Cloud storage, Heart diseases, Personal Electronic Health Record, Robotics, Telehealth

1. Introduction. In this Contempo era, the healthcare industry is growing fast to provide better cure to the patients as well as health related improvement. From surveys, it has been observed that the world is drifting towards more digitization and technological involvement. According to several prior industrial reports it has been predicted that [1, 2] by 2020, the year will be the year of the technologist such as involvement of cloud in the healthcare industry.

The innovative and technological advancements in the information technology (IT) sector have led to adaptation of the technologies in the healthcare sector and considerable positive growth has also been observed that is considered as technology-enhanced healthcare (TEH). It has been expected and witnessed that there will be surprising growth in the use of advanced technologies by 2021 with 40% rate from $6.6 billion (Dh24.2 billion). It has also been predicted that that by 2025, there will be an unexpected growth in the IT-associated Extended Reality (XR) with the rate of $5.1 billion through the worldwide [3]. TEH is rapidly emerging as a prominent technological sector. The hospitals and certain medical centers are still following the comprehensive and traditional offline record maintaining procedures on a routine basis. With contrast to traditional healthcare provisioning, TEH is widely engaged in improving the capabilities and quality of healthcare services.

The TEH sector is witnessing a boost owing to the facility for proper utilization, management and collection of health information that plays a vital act in the detection of several medical problems [4] whilst identifying novel and innovative technological solutions for treating the patients.

*Lovely Professional University, Jalandhar, Punjab 144411, India
†Lovely Professional University, Jalandhar, Punjab 144411, India
‡School of Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab 144411, India
§School of Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab 144411, India
¶Department of Computer Science and Engineering, School of Engineering, Kathmandu University, Banepa, Kathmandu, Nepal; Lovely Professional University, Jalandhar, Punjab 144411, India
‖Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, P. O. Box. 151, Alkharj 11942, Saudi Arabia; University Center for Research and Development (UCRD), Department of Computer Science and Engineering, Chandigarh University, Gharuan, Mohali 140413, Punjab, India (Corresponding author, s.alisher@psau.edu.sa)
**Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, P. O. Box. 151, Alkharj 11942, Saudi Arabia

Traditional Healthcare system (THS) is categorized into two services such as personal and public healthcare services. Traditional Healthcare system also provides the services related to teaching and research. Personal Healthcare services are basically for hospitals, homes as well as organizations whereas public healthcare services offer guidelines for drugs, food and various safety policies etc. Research and teaching services are related to disease prevention, detection, tracking of disease and treatment.

The evolution of technologies has brought unbelievable innovation in the healthcare industry that provides the facilities from ground level to higher level, so healthcare has become important as it provides healthcare-as-a-service (HaaS). Enhanced technology also provides the facility to monitor and record the status of various patients' issues. Cloud computing as an emerging technology provides the platform for the effective use of resources in a secure manner to improve the quality, safety as well as efficiency of healthcare data [82]. As per the survey by Health Information and Management System Society (HIMSS) analytics and ISO/TS 18308 standards, the patient medical records are maintained electronically in the form of Electronic Medical Record (EMR), Electronic health record (EHR) and personal health record (PHR). EMR organizes the relevant information about the patient such as the patient's history, test results centrally for the purpose of further use.

1.1. Motivation for the Work. The paper presents a deep and extensive study about the significance of cloud computing and the Internet of Things (IoT) in the healthcare era. This survey highlights the role of IoT and cloud computing in the healthcare industry, focusing on the structure of IoT and cloud. The capabilities of these platforms in providing and supporting healthcare applications have been discussed. Additionally, the paper reviews different existing digital healthcare frameworks based on cloud, IoT, as well as CoT. The major contributions of the paper include:

1. In-depth analysis of the role of IoT and cloud computing in the field of healthcare.
2. Detailed description of IoT frameworks related to the healthcare industry, such as typologies, platforms, and structure.
3. Survey related to cloud computing frameworks, standards, and architecture integrated with the healthcare industry.
4. An overview of e-healthcare services and m-healthcare.

2. Research methodology. This Section highlights the objectives of the review article and discuss the factors that motivate for conducting this review article in depth and also elaborate the methodology used in detail. In this paper, an analysis of the advancements of the technologies such as cloud computing, Internet of things and Cloud of things in the healthcare sector has been discussed. A systematic approach has been followed to identify, evaluate and summarizing the survey related to HaaS. The Research methodology used in the review is as shown in the figure 2.1.

2.1. Review Plan. The review plan defines the steps involvement in the formulation of the research related queries, exploring of the relevant databases and identification of diverse resources such as springer, ScienceDirect, ACM Digital Library, IEEE Xplore and Google Scholar. Further refinement of the primary study is followed by the specific inclusion criteria for the final findings.

2.2. Review Sources and Distribution. This section signifies the findings related to the review analysis based on the taxonomy. The following figure 2.2 gives the distribution analysis of the different database sources.

3. Role of Cloud Computing and IoT Technologies in Delivering Healthcare. Technologies have greatly impacted the surroundings and have led to high-level digitization from the communication sector to the health sector as well as the entertainment sector. The latest technologies are used in various sectors such as education, the healthcare industry, organizations, and IT companies, etc. This section highlights the role of cloud computing, IoT, and Cloud of Things (CoT) in the healthcare sector.

3.1. Cloud Computing in Healthcare Sector. Cloud computing is considered as internet-based computing that helps to provide resources, software, as well as required information on-demand to computers. Cloud computing provides numerous cloud services such as database, storage, networking, and data processing through the internet instead of hard drives. All these services are also termed as "cloud," where cloud means the metaphor of the internet. In cloud computing, there is only a need to pay for the resources that are actually used, which is termed as a pay-per-use model [5, 6].

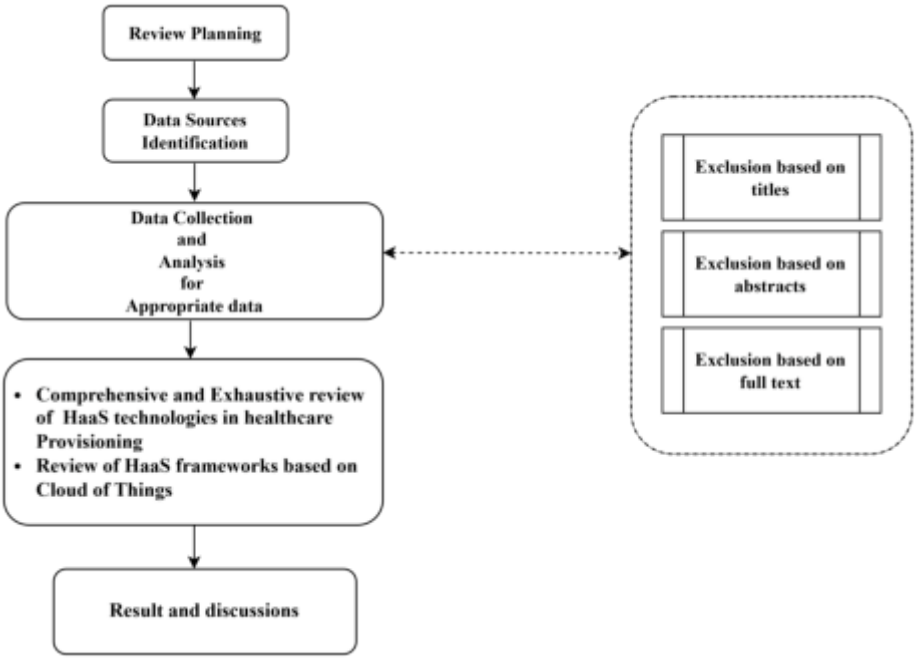


Fig. 2.1: Research Methodology

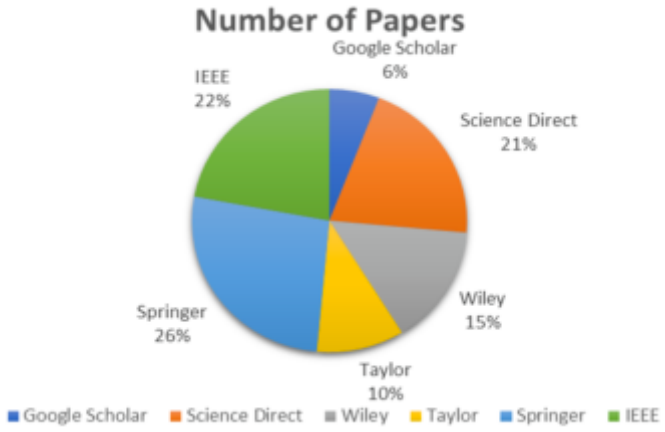


Fig. 2.2: Article Distribution based on scientific Sources

There are different service models which include Infrastructure as a Service (IaaS), Software as a Service (SaaS), Platform as a Service (PaaS), and Security as a Service (SecaaS), etc., and different delivery models such as public, private, and hybrid cloud models. In Software as a Service (SaaS), the cloud provides on-demand services for the healthcare industry such as quick access to business applications and customer relationship management (CRM). In IaaS, the cloud provides an on-demand storage facility for medical records, and lastly, as PaaS, it provides a secure environment for the deployment of web applications [7].

In the medical field, cloud computing is associated with remote servers for medical data storage, proper management, and processing. Cloud storage provides secure facilities for storage to healthcare professionals as

Table 3.1: Cloud-Based Tools in Healthcare Sector

Ref.No	Tools	Description
[10]	HIPAA-Compliant Cloud	HIPAA Cloud tool is becoming more prominent as service providers in the healthcare field, such as nimble and low-cost IT-related solutions and services. The tools integrated with HIPAA-Compliant Cloud provide benefits such as file-sharing, cost-saving facilities, expanded storage facilities, and dynamic as well as future-proof infrastructure creation.
[11]	NetApp	NetApp is a hybrid cloud service company that provides cloud-based services as a platform for healthcare management. This tool helps to provide real-time patient data for faster delivery and quick efficiency by creating backups, reducing EHR latency, data restoration, and easy access to data.
[12]	Google Cloud Healthcare API	This is a secure development environment that helps to build solutions for medical practices. It helps to make the data shareable between relevant applications on Google Cloud. This tool provides the facility of data processing, analytics, and a machine learning platform.
[12]	Medscape	Medscape is a cloud app used globally by professionals and physicians. It helps in delivering the latest information related to medical treatment and clinical innovations.
[12]	VisualDX	VisualDX is an innovative tool that improves the accuracy of diagnosis and patient care. It provides an alternative way for diagnosis, patient treatments, and helps in identifying the best treatment options.

well as medical sectors [8].

Subsequently, the cloud-based healthcare field has emerged with basic necessary requirements such as on-demand access for online storage, handling a huge amount of data in terms of EHR, radiology images related to patients, information related to patients' lifesaving, patients' treatment and diagnosis improvement, and proper analysis of patient data [9].

Cloud-based EHRs reduce the use of storage areas such as desktop computers, etc., and provide a centralized system that facilitates easier and more efficient use of records. Not only doctors but also patients can access their records for healthcare purposes. Secondly, EHR data analytics provides centralized access to the data and its functionality. EHR Analytics uses algorithms to provide effective quality to patient data.There are various existing Cloud-based tools as discussed in table 3.1 used in the healthcare sector to provide cloud-based services.

Summarily, the key highlights as inferred from Table 3.1 include:

- HIPAA-Compliant Cloud and NetApp are most relevant to Data Management. It gives you the protection of your data and expansionary storage for proper data management.
- Advanced Analytic Google Cloud Healthcare API is used that provides unique features for utilizing the concept of machine learning and data processing.
- Medscape Application help in sharing of knowledge and making sure that the Healthcare professionals know as well as updated on the latest trends and treatment in the medical field.
- VisualDX improve the accuracy and patient-centered decisions involved in or pertaining to a particular patient's care. Besides, promotes accurate and proper identification of a disease and the subsequent management of the same.

3.2. Internet of Things (IoT) in Healthcare Sector. The Internet of Things (IoT) includes the utilization of objects or "things" that are interconnected with the internet instead of computers. IoT is considered an innovative computing model in which real-world objects are integrated with the internet. These IoT objects are collaborated with sensors and innovative technologies. Such sensing objects have the capability of extracting and collecting data to perform analysis without human association to perform any function [16]. In actuality, the "things" in IoT can consist of multiple devices such as sensors forming a Sensor Network (SN), radio-frequency equipment, etc. In the IoT standard, SN constitutes an important part that interconnects the SN either using wired or wireless techniques [13].

In IoT, there exists an interconnection of billions of devices throughout the internet that are collecting and sharing data. Specifically, IoT is the connection of various devices ranging from sensors to smartphones and wearable devices. When these devices are connected with an automated system, it becomes easy to collect the required information, analyze it, and take action [6]. IoT as a technology provides better opportunities for building standards that are useful for exploring the work to be done, time efficiency, as well as money-saving.

Table 3.2: IoT Tools Used in Healthcare

Sr.No	Tools	Description
[18]	Pulse-Rate Sensor	Pulse-Rate sensor is used to know the heartbeat rate information. Another name for this sensor is heart rate sensor. The sensor helps to monitor patient health, track sleep, and monitor anxiety.
[19]	Blood Pressure Sensor	Blood Pressure sensor is used to measure the blood activity of patients, including systolic, diastolic, and mean arterial pressure levels using the oscillometer method.
[20]	Temperature Sensor	A sensor used to measure the temperature of the patient's body. Such sensors are also installed to monitor air sterility.
[21]	Sugar Level Sensor	A sensor used to measure glucose levels in the blood of patients to monitor diabetes type I or II.
[22]	Motion Sensor	A sensor used to measure the motion activity of the patient. It helps to identify physical movements or actions related to the patient to measure the BMI level.

IoT is not only a collection of things but also considered an ecosystem of interrelated resources and other technologies. Apart from interconnected devices, IoT defines resources such as hardware, software, connectivity, protocols, and middleware. It is characterized by seven prominent features: Connectivity, Things, Data, Communication, Intelligence, Action, and Ecosystem [14]. Connectivity provides the opportunity to bring the objects together, improves communication between connected devices, whilst intelligence saves time by transferring the data over the network as well as reducing human intervention. These all together in action maintain the ecosystem [15].

Predominantly, the healthcare industry is based on five major factors: electronic databases, doctors embracing mobility, testing and imaging, online database prediction, and errorless healthcare technology. Due to Technology-Enhanced Healthcare (TEH), major technological changes have been observed in the healthcare system. Electronic databases basically handle the patient health data obtained from sensors that help to generate the Electronic Health Record (EHR). IoT with EHR helps to enhance patient healthcare by providing services to monitor the patient's health anywhere and anytime [79]. The doctor embracing mobility is associated with the security of patient data as well as improvement in health data. The latest IoT technologies help with end-to-end connectivity and affordability for automating patient care with the help of mobile devices. Smartphones have become one of the ways to communicate with doctors either by voice calls or video calls throughout the world [80]. With IoT and medical imaging, it has become one of the best ways to represent patient records effectively, allowing doctors, caregivers, and staff to make appropriate decisions for the patients [81].

Data analytics and IoT for healthcare records act as a powerful tool in the transition of the healthcare system from profit to value-based care. The mobile medical devices using IoT sensors, such as wearable gadgets to examine diabetic patients and monitor their glucose levels, are gaining attention. For example, the Apple Watch can be considered a product of TEH and helps predict and measure the diet, fitness, and sleeping data of patients [4].

Compared to IoT, cloud computing is considered a computing technology with the capability of accessing a variety of resources and services. This technology has a feature of resource sharing through a virtualization process from anywhere. In addition to this, cloud computing has the ability to maintain physical resources.There are various existing IoT based tools as discussed in Table 3.2 used in the healthcare sector to provide the IoT-based services.

Summarily, the key highlights associated with various IoT tools in healthcare involve the following:

- Comprehensive Monitoring: These sensors use the holistic approach for monitoring the overall patient's health and the physiological parameters such as blood pressure, glucose levels, heart rate, and physical activity.
- Versatility: It is used in health care as wearables by the patients such as fitness trackers, CGMs and in highly technical hospitals as sterilization tools, fall detection devices etc.
- Personalized Care: These sensors help in the collection of real-time data from that is used for the proactive and personalized healthcare management It reduce the risk of chronic diseases.

3.3. Cloud of Things in Healthcare Sector. Cloud computing and IoT technology complement each other. On the flip side, cloud computing has the efficiency to address the limitations of IoT technology. To

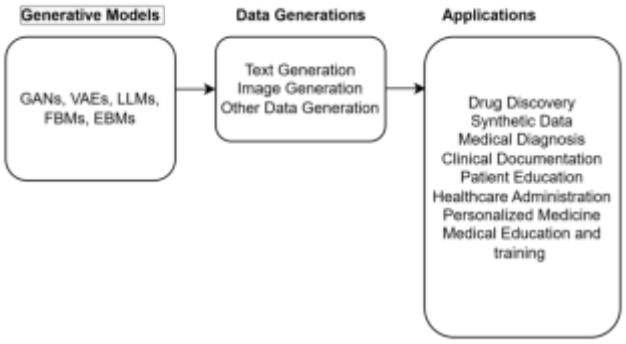


Fig. 3.1: Generative AI Applications

respond more effectively to the digital world, cloud computing and IoT are collaborated, which is referred to as Cloud of Things (CoT). With the help of IoT and Cloud, access to the medical field as well as data has become efficient at a global level and delivers better solutions based on real patient monitoring and health analysis [17].

Additionally, from a security perspective, public cloud in the healthcare field provides the facility of resource sharing, while IoT enables safe and efficient monitoring of patient health. The tools for CoT are collectively based on cloud and IoT tools and methodologies.

3.4. Generative AI in healthcare. Artificial intelligence (AI) has been emerged as transformative technology in healthcare sector that enhances the patient-care, provides effective disease diagnosis and treatment [84] [85]. Generative AI reduces the cost and time for the effective delivery of healthcare services and improve the patient care [86] [87] [88]. The applications of Generative AI are as shown in the figure 3.1.

Generative AI models such as Generative Adversarial Networks, Variational Autoencoders, Generative Pre-trained Transformer etc helps in the generation of the modalities such as Text, image and other data generation formats. This generated data used for various scenarios such as drug discovery, data synthetic, medical diagnosis and healthcare administration.

4. Research Contributions of the Work. The review focuses on the analysis and review of healthcare provisioning through cloud computing. It involves addressing various research questions as listed in Table 4.1.

4.1. Analysis and Comparison of Primary Studies Focusing on Review of HaaS. Various healthcare-related existing review papers and survey papers have been reviewed to analyze the comparison in different years. Table 4.2 lists the comparative review and survey papers in the healthcare sector associated with Cloud Computing, IoT, and CoT.

5. Existing Healthcare-as-a-Service (HaaS) Frameworks. Healthcare facilitated through technology is a novel notion that has emerged recently and is significantly rising. In recent years, extensive research has been carried out in this area. This section discusses the basic frameworks designed for providing HaaS, highlighting key findings and features of existing HaaS frameworks. A detailed and comprehensive survey has been conducted on IoT and cloud computing in the healthcare industry, exploring the standard structures of IoT and cloud computing, and the frameworks and platforms that facilitate communication.

5.1. Layered HaaS Provisioning Framework. The National Academy of Engineering (US) and the Institute of Medicine (US) Committee on Engineering and the Health Care System [30] have adopted a four-level model to understand the structure of the healthcare system. Figure 5.1 provides a hierarchical ordering that describes a four-level healthcare system.

The first level is patient-centric where individual patient needs and preferences are defined. At this level the responsibilities differ from patient to patient. The patient and his/her family members have access to educational, decision support, information management as well as tools that help to integrate the information

Table 4.1: Different Research Questions and Corresponding Motivations

Sr.No	Questions	Motivations	Sections Contribution
RQ1	What is the significance of healthcare delivery?	Healthcare delivery helps to understand the role of digital healthcare in the e-world. This research question studies the various technologies' roles in the healthcare sector such as cloud computing, IoT, and CoT.	Section 2
RQ2	What is healthcare as a service model?	The role of this question is to explore healthcare services such as SaaS, IaaS, and PaaS related to cloud computing.	Section 4
RQ3	What is the role of technology in facilitating Healthcare as a Service (HaaS)?	This question helps to analyze how various technologies are playing a role in the healthcare sector to provide better services to facilitate HaaS.	Subsection 2.1, 2.2, and 2.3
RQ4	How has ICT impacted healthcare delivery and provision?	This type of study carries out the use of ICT in healthcare delivery services and provides the provision of digital healthcare.	Section 2
RQ5	Why use Cloud computing and IoT for offering Healthcare as a Service?	This question defines how cloud computing and IoT are used to deliver healthcare-related services by proposing healthcare monitoring frameworks and models.	Section 4
RQ6	What is the need for HaaS?	This question defines the needs of HaaS in real-time scenarios with the involvement of various factors and tools.	Section 4
RQ7	What are the existing proposed frameworks?	It helps to perform an in-depth survey of proposed frameworks associated with the digital healthcare sector.	Subsection 4.3
RQ8	How does this contemporary healthcare system differ from the traditional healthcare system?	This question helps to identify how the healthcare sector has become more advanced compared to traditional healthcare.	Section 2 and Section 4

Table 4.2: Comparisons of Healthcare Related Existing Review Papers

Sr No	Ref.	Year	Review	Taxonomy	Cloud-Based	IoT-Based	Focus of Review
1	[23]	2012	✓		✓	✓	ECG
2	[29]	2014	✓			✓	Smart healthcare
3	[25]	2014	✓		✓		Healthcare facility such as storage
4	[26]	2014	✓	✓	✓		Cardiovascular disease
5	[27]	2014	✓	✓	✓		Healthcare industry such as data access and sharing
6	[24]	2021	✓		✓		Diabetics
7	[28]	2022	✓	✓	✓		Cardiovascular disease
8	[72]	2022	✓		✓	✓	Covid-19
9	[32]	2015	✓		✓		Security and Privacy of Patient data
10	[34]	2016	✓		✓		Healthcare Services such as storage, data access
11	[36]	2018	✓	✓		✓	Healthcare services such as storage, data access and privacy
12	[44]	2017	✓			✓	Smart Healthcare
13	[51]	2018	✓			✓	Healthcare facilities
14	[31]	2018	✓	✓		✓	Health Monitoring
15	[73]	2020	✓		✓		Cardiovascular disease
16	[74]	2022	✓		✓		Cardiovascular disease
17	[75]	2022	✓		✓	✓	Cardiovascular disease
18	[76]	2021	✓			✓	Cardiovascular disease
19	[78]	2022	✓		✓		Diabetic
20	[82]	2023			✓	✓	Diabetic

from different sources. The second level depicts care providers that directly impact each subsequent level such as the patients, organization, and environment. At the organizational level, the care providers help for the improvement in communication as well as facilities. At the environmental level, the care providers coordination leads with the benefits for insurance companies of both public as well as private sectors.

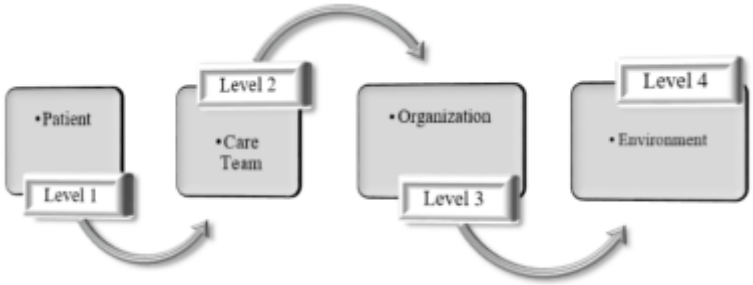


Fig. 5.1: 4-level Healthcare System

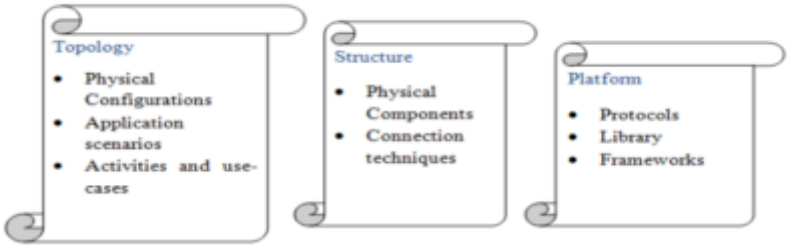


Fig. 5.2: Three Basic Components and their Main Functions in the IoT Framework for Healthcare.

5.2. General Framework for Offering Healthcare as a Service (HaaS). A framework provides protocols that facilitate the communication and broadcasting of fresh as well as raw medical data from smart devices and various sensors. Principally, IoT frameworks consist of three components: topology, structure, and platform, as shown in Figure 5.2. Each component of the framework has specific functions in the healthcare industry [31].

Figure 5.2 depicts the components with functionality related to the IoT frameworks used in the healthcare sectors. The first component is Topology that define the infrastructure related to the healthcare sector and defines physical configurations in the form of doctors, patients and protocols, activities such as EHR. Second component is Structure that defines the physical components related to the healthcare and connection techniques. Third component is healthcare frameworks.

6. Results and Discussions. This section depicts various types of HaaS frameworks deployed with the collaboration of innovative technologies. It elaborates on the classification and taxonomy of HaaS frameworks integrated with technologies such as cloud computing, IoT, etc. Additionally, the essential key aspects, such as methodologies, implementation environments, and resource utilization are discussed. The subsections provide insights into the classification and taxonomy of HaaS frameworks.

6.1. Classification and Taxonomy of HaaS Frameworks. HaaS frameworks have been categorized based on implemented technologies, delivered services, and focused diseases (as depicted in Figure 6.1).

6.1.1. Classification on the basis of Technologies Implemented. The healthcare field has become one of the innovative sectors due to the involvement of digital technologies. Various frameworks have been proposed using technologies such as cloud computing, Internet of Things (IoT), and Cloud of Things (CoT). The classification and taxonomy are illustrated in Figure 6.2.

I. **Healthcare as a Service (HaaS) Frameworks based on Cloud Computing**
The healthcare sector is boosting due to the integration of cloud based services. Various cloud-based frameworks have been proposed for better facilities in the healthcare sector [83]. This section highlights

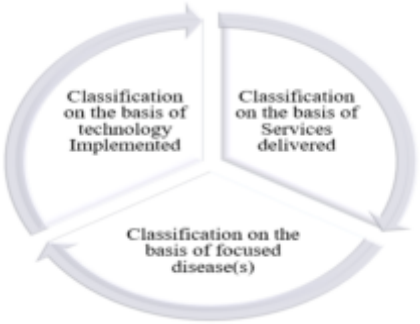


Fig. 6.1: Classification of HaaS Frameworks.

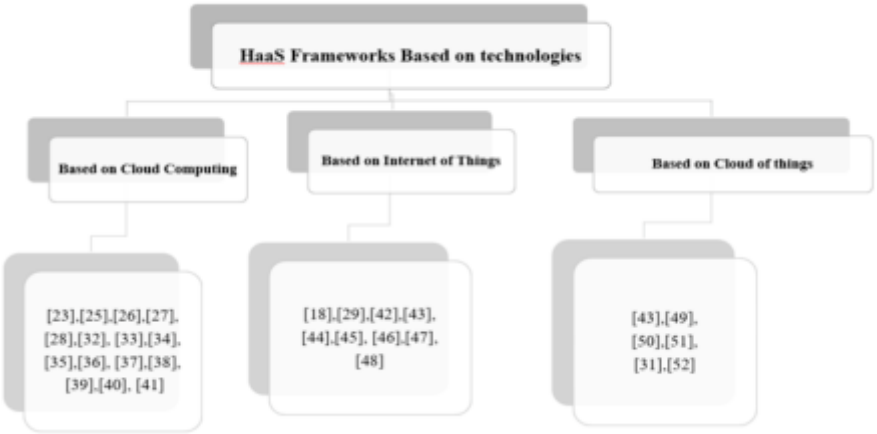


Fig. 6.2: Taxonomy of HaaS Frameworks based on Technologies.

the various existing HaaS frameworks based on cloud computing as shown in Table 6.1.

II. **Healthcare as a Service (HaaS) Frameworks based on IoT**
IoT has been broadly integrated in the healthcare field for real time health monitoring and report analysis. Various frameworks and models have been proposed for doctors, patients, and various health related fields. This section elaborates the frameworks in the Table 6.2.

III. **Healthcare as a Service (HaaS) Frameworks based on Cloud of Things (CoT)**
Cloud and IoT has been integrated as collaboratively in healthcare field for storage and health monitoring. Various advanced healthcare models have been proposed for providing such facilities. This section elaborates the various health related frameworks in Table 6.3.

6.1.2. Classification on the Basis of Services Offered. Healthtech or healthcare technologies defines the technologies used for the development of the healthcare system [53]. From telehealth to robotic-assisted surgery related services has brought a drastic change in the healthcare sector. Technology innovation advancement has delivered better services in digital format such as wearable devices, telehealth services, mhealth and so on. Table 6.4 describes the services provided by various technologies involved in the healthcare sector.

6.1.3. Classification on the basis of Focused Diseases. The effectiveness of the existing frameworks has been accessed to explore the specific disease or problem handled by them. Table 6.5 lists various existing technologies and algorithms presented in the frameworks that can handle various healthcare problems.

Table 6.1: Healthcare as a Service (HaaS) Framework Based on Cloud Computing

Sr. No	Reference	Implemented Environment	Platform	Description
1	[23]	Microsoft.NET framework	Aneka Platform	– The prototype system is developed for system design. – Stores the patient data in a cloud-based repository.
2	[25]	Matlab	iFit platform	– Cloud-based diagnosis system is proposed. – Determine the health issues based on the physiological data of the patient.
3	[26]	Java Android SDK, MySQL 5.5	AliveCor mobile app	– Identify the issues related to security and privacy in terms of cloud computing in the medical industry. – To ensure the patient about the health data shared with doctors are secure and confidential.
4	[33]	TCP/IP and Zigbee	Cloud-assisted Wireless Body Area Network (WBAN)	– Healthcare data sharing network is proposed. – Integration of Zigbee with TCP/IP for reliable interaction.
5	[34]	ICT	Cloud Network	– Cloud-based network is proposed. – Enable the communication and interaction between patients and doctors.
6	[35]	Cloud stack	SAMI healthcare platform of Samsung	– Mobile healthcare network is proposed. – Mobility cloud control software is proposed for the management of networks.
7	[36]	-	Gateways	– Indicates the opportunities and challenges of fog computing in the healthcare industry. – Proposed 3-layer architecture for the healthcare industry for the purpose of real-life applications.
8	[37]	C++	OpenMP	– Fog computing-based framework is proposed to give the response on patient mobile. – Apply the framework to reduce the response time of the prototype.
9	[38]	Cloud computing, health IoT, digital twin, and big data	CloudDTH platform	– Cloud healthcare framework is proposed based on digital twins. – Provide the services for the management as well as operations of the healthcare system.
10	[39]	C# programming and JSON	Relational Database Service and Amazon Web Service	– Privacy-preserving mobile healthcare framework is proposed based on the location. – Framework integrated with the capability of Context-based decision making.
11	[40]	Cloud Computing	SecHC framework	– Cloud-based Secure Healthcare Framework was proposed to provide the security and integrity features in healthcare and medical records. – Proposed framework modules are setup, key generation, encryption, and decryption.
12	[28]	CoreML package	Google Cloud Firebase	– HealthCloud system is proposed to monitor the health status of patients. – Machine learning algorithms such as Super-Vector classifier, KNN, Neural Network, Logical Regression, Gradient Boosting Trees are used.
13	[41]	Java and MATLAB	IoMT Cloud Environment	– Heart Disease prediction system is proposed using modified salp swarm optimization (MSSO) and an adaptive neuro-fuzzy inference system (ANFIS). – An accuracy of the prediction model is 99.45 with a precision of 96.54.
14	[73]	MATLAB	Supervised machine learning	– Heart disease is predicted using SVM and machine learning. – An accuracy of the model is 95.31%.
15	[74]	Python	– Random Forest – K-Neighbors classifier – Support vector machines – Decision Tree	– Healthcare applications are developed to detect heart diseases. – Diagnosis system is developed for heart disease.

Table 6.2: Healthcare as a Service (HaaS) Framework based on IoT

Sr. No	Reference	Implemented Environment	Platform	Description
1	[29]	Arduino	Ell-I arduino	– The proposed system is based on the IoT for the healthcare industry that is concerned with effective data transmission and IoT sensors with power. – Improves performance as well as energy efficiency.
2	[42]	PHP, MySQL	Clinical environment	– Smart gateway system is proposed for an e-health monitoring system. – Facilitates various services such as compression, storage, and security.
3	[43]	Microcontroller Unit	Mobile App	– Determine the short term as well as long term communication protocols for the purpose of communication between machines. – Identify the key findings and scope of IoT in the Medicare area and healthcare system.
4	[44]	Sensors	Machine Learning enabled cloud	– Determine the IoT basic elements in the healthcare system. – Focusing on the various types of sensors and communication methods.
5	[45]	XML	ATmega32U4 micro-controller	– Identify the consequences and future scope in the healthcare system. – Pointing to the security challenges in IoT devices as well as in networks.
6	[46]	MATLAB 2020a	FogBus, Healthcare Sensors	– A smart heart disease prediction system is proposed using IoT and Fog Computing. – Optimization algorithms are used to check the accuracy of the dataset.
7	[47]	MATLAB	Cuckoo enables classifier IoT based framework	– A heart disease prediction model is proposed using Cuckoo Search-based Conv LSTM Classifier. – Cuckoo-based deep ConvLSTM model is analyzed based on different evaluation metrics such as accuracy, specificity, and sensitivity.
8	[48]	Keras and Tensorflow	smart IoT-based framework	– Smart healthcare model is proposed using IoT with Cloud computing technology for diagnosis of heart failure. – Proposed model helps to investigate the classification of the heart diseases patient status such as alive or deceased.
9	[76]	Python language	IoT platform based on Machine Learning	– 4-tier CDPS-IoT model is proposed using machine learning for heart disease prediction. – Better accuracy is achieved with low error rates.

Table 6.3: Healthcare as a Service (HaaS) Frameworks based on Cloud of Things (CoT)

Sr. No	Reference	Implemented Environment	Platform	Description
1	[49]	XML	Machine-to-Machine platform	Proposed on Cloud-based IoT platform. Identify the efficiency of the framework based on factors such as efficiency and energy consumption.
2	[43]	Microcontroller Unit	Mobile App	IoT-based ECG monitoring system is proposed which consists of three main parts such as IoT, cloud, and GUI. This proposed system is reliable to collect real-time ECG data. Proposed system supports a high data rate and wide area.
3	[50]	Scyther	Healthcare System	A new healthcare system is proposed that contains a storage stage and data retrieving stage. Provides the facility of high data storage and security to the data.
4	[52]	Cloud of medical things	Cloud and IoT	The proposed system is the Cloud of Medical Things (CoMT) based Smart Healthcare Framework. Associated factors are resource Co-allocation, security, and energy efficiency.
5	[75]	Tensorflow, Apache Spark, and Cassandra	i2k2 Cloud platform	Risk of heart disease is predicted using Bi-LSTM. Alerts to notify the clinicians and caretakers of the patient's condition.

Effectiveness Analysis of Existing HaaS Frameworks. Various contextual frameworks identify the effectiveness of the frameworks in specific scenarios such as data complexity, real-time data analysis, and scalability. There are certain characteristics related to the frameworks that are as follows.

Table 6.4: Classification on the Basis of Services Offered

Sr No	References	Services	Description
1	[54]	EHR	EHR is a digital version that is a real-time and patient-centric record. EHR is available instantly at both global and local levels. EHR contains data related to patients' history to provide better treatment facilities and care. EHR helps to provide patient information to the hospital service providers for care.
2	[55]	Wearable Technology	Such technologies provide services to track daily activities such as steps taken, sleeping duration, heart rate, and other physical activities of the patient. Wearable technologies also help in health monitoring by observing blood sugar and oxygen levels in patients.
3	[56]	mHealth	mHealth is a mobile application used for health monitoring, digital assistance, and real-time disease analysis. mHealth service is used for phone calls with health professionals.
4	[57]	Telehealth	Telehealth provides digital ICT facilities for remote and global-level services for health care. This healthcare service provides various facilities such as patient safety, primary care, easy accessibility to health-related specialists, and improvement in communication, etc.

- Existing frameworks impacting the data quality as well as volume on the frameworks such as Machine learning-based proposed CVD-prediction frameworks utilize the cloud-based platform for the data storage more effectively and efficiently.
- Proposed frameworks are utilized for specific needs such as CVD predictions, Diabetes monitoring and Liver prediction.
- ML based algorithms are effective for patient-health monitoring and hospital-centric frameworks are focusing on patient-centric services such as EHR and high computational energy.

6.2. Technical and logistical challenges in Cloud and IoT-based Healthcare Frameworks. An amalgamation of the innovative technologies such as cloud computing and Internet of Things has empirical impact in the healthcare sector that enhance the patient-care and provide the real-time health monitoring. But, involvement of such frameworks has several technical and logistical challenges that are discussed in Table 6.6.

7. Conclusion. With the advancement in the IT sector, the healthcare field has become one of the curious fields at the global level. Various IT technologies such as cloud computing, big data, edge computing, and fog computing are playing significant roles in facilitating technology-enhanced healthcare or Healthcare-as-a-service. The paper highlights a detailed study about the role of ICT technologies such as cloud computing and IoT in delivering Healthcare. A comparative analysis of the existing technological involvement in the healthcare sector has been discussed towards the digital sector. The paper lists the various proposed frameworks pertaining to technologies such as cloud computing, IoT, and cloud-of-things. It also presents related taxonomies for Healthcare-as-a-Service based on diseases, services, and technologies. While various frameworks based on cloud computing, IoT and CoT has been proposed but also there is need for the involvement of the advanced and innovation technologies such as generative AI. Future study can analyze the impact of the advanced and informative technologies in the healthcare sector.

Acknowledgment. We thank the Deanship of Scientific Research, Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia for help and support. This study is supported via funding from Prince Sattam Bin Abdulaziz University project number (PSAU/2025/R/1446).

Funding. This study is supported via funding from Prince Sattam Bin Abdulaziz University project number (PSAU/2025/R/1446).

Author Contributions. All authors contributed to design and development of the system as well as the manuscript. All authors have read and approved the final manuscript.

Data availability. Data sharing not applicable to this article as no datasets were generated during the current study.

Table 6.5: Classification Based on Focused Diseases

Sr.No.	Reference	Focused Disease(s)	Techniques/Algorithm Used
1	[58]	– Heart Diseases – Breast Cancer – Diabetes – Spect_heart – Thyroid – Dermatology – Liver disorders	– K-NN – Support Vector Machine – Decision Trees – Random Forest – MLP
2	[59]	Breast Cancer	– Basic classifier – Ensemble method – Cost-sensitive method
3	[60]	Kidney	– Naïve Bayes – Support Vector Machine (SVM)
4	[61]	Thyroid	– NB tree – J48 – LADTree – BFTree – LMT – Random Forest
5	[62]	Diabetes	Enhanced class outlier
6	[63]	Diabetes	SVM, K-means
7	[64]	Cancer	SVM
8	[65]	Cardiovascular disease	Ensemble deep learning
9	[66]	Diabetes mellitus	K-means clustering
10	[67]	Cardiovascular disease	Naive Bayes
11	[68]	Chronic diseases such as diabetes, blood pressure, cardiac arrhythmia	IoT
12	[69]	Cardiovascular disease	– Deep learning – Machine learning
13	[70]	COVID-19	– Cloud Computing – IoT
14	[71]	Cardiovascular diseases	– A stacked ensemble classifier – Machine Learning algorithms such as Random Forest and XGBoost
15	[73]	Cardiovascular disease	Machine Learning
16	[74]	Cardiovascular disease	– Random Forest – K-Neighbors classifier – Support Vector Machines – Decision Tree
17	[75]	Cardiovascular disease	– RNN – Long short-term memory
18	[76]	Cardiovascular Disease	– Random Forest – Decision Tree – Naïve Bayes – K-nearest neighbors – Support Vector Machine
19	[78]	Diabetic	– KNN – SVM – ANN – Deep Neural Network
20	[82]	Diabetic	– Random Forest – Logistic regression

REFERENCES

[1] Bello, O., & Zeadally, S. (2014). Intelligent device-to-device communication in the internet of things. *IEEE Systems Journal*, **10**(3), 1172-1182.

[2] Healthcare in 2020: the era of digitization. Retrieved from `https://health.economictimes.indiatimes.com/news/health-it/healthcare-in-2020-the-era-of-digitization/73045907`

Table 6.6: Technical and Logistical Challenges in Cloud and IoT-based Healthcare Frameworks

Sr.No	Challenges	Description
1	Technical Challenges	– Data Security and Privacy: Attackers make the target mostly Cloud-based systems that expose the patient health records such as PHR or EHR for unauthorized activities. – Insufficient and Expensive Data Security
2	Logistical Challenges	– Infrastructure Gaps: Not accurate deployment of Infrastructure – High Initial Cost

[3] New technologies to transform healthcare in 2020. Retrieved from https://gulfnews.com/technology/ new-technologies-to-transform-healthcare-in-2020-1.1576486067662

[4] Yang, J.-J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., Wang, Q., & Pan, H. (2015). Emerging information technologies for enhanced healthcare. *Computers in Industry*, **69**, 3-11.

[5] Ramgovind, S., Eloff, M. M., & Smith, E. (2010, August). The management of security in cloud computing. In *2010 Information Security for South Africa* (pp. 1-7). IEEE.

[6] Al-Issa, Y., Ottom, M. A., & Tamrawi, A. (2019). eHealth Cloud Security Challenges: A Survey. *Journal of Healthcare Engineering*, 2019.

[7] Healthcare and Cloud Computing. Retrieved from https://www.innovativearchitects.com/KnowledgeCenter/industry-specific/healthcare-and-cloud-computing.aspx

[8] Cloud Computing in Healthcare. Retrieved from https://www.antino.io/blog/cloud-computing-in-healthcare/

[9] Cloud Computing in Healthcare. Retrieved from https://www.leewayhertz.com/cloud-computing-in-healthcare/#hccb

[10] Understanding HIPAA Compliant Cloud Options for Health IT. Retrieved from https://hitinfrastructure.com/features/understanding-hipaa-compliant-cloud-options-for-health-it

[11] Importance of Cloud Computing and IoT in Healthcare and Life Sciences. Retrieved from https://quokkalabs.com/blog/importance-of-cloud-computing-and-iot-in-healthcare-and-life-sciences/

[12] Healthcare Cloud Tools. Retrieved from https://www.patientcalls.com/blog/healthcare-cloud-tools/

[13] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys and Tutorials*, **17**(4), 2347–2376.

[14] Cvitić, I., & Vujić, M. (2015). CLASSIFICATION OF SECURITY RISKS IN THE IOT ENVIRONMENT. *Annals of DAAAM & Proceedings*, **26**(1).

[15] Islam, S. M. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K.-S. (2015). The internet of things for health care: a comprehensive survey. *IEEE Access*, **3**, 678–708.

[16] Khodadadi, F., A.V., D., & R., B. (2016). Internet of things principles and paradigms. Elsevier.

[17] How Cloud with IoT is Helping Healthcare Sector. Retrieved from https://www.jainuniversity.ac.in/blogs/how-cloud-with-iot-is-helping-healthcare-sector

[18] Pulse Sensor. Retrieved from https://www.rohm.com/electronics-basics/sensor/pulse-sensor

[19] Blood Pressure Sensor: Working & Applications. Retrieved from https://www.elprocus.com/blood-pressure-sensor-working-applications/

[20] Body Temperature Measurement. Retrieved from https://www.te.com/usa-en/trends/connected-life-health-tech/medical-sensor-technology-and-applications/body-temperature-measurement.html: :text=Infrared (IR) temperature sensors enable,forehead temperature or skin temperature.

[21] Glucose Continuous Glucose Monitoring. Retrieved from https://my.clevelandclinic.org/health/drugs/11444-glucose-continuous-glucose-monitoring

[22] Sensors Facilitate Health Monitoring. Retrieved from https://www.fierceelectronics.com/components/sensors-facilitate-health-monitoring

[23] Pandey, S., Voorsluys, W., Niu, S., Doker, A., & Buyya, R. (2012). An autonomic cloud environment for hosting ECG data analysis services. *Future Generation Computer Systems*, **28**, 147–154.

[24] Renugadevi, G., Sasi Kala Rani, K., Gowthamani, R., & Saranya, N. (2022). Review of Mobile Cloud Computing in Healthcare for Diabetics Patients Using Machine Learning Techniques. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, **9**(XII), 286-290.

[25] Tseng, K. C., & Wu, C. C. (2014). An Expert Fitness Diagnosis System Based on Elastic Cloud Computing. *The Scientific World Journal*. Retrieved from http://dx.doi.org/10.1155/2014/981207

[26] Thilakanathan, D., Chen, S., Nepal, S., Calvo, R., & Alem, L. (2014). A platform for secure monitoring and sharing of generic health data in the Cloud. *Future Generation Computer Systems*, **35**, 102–113. doi:10.1016/j.future.2013.09.011

[27] Abbas, A., & Khan, S. U. (2014). A Review on the State-of-the-Art Privacy Preserving Approaches in the e-Health Clouds. *IEEE Journal of Biomedical and Health Informatics*, **18**(4), 1431–1441.

[28] Desai, F., Chowdhury, D., Kaur, R., Peeters, M., Arya, R. C., Wander, G. S., & Buyya, R. (2022). HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing. *Internet of Things*, **17**, 100485.

[29] Granados, J., Rahmani, A. M., Nikander, P., Liljeberg, P., & Tenhunen, H. (2014, November). Towards energy-efficient HealthCare: An Internet-of-Things architecture using intelligent gateways. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)* (pp. 279-282). IEEE.

[30] National Academy of Engineering (US) and Institute of Medicine (US) Committee on Engineering and the Health Care System; Reid, P. P., Compton, W. D., Grossman, J. H., et al., editors. (2005). *Building a Better Delivery System: A New Engineering/Health Care Partnership*. Washington (DC): National Academies Press (US). Available from: https://www.ncbi.nlm.nih.gov/books/NBK22878/

[31] Mutlag, A. A., Ghani, M. K. A., Arunkumar, N., Mohamed, M. A., & Mohd, O. (2019). Enabling technologies for fog computing in healthcare IoT systems. *Future Generation Computer Systems*, **90**, 62–78.

[32] Calabrese, B., & Cannataro, M. (2015). Cloud computing in healthcare and biomedicine. *Scalable Computing: Practice and Experience*, **16**(1), 1-18.

[33] Hassan, M. M., Lin, K., Yue, X., & Wan, J. (2016). A multimedia healthcare data sharing approach through cloud-based body area network. *Future Generation Computer Systems*, **66**, 48–58. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X15004070

[34] Miah, S. J., Hasan, J., & Gammack, J. G. (2017). On-cloud healthcare clinic: An e-health consultancy approach for remote communities in a developing country. *Telematics and Informatics*, **34**(1), 311–322. Retrieved from http://www.sciencedirect.com/science/article/pii/S0736585315300897

[35] Chung, K., & Park, R. C. (2017). Cloud based u-healthcare network with QoS guarantee for mobile health service. *Cluster Computing*, Oct. Retrieved from https://doi.org/10.1007/s10586-017-1120-0

[36] Kumari, A., Tanwar, S., Tyagi, S., & Kumar, N. (2018). Fog computing for Healthcare 4.0 environment: Opportunities and challenges. *Computers and Electrical Engineering*, **72**, 1–13.

[37] García-Valls, M., Calva-Urrego, C., & García-Fornes, A. (2018). Accelerating smart eHealth services execution at the fog computing infrastructure. *Future Generation Computer Systems*. doi:10.1016/j.future.2018.07.001

[38] Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., ... & Deen, M. J. (2019). A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access*, **7**, 49088-49101.

[39] Abdo, M. A., Abdel-Hamid, A. A., & Elzouka, H. A. (2020). A Cloud-based Mobile Healthcare Monitoring Framework with Location Privacy Preservation. In *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*. doi:10.1109/3ict51146.2020.931199

[40] Satar, S. D. M., Hussin, M., Hanapi, Z. M., & Mohamed, M. A. (2021). Cloud-Based Secure Healthcare Framework by using Enhanced Ciphertext Policy Attribute-Based Encryption Scheme. *International Journal of Advanced Computer Science and Applications*, **12**, 393-399.

[41] Khan, M. A., & Algarni, F. (2020). A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. *IEEE Access*, **8**, 122259-122269.

[42] Rahmani, A. M., et al. (2018). Exploiting smart e-health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. *Future Generation Computer Systems*, **78**, 641–658. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X17302121

[43] Yang, Z., Zhou, Q., Lei, L., Zheng, K., & Xiang, W. (2016). An IoT-cloud based wearable ECG monitoring system for smart healthcare. *Journal of Medical Systems*, **40**(12), 286.

[44] Baker, S. B., Xiang, W., & Atkinson, I. (2017). Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities. *IEEE Access*, **5**, 26521–26544.

[45] Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., & Mankodiya, K. (2018). Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare. *Future Generation Computer Systems*, **78**, 659–676.

[46] Butchi Raju, K., Dara, S., Vidyarthi, A., Gupta, V. M. N. S. S. V. K. R., & Khan, B. (2022). Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model. *Computational Intelligence and Neuroscience*, **2022**, Article ID 1070697. Retrieved from https://doi.org/10.1155/2022/1070697

[47] Raju, K. B., Dara, S., Vidyarthi, A., Gupta, V. M. N. S. S. V. K. R., & Khan, B. (2022). Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model. *Computational Intelligence and Neuroscience*, **2022**, Article ID 1070697. Retrieved from https://doi.org/10.1155/2022/1070697

[48] Kumar, A., Reddy, S. S., Baig, M., Khan, B., & Sharma, R. (2022). Smart Healthcare: Disease Prediction Using the Cuckoo-Enabled Deep Classifier in IoT Framework. *Scientific Programming*, **2022**, Article ID 2090681. Retrieved from https://doi.org/10.1155/2022/2090681

[49] Umer, M., Sadiq, S., Karamti, H., Karamti, W., Majeed, R., & Nappi, M. (2022). IoT Based Smart Monitoring of Patients' with Acute Heart Failure. *Sensors*, **22**(7), 2431.

[50] Papageorgiou, A., Zahn, M., & Kovacs, E. (2014). Efficient auto-configuration of energy-related parameters in cloud-based IoT platforms. In *Proc. IEEE 3rd International Conference on Cloud Networking (CloudNet)* (pp. 236-241), Oct.

[51] Alkeem, E. A., Shehada, D., Yeun, C. Y., et al. (2017). New secure healthcare system using cloud of things. *Cluster Computing*, **20**, 2211–2229. Retrieved from https://doi.org/10.1007/s10586-017-0872-x

[52] Ahmadi, H., Arji, G., Shahmoradi, L., Safdari, R., Nilashi, M., & Alizadeh, M. (2018). The application of internet of things in healthcare: A systematic literature review and classification. *Universal Access in the Information Society*, 1–33. doi:10.1007/s10209-018-0618-4

[53] Surendran, R., & Tamilvizhi, T. (2020). Cloud of medical things (CoMT) based smart healthcare framework for resource allocation. In *3rd Smart Cities Symposium (SCS 2020)* (pp. 29-34). doi:10.1049/icp.2021.0855

[54] Jha, S., Seo, C., Yang, E., et al. (2021). Real time object detection and tracking system for video surveillance system. *Multimedia Tools and Applications*, **80**, 3981–3996. Retrieved from https://doi.org/10.1007/s11042-020-09749-x

[55] Technology in Healthcare Transformation. Retrieved from https://www.rasmussen.edu/degrees/health-sciences/blog/technology-in-healthcare-transformation/

[56] What is Wearable Tech? Retrieved from https://www.wareable.com/wearable-tech/what-is-wearable-tech-753

[57] Jha, S., Prashar, D., Long, H. V., & Taniar, D. (2020). Recurrent neural network for detecting malware. *Computers &*

Security, **99**, 102037.

[58] Telehealth. Retrieved from https://www.mayoclinic.org/healthy-lifestyle/consumer-health/in-depth/telehealth/art-20044878

[59] Kaur, P., Kumar, R., & Kumar, M. (2019). A healthcare monitoring system using random forest and internet of things (IoT). *Multimedia Tools and Applications*, **78**(14), 19905-19916.

[60] Hsu, J. L., Hung, P. C., Lin, H. Y., & Hsieh, C. H. (2015). Applying under-sampling techniques and cost-sensitive learning methods on risk assessment of breast cancer. *Journal of Medical Systems*, **39**(4). Retrieved from https://doi.org/10.1007/s10916-015-0210-x

[61] Vijayarani, S., & Dhayanand, S. (2015). Data mining classification algorithms for kidney diseases prediction. *International Journal on Cybernetics & Informatics*, **4**(4), 13–25.

[62] Turanoglu-Bekar, E., Ulutagay, G., & Kantarc-Savas, S. (2016). Classification of thyroid disease by using data mining models: A comparison of decision tree algorithm. *Oxford Journal of Intelligent Decision and Data Science*, **2**, 13–28.

[63] Jahangir, M., Afzal, H., Ahmed, M., Khurshid, K., & Nawaz, R. (2017). An expert system for diabetes prediction using auto tuned multi-layer perceptron. In *Proceedings of the Intelligent System Conference* (pp. 722–728).

[64] Osman, A. H., & Aljahdali, H. M. (2017). Diabetes disease diagnosis method based on feature extraction using ksvm. *International Journal of Advanced Computer Science and Applications*, **8**(1), 236–244.

[65] Zhang, L., Zhou, W., Wang, B., Zhang, Z., & Li, F. (2018). Applying 1-norm svm with squared loss to gene selection for cancer classification. *Applied Intelligence*, **48**(7), 1878–1890.

[66] Jha, S., Nkenyereye, L., Joshi, G. P., & Yang, E. (2020). Mitigating and monitoring smart city using internet of things. *Computers, Materials & Continua*, **65**(2), 1059-1079.

[67] Barik, R. K., Priyadarshini, R., Dubey, H., Kumar, V., & Mankodiya, K. (2018). FogLearn: Leveraging fog-based machine learning for smart system big data analytics. *International Journal of Fog Computing (IJFC)*, **1**(1), 15-34.

[68] Gupta, N., Ahuja, N., Malhotra, S., Bala, A., & Kaur, G. (2017). Intelligent heart disease prediction in cloud environment through ensembling. *Expert Systems*, **34**.

[69] Gómez, J. E., Oviedo, B., & Zhuma, E. (2016, December). Patient monitoring system based on Internet of Things. In *ANT/SEIT* (pp. 90-97).

[70] Elwahsh, H., El-Shafeiy, E., Alanazi, S., & Tawfeek, M. A. (2021). A new smart healthcare framework for real-time heart disease detection based on deep and machine learning. *PeerJ Computer Science*, **7**, e646. doi: 10.7717/peerj-cs.646

[71] Nasser, N., Emad-Ul-Haq, Q., Imran, M., Ali, A., Razzak, I., & Al-Helali, A. (2021). A smart healthcare framework for detection and monitoring of COVID-19 using IoT and cloud computing. *Neural Computing and Applications*. doi: 10.1007/s00521-021-06396-7

[72] Tiwari, A., Chugh, A., & Sharma, A. (2022). Ensemble framework for cardiovascular disease prediction. *Computers in Biology and Medicine*, **105624**.

[73] Darbandi, M., Alrasheedi, A. F., Alnowibet, K. A., et al. (2022). Integration of cloud computing with the Internet of things for the treatment and management of the COVID-19 pandemic. *Information Systems and E-Business Management*. Retrieved from https://doi.org/10.1007/s10257-022-00580-5

[74] Khan, M. A., Abbas, S., Atta, A., Ditta, A., Alquhayz, H., Khan, M. F., & Naqvi, R. A. (2020). Intelligent cloud based heart disease prediction system empowered with supervised machine learning.

[75] Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, **2**, 100016.

[76] Nancy, A. A., Ravindran, D., Raj Vincent, P. M. D., Srinivasan, K., & Gutierrez Reina, D. (2022). IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning. *Electronics*, **11**(15), 2292. Retrieved from https://doi.org/10.3390/electronics11152292

[77] Ahamed, J., Koli, A. M., Ahmad, K., Jamal, A., & Gupta, B. B. (2022). CDPS-IoT: Cardiovascular disease prediction system based on IoT using machine learning.

[78] Tarawneh, M., AlZyoud, F., AlSharrab, Y., & Khatatneh, K. (2022, November). Cloud-based Diabetic Prediction Framework: Deep Learning Approach. In *2022 International Conference on Emerging Trends in Computing and Engineering Applications (ETCEA)* (pp. 1-6). IEEE.

[79] Alamri, A. (2018). Ontology middleware for integration of IoT healthcare information systems in EHR systems. *Computers*, **7**(4), 51.

[80] Solutelabs. (2020). IoT in healthcare. Retrieved from https://www.solutelabs.com/blog/iot-in-healthcare

[81] Chandy, A. (2019). A review on IoT based medical imaging technology for healthcare applications. *Journal of Innovative Image Processing (JIIP)*, **1**(01), 51-60.

[82] Shilpa, Kaur, T. (2022). Digital Healthcare: Current Trends, Challenges and Future Perspectives. In Arai, K. (Ed.), *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 2* (pp. 123-133). Springer. Retrieved from https://doi.org/10.1007/978-3-030-89880-9_48

[83] Shilpa, Kaur, T. (2022). Blockchain and Cloud Technology: Leading the ICT Innovations. In Tuba, M., Akashe, S., & Joshi, A. (Eds.), *ICT Systems and Sustainability* (pp. 345-355). Springer. Retrieved from https://doi.org/10.1007/978-981-16-5987-4_41

[84] Kothari AN. (2023). large language models, and generative ai as future augments of surgical cancer care. *Ann Surg Oncol*,30(6):3174–6.

[85] Reddy, S. (2024). Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Sci* 19, 27. retrieved from https://doi.org/10.1186/s13012-024-01357-9

[86] Uprety D, Zhu D, West HJ. (2023) ChatGPT-a promising generative AI tool and its implications for cancer care. *Cancer*;129(15):2284–9.

[87] Haupt CE, Marks M.(2023) AI-generated medical advice-GPT and beyond. *JAMA*.;329(16):1349–50.
[88] Korngiebel DM, Mooney SD.(2021) Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *NPJ Digit Med*.4:93.

Edited by: Prakash Mohan
Special issue on: Advancing Healthcare through Scalable Machine Learning: Overcoming Challenges and Embracing Innovations
Received: Sep 3, 2024
Accepted: Feb 8, 2025


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2231–2243, DOI 10.12694/scpe.v26i5.4738

ONLINE EDUCATION STUDENT COGNITIVE STATE RECOGNITION BASED ON IMPROVED MULTI-TASK CONVOLUTIONAL NEURAL NETWORK

WEIJUAN AN *, LI SHEN† AND YALI YUAN‡

Abstract. With the widespread application and development of internet technology in public education scenarios in China, the application of deep learning technology on online learning platforms is becoming increasingly widespread. This study aims to address the difficulty in determining students' cognitive state under the current network learning mode, and proposes a face recognition algorithm for student cognitive state detection using multitask convolutional neural network image recognition technology. At the same time, research is conducted on extracting two-dimensional feature points of facial color images through cascading regression tree localization methods. In the practical application experiments of the method, the research method can effectively detect images with facial offset angles greater than 15° for students, and the cognitive state of online learning can be analyzed from the frequency detection of students' blinking and yawning. From the results of image simulation experiments, it can be seen that this study proposes a cascaded regression tree localization optimization multitask convolutional neural network face recognition method, which has the highest image recognition accuracy of 86%, a recall rate of 0.85, and an f1 value of 0.855. The experimental results show that online learning state recognition based on image analysis can effectively monitor abnormal states during students' learning process, improve students' online learning efficiency, and provide necessary student state information support for teachers, promoting the improvement of online teaching quality.

Key words: Multi-task convolutional neural network; Cascaded regression tree; Network teaching; Cognitive load; Teaching link

1. Introduction. The profound impact of the epidemic on social, economic and cultural aspects has also had an impact on the education cause. In the past few years, online teaching situations have been less applied and gradually normalized. However, it is difficult to identify and improve students' learning status through face-to-face supervision and communication. Therefore, scholars at home and abroad have begun to apply machine algorithms and video image recognition technology in online teaching situations [1]. Based on this, the research uses the Multi-task convolutional neural network (MTCNN) algorithm to detect and recognize the face state of students. In this study, a depth-separable convolution structure is introduced for the case that the parameters of the traditional convolution layer are too large, and the median filtering method is used to filter the image [2]. In the recognition of students' cognitive load state, the research analyzes the degree of deviation of the face, the blinking frequency of the students' eyes, and the degree of opening and closing of the mouth. At the same time, the Ensemble of Regression Trees (ERT) algorithm is used to locate the key feature points of the face [3]. In the feature extraction stage, in order to integrate the camera applications in different scenarios in network teaching, the RGB-D image acquisition method of the depth camera is used to analyze the three-dimensional pixel relationship between the depth image and the color image. Different from the current conventional online video communication technology in facial expression recognition, in order to increase the adaptability of image recognition technology to classroom education, research has added indicators such as facial offset and blink frequency to image expression recognition to identify students' classroom learning situation [4]. The method proposed in the study is also different from traditional resource recommendation and learning needs identification models that are built around user data. This method is basically applied to learning scenarios in classroom teaching [5]. The innovation of this experiment lies in the combination of image recognition technology and the characteristics of students' cognitive state, so as to analyze the students' learning cognitive state. It is expected that the method proposed by the research can effectively improve the cognitive

*Shijiazhuang Information Engineering Vocational College, China (anweijuan80@126.com)
†Shijiazhuang Information Engineering Vocational College, China (Corresponding author, anweijuan80@126.com)
‡Shijiazhuang Information Engineering Vocational College, China (yuanyali2023@126.com)

load state of students and optimize the cognitive behavior of online teaching scientifically and effectively.

2. Literature review. In the process of technology driving the development of education, CNN methods that have been applied to image and video monitoring have gradually become more abundant. Minghui et al. used convolutional neural networks to perform privacy and confidentiality work on data communication in outsourcing environments, and optimized the pooling layer by linear averaging. Experiments show that this method can not only improve the information security of outsourced data, but also reduce the risk of privacy leakage [6]. After determining the parameter structure of the convolution layer, pooling layer and classification layer of the CNN network, Atik, I applied it to the classification of electronic components. After the simulation experiment, the research shows that the method has the highest performance of electronic component classification accuracy, and its accuracy rate is 98.99% [7]. Kiki et al. constructed a neurobiologically inspired CNN model and applied the model to the auditory spatial localization of human voice. Experiments show that this method is an effective method for human spatial hearing [8]. El-Shafai et al. proposed a spectrum sensing model for colleges and universities based on CNN network. Compared with the traditional SS method, the optimized spectrum sensing model improves the detection accuracy by up to 17%, and reduces the sensing time by 16.6ms [9]. Thanammal et al. designed a CNN network model based on cross-wind-driven optimization, and applied the model to the detection of tomato pests. Experiments show that the recognition accuracy of this method is 99.86%, and the iteration time is only 12.3s. The data proves that this method can help tomato agricultural production and planting for real-time detection of pests and diseases [10]. Ben et al. proposed a randomly initialized network architecture of RND-CNN and applied the method to detect COVID-19 in chest X-ray images. The experimental results show that the detection accuracy of the optimized method in the COVIDx dataset reaches 94% [11]. Rahimilarki et al. combined time series analysis technology with CNN network and applied it to fault identification and detection of wind turbines. Experiments show that the method is used in simulation experiments, and its fault detection classification accuracy is higher than 95%, and the detection performance is stable [12].

Aiming at the problems of low resolution and poor sharpness of lens less cameras, Zhang et al. proposed an image reconstruction technique that fuses text proposal network and Convolutional Recurrent Neural Network(CRNN). Experiments have confirmed its applicability [13]. Fernandes et al. proposed a variable convolution backbone-based cascaded architecture to address text extraction difficulties in electronic documents. Experiments show that the extraction accuracy of the method in the two datasets is 99.3% and 95.1% respectively [14]. Shuai TENG et al. proposed a method based on CNN image detection for the structural stability of steel frame buildings. By fusing feature extraction and classification blocks into an intelligent learning system, the structural state of steel frames is detected. Experiments show that the classification accuracy of this method is as high as 99%, while the time used is only 19% of the traditional BP method [15]. MohsinNadia combines Canny and Prewitt edge detection techniques to construct a method for improving image embedding capacity in edge regions. Experiments show that this method has a high embedding capacity, which means that the imperceptibility of steganographic images is maintained [16]. Kamil et al. proposed a face detection method for attendance scenarios, which utilizes SVM algorithm and OpenCV software for image processing. In the experiment, this method can achieve an accuracy of about 81.8%. Compared to the proposed method, the research constructed method can achieve an accuracy of approximately 86% [17]. Andrejevic et al. applied facial recognition technology in campus environments, with the main purpose of addressing issues such as offline campus safety, automatic registration, and student emotion detection. This study mainly considers the feasibility and reasons for the application of technology. The research and construction methods are mainly used for student state identification in online classrooms [18].

To sum up, the CNN network is widely used in image recognition, indicating that the image detection technology has become mature. However, it is rare to combine image detection technology with cognitive state detection of students. Therefore, the research will use the optimized MTCNN image recognition technology to locate and feature extraction of key points such as students' faces, mouths, and eyes to distinguish students' cognitive behaviors and cognitive load states. The innovative contribution of the research lies in the analysis of details such as changes in students' eyes, mouth, and head displacement through image detection technology, and the construction of a detection system for cognitive fatigue in online education for students. The system uses mouth opening and closing, blink frequency, and head offset angle as quantitative indicators to measure

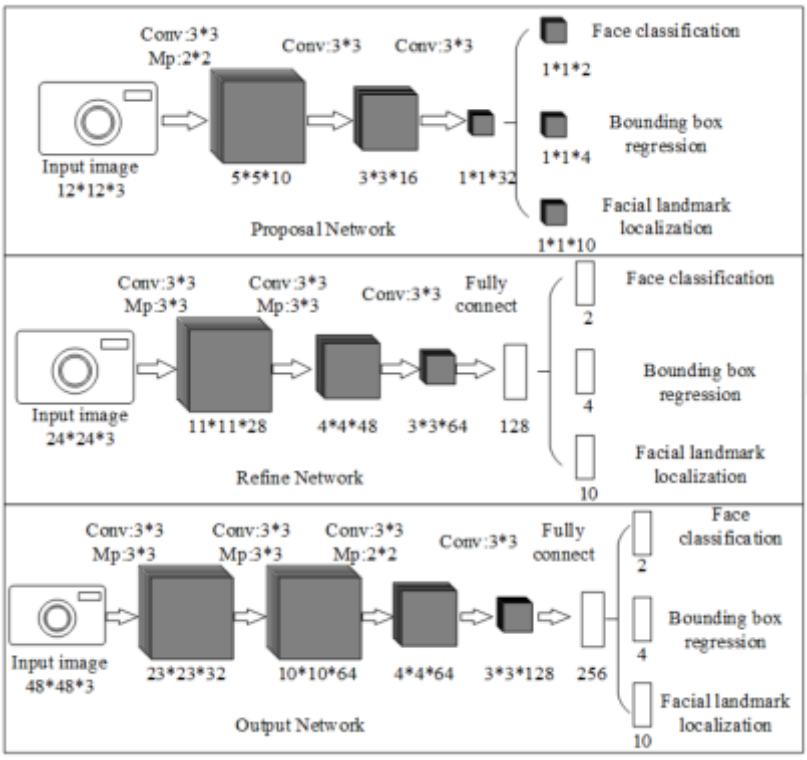


Fig. 3.1: MTCNN's three-layer structure

students' cognitive fatigue status, thereby improving the data sparsity problem in current cognitive fatigue detection. The purpose of the research is to provide a scientific and effective monitoring method of student status for the gradually normalized network teaching.

3. Research on the state recognition algorithm of students in online teaching.

3.1. Construction of online teaching face detection method based on convolutional neural network. Multi-task convolutional neural network (MTCNN) can combine face region detection and face key point detection. At the same time, the algorithm has the idea of candidate frame and classifier, which can detect faces quickly and accurately [19]. Therefore, the core algorithm of cognitive state detection of students in online teaching proposed in this study is MTCNN. MTCNN consists of three layers of network structure, among which the Proposal Network structure is responsible for quickly generating candidate boxes that may contain faces. The Refine Network network structure performs high-precision filtering and selection on the generated candidate boxes to improve the accuracy of detection. The final output network structure is responsible for generating the final facial bounding boxes and feature regression of keypoints, that is, determining the specific position of the face and the precise coordinates of keypoints. The three-layer structure of MTCNN is shown in Figure 3.1.

As can be seen from Figure 3.1, in the above MTCNN standard convolution layer, after obtaining the size of the input feature image, the number of channels and the number of convolution layers, the parameters of the convolution layer can be calculated. The calculation formula is as follows Formula (3.1) is shown.

$$n_p = W_F \times H_F \times D_{im} \times N \tag{3.1}$$

In formula (3.1), it n_p represents the amount of convolution parameters, W_F, , $H_F$and , D_{im}represent the width, height, and number of channels of the feature map of the convolution layer, which are the number $N$of layers of the convolution layer. It can be seen from the above calculation formula that when the size of the image

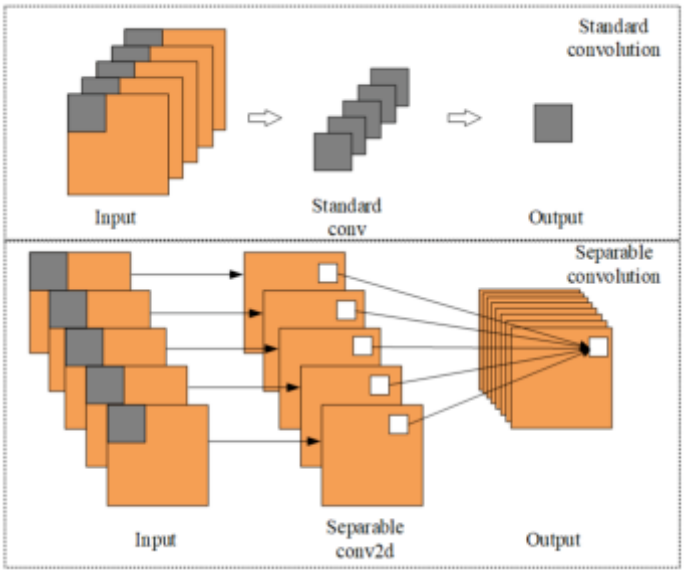


Fig. 3.2: MTCNN Standard Convolution Process and Depth Separable Convolution Process

is large, the total amount of parameter calculation data of the convolution layer is huge, so the research will optimize the standard convolution of traditional MTCNN through depth wise separable convolution, aiming to reduce the volume The amount of product parameters increases the processing efficiency of feature maps [20]. The standard convolution process of MTCNN and the depth wise separable convolution process are shown in Figure 3.2.

It can be seen from Figure 3.2 that the calculation of the parameter amount of the depth wise separable convolution is expressed as $n_p = W_F \times H_F \times D_{im} + D_{im} \times N$, comparing the parameter amount of the standard convolution and the depth wise separable convolution, the result is shown in formula (3.2).

$$\frac{W_F \times H_F \times D_{im} + D_{im} \times N}{W_F \times H_F \times D_{im} \times N} = \frac{1}{N} + \frac{1}{W_F \times H_F} \tag{3.2}$$

It can be seen from equation (3.2) that the parameter amount of the depth wise separable convolution is smaller than the standard convolution parameter amount. In the online teaching environment, not all images have faces displayed in the image detection stage. Therefore, the study introduces the Focal Loss loss function for uneven sample classification, and its optimization principle is to increase the weight according to the difficulty of sample classification. First, the loss function of the MTCNN algorithm in face detection consists of classification, boundary regression and key point feature regression loss functions. Therefore, in the image detection and classification, the face classification and other classifications are expressed as a two-class cross entropy loss function, and its formula is shown in formula (3.3).

$$L^c = -(y^c \log(p) + (1 - y^c)(1 - \log(p))) \tag{3.3}$$

In formula (3.3), it L^crepresents the loss function of face classification, which is $y^c \in \{0, 1\}$the face classification of the image, 0 represents the non-face category, and 1 represents the face category. pRepresents the probability that the image sample is a face. In the traditional cross-entropy loss function, the classification does not consider the difficulty of image classification. Once non-face images account for a large proportion of image samples, it is easy to cause imbalance of sample categories. Therefore, the weight and modulation factor of the Focal Loss loss function are introduced to optimize the image classification loss function as formula (3.4).

$$L^c = -\sigma(1 - p)^\rho \log(p) \tag{3.4}$$

In formula (3.4), $\sigma \in [0,1]$represents the weight factor, which is used to measure the proportion of positive and negative samples, $(1-p)^{\rho}$and represents the modulation factor, which is used to distinguish the difficulty of sample classification. The face frame regression loss function represents the distance between the predicted result of the face regression frame and the actual face frame, so the loss function calculation is completed by the Euclidean distance calculation formula. In the same way, the loss function of the key point of the face feature represents the difference between the pre-positioning of the key point and the actual position, so it is the same as the frame regression loss, which is represented by the Euclidean distance. The specific loss functions of the two are shown in formula (3.5).

$$\begin{cases} L^b = \left\| \widehat{y}^b - y^b \right\|_2^2 \\ L^l = \left\| \widehat{y}^l - y^l \right\|_2^2 \end{cases} \tag{3.5}$$

In formula (3.5), $L^b$and L^lrepresent the face frame regression loss function and the face key point feature regression loss function, respectively. $\widehat{y}^b$ If it is the predicted coordinates of $y^b$the frame, it is the actual coordinates of the frame; if it is the $\widehat{y}^l$ pre-positioning coordinates $y^l$of the key points of the face, it is the actual coordinates of the key points. Combining the classification, boundary regression loss function and face key point regression loss function by introducing weights, the overall loss function of MTCNN face detection is obtained as shown in formula (3.6).

$$L^t = \min \sum_{i=1}^{N} \sum_{j \in \{c,b,l\}} \alpha_j \beta_i^j L_i^j \tag{3.6}$$

In formula (3.6), $L^t$is the overall loss function, $N$is the total number of image samples, $i$represents the number of the image samples, is the number of $j$the three tasks of classification, frame regression, and key point regression, $\alpha_j$represents the task weight, $\beta_i^j \in \{0,1\}$is the sample label, and $L_i^j$is the loss function of the task . When the image enters the convolutional layer for calculation, due to the different image shooting environments of online learning, there are a lot of interference factors in the image, so the research will optimize the image filtering operation. The research and application method are the median filter method for smoothing. The pixels in the image window are sorted according to the gray value, and the median value is taken to replace the pixel value. The formula is expressed as formula (3.7).

$$g(x,y) = mid\{f(x-a, y-b)\}; (a,b) \in S \tag{3.7}$$

In formula (3.7), it $g(x,y)$represents the planting filter pixel, which $f(x,y)$is the original noisy pixel, which $f(x-a, y-b)$represents the domain pixel of the original pixel, which Srepresents the filter. Finally, MTCNN detects the facial situation of online learning students, including image denoising, generating image pyramid structure; determining face candidate frame; deleting unselected candidate frame; using sink to delete candidate frame with high degree of overlap [21]. The applied parameters include the image scaling factor, the minimum detectable face image size, the selection threshold of the human frame, and the screening and intersection ratio (Intersection over Union, IoU) of the Non-Maximum Suppression (NMS) algorithm.) threshold. Its calculation formula is shown in formula (3.8).

$$IoU = \frac{[(Cx2 - Cx1) \times (Cy2 - Cy1)] \cap [(Gx2 - Gx1) \times (Gy2 - Gy1)]}{[(Cx2 - Cx1) \times (Cy2 - Cy1)] \cup [(Gx2 - Gx1) \times (Gy2 - Gy1)]} \tag{3.8}$$

In formula (3.8), $[(Cx2 - Cx1) \times (Cy2 - Cy1)]$the area of the region C is expressed, and the area of $[(Gx2 - Gx1) \times (Gy2 - Gy1)]$the region G is expressed.

3.2. Facial feature analysis of students' cognitive load status. The cognitive load state of students refers to the physical exhaustion caused by students investing their own cognition and energy resources to receive information and process them in the learning process. The measurement of cognitive load in academics is difficult to achieve through network technology. Therefore, the study summarizes the facial performance in the academic theory of cognitive load state, and uses the specific facial key state features to detect the MTCNN algorithm to identify the academic fatigue state. and cognitive load [22]. In this study, the two-dimensional feature points of the color image of the face are extracted by the feature point location method of the cascade regression tree (Ensemble of Regression Trees, ERT) algorithm. The principle of the ERT algorithm to extract feature points is to connect the strong regressors in series through two-level cascaded regression to build a feature mathematical model. The iterative formula is shown in Equation (3.9).

$$\begin{cases} \hat{S}^{t-1} = \hat{S}^t + \psi_t(I, \hat{S}^t) \\ \Delta S_i^t = S_i + \hat{S}^t \end{cases} \tag{3.9}$$

In formula (3.9), it $\hat{S}^t$represents $t$the predicted shape vector of the stage regressor, which $\psi_t$represents the $t$stage regressor, $I$is the input image, ΔS^trepresents $t$the difference between the prediction of the stage regressor and the actual result, and $i$represents the data sample number. The shape vector of facial feature points is updated by an iterative formula, and the number of regression cascade layers in the first layer is the number of regressors. In the regression calculation of the second layer, the regressor is trained by the method of enhancing the gradient, so the initialization mathematical model of the regressor is shown in formula (3.10).

$$f_0(I, \hat{S}_i^t) = \arg\min \sum_{i=1}^{N} \left\| \Delta S_i^t - \psi \right\|^2 \tag{3.10}$$

In Equation (3.10), it $f_0(I, \hat{S}_i^t)$represents the fitting of the residual regressor of the initial regressor, and the value of the gradient update is set r_{ik}, and its calculation formula is shown in Equation (3.11).

$$r_{ik} = \Delta S_i^t - f_{k-1}(I_i, \hat{S}_i^t) \tag{3.11}$$

In formula (3.11), it $f_{k-1}(I_i, \hat{S}_i^t)$represents $k-1$the fitting of the first-level strong regressor to the residual regressor. $k$represents the number of regressors. Finally, a strong regressor is constructed through a weak regressor, and the final output of the iterative convergence is expressed as Eq. (3.12).

$$\begin{cases} f_k(I, \hat{S}^t) = f_{k-1}(I, \hat{S}^t) + v g_k(I, \hat{S}^t) \\ R(I, \hat{S}^t) = f_K(I, \hat{S}^t) \end{cases} \tag{3.12}$$

In Equation (3.12), $v \in [0, 1]$denotes the learning rate, and $g_k$denotes the weak regressors that constitute the strong regressor, Kdenotes the number of weak regressors. After the two-dimensional positioning of the feature points by the ERT algorithm, the study analyzes and models the cognitive load characteristics. Firstly, students' learning cognitive load status is divided into mouth features and eye features [23]. After obtaining the two-dimensional positioning of the mouth and eyes through the ERT algorithm, the distance change between the two points is calculated to detect the frequency of students blinking and yawning. Finally, in the extraction of 3D feature points of the face, the RGB-D image acquisition method of the depth camera is studied. Corresponding the information of the color image and the depth image, the relationship change formula is shown in formula (3.13).

$$\begin{cases} p_{de} = H_{de} P_{de} \\ p_r = H_r P_r \end{cases} \tag{3.13}$$

Formula (3.13), $p_{de}$is the projection coordinate of the object on the depth image, is the H_{de}internal parameter of the depth camera, and $P_{de}$is the spatial coordinate of the object in the depth image. $p_r$is the projected coordinate of the object on the color image, is the $H_r$internal parameter of the color camera, and $P_r$is the

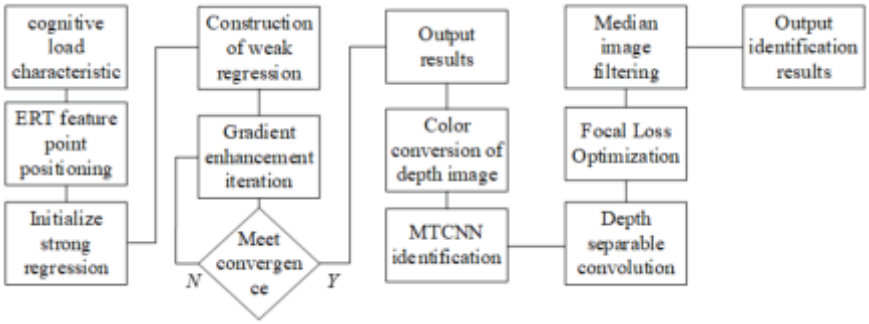


Fig. 3.3: MTCNN algorithm identifies the flow of academic cognitive load status

spatial coordinate of the object on the color camera. Therefore, for the point coordinates of the object in space P, there is a relational expression as shown in Equation (3.14).

$$\begin{cases} P_{de} = R_{de}P + T_{de} \\ P_r = R_r P + T_r \end{cases} \tag{3.14}$$

In formula (3.14), it (P, R, T)represents the three-dimensional coordinates of the object, and the coordinates of the depth image can be converted into the coordinates of the color image through the formula. Finally, the calculation formulas such as precision, recall, and accuracy verify the image detection performance of the MTCNN algorithm for students' online classrooms. The specific formulas are shown in formula (3.15).

$$\begin{cases} pre = \frac{TP}{TP+FP} \\ Re = \frac{TP}{TP+FN} \\ Acc = \frac{TP+TN}{TP+TN+FP+FN} \\ f1 = \frac{2pre \cdot Re}{pre+Re} \end{cases} \tag{3.15}$$

In formula (3.15), it TPrepresents the number of correctly identified samples, the number of FPincorrectly recognized samples, the FNundetected wrong samples, and $TN$the undetected correct samples. The core idea of this research is to firstly summarize the common facial manifestations such as blink frequency, eye closure degree, yawn frequency, etc. under students' cognitive load state. Secondly, the ERT algorithm is used to determine the two-dimensional positioning of key points such as the mouth or eyes. Thirdly, the spatial coordinates of the depth image and color image are converted by the depth camera RGB-D image acquisition method, and finally the image is recognized by the MTCNN algorithm model. Classification. Therefore, the overall flow of the algorithm is shown in Figure 3.3.

As can be seen from Figure 3.3, the optimization of the algorithm model in this experiment includes the depth wise separable convolution of the MTCNN model, which is used to reduce the number of parameters in the standard convolution. At the same time, the Focal Loss method is also introduced to optimize the model loss function, which is expected to increase the performance of MTCNN in student cognitive state monitoring through optimization.

4. Application Analysis of MTCNN State Recognition Algorithm for Cognitive Load in Online Teaching.

4.1. Simulation Experiment of MTCNN Image Recognition. In order to verify the image detection performance of the MTCNN network teaching student state recognition algorithm constructed in the research, the research will carry out simulation experiments from the open network face image data set. This simulation training experiment is carried out on the TensorFlow platform. The three-layer structure parameters of the MTCNN algorithm optimized by filtering optimization and depth wise separable convolutional network parameters are shown in Table 4.1.

Table 4.1: Three-layer structure parameters for optimizing MTCNN algorithm

P-Net		R-Net		O-Net	
base-Ir	0.001	base-Ir	0.001	base-Ir	0.001
batch-size	384	batch-size	384	batch-size	384
epochs	30	epochs	twenty-two	epochs	twenty-two
momentum	0.9	momentum	0.9	momentum	0.9
Lr - factor	0.1	Lr - factor	0.1	Lr - factor	0.1
Lr - epoch	6	Lr - epoch	14	Lr - epoch	20

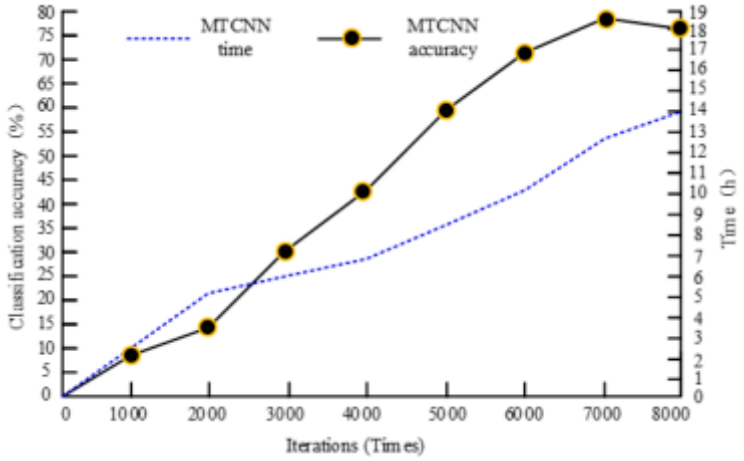


Fig. 4.1: Optimizing the overall training recognition accuracy and time consumption of MTCNN model

In Table 4.1, the size of the training image is 12*12*3, and its main task is to classify the image as containing or not containing faces, so its classification loss function is used for R-Net 's bounding box regression loss and O-Net The key point regression loss is not required, so the weight parameters of the overall loss function of the α_jP-Net simulation training are set to 1, 0.5, and 0.5 in the three tasks, respectively. At the same time, the maximum number of training samples is set to 8000 times, and the overall recognition accuracy and time-consuming of the model are shown in Figure 4.1.

It can be seen from Figure 4.1 that when the image size is 12*12*3, the overall recognition accuracy of the MTCNN algorithm in this simulation increases with the number of iterations. After the number of iterations exceeds 7 000, the classification accuracy increases. The trend turned to decline. In the simulation training, the highest classification accuracy in the MTCNN algorithm model is 77.8%, and the time to complete 8000 trainings is 14.1h. The experimental results show that the MTCNN algorithm has high classification accuracy performance in face image recognition, and at the same time, due to the optimization of the depth wise separable convolution, the simulation training takes less time. After the simulation training, the research will validate the MTCNN algorithm after training in the Emotional State Dataset for Online Education (DAiSEE) and the 2018 Emotion Recognition Challenge Dataset (EmotiW2018). The sample size distribution of the dataset and the performance analysis of the MTCNN algorithm are shown in Table 4.2.

It can be seen from the table that there is a large amount of data in the emotional state data set of online education. Among the 1813 sample identifications in the validation set, the precision, recall rate and f1 value of the MTCNN algorithm are 0.9076, 0.6923, and 0.7855, respectively. In the 48 data sets of the 2018 Emotion Recognition Challenge dataset, the precision, recall and f1 value of the MTCNN algorithm were 0.69, 0.89, and 0.78, respectively. Experiments show that the recognition accuracy performance of MTCNN is higher than 75% in data sets with large and small data samples, indicating that the recognition ability of MTCNN algorithm

Table 4.2: Sample size distribution of data sets and performance analysis of MTCNN algorithm

Data set	Sample distribution		MTCNN effect	
DAiSEE	train	5441	pre _	0.9076
	verification	1813	Re	0.6923
	test	1814	f 1	0.7855
EmotiW2018	train	147	pre _	0.69
	verification	48	Re	0.89
	test	64	f 1	0.78

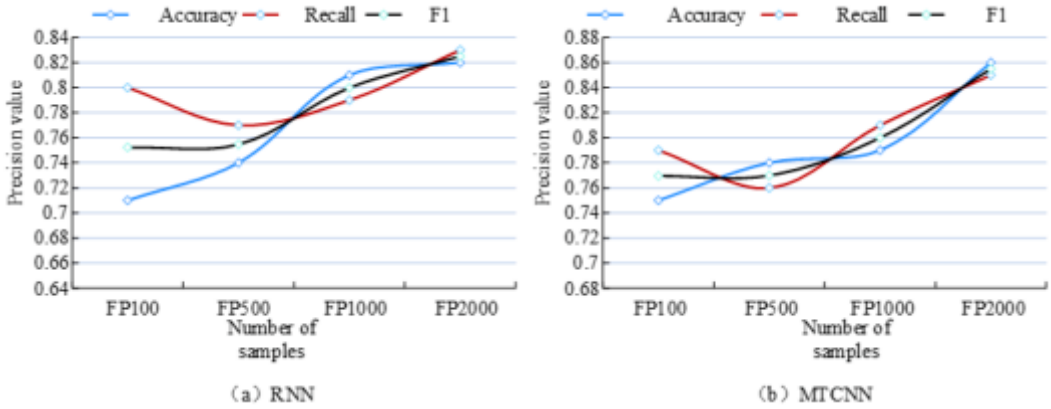


Fig. 4.2: Image recognition performance of RNN and MTCNN algorithms with different dataset sizes

is stable. Finally, the research will compare the image recognition performance of the recurrent convolutional neural network and the multi-task convolutional neural network. The number of samples in this test set is 100, 500, 1000, and 2000 four stages to compare the recognition performance of the two algorithms. The result is shown in Figure 4.2.

The RNN in the figure represents a recurrent convolutional neural network, while MTCNN is a multi-task convolutional neural network optimized for research. In the comparison of the results in Figure 4.2, it can be seen that the overall image recognition performance of the MTCNN algorithm is higher than that of the RNN algorithm. When the number of samples in the dataset increases, the recognition performance of the two algorithms is slightly improved. When the number of samples in the test dataset is 100, the accuracy of the recurrent convolutional neural network is 0.71, while the recognition accuracy of the MTCNN algorithm is 0.75; the recall rate and f1 value of the RNN algorithm are 0.8 and 0.752, respectively, but the recognition accuracy of the MTCNN algorithm is 0.8 and 0.752, respectively. In the test, the recall and f1 value of the optimized algorithm were 0.79 and 0.769, respectively. When the maximum number of samples in the test dataset is 2000, the recognition accuracy of the recurrent convolutional neural network is the highest 0.82, the recall rate is 0.83, and the f1 value is 0.825. The optimized multi-task convolutional neural network achieves the highest recognition accuracy of 0.86, recall of 0.85, and f1 value of 0.855. Experiments show that the performance of the optimized MTCNN algorithm is higher than that of the recurrent convolutional neural network in network image recognition.

4.2. Application Analysis of Cognitive Regulation Strategies in Online Teaching Based on MTCNN. The practical application of the MTCNN optimization algorithm takes a social science course in the MOOC forum as the experimental object. In the online teaching video, the course implementation process is divided into two categories: student-led and teacher-led, and the course content is divided into teaching platforms and other topics. 6 content categories: questioning, discussion, discussion results integration,

Table 4.3: The influence of different links on students' cognitive behavior

link	s1	s2	s3	s4	s5	s6
t1	11.98	2.69	-2.83	-0.52	-0.81	-0.64
t2	1.12	17.48	0.94	-1.46	-2.29	-5.6
t3	-0.94	0.19	61.22	-22.94	-6.42	-39.77
t4	-1.63	0.01	-21.97	67.48	-17.9	-35.44
t5	-0.78	-3.23	-8.04	-16.06	54.36	1.31
t6	0.86	-5.04	-39.88	-35.53	1.03	87.27
link	s1	s2	s3	s4	s5	s6
s1	34.06	-0.21	6.46	-0.02	-0.06	5.1
s2	-0.42	5.29	0.79	-0.17	-0.46	-2.28
s3	-1.93	1.98	-1.07	-0.79	-2.08	-11.61
s4	-1.6	-3.74	-3.44	1.53	-1.73	-9.93
s5	-0.55	0.7	-0.68	-0.22	1.18	-3.46
s6	1.58	-1.81	-2.33	-0.55	1.25	-2.98

solution narrative, and social background narrative. The cognitive concentration degree of different course links is quantified from the students' learning cognitive behavior, and the moderating effect of the arrangement of different course links on students' cognitive load is analyzed. The students' teaching platform and other topics, questioning, discussion, discussion results integration, solution narrative, and social background narrative are classified as s1-s6, and the six teacher-led links are divided into s1-s6. From t1 to t6, the following sequence analysis method is used to calculate the influence of different links dominated by different subjects on students' cognitive behavior. The specific results are shown in Table 4.3.

It can be seen from the table that the teaching platform guided by teachers and other topic links, questioning link, discussion link, discussion result integration link, solution method narrative link, and social background narrative link all mobilize students' cognitive behavior. It has a positive impact, and the effect is most significant in the results integration link and the social background narrative link. At the same time, in the process of student-led communication with teachers, the frequency of students' cognitive behavior is higher than other states. To sum up, the experiments show that teacher-guided discussions and summaries have a scheduling effect on students' cognitive behavior. Teachers can effectively improve students' cognitive load by asking questions, exchanging information, and by expressing questions and showing ideas. Therefore, after the optimization of the implementation sequence through the course links, the effect of improving students' cognitive behavior can be analyzed through the MOOC teaching videos, and 80 videos showing the student's status with a length of 10s and the students after the adjustment of the teaching implementation link are displayed. Ease to perform MTCNN image recognition analysis. The cognitive state of students is analyzed through the eye, mouth and face recognition proposed in the research method. The specific results of the face part are shown in Figure 4.3.

The ordinate in FIG. 4.3 is the student's face offset angle, and the student's face offset represents a negative angle to the left and a positive angle to the right. When the student's face offset angle is between -15° and 15°, it indicates that the student is facing the front screen and his cognitive behavior is in a state of concentration, while the student's face is greater than 15° and less than -15°, it indicates that the students are in a state of cognitive load. It can be seen from the dot matrix distribution diagram in Figure 6(a) that 34 of the top 80 students in the teaching process are in a state of cognitive load with excessive facial offset. In Figure (b), only 5 of the 80 students have a facial offset of less than -15° after the optimization and adjustment of the teaching process. Experiments show that the improved regulation of online teaching links proposed in the study has a significant positive impact on students' cognitive behavioral attention. Finally, the changes of the students' mouth and eyes are also compared, and the specific results are shown in Figure 4.4.

Figure 4.4 shows the changes in the eyes and mouth of middle school students during video monitoring. In the online teaching, this study expressed the students' blink rate of more than 25 times and less than 10 times per minute as the cognitive load state. In the change of the mouth, the ratio of the vertical and horizontal distance of the mouth, which represents the opening and closing degree of the mouth, is used to distinguish

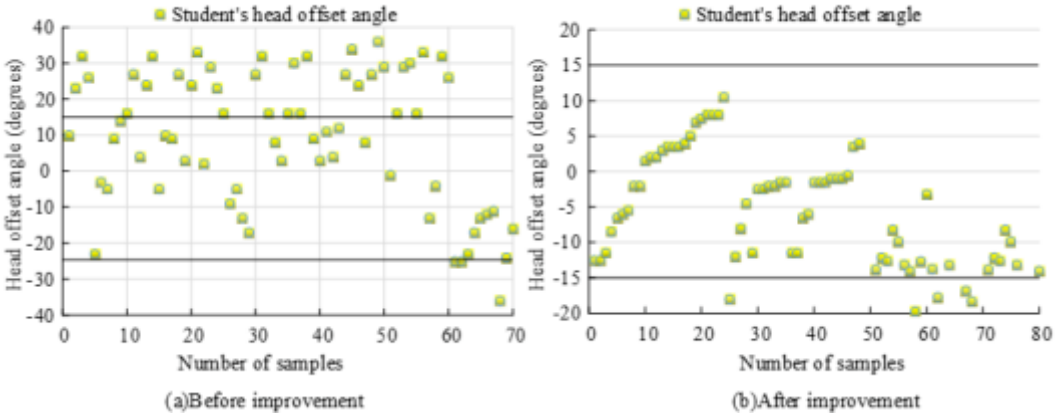


Fig. 4.3: Comparative analysis of students' facial deviation before and after teaching optimization

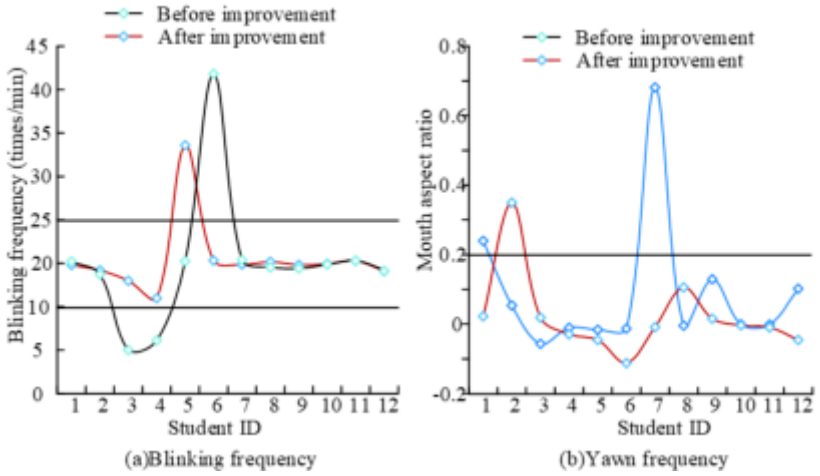


Fig. 4.4: Comparative analysis of changes in students' mouth and eyes before and after teaching optimization

the speaking state and the yawning state of the students. The case where the mouth aspect ratio is greater than 0.2 is classified as yawning, and the case where the ratio is less than 0.2 is classified as speaking or the mouth is tightly closed. It can be seen from the sub-figure (a) that among the blink frequencies in the tired and dazed states, 3 of the 12 students' video samples are in a state of cognitive load, but after the adjustment of the online teaching link, only A classmate's blink rate is 33 times per minute. In sub-figure (b), among the 12 students before the course optimization, there were 2 students whose mouth aspect ratio was greater than 0.2, but after the course optimization, the students' cognitive load status improved, and only one classmate's mouth was Aspect ratio greater than 0.2. Experiments show that after the optimization and adjustment of the course link of online teaching, the blinking frequency of students is gradually in a normal state, rather than a state of fatigue that is too fast and a state of daze that is too slow; at the same time, the situation of students yawning is also reduced.

5. Conclusion. In the context of online teaching, it is difficult for students to understand their cognitive behaviors and cognitive states through communication. Therefore, the research proposes a cognitive feature extraction method combined with MTCNN image recognition, and conducts simulation experiments and prac-

tical applications. In the simulation experiment, the highest classification accuracy of the MTCNN algorithm model is 77.8%, and the time to complete 8000 training sessions is 14.1h. At the same time, when comparing the RNN algorithm, its accuracy and f1 value are respectively 4% and 3% higher. In the practical application experiment, the study proposes to improve the cognitive load state of students and mobilize students' cognitive behavior by adjusting the links of online teaching courses. Through sample case analysis, it is found that 34 of the first 80 students have facial deviations. The angle is greater than 15° or less than -15°, and after adjustment, only 5 students' faces are not concentrated in front. In the video recognition of blink frequency and yawn state, 3 students in the video samples of the first 12 students are in a state of cognitive load, but after the adjustment of the online teaching link, only 1 student has a blink frequency of 33 times per class. minute. Before the course optimization, 2 of the 12 students had a mouth aspect ratio greater than 0.2, but after the course optimization, the students' cognitive load status improved, and only one student had a mouth aspect ratio greater than 0.2. Experiments show that the optimized method of MTTCNN network identification and students' cognitive feature extraction can help optimize and adjust courses in network teaching and improve students' learning status. The shortcomings of the research are that the detection system has high hardware requirements and the production cost is relatively expensive. In future research, efforts will be made to adjust the adaptability of the detection model to other hardware, thereby reducing device costs and improving the applicability of the method.

REFERENCES

[1] Xiangyu Liu. (2021). Validation research on the application of depthwise separable convolutional facial expression recognition in non-pharmacological treatment of BPSD. *Clinical Nursing Research*, 5(4), 31–37.

[2] Fu, H., Guan, J., Jing, F., …. (2021). A real-time multi-vehicle tracking framework in intelligent vehicular networks. *China Communications*, 18(6), 89–99.

[3] Liu, J., Xu, X., Shi, Y., Deng, C., Shi, M. (2022). RELAXNet: Residual efficient learning and attention expected fusion network for real-time semantic segmentation. *Neurocomputing*, 474(Feb.14), 115–127.

[4] Ye, F. (2022). Emotion recognition of online education learners by convolutional neural networks. *Computational Intelligence and Neuroscience*, 2022(1), 4316812.

[5] Kaur, R., Gupta, D., Madhukar, M., Singh, A., Abdelhaq, M., Alsaqour, R., Goyal, N. (2022). E-Learning Environment Based Intelligent Profiling System for Enhancing User Adaptation. *Electronics*, 11(20), 3354.

[6] Li, M., Chow, S. M. S., Hu, S., Yan, Y., Shen, C., Wang, Q. (2022). Optimizing privacy-preserving outsourced convolutional neural network predictions. *IEEE Transactions on Dependable and Secure Computing*, 19(3), 1592–1604.

[7] Atik, I. (2022). Classification of electronic components based on convolutional neural network architecture. *Energies*, 15(7), 2347–2361.

[8] Kiki, V., Mehrkanoon, S. (2022). Goal-driven, neurobiological-inspired convolutional neural network models of human spatial hearing. *Neurocomputing*, 470(22), 432–442.

[9] El-Shafai, W., Fawzi, A., Sedik, A., Zekry, A. M., El-Banby, G. M., Khalaf, A. A. M., Abd El-Samie, F. E., Abd-Elnaby, M. (2022). Convolutional neural network model for spectrum sensing in cognitive radio systems. *International Journal of Communication Systems*, 35(6), e5072.1–e5072.22.

[10] Indu, V. T., Priyadharsini, S. (2022). Crossover-based wind-driven optimized convolutional neural network model for tomato leaf disease classification. *Journal of Plant Diseases and Protection*, 129(3), 559–578.

[11] Ben Atitallah, S., Driss, M., Boulila, W., Ben Ghezala, H. (2022). Randomly initialized convolutional neural network for the recognition of COVID-19 using X-ray images. *International Journal of Imaging Systems and Technology*, 32(1), 55–73.

[12] Rahimilarki, R., Gao, Z., Jin, N., Zhang, A. (2022). Convolutional neural network fault classification based on time-series analysis for benchmark wind turbine machine. *Renewable Energy*, 185(Feb.), 916–931.

[13] Zhang, Y., Wu, Z., Lin, P., Wu, Y., Wei, L., Huang, Z., Huangfu, J. (2022). Text detection and recognition based on a lensless imaging system. *Applied Optics*, 61(14), 4177–4186.

[14] Fernandes, J., Simsek, M., Kantarci, B., Khan, S. (2022). TableDet: An end-to-end deep learning approach for table detection and table image classification in data sheet images. *Neurocomputing*, 468(Jan.11), 317–334.

[15] Teng, S., Chen, G., Wang, S., Zhang, J., Sun, X. (2022). Digital image correlation-based structural state detection through deep learning. *Frontiers of Structure and Civil Engineering*, 16(1), 45–56.

[16] Mohsin, N. A. (2021). A hybrid method for payload enhancement in image steganography based on edge area detection. *Cybernetics and Information Technologies*, 21(3), 97–107.

[17] Kamil, M.H.M., Zaini, N., Mazalan, L., Ahamad, A.H. (2023). Online attendance system based on facial recognition with face mask detection. *Multimedia Tools and Applications*, 82(22), 34437–34457.

[18] Andrejevic, M., Selwyn, N. (2020). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology*, 45(2), 115–128.

[19] Adelson, C., Jordan, M. I., Muller, R. (2022). SOUL: An energy-efficient unsupervised online learning seizure detection classifier. *IEEE Journal of Solid-State Circuits*, 57(8), 2532–2544.

[20] Li, Q., Xu, L., Yang, X. (2022). 2D multi-person pose estimation combined with face detection. *International Journal of*

Pattern Recognition and Artificial Intelligence, 36(2), 2256002.1–2256002.23.
[21] Iqbal, K., Abbas, S., Khan, M. A., Ather, A., Khan, M. S., Fatima, A., Ahmad, G. (2021). Autonomous parking-lots detection with multi-sensor data fusion using machine deep learning techniques. *Computers, Materials and Continuum*, 67(2), 1595–1612.
[22] Aria, M., Agnihotri, V., Rohra, A., Sekhar, R. (2020). Secure online payment with facial recognition using MTCNN. *International Journal of Applied Engineering Research*, 15(3), 249–252.
[23] Sharma, N., Gupta, S., Mohamed, H.G. (2020). Siamese convolutional neural network-based twin structure model for independent offline signature verification. *Sustainability*, 14(18), 11484.

Edited by: Nasrullah Sheikh
Special issue on: Transformative Horizons:
The Role of AI and Computers in Shaping Future Trends of Education
Received: Aug 27, 2024
Accepted: Feb 5, 2025

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2244–2252, DOI 10.12694/scpe.v26i5.4771

RESEARCH ON AI AND COGNITIVE DECISION-MAKING IN ARTISTIC DESIGN INNOVATION AND APPLICATION

HUISAN WANG*

Abstract. A revolutionary change in the creative environment is represented by the incorporation of Artificial Intelligence (AI) and cognitive decision-making in creative design. This study investigates how cognitive psychology and artificial intelligence (AI) connect with creative design development and implementation. The study explores how artificial intelligence (AI) might enhance human creativity by utilizing machine learning techniques, neural networks, and computational imagination to provide new opportunities in creative thinking. The study explores models of cognitive decision-making and looks at how artificial intelligence (AI) systems might mimic and improve the cognitive processes that artists use to create their works. It draws attention to how AI can produce original design ideas, streamline design procedures, and offer individualized design alternatives. This project intends to investigate the consequences of AI-driven cognitive decision-making on the evolution of creative design through an integration of mathematical modeling and real-world experience, offering insights into the cooperative symbiosis among human creators and artificially intelligent systems. The results point to an important change in creative endeavors wherein AI collaborates with artists to explore the frontiers of design innovation and creativity.

Key words: Artificial Intelligence, Cognitive Decision-Making, Artistic Design, Innovation, Application

1. Introduction. The field of artistic design has undergone a radical transformation in recent times due to the convergence of artificial intelligence (AI) and cognitive decision-making. This study explores the complex interplay between artificial intelligence (AI) technology and the imaginative procedures that characterize artistic creativity[18]. AI's ability to replicate and even enhance human intellect has created new opportunities for investigating methods of design that were previously limited to the world that humans fantasy as it develops. AI is now able to help architects create fresh ideas, streamline design procedures, and forecast trends by utilizing algorithms that train neural networks, and data-driven recommendations[1]. In the framework of creative design, this study intends to investigate the applications and consequences of AI-driven cognitive decision-making, emphasizing its ability to improve innovation, efficacy, and the overall standard of artistic products[15]. The study aims to offer a sophisticated knowledge of how AI is transforming the prospects of creative development and its real-world uses across numerous sectors through a thorough review of existing technology and instances[6].

A revolutionary frontier in technology and creativity lies at the nexus of artificial intelligence (AI) and cognitive decision-making in artistic design. Artificial intelligence (AI) systems are becoming more and more like human cognitive processes[16]. This means that machines can now aid humans or even create art on their own. This combination creates new opportunities for artistic expression and innovation while also challenging conventional ideas of innovation[20]. Creators and artists may extend the limits of artistic practice by utilizing AI algorithms to create new ideas, explore with new media, and investigate complicated patterns. The potential and ramifications of AI-driven cognitive decision-making in decorative design are explored in this study, which also looks at the way these advances will affect the creative process, improve artistic results, and change the field of modern art[12, 13]. This study uses an interdisciplinary approach to investigate the possibilities and constraints of artificial intelligence in the creative domain, providing perspectives on the potential future of human-machine cooperation in artistic pursuits.

The creative business is witnessing a transformational threshold as artificial intelligence (AI) and artistic design come together[8]. The incorporation of artificial intelligence (AI) tools in the creative design processes is investigated in this study, with a focus on the ways that generative models, neural networks, and machine

*Art School, Zhengzhou University of Science and Technology, Zhengzhou, Henan, 450000, China (huisanwangreseas@rediffmail.com)

learning algorithms are changing the production, usage, and curation of art. The goal of the study is to comprehend how AI functions as a partner and tool in the creative process, from ideation to final creation[19]. Creative professionals can test the limits of imagination, investigate new styles, and tailor encounters for a variety of audiences by utilizing AI's powers in recognition of patterns, data visualization, and imaginative suggestions. The main contribution of proposed method is given below:

1. This research's main contribution is the way it brings cognitive decision-making and artificial intelligence (AI) together in the context of innovative and applied artistic design.
2. The project investigates the ways in which AI-powered tools can enhance human innovation by providing innovative methods for designing that combine creative intuition with computational capability.
3. The research suggests foundations that allow intelligent machines to operate in real-time, cooperatively with artists and designers, by utilizing theories of cognitive decision-making. This promotes a mutually beneficial connection among human imagination and artificial intelligence (AI).
4. This cross-disciplinary strategy broadens the scope of what is feasible in the realm of design by improving the effectiveness and productivity of the design method as well as creating new opportunities for individualized and flexible expression of art.

The rest of our research article is written as follows: Section 2 discusses the related work on various Artificial Intelligence, Cognitive Decision-Making, Artistic Design, Innovation, Application. 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. In recent years, there has been a lot of interest in the relationship between artificial intelligence (AI) and cognitive reasoning in creative design[5]. The second part examines the condition of the field at the moment, emphasizing significant developments and techniques. Artificial Intelligence (AI) has become more prevalent in the field of creative design[17]. It uses machine learning techniques and algorithms to provide innovative results. The author's research unveiled "Creative Adversarial Networks" (CAN), a GAN-based methodology that investigates how AI might be used to produce original artistic forms. Convolutional neural networks (CNNs) are also used in style transfer, as the author demonstrated, enabling the merging of many kinds of art in computerized paintings[10].

In architecture, cognitive thinking refers to the innovative problem-solving and decision-making processes of architects. The author has explored the cognitive mechanisms that underlie creative thinking, highlighting the significance of deductive reasoning[3]. The author also looked at how designers use mental images and visual reasoning, emphasizing how they employ cognitive techniques to produce and refine ideas. By giving designers new tools and techniques, the combination of AI with cognitive decision-making seeks to enhance the creative process[7]. The author talked about how AI may enhance human inventiveness and provide fresh viewpoints as a cooperative partner in the creative process. Additionally, the author's research looked at how AI might be used to automate tedious design processes so that designers might concentrate on more intricate and imaginative details[14].

AI has a wide range of uses in artistic layout, from interactive displays to generative art. The author gave an example of how AI is used to create dynamic art displays that react to user input[11]. The author's discussion of the use of AI in game development demonstrates how this technology may be used to improve gameplay and story. The integration of AI and cognitive decision-making in decorative design still faces difficulties, despite encouraging developments[4]. The application of art produced by artificial intelligence must take into account ethical factors like authorship and uniqueness, as the author discusses[9, 2]. In addition to investigating new paradigms in AI-facilitated cognitive decision-making, subsequent study is anticipated to concentrate on creating simpler artificial intelligence instruments that can work smoothly with human designers.

This section presents the need to conduct some research.

At the core of this paradigm shift is the invocation of Artificial Intelligence (AI) within the field of artistic creation more so with Cognitive Decision-Making models. As in other areas of creativity, paradigms are made in the conventional way of performing tasks with considerable reliance on the faculties and experience of the creators. However, with a deep learning approach, AI has the potential to revolutionize the creative industry by relieving the artist of monotonous tasks, following design tools and approaches that have been indigenously used, and offering customized targeting solutions that stand out. This research is critical in order to appreciate

better such intricate interaction between the mind and artificial intelligence with regards to the artistic aspect which is predominant considering that creativity is the driving force. With the growing global integration of creative AI into different traditional creative environments, further investigation about the central role of a cognitive model of creativity in the processes of creation needs to be done to avoid putting an end to the artist in the new art-making process. Towards the end of the thesis, the focus will involve elaboration on the reasons why there are gaps in knowledge and how the phenomenon of artists and AI in the creative processes should be initiated.

Existing Research Gaps.

1. Although AI has been employed in many areas as an effective tool, how AI systems can aid and improve artists' cognitive decision making is a domain which lacks extensive research. Most of the studies investigate the outcomes provided by AI, but there is no focus on the sophisticated and rather interesting questions or processes of how AI models can provide or assist the high order functions such as imaging, thinking, recognizing patterns, solving problems etc. which are normally executed by human beings and dubbed creative works.
2. like Many other terms, artificial imagination is novel in the field of Application of AI, September, 2020 and its ability to generate novel inefficiencies remains under researched. It is not clear how else the AI systems can be utilized to transform design from just retaining and rearranging data to a more complex process which incorporates the ingenuity and surprise of the human mind. There is a need to examine the forces of computational models that can reproduce the neurophysical process where the human brain creates jumps in concepts and thinks out of the conceptual box.
3. Moreover, the degree of co-creation between the designers and AI creators has also been left under researched. Most of the available literature either centers on the design techniques utilized by the AI, or those that depend on human designers but, minimal attention is paid to how the two groups of designers i.e., the AI and the human designers can be combined in the design process. More specifically, there are some issues that have not drawn much attention whereby, it is expected that research will be conducted into the ways of integrating AI into the creative processes in practice and taking into consideration how the background of the users is reflected in the AI system generated designs.

3. Proposed Methodology. The purpose of the proposed study is to investigate how artificial intelligence (AI) functions in cognitive tasks related to creative decision-making. In order to better understand how AI technologies may support creative professionals like artists and designers, this research will look at how cognitive AI models can be incorporated into the process of creativity and how that can affect the final product of design. The goal of the suggested technique is to offer an organized way to look into how AI and cognitive reasoning interact with artistic design. The results will add to the expanding corpus of research on AI-driven imagination and provide useful guidance for creative people. In figure 3.1 shows the architecture of proposed method.

The graphic seems to be a diagram that illustrates a suggested technique for a study on artificial intelligence (AI) and cognitive decision-making in the creation and implementation of artistic design. Pre-processing is done on the information collected, in both qualitative and quantitative ways. The information must be cleaned, transformed, and made ready for examination in this step. This could involve processing missing values, classifying, or normalizing data. This stage, which analyzes and models cognitive processes, serves as the study's central focus. It entails comprehending the decision-making process involved in artistic design and may make use of cognitive theories and paradigms.

Textual data is analyzed using techniques from Natural Language Processing. To derive significant insights and trends, this may entail examining artist declarations, design explanations, or any pertinent text data. When creating AI systems, cognitive decision-making and natural language processing provide valuable insights. These artificial intelligence (AI) tools are intended to support or mimic the creative designer's process of decision-making. The ultimate objective is to attain the best possible outcome, which could include creatively stimulating designs, improved artistic procedures, or useful AI tools that assist artists in their artistic endeavors.

3.1. Data Collection and Analysis. Interviews: To learn more about AI's place in artistic design, undertake deep conversations with creative professionals, designers, and AI specialists. Case Studies: Examine particular examples of AI-assisted creative projects to comprehend the selection procedure. Examine scholarly articles, business reports, and case studies about artificial intelligence (AI) in design and cognitive decision-

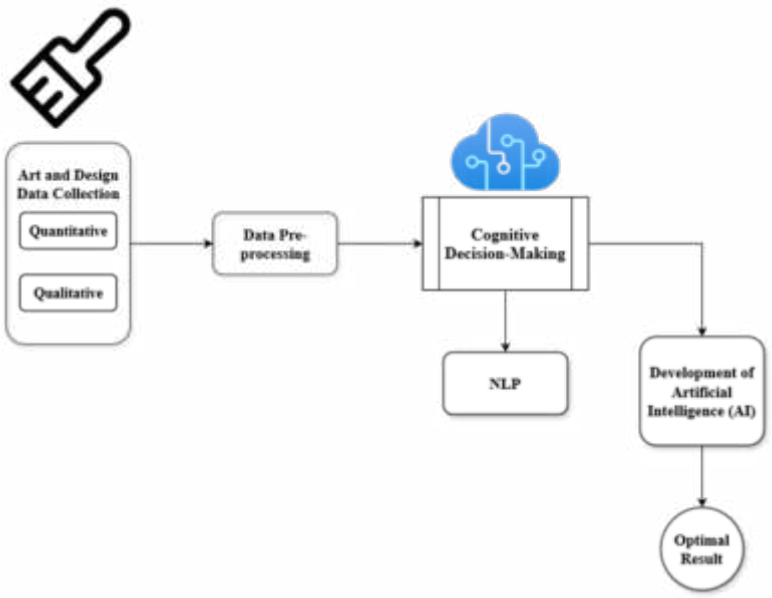


Fig. 3.1: Architecture of Proposed Method

making. Datasets: Make advantage of AI-generated works of art, design venture information sets, including user reviews that are all available to everyone.

Data Analysis. Thematic investigation involves finding recurring themes and trends in instances and information from interviews. Content Analysis: Examine the qualitative information to glean understanding of how AI and cognitive reasoning contribute to creative creation.Algorithm Testing: Use and evaluate artificial intelligence (AI) algorithms to provide design outputs, such as neural networks and generative adversarial networks. Cognitive Model Modeling: Construct and run models of cognitive decision-making to determine how they affect decisions related to design. Statistical Analysis: For assessing the efficacy as well as productivity of AI-driven method of design to conventional ones, use statistical techniques.

3.2. AI method for Training. To imitate cognitive processes for making decisions in design, choose and apply a variety of AI techniques, including machine learning methods, generative adversarial networks (GANs), and natural language processing (NLP) models. Give participants particular designing assignments to complete using or without AI support to assess their decision-making and inventiveness.

Generative Adversarial Networks (GANs). Deep learning models called Generative Adversarial Networks (GANs) are made up of two distinct neural networks—a discriminator and a generator—that fight with one another in a manner akin to a game.

Generator: Particularly, the generator aims at generating synthetic data, such as images, which are easy to associate with actual data on the original dataset. It commences with complete chaos and, via a series of operations through deep neural networks, aspires to create information that is as realistic as possible. At first, outputs from the generator are most of the time low quality, which means that the discriminator is capable of easily classifying the same as forgeries, but as the training goes on, the generator improves the outputs. Feedback from the discriminator provides further impetus in the learning of the generator as it always looks to produce more plausible data. The sole purpose of the generator is also to maximize the output by 'tricking' the discriminator into believing that the generator 's output s' ability is being diverted in progressively more realistic directions.

In a more practical context such as creative design, the generator can be said to take the design patterns learned from the training data, put them together in such a new way, replace some of the parts, and expand the design. Such an ability makes it possible to obtain not just variations of preexisting designs

but completely new high quality designs that is why GANs are effective in fields that require art.

Discriminator: As Gani and Sìmik stated, the discriminator's function which consists of two parts (discriminator and generator), is to recognize what is real and what is fake. It performs as a binary classifier which is fed with the both real data from the provided data set and the virtual sample generated by the generator. The discriminators perceive the information and give a score with probabilities judging whether the information is real or generated by some program. To begin with, the true instances towards the output of the discriminator which are the generator's outputs are fake, but as time progresses, and the generator is trained to be better, the discriminator is expected to get better at telling real instances from fake ones.

This relationship of anger between the two networks facilitates that both networks goes under constant progress. The effect of this relationship is that the discriminator becomes better at noticing features and patterns that bring a separation to the two data – real and generated. On the other hand, the generator gets better at generating data that is more and more similar to the real one. This response of moving to adjust to the changes that occur in the GAN system results in the enhancement of the GAN system.

The generator's objective is to generate information that is sufficiently authentic that a discriminator is unable to distinguish it from actual information. The two networks get better with time, which makes the generator provide more and more credible data. In many creative applications, including as picture production and video synthesis, GANs are employed extensively.

Natural Language Processing (NLP). The study of artificial intelligence's Natural Language Processing (NLP) area is concerned with how computer and human speech communicate. It entails creating models and algorithms that let machines understand, decode, and produce language that humans use.Work along with AI developers to build or modify AI tools specifically for applications in artistic design. Start an initial evaluation phase in which a limited number of participants use the tools that have been built for their design activities. Gather comments on performance and indicators. Peer a briefing, member checking, and triangulate are methods used to guarantee the validity and reliability of data that is qualitative as well as quantitative.Conduct pilot investigations to improve study tools and techniques. This suggested methodology offers a methodical way to look into how AI and cognitive decision-making interact with innovative artistic design. It seeks to make a significant contribution towards both use in the design sector and academic study.

4. Result Analysis. The study concentrated on the relationship between AI and thinking about decisions in the setting of innovative creative design. The main goals were to investigate how artificial intelligence (AI) might improve the creation process, assess how well cognitive models for decision-making work in creative architecture, and pinpoint possible uses for these advances across a range of artistic disciplines. The study effectively illustrated how artificial intelligence (AI) and cognitive decision-making might boost creative design development. In addition to increasing the effectiveness and caliber of design processes, the incorporation of these innovations created new channels for creative communication. To guarantee the appropriate and advantageous application of AI in the arts, nevertheless, rigorous evaluation of the scientific and moral problems is required.

The image 4.1 that was uploaded shows a bar chart that contrasts the effectiveness measurements of cognitive models and AI algorithms in relation to innovative artistic design. The metrics Accuracy, Precision, Recall, and F1-Score are displayed in the graph. Accuracy: When compared to conventional cognitive models, the accuracy of AI algorithms is marginally higher. Precision: Compared to cognitive models, AI algorithms have a marginally superior precision. Recall: Compared to cognitive models, AI algorithms have a greater recall measure. F1-Score: AI algorithms and neural networks both display a high F1-Score, yet AI techniques slightly outperform the latter.

The graph shows that, overall, algorithms using AI perform better than cognitive models on all parameters, but not significantly. The results show that neural networks and AI perform comparably, with AI doing somewhat better.

A line graph 4.2 showing the qualitative evaluation of AI-assisted ideas throughout five distinct design endeavors is included in the picture. The positive feedback percentages from critics and artists are compared in this chart. Artist Comments: Throughout the course of the five projects, the graph indicates a steady rise

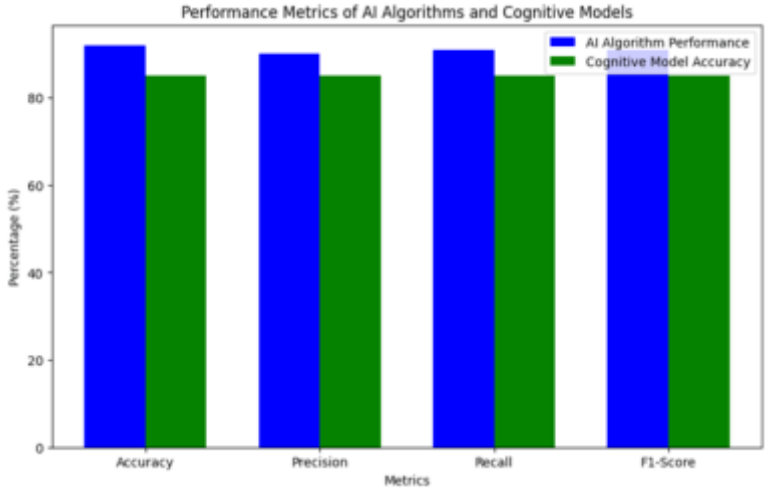


Fig. 4.1: Result of Different Performance Metrics

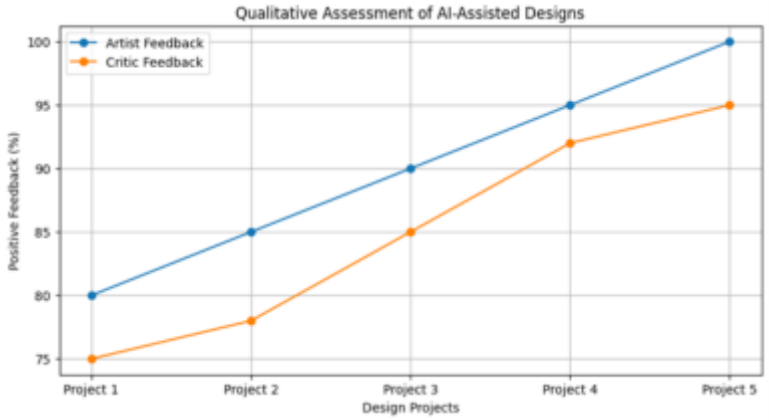


Fig. 4.2: Qualitative Assessment of AI-Assisted Design

in the artists' positive comments. For Project 1, the response is approximately 80%, and by Project 5, it has reached 100%. Critic Responses: The negative responses likewise exhibit an increasing pattern, although they begin at a lower percentage—roughly 75% for Project 1—and rise to a higher percentage—roughly 95%—for Project 5. The graph shows that both critics and artists are beginning to accept AI-assisted designs more and more. During the assignments, artists tended to provide more positive feedback than critics, which suggests that AI's influence on design may more closely match artists' imaginative aspirations. The higher trend for the two categories indicates a steady improvement in the quality and acceptability of AI-assisted designs throughout the course of the projects.

The evaluation period spans multiple a period of time from January to August. Diverse Design choices: In January, the response score is roughly 70, and by August, it has risen to 90. This suggests that user happiness with the range of design alternatives has significantly improved, indicating that the AI's capacity to provide a varied range of designs has become more sophisticated and valued over time. Effectiveness in Making Decisions: In January, the response score is approximately 65, and by August, it is approximately 85. This increasing tendency suggests that users are becoming more satisfied with how well cognitive frameworks

Fig. 4.3: User Experience Feedback Over Time

facilitate design-related decision-making, which expedites the process of innovation.

The graph 4.3 shows that over the course of the months, favorable feedback has increased for both a wide range of creative possibilities and the effectiveness of decision-making. Users' appreciation of AI-assisted tools' contributions in these domains grew over time, indicating improved customer service and a broader recognition of AI's place in the process of creation. The steady increases imply that technological advances were always advancing and becoming more in line with the requirements and expectations of users.

The graphic 4.4 classifies AI's relevance or influence in several artistic domains as well as its possibilities for advancement in the future. Graphic Design (85%): AI had a big influence on this field, and a lot of people think it's important for automating jobs and coming up with creative design ideas. Fashion (75%): Given the particular difficulties and creative processes associated with fashion design, AI's impact in this field is noteworthy but marginally smaller than in graphic arts and digital artwork. AI is also essential to digital art (80%), as it offers instruments for producing and modifying digital artworks and expands the possibilities for creativity.

Expectations for the Future (90%): The largest percentage (90%) points to a promising future for AI research and applications in innovative design. This implies that AI has a great deal of potential to further transform the industry with cutting-edge instruments and technology. The graph shows that artificial intelligence (AI) has a wide range of real-world uses in the arts, with a special emphasis on the creative industries. The steady and comparatively significant effect ratios in every category show that artificial intelligence (AI) is increasingly playing a crucial role in artistic creation by bringing novel abilities and boosting the process of creation. The opportunity for AI to revolutionize the arts is highlighted by the expectation of even bigger breakthroughs in the years to come.

Generative adversarial networks (GANs) tap the creative design space with notable breakthrough benefits. For example, they can learn and synthesize new designs without difficulty using existing data, and that makes them highly innovative in artistic endeavors. Such networks give the opportunity to the designers and allow them to diverge from the conventional approached which opens up other avenues to address the problem in hand. This fosters creativity, saves time during the design process, and speeds up the design process by utilizing the automated ideation and iteration stages. On top of that, GANS can generate various styles of designs that extend beyond the capabilities of human designers and can easily fit in diverse creative industries such as visual design, fashion or architecture.

Not overestimating the effectiveness of these models, it is evident that there's a number of issues that need to be tackled. The first limitation has to do with the fact that GANs are designs that have a high instability during training leading to problems such as mode collapse where the generator only produces a few outputs instead of a range that would be expected. Also, GANs are limited by the sustained processing speed and time if high quality results are to be produced due to complex and high-resolution designs. One more limitation of this

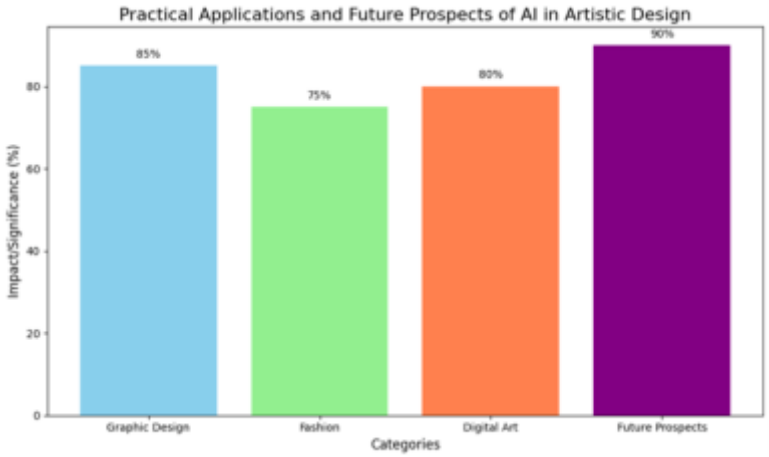


Fig. 4.4: Future Prospects of AI in Artistic Design

technology has to do with the understanding of context - even if GANs can yield good looking graphics, they can be unfit for the underlying design problem. Thus, human monitoring and control for the generated designs is required to attain the design or artistic objectives especially when some constraints were predetermined.

5. Conclusion. The investigation of AI and cognitive decision-making in innovative and applied artistic design highlights the groundbreaking possibilities of these advances in transforming the artistic environment. The creation of original creative works is made easier and design workflow effectiveness is increased by the use of AI with cognitive processes. According to our research, AI-powered tools can help designers explore new ideas, make the best decisions, and customize creative products to suit the tastes of the audience. Furthermore, an environment of collaboration wherein human imagination and machine intelligence collaborate together to expand the limits of creativity is fostered by the confluence of AI and cognitive decision-making. It is crucial to tackle moral problems as the area develops and make certain that these innovations are used in manners that uphold the authenticity of the artistic method. Subsequent studies ought to concentrate on improving AI algorithms in order to comprehend and mimic human cognitive processes more accurately, which will ultimately result in more impactful and intuitive design solutions. This investigation lays the groundwork for future research into the unexplored potential of this dynamic intersection while also adding to the expanding corpus of information on AI's role in artistic design.

Gazing into the future, the scope of future work in this regard could aim at enhancing the stability and control of GANs within a research context. Work on conditional gan, cgan in modified conditions and StyleGAN has the potential to give designers more control over what is generated, helping in the generation of designs that are more detailed and appropriate. As well, in the case of the introduction of other models of AI into GANS, that is reinforcement learning, or domain knowledge to GANS could improve the quality of the designs produced to become not only novel but also useful. In a similar vein, future works could also investigate how GANs could work better with human designers in a more interactive fashion allowing AI to integrate its capabilities and be more creative in assisting human creativity for better and faster design work at the end.

REFERENCES

[1] M. A. Ali Elfa and M. E. T. Dawood, *Using artificial intelligence for enhancing human creativity*, Journal of Art, Design and Music, 2 (2023), p. 3.

[2] L. Fu, J. Li, and Y. Chen, *An innovative decision making method for air quality monitoring based on big data-assisted artificial intelligence technique*, Journal of Innovation & Knowledge, 8 (2023), p. 100294.

[3] F. Gama and S. Magistretti, *Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of ai applications*, Journal of Product Innovation Management, (2023).

[4] L. Grilli and M. Pedota, *Creativity and artificial intelligence: A multilevel perspective*, Creativity and Innovation Management, 33 (2024), pp. 234–247.
[5] L. Huang and W. Peissl, *Artificial intelligence—a new knowledge and decision-making paradigm?*, in Technology Assessment in a Globalized World: Facing the Challenges of Transnational Technology Governance, Springer International Publishing Cham, 2023, pp. 175–201.
[6] S. Kaggwa, T. F. Eleogu, F. Okonkwo, O. A. Farayola, P. U. Uwaoma, and A. Akinoso, *Ai in decision making: transforming business strategies*, International Journal of Research and Scientific Innovation, 10 (2024), pp. 423–444.
[7] M. F. Ke, *Applications and challenges of artificial intelligence in the future of art education*, Pacific International Journal, 6 (2023), pp. 61–65.
[8] Z. Kowalczuk and M. Czubenko, *Cognitive motivations and foundations for building intelligent decision-making systems*, Artificial Intelligence Review, 56 (2023), pp. 3445–3472.
[9] D. R. Lukkien, N. E. Stolwijk, S. I. Askari, B. M. Hofstede, H. H. Nap, W. P. Boon, A. Peine, E. H. Moors, and M. M. Minkman, *Ai-assisted decision-making in long-term care: Qualitative study on prerequisites for responsible innovation*, JMIR nursing, 7 (2024), p. e55962.
[10] J. Milošević, L. Đukanović, M. Živković, M. Žujović, and M. Gavrilović, *Automated compositions: artificial intelligence aided conceptual design explorations in architecture*, in Proceedings: Geometry, graphics and design in the digital age, The 9th International Scientific Conference on Geometry and Graphics MoNGeometrija 2023, Faculty of Technical Sciences, University of Novi Sad, 2023, pp. 103–115.
[11] K. Mullangi, N. Dhameliya, S. K. R. Anumandla, V. Yarlagadda, D. Sachani, S. C. R. Vennapusa, S. S. Maddula, and B. Patel, *Ai-augmented decision-making in management using quantum networks*, Asian Business Review, 13 (2023), pp. 73–86.
[12] K. Patel, D. Beeram, P. Ramamurthy, P. Garg, and S. Kumar, *Ai-enhanced design: Revolutionizing methodologies and workflows*, Development (IJAIRD), 2 (2024), pp. 135–157.
[13] L. S. Samayamantri, *Transforming industry through innovation: A comprehensive study of cognitive-first digital factory implementations and their impact on manufacturing efficiency*, International Journal of Creative Research In Computer Technology and Design, 5 (2023), pp. 1–19.
[14] M. Shin, J. Kim, B. Van Opheusden, and T. L. Griffiths, *Superhuman artificial intelligence can improve human decision-making by increasing novelty*, Proceedings of the National Academy of Sciences, 120 (2023), p. e2214840120.
[15] D. D. Valluri et al., *Exploring cognitive reflection for decision-making in robots: Insights and implications*, International Journal of Science and Research Archive, 11 (2024), pp. 518–530.
[16] J. von Thienen, O. Kolodny, and C. Meinel, *Neurodesign: the biology, psychology, and engineering of creative thinking and innovation*, in Brain, Decision Making and Mental Health, Springer, 2023, pp. 617–659.
[17] T. Wang, Z. Ma, and L. Yang, *Creativity and sustainable design of wickerwork handicraft patterns based on artificial intelligence*, Sustainability, 15 (2023), p. 1574.
[18] X. Wenjing and Z. Cai, *Assessing the best art design based on artificial intelligence and machine learning using gtma*, Soft Computing, 27 (2023), pp. 149–156.
[19] C. Wu, R. Zhang, R. Kotagiri, and P. Bouvry, *Strategic decisions: survey, taxonomy, and future directions from artificial intelligence perspective*, ACM Computing Surveys, 55 (2023), pp. 1–30.
[20] A. Zhang, O. Walker, K. Nguyen, J. Dai, A. Chen, and M. K. Lee, *Deliberating with ai: improving decision-making for the future through participatory ai design and stakeholder deliberation*, Proceedings of the ACM on Human-Computer Interaction, 7 (2023), pp. 1–32.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: Aug 28, 2024
Accepted: Oct 16, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2253–2264, DOI 10.12694/scpe.v26i5.4879

EXPLORING FOREIGN LANGUAGE EDUCATION USING PERSONALIZED LEARNING ALGORITHMS AND DISTRIBUTED SYSTEMS BASED ON BIG DATA ANALYTICS

JIAJUN MA*

Abstract. The limitations of traditional methods result in a lack of personalization and real-world immersion in language learning, particularly in foreign language learning skills. It leads to limited conversational practices and slower development. Also, traditional methods fail to adjust the respective needs of every individual learner and varying proficiency levels. To address these kinds of problems and to enhance foreign language education, the present study supports with its unique novel framework called the Cognitive Collaborative Language Optimizer (CCLO) model, which combines the benefits of collaborative filtering, fuzzy cognitive systems, and neighborhood-based recommendation algorithms based on distributed systems and big data analytics (BDA). By using real-time data based on every learner's performance, preferences, and progress, the CCLO framework adjusts learning experiences for each individual and provides an effective path for language development. While neighborhood-based algorithms form clusters of learners with shared learning paths, on the other hand, collaborative filtering identifies patterns in learner behavior. It suggests education materials that work well for learners with comparable backgrounds. This guarantees a personalized learning experience. Managing uncertainty is an important function in CCLO, and this was performed using fuzzy cognitive systems, where learners commonly prove their partial knowledge of learning. CCLO provides the best experience by adjusting every learner's needs and offers the benefit of vocabulary, grammar, and conversation practice at the correct time. Finally, the simulation of the suggested CCLO is conducted using a seven-week study based on the Chinese as a Foreign Language (CFL) dataset. The efficacy of the suggested model shows a notable improvement in conversation proficiency. The detailed illustrations and experiments are discussed in the following sections.

Key words: Cognitive systems, collaborative filtering, foreign language learning, personalized learning, fuzzy logic, recommendation algorithms, immersive learning and AI based education.

1. Background.

1.1. Challenges in Traditional Foreign Language Learning. The effectiveness and adaption are required for successful language development, these are limited due to the number of issues with traditional foreign language learning techniques. The inability to handle different learning methods and skill levels creates a major problem. A lot of traditional techniques take standard strategy that does not consider the needs of different learners, which commonly results in interest decrement and slower language gaining progress [10, 3]. Furthermore, learners find it challenging to practice traditional skills in real-world conditions and traditional settings due to a lack of real-world involvement [15, 6]. The limited use of technology limits the ability to design effective, attractive learning environments, which is another major obstacle. Traditional approaches commonly highlight textbook-based instruction, which cannot encourage recent advancements and does not promote critical thinking skills and active learning engagement [19]. Furthermore, the lack of effective learning pathways and personalized feedback makes it difficult for learners to concentrate on their areas of weakness (SpringerLink). Because of this, learners commonly find it difficult to become confident in everyday chats, particularly in situations involving other languages where cultural details are crucial [11]. Many of these issues are addressed by the move toward digital tools and customized algorithms, which helps to encourage flexibility and practical usage

In this era of globalization, knowledge of more than one foreign language has become a key skill that aids communication in personal, academic, and professional spheres. Conventional systems used in the teaching of the foreign languages are generally characteristic of a uniform model where all learners are treated the same regardless of their variations in learning styles and techniques. Such approaches often result in ineffective learning habits, disinterest, and poor performance in oral skills and pragmatics. In the same way, students

*European College of Xi'an Foreign Studies University, Xian,710128, China, email: jiajunmainfirm12@rediffmail.com

are also unable to fully maximize their efforts because such forms of education are also void of feedback or possibilities of shifting the parameters and thus stimulating the students.

New technologies related to the active usage of big data and personalized learning algorithms bring radical improvement to the approaches taken in learning foreign languages. With the use of these technologies, processing a great deal of data and recognizing the features of the language learners and the concerns they are facing, enables each student to be provided with a learning plan that is tailor-made for him/her. The enhancement of the process consists in the connection of uniformly structured distributed systems for acquisition on-line, processing and returning the information about the learner‘s state.

The need to overcome the known disadvantages of conventional foreign language education systems is the motivation underlying this study. This study also examines and tries to identify how personalized learning algorithms coupled with big data and distributed systems can improve the process of language study. What is required is creating a modular and configurable system enabling the achievement of the stated objectives, where improving the speed and efficiency of the foreign language acquisition will not be the only goal, but enhancing a conversational experience, gaining cultural exposure, and improving comprehensively in order to meet the learners' demands as well.

1.2. Advancements in Personalized Learning Algorithms. By adapting learning opportunities for each learner, advanced personalized learning algorithms completely changed the nature of foreign language learning. These algorithms develop personalized learning ways that help support each learner's individual needs by analyzing massive volumes of data, such as learner behaviors, preferences, and progress [5, 7]. This approach guarantees that learners receive resources suitable for their current ability level and highlights the areas that need to be improved. By continuously adjusting student's growing vocabulary, these systems provide a more efficient and engaging way to learn a language. The main advantage of the personalized learning algorithm is its ability to provide real-time adjustments to the learning process [13]. While modern algorithms continuously adjust content, tempo, and difficulty according to the learner's development, traditional language learning methods commonly accept a static approach. For example, the system can highlight vocabulary or grammar rules that students find difficult and then offer exercises and materials that specifically target these areas and fill in their knowledge gaps. This makes learners by creating a highly engaged environment for learning [17, 12]. Furthermore, to improve communication skills, personalized learning systems include interactive components like dialogue simulations that imitate real-world language use. These characteristics allow learners to practice in reliable situations. The incorporation of adaptive learning technologies improves learners' language and cultural knowledge by allowing them to understand refinements that are important to learning a foreign language.

1.3. Suggested CCLO and its Contributions. By considering the impacts and benefits of the existing and current techniques, we present the advanced solution for improving foreign language proficiency, called CCLO, developed by using CF and fuzzy cognitive diagnosis and neighborhood recommendation algorithms [4, 18, 16]. With the help of various integrated, personalized learning approaches, the CCLO framework provides learners with customized, adaptive learning experiences that change in real time. CCLO guarantees that learners are guided through a customized path of language knowledge and focusing on areas that need to be improved. CCLO offers particular suggestions, which is one of its main contributions. It can leverage data from each learner's behavior and provide resources for their learning needs. For example, when a learner is strong in vocabulary but weak in sentence structure, CCLO modifies its recommendations to highlight sentence-building activities. Moreover, the problem of managing incomplete knowledge is addressed by CCLO. Learners commonly lack proficiency in specific areas of language learning, such as grammar or pronunciation. By handling this situation, we promote this suggested approach of CCLO.

2. Literature Discussions. This study [14] highlights using real materials and interactive exercises can help students become more proficient listeners in language learning environments. With the use of technology, educators can create personalized, interesting listening activities that meet each student's needs and help increase motivation and engagement. To improve English listening skills, a personalized learning path system is presented in [8], which adjusts learning paths according to the user's preferences and level of knowledge. The system recommends relevant listening materials, such as podcasts or audiobooks, using natural language processing (NLP) approaches. This dramatically improves expertise and student engagement. The study

displays student satisfaction levels and the system's recommendation accuracy, from 85% to 92%. This study [9] compares collaborative filtering with content-based filtering algorithms to investigate how recommender systems are used to improve English language learning resources in higher education. The outcomes show the balance between recall and precision, with content-based methods providing more complete material coverage. The limitations of traditional e-learning systems, which commonly fail to adjust the unique needs of each learner, are clearly described in study [1]. The novel personalized recommender system NPR_eL is suggested to adapt learning materials according to a student's background, preferences, and memory capacity, which is introduced in this work. The study [2] introduces a customized intelligent mobile learning system (PIMS) that uses a learner's reading proficiency to suggest English news articles. Using a fuzzy item response theory, the system assesses learners' abilities and suggests new words to improve vocabulary acquisition.

3. Methodology.

3.1. Suggested CCLO design. The suggested CCLO structure is developed through three important key features: Fuzzy cognitive systems, neighborhood-based recommendation algorithms, and collaborative filtering.

The Learner Profile Module is the first module of the design, it helps to collects real-time data based on vocabulary, comprehension and conversational performance for every user. The Collaborative Filtering Engine uses this data to find trends and preferences between learners who are similar to the user and recommends learning resources such as vocabulary exercises or conversation scenarios, that are suitable for the user's level of proficiency.

At the same time, learners' confusing inputs are processed using the Fuzzy Cognitive System, which effectively modifies the difficult level and target areas of learning. For example, if the learner finds it difficult to understand grammar but performs well in vocabulary, the fuzzy system balances this by providing more grammar exercises and maintaining vocabulary at the same time. Neighborhood-Based Recommendation is the third component that helps cluster students regarding comparable performance courses. It generates a content recommendation based on how learners in the similar neighborhood are processed, and ensuring an improved learning path.

The AI Interaction Module also presents interactive conversational agents to make learners practice speaking, get real-time feedback, and modify their responses to real-world application. Finally, the Evaluation Module provides an exact outcome of suggested CCLO. Figure 3.1 displays the exact representation of suggested CCLO.

Further justification for the need for this method is the lack of any system that can efficiently customize language learning for individual students as this is always done in small magnitudes. Given that there are learners who have different command over the vocabulary or grammar, or even their conversational skills, the CCLO framework makes sure that each learner is taken along a unique, appropriate learning path that meets their needs. How the CCLO is implemented calls for the grouping of learners who have similar patterns of learning and through collaborative filtering, effective teaching aids are brought forth to ensure that the learning process is efficient. Fuzzy cognitive systems mitigate barriers arising from incomplete language learning by attending to spatial partial information and uncertain objectives.

Such efficacy of the CCLO framework is primarily due the focus on extending the boundaries of conversational capabilities as all language learners want to reach the goal. Due to the shift of focus in language learning towards the interactions over the internet, methods like CCLO are very crucial in opening up better and faster ways of learning that deal with the problem of difficulties each learner experiences.

3.2. Suggested CCLO Implementation.

3.2.1. Recommendation model for foreign language learning (CFL) based on fuzzy cognitive diagnosis. Binary scoring is mainly applicable to objective questions in traditional cognitive diagnostic models (CDMs), but it is not enough for foreign language learning, speaking, and writing. A fuzzy cognitive diagnostic model is presented to overcome these limitations. Fuzzy set theory is used mainly in CCLO to handle partial knowledge and inaccuracy. A learner's ability is denoted by a continuous value between 0 and 1, which indicates the student's partial or complete knowledge of particular language abilities rather than a binary approach. In CFL learning, every knowledge point k denotes each of the vocabulary, grammar, and syntax in a corresponding fuzzy set $(sl, \mu_k)$where μ_k is the membership function that specifies each learner's proficiency level, and sl is the

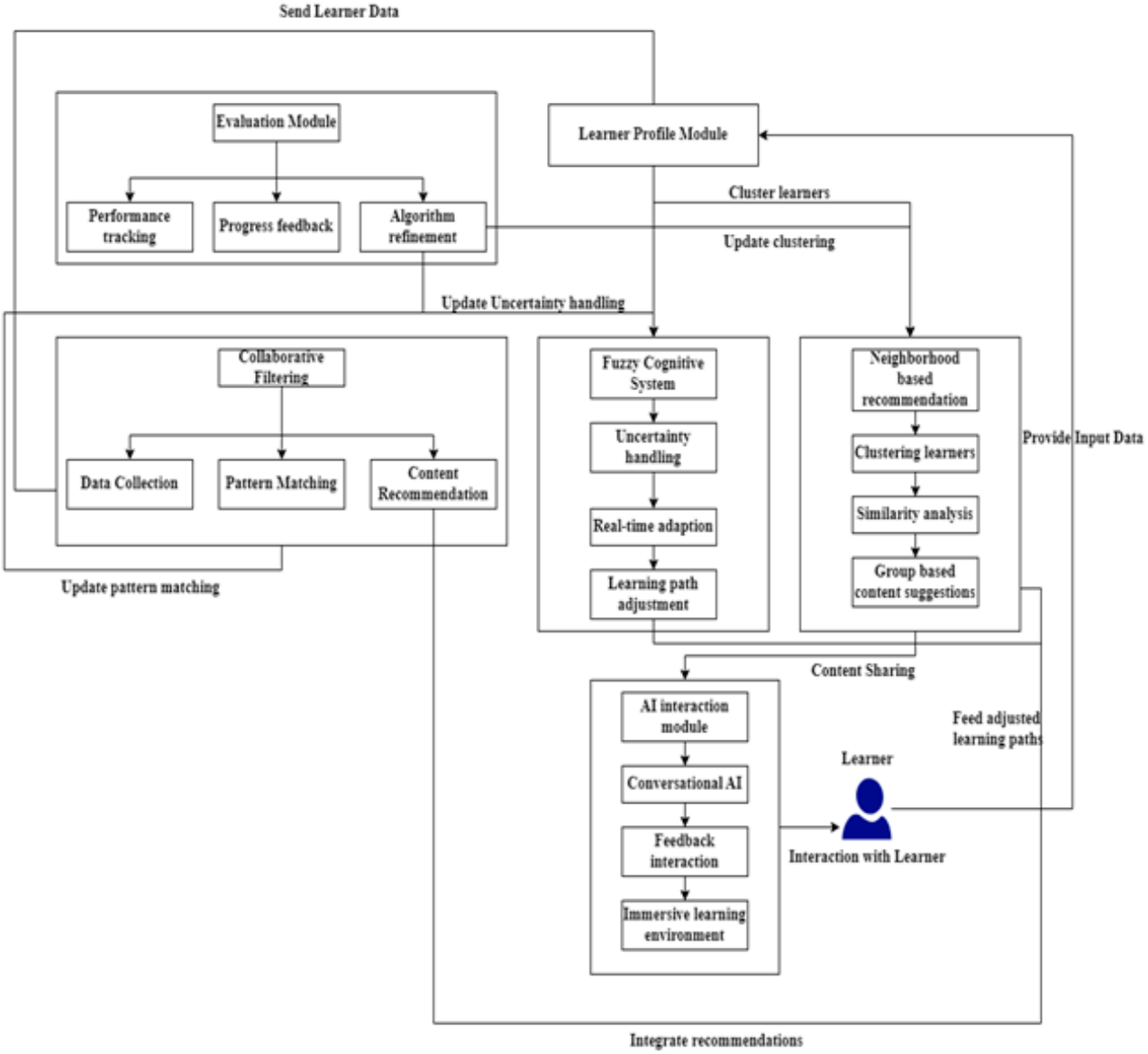


Fig. 3.1: Suggested CCLO Workflow

learner set. The logistic function is used to calculate the proficiency of the Fuzzy Cognitive Diagnosis Model pa_{slk}, which is comparable to the two-parameter logistic model in Item Response Theory (IRT).

$$pa_{slk} = \frac{1}{1 + \exp[-1.7\,(pa_{slk}\,(\vartheta_m - db_{slk}))]} \tag{3.1}$$

Here, pa_{slk} is the proficiency of the learner, ϑ_{sl} denotes the learner latent ability and db_{slk} highlights the difficulty of knowledge point k.

In the fuzzy cognitive model, we consider the objective and subjective questions. Students must observe all relevant data points to answer objective questions, whether they are correct or incorrect. We apply the linkage-type concept for this, where the student's minimum proficiency in each assessed knowledge point is

used to calculate the proficiency $p\eta_{sln}$ of question n by learner sl.

$$p\eta_{sln} = \min_{1\leq k\leq K}(\mu_{sl}(k)) \tag{3.2}$$

According to subjective questions, speaking and writing are examples of scores that depend on partial knowledge. The adequate assumption is applicable in this situation, and the maximal proficiency across knowledge points is used to calculate the level of proficiency $p\eta_{sln}$.

$$p\eta_{sln} = \max_{1\leq k\leq K}(\mu_{sl}(k)) \tag{3.3}$$

The approach additionally handles guessing and errors students make when answering subjective and objective questions. The probability that a learner sl will correctly respond to question n in an objective question which is expressed as

$$pr(r_{sln} = 1) = gp_n + (1 - gp_n)p\eta_{sln} \tag{3.4}$$

Here gp_n is the guessing parameter and $p\eta_{sln}$ is the learner's knowledge of the question. In subjective questions, a Gaussian distribution is used to simulate the probability distribution of learner scores.

$$pr(r_{mn} = s) = N + (s|p\eta_{mn}, \sigma^2) \tag{3.5}$$

$N + (s|p\eta_{sln}, \sigma^2)$ is a normal distribution centered at $p\eta_{sln}$ with variance σ^2, sl is the learner score on the subjective question.

The CCLO system incorporates several real-time information and enhances the learning by making it more personal for the users. In the case of performance data, it may consist of results of quizzes and exercises as well the time and accuracy of responses during language activities. Other measures focus on the behavioral data, such as how much idle time users have on the task and which specific areas they seem to perform less well, and the duration of utility of the platform. Learning data consists of how learners interact with other people by way of role plays, conversations, and responses to feedback given by the trainer in person or electronically. In effect, the system demands preference data from users and in this case, learners outline style of learning they are comfortable with, content they will require and feedback they expect. It might also entail gathering social and collaborative data by performing group activities and group tasks and as for the uncertainty facet, this relates to fuzzy cognitive systems that recognize partial understanding and areas of confusion that learners encounter.

After the information is collected, several algorithms are undertaken to process the information in the database. Collaborative filtering establishes associations among the learners to suggest suitable learning resources. Fuzzy cognitive systems work in situations of uncertainty by modifying task difficulty depending on the performance of the learner. Learner progress and preferences are used in neighbourhood recommendation systems to enrol learners with such characteristics who would benefit from learning materials and peer interaction. The system embeds feedback in the form of continuous updating of the learning trajectory as the learner progresses with an aim of enhancing the learning experience actively rather than passively.

3.2.2. Neighborhood-Based Recommendation Algorithms. Based on their performance and proficiency levels, the neighborhood-based recommendation algorithm identifies similar learners and suggests individualized learning materials. Calculating the similarity between learners based on their proficiency between knowledge points is an essential step.

The similarity between two learners sl_1 and sl_2 based on the proficiency in knowledge points can be calculated using

$$sim(sl_1, sl_2) = \frac{\sum_k a_{sl_1k} \bullet a_{sl_2k}}{\sqrt{\sum_k a_{sl_1k}^2} \bullet \sqrt{\sum_k a_{sl_2k}^2}} \tag{3.6}$$

where a_{sl_1k} and a_{sl_2k} are the proficiency of learners sl_1 and sl_2 on knowledge point k.

According to Pearson's correlation coefficient-based similarity calculation, it considers the mean proficiency of students across all knowledge points, was expressed as

$$sim\left(sl_1, sl_2\right) = \frac{\sum_k \left(a_{sl_1k} - \overline{a}_{sl_1}\right)\left(a_{sl_2k} - \overline{a}_{sl_2}\right)}{\sqrt{\sum_k \left(a^2_{sl_1k} - \overline{a}_{sl_1}\right)^2} \bullet \sqrt{\sum_k \left(a^2_{sl_2k} - \overline{a}_{sl_2}\right)^2}} \tag{3.7}$$

Here $\overline{a}_{sl_1}$ and $\overline{a}_{sl_2}$are the mean proficiency of students sl_1 and sl_2 across all knowledge points.

3.2.3. Fuzzy cognitive and neighborhood-based learning model for CFL. Once the student similarity has been identified, the model uses proficiency and similar learner performance to select questions and learning resources for each student.

For each learner, sl, the system suggests the knowledge points of k based on their current proficiency and the difficulty of learning materials. The cosine similarity between learners' proficiency and difficulty of the test question is calculated by

$$dist\left(sl, n\right) = \frac{\sum_k \emptyset_{slk} \bullet \diagup_{nk}}{\sqrt{\sum_k \emptyset^2_{slk}} \bullet \sqrt{\sum_k \diagup^2_{slk}}} \tag{3.8}$$

Here $\emptyset_{slk}$ denotes the student proficiency of knowledge point k and $\diagup^2_{slk}$ denotes the difficulty of knowledge point k in the recommended content n.

Fuzzy cognitive systems (FCS) are excellent when it comes to overcoming the problems of uncertainty in language learning, those are appropriate for working with undefined or uncertain, incomplete information typical for such situations as education. In language learning, students do not have absolute knowledge of the different aspects of grammar, vocabularies, and conversation, and can therefore have different fluency levels. This partial understanding may originate from the fact that the learner does not recall a specific grammar rule or cannot remember when a given word seems contextually appropriate. Binary or deterministic (this is right or this is wrong) approaches to the problem do not comprehend such insights whereas however, this is where fuzzy systems come into its consideration and treatment of such vagueness.

FCS does this by treating the different variables as belonging to certain degrees of membership as opposed to simply categorizing them as either true or false. For instance, with regard to learning a language, rather than using a flip chart with a straight indication of right and wrong for a learner's grasp of a certain concept, a fuzzy logic diagram would show how much a learner has mastered the concept probably on a scale of between 0 and 1. Thus learning system determines such optimal responses or exercises that are not overboard difficult or so easy for that matter but are just right basing on his/her current level of competence.

Fuzzy systems resolve this by permitting the learner to possess or be assessed to have a "good understanding" or "needs improvement" range of language proficiency allowing partial membership in different language knowledge categories. It is also true that learners might not have a full-fledged picture of things all the time. Fuzzy systems are designed to use part of information thereby making their application conducive when giving feedback based on the level of knowledge and performance of the learner which is often in a state of flux hence bridging the existing learning gaps more efficiently.

Personalization and Adaptation. In contrast to a more rigid approach, fuzzy systems facilitate the treatment of the learners' learning behaviors and performance with some leeway allowing for relative decomposition of the learning structure. Just as it is typical with all learners, there are various levels of mastery over grammar, praction, or vocabulary among learners which often times are fuzzy logical developmental and hence improveover time.

Understanding how a human thinks. The human logic or reasoning is complex and is followed up with some degree of uncertainty or consistent with incomplete data and so fuzzy systems are built around this concept. Thus, they are particularly useful in teaching situations as students usually have partial information about the right answer and tend to be confused.

3.3. Hybrid Collaborative Filtering (HCF) Algorithm under CCLO. Collaborative filtering technique, which helps to continuously monitor learner behavior and generate recommendations based on similarities

across users and resources, is one of the most popular recommendation approaches, particularly in personalized learning platforms.

Following the execution of Neighborhood-Based Recommendation Systems and Fuzzy Cognitive Diagnostic Models, CF algorithms are the additional advantage contributed by CCLO. The neighborhood-based and fuzzy cognitive models evaluate each learner's proficiency and recommend resources according to their observation of certain knowledge points. Again, the CF algorithm adds another level of personalization by considering similar learners' more general learning habits, which is a further effective refinement process performed under CCLO. The hybrid CF depends on the integration of two classifications of filtering. User-based filtering and item-based filtering.

Real time data accessibility for working with learners in a learning environment is one of the elements within the Cognitive Collaborative Language Optimizer (CCLO) framework that makes the learning process more personalized. The system utilizes big data analytics in a distributed fashion and captures, processes and analyzes learner engagement with the language learning activities in real-time.

All these activities go on in a real time environment while the system collects and captures data as learners work on completing exercises, partaking in discussions, working with learning content and providing as well as receiving feedback and comments. The content of the learning path and the tasks being faced by the learner are monitored so that changes are made actively and timely based on the so that the particular content and tasks suited for the learners are appropriate relative to their current state of progression in learning.

In the CCLO framework collaborative filtering and fuzzy cognitive systems are applied using this real time data as a way of individualizing the learning process. These algorithms work hand in hand to make accurate predictions on what materials and activities will be most suitable for the learner and the same algorithms will reposition with respect to the learners' proficiency level.

3.3.1. User-based filtering. Based on the performance and interactions with learning materials, similar learners are identified by the User-Based CF algorithm. By comparing students' results from a variety of tasks, such as grammar, vocabulary, and listening comprehension, the algorithm can calculate the similarity between learners. To predict the interest of learners in a specific source it can be expressed as

$$pr_{u,j} = \overline{ar}_u + \frac{\sum_{v \in knn(u)} siml(u,v) \bullet (ar_{v,j} - \overline{ar}_v)}{\sum_{v \in knn(u)} siml(u,v)} \tag{3.9}$$

Here $pr_{u,j}$ is the predicted score of learners u for resource j, whereas $\overline{ar}_u$ and $\overline{ar}_v$ are the average ratings of learners u and v and $ar_{v,j}$ is the actual ratings of resource j by learner v. $siml(u,v)$ highlights the similarity between learners u and v based Pearson's.

3.3.2. Item-based CF algorithm. The calculation of similarity between various learning resources such as vocabulary tests and reading comprehension activities is done by item-based CF. For learners who have showed proficiency in a particular exercise, this is helpful because it allows the system to suggest similar materials that have assisted other learners to improve. The similarity between two resource i and j is expressed as

$$siml\,(i,j) = \frac{\sum_u ar_{u,i} \bullet ar_{u,j}}{\sqrt{\sum_u ar_{u,i}^2} \bullet \sqrt{\sum_u ar_{u,i}^2}} \tag{3.10}$$

With the use of this technique, instructors can provide recommendations for activities or courses which are effective for other learners with similar learning needs.

4. Evaluation and Experiments.

4.1. Experimental Setup. The present study is evaluated based on CFL dataset, inspired from [4]. According to the features, we perform evaluation for the suggested CCLO model. Table 4.1 presents the important details about dataset features.

Table 4.1: Dataset Details

Category	Details
Participants	10 students, aged 18-22, no prior experience with Chinese
Participant Information	Level-1 Chinese course, minimal outside exposure
	Lesson on shopping (32 vocab words, 6 sentences)
Learning Session	4 sessions in Panoramic Scenes and Virtual World
	Vocabulary and conversation via task-based teaching
	AI role-play for listening, comprehension, speaking

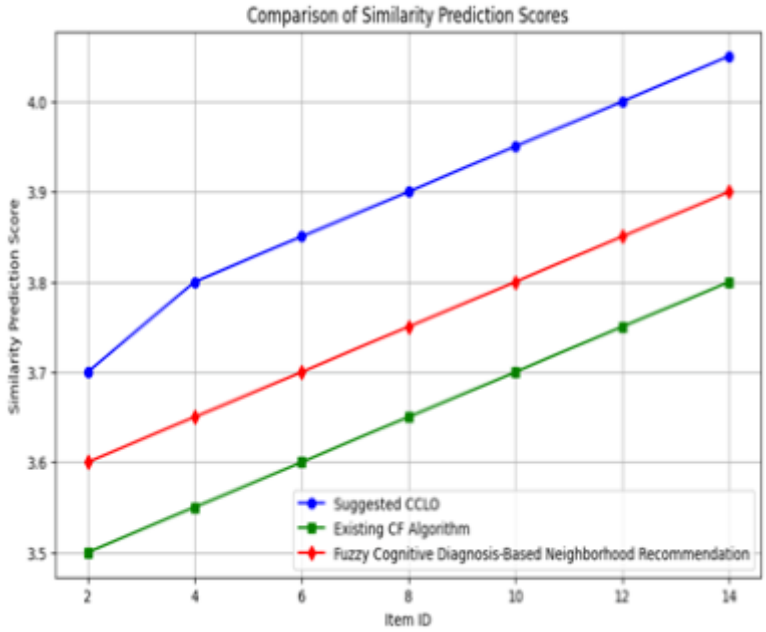


Fig. 4.1: Similarity Prediction Comparison

4.2. Evaluation Criteria. Figure 4.1 presents the efficacy of suggested CCLO in terms of similarity prediction scores across different item id. Due to the combining strength of both CF and fuzzy logic, which considers the partial knowledge and learning gaps, this efficacy is possible. The suggested model provides more effective learning experience by perfectly adjusting to the user behavior and individual performance. As a result, CCLO achieves best scores when compared with existing techniques in maintaining the prediction stability and improving overall recommendation accuracy.

Fig 4.2 presents the CCLO in improving the personalization recommendations by considering both the previous interactions and partial understanding of learners. Here the modified cosine similarity indicates the similar recommended resources. Through the maintenance of uncertainty and effective adaption of learners, performance the suggested CCLO obtains better language proficiency outcomes.

Figure 4.3 indicates the MAE scores of nearest neighbors n. According to the figure, suggested CCLO shows the low MAE scores, which highlights that the CCLO prediction of learner's performance or learning progress are more accurate. Similarly, Figure 4.4 shows the RMSE scores comparison of various models. Based on the figure we observe that, the suggested CCLO predicts few large errors, which means it provides more reliable outcomes of learning item effectiveness for learners.

Figure 4.5 presents the index comparison scores. The suggested CCLO considerably outperforms the existing CF, and fuzzy models. It shows its effectiveness in terms of high precision, recall, coverage and

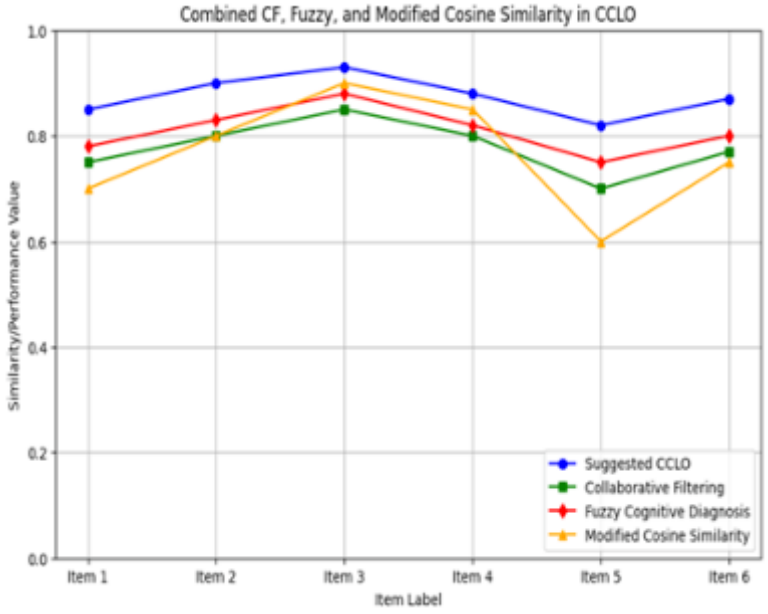


Fig. 4.2: Cosine Similarity Value

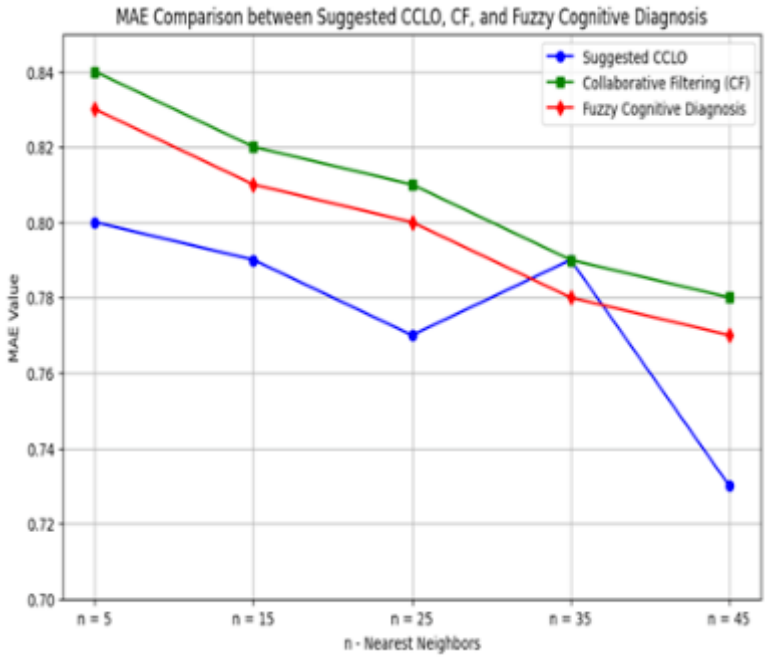


Fig. 4.3: MAE Comparison Scores

popularity.

5. Conclusion. The suggested study presents the effective fusion model called CCLO which combines fuzzy cognitive diagnosis and neighborhood recommendation algorithm with collaborative filtering algorithm

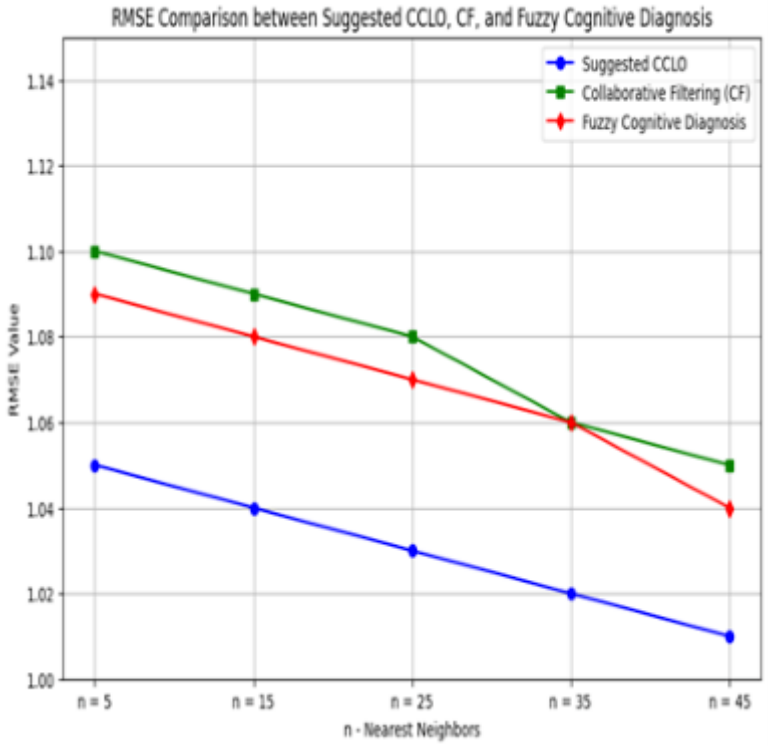


Fig. 4.4: RMSE Comparison

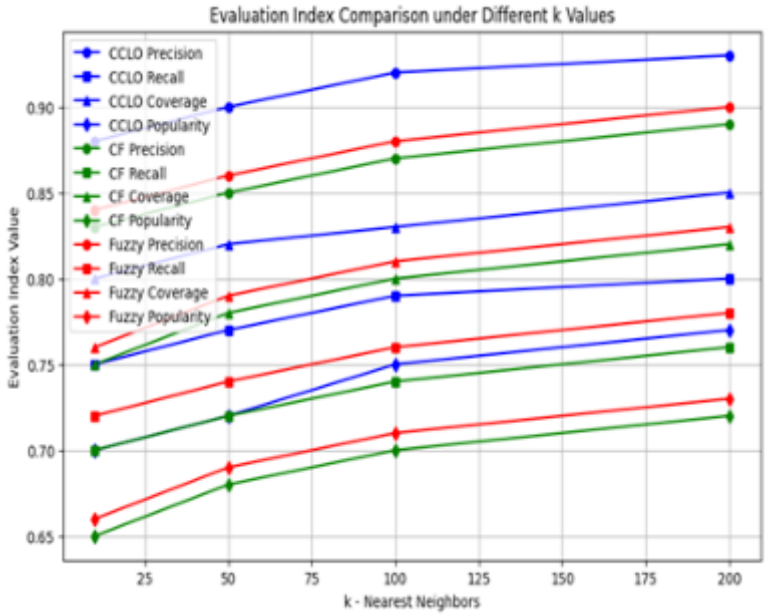


Fig. 4.5: Evaluation Index Scores Comparison

for improving foreign language education. According to this, we use CFL dataset to evaluate the suggested CCLO. The present model is a combination of both fuzzy and CF algorithms, so the results of the suggested model is effectively compared with both CF and fuzzy models. The experiments proves that the effectiveness of proposed model in improving the expectable proficiency of learners. But however, this type of effective fusion also leads to some common drawbacks, such as the fusion model is just a novel term; to implement this in real world application, we need more refined algorithms and the adaption of advanced techniques to execute. During the language processing the system feels more complexity to process this due to the advanced filtering and recommendation combinations. Apart from that, it is a most powerful contribution to the field of language learning and personalized recommendation systems. Future research is most necessary to rectify the limitation of this suggested model.

Major limitation of CCLO framework is forming appropriate learner clusters using neighborhood-based recommendation algorithms. In those cases when a dataset was sparse or learners had different styles and level of progress, it was difficult to group learners into cliques. Because of this, the particular learning paths created for these clusters were sometimes ineffective, causing lesser engagement or slower progress of these learners. In future research, Given these obstacles, it may prove worthwhile to examine new clustering techniques in subsequent studies, including, for example, deep learning methods or k m-hybrid designs, which practice several clustering strategies together. That would increase the efficiency of grouping learners and provide for more targeted and successful learning environments.

Funding Statement. Research on the Discourse Construction Characteristics of "Shaanxi Image" in Spanish Media from the Perspective of Corpus in the Special Research Program of Shaanxi Provincial Department of Education, Project No. 21JK0303

REFERENCES

[1] S. Benhamdi, A. Babouri, and R. Chiky, *Personalized recommender system for e-learning environment*, Education and Information Technologies, 22 (2017), pp. 1455–1477.
[2] C.-M. Chen, S.-H. Hsu, Y.-L. Li, and C.-J. Peng, *Personalized intelligent m-learning system for supporting effective english learning*, in 2006 IEEE International Conference on Systems, Man and Cybernetics, vol. 6, IEEE, 2006, pp. 4898–4903.
[3] B. Czerkawski and M. Berti, *Language learning in the 21st century: current status and future directions*, Language learning and professionalization in higher education: pathways to preparing learners and teachers in/for the 21st century, 11 (2020).
[4] R. R. Divekar*, J. Drozdal*, S. Chabot*, Y. Zhou, H. Su, Y. Chen, H. Zhu, J. A. Hendler, and J. Braasch, *Foreign language acquisition via artificial intelligence and extended reality: design and evaluation*, Computer Assisted Language Learning, 35 (2022), pp. 2332–2360.
[5] S. Hashim, M. K. Omar, H. Ab Jalil, and N. M. Sharef, *Trends on technologies and artificial intelligence in education for personalized learning: systematic literature*, Journal of Academic Research in Progressive Education and Development, 12 (2022), pp. 884–903.
[6] R. M. Hein, C. Wienrich, and M. E. Latoschik, *A systematic review of foreign language learning with immersive technologies (2001-2020)*, AIMS Electronics and Electrical Engineering, 5 (2021).
[7] M. Jian, *Personalized learning through ai*, Advances in Engineering Innovation, 5 (2023).
[8] X. Li and B. Zhang, *Personalized learning path recommendation algorithm for english listening learning*, Journal of Electrical Systems, 20 (2024), pp. 2188–2199.
[9] T. Liu, Y. Huang, and X. Tong, *Optimizing personalized recommendation of college english learning resources using recommender system algorithms*, Journal of Electrical Systems, 20 (2024), pp. 640–647.
[10] E. Piechurska-Kuciel, E. Szymańska-Czaplak, and M. Szyszka, *At the crossroads: challenges of foreign language learning*, Springer, 2017.
[11] R. A. Schulz, *The challenge of assessing cultural understanding in the context of foreign language instruction*, Foreign language annals, 40 (2007), pp. 9–26.
[12] D. Shawky and A. Badawi, *Towards a personalized learning experience using reinforcement learning*, Machine learning paradigms: Theory and application, (2019), pp. 169–187.
[13] S. Tatineni, *Recommendation systems for personalized learning: A data-driven approach in education*, Journal of Computer Engineering and Technology (JCET), 4 (2020).
[14] L. Uzun, *Original paper enhancing foreign language learners' listening skills through technology: A sample lesson.*
[15] S. Vijayakumar, *Foreign language learning then, now and after covid-19: An exploration of digital tools to augment the receptive and productive skills of language learners*, in Technologies, Artificial Intelligence and the Future of Learning Post-COVID-19: The Crucial Role of International Accreditation, Springer, 2022, pp. 283–302.
[16] K. Yang, *Research on personalized english language learning based on artificial intelligence*, Applied Mathematics and Nonlinear Sciences, 9.

[17] L. Zhang, J. D. Basham, and S. Yang, *Understanding the implementation of personalized learning: A research synthesis*, Educational research review, 31 (2020), p. 100339.
[18] Q. Zhang, *Construction of personalized learning platform based on collaborative filtering algorithm*, Wireless Communications and Mobile Computing, 2022 (2022), p. 5878344.
[19] B. Zhao, *The role of classroom contexts on learners' grit and foreign language anxiety: online vs. traditional learning environment*, Frontiers in Psychology, 13 (2022), p. 869186.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: Sep 11, 2024
Accepted: Sep 17, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2265–2274, DOI 10.12694/scpe.v26i5.5022

DISTRIBUTED SYSTEMS FRAMEWORK FOR PACKAGING DESIGN INNOVATION USING VISUAL PERCEPTION AND ALGORITHM OPTIMIZATION

YING HUANG*

Abstract. Food consumed by humans is becoming more and more customized to fit each person's unique demands, with a vast array of product labels readily available. As a result, many companies are beginning to concentrate on enhancing the practicality of contemporary packaging. Throughout the lifetime of a user-product engagement, sensory paradigms and emotional responses may shift. Traditional product packaging layout is largely based on the designer's emotional imagination and prior events; however, it is limited by uncontrollable content and a lack of expert advice; most earlier studies involving mental analysis of images focused on predicting the most prevalent viewers feelings. There are situations when an image's general impression is insufficient for practical purposes since the emotions it arouses are very subjective and differ from viewer to viewer. The proposed methodology uses Genetic Algorithm based Multi-Layer Ant Colony Optimization for analysing visual perception and emotion perception to identigy the senses of human being. A significant set of images called Image-Emotion-Social-Net is utilized to assess categorized and multivariate attitude representations. The collection, which comes from Flickr, has more than a million photos uploaded by more than 9,000 members.The results of this dataset's research indicate that the suggested approach performs better in personalized emotional identification than several contemporary methods. In comparison to other current methods, the experimental findings demonstrate that the suggested approach obtains a high packaging layout excellence rate of 95.1%, a performance success rate of 98.5%, and a mean square error rate of 1.5%.

Key words: Distributed Systems Framework, Packaging Design,Visual Perception, Algorithm Optimization

1. Introduction. In essence, packaging layout is the translation of symbols into images. It is a useful tool for promoting goods and enhancing the way they look due to its logic and useful characteristics. Pictures can express a wide range of feelings and intricate interpretations [24].Earlier research on emotional perception analysis have mostly focused on predicting the most prevalent feelings among viewers [4].This general emotion isn't always sufficient for practical purposes because the sentiments a picture inspires are very subjective and differ from viewer to viewer [9]. The main challenges facing picture emotion analysis are subjective assessment and emotional perception. Multimedia methods are becoming more and more common and showing varied growth [6]. The packaging of a good is one of the best ways to market it. Because of consumers' inescapable drive for material prosperity and their more accepting attitudes toward purchase, packaging layout has become an increasingly important tool in the marketing of goods [14]. Visual communications must therefore be incorporated into packaging designs. Communicating visually delivers visual information quickly and efficiently, forming buyers' first impression of the product.

The main objective of this research is to develop an advanced system applying distributed systems to optimize packing design which seeks to personalize and dynamically engage the user and product through our designed sensory paradigms and emotions. Genetic Algorithm-Multi-Layer Ant Colony Optimization (GA-MLACO) will be used to evaluate emotional and visual perception generating personalized insights on human sensory perception and emotional behaviours.

The primary objective of this method is to use visualization conversion to change the initial information into a format that models trained on deep learning can use.This technique increases the model's resilience and can withstand adversarial perturbations to some degree [11, 21]. The exterior appearance of brand packaging in today's market climate has a significant influence on consumer perceptions of items as well as their market performance. Any superficial flaw has the potential to harm a brand's reputation and decrease consumers' propensity to buy. As a result, it's critical to identify surface flaws in brand packaging promptly and precisely

*Public Courses Department, Xi'an Traffic Engineering Institute, Xi 'an Shaanxi, 710300 China (yinghuangview12@rediffmail.com)

[18, 22]. A two-stage attention-based feature integration net was presented by author [17, 20] to identify surface flaws in trademark package. It first emphasizes how crucial it is to find surface flaws in product marketing.As customers' expectations for good quality and appearance grow, brand packaging's surface quality has a significant influence on what people decide to buy.

Food products need to be packaged in a way that wraps and protects them in order to protect them against ecological, microbiological, and transit dangers. An essential safety factor for packaging materials that come that come into contact with food is being as passive as possible with minimal food-material interactions. Additionally, shipping, logistics, advertising, and environmental advertising are all impacted by packaging [1, 12]. Adopting packaging that is environmentally friendly will need more work. First things first: when designing container for an atmosphere of comfort, remember the functions that it fulfills. The main goals of packaging are security, preservation, launching and transportation in addition to purchase, marketing, service, and guarantee.

In anticipation of this trend toward acceptance, a brand-new area of computer vision known as computable image aesthetically has emerged.Relying on precise assessments and computations of picture aesthetics, human-computer communication success depends on users' visual interactions with image systems that communicate [27, 2]. As a result, designers are better able to assess and produce aesthetically pleasing statements that complement users' feelings. This study proposes LSTM and hypergraph machine learning techniques that, through combining audio, visual, written, and affective input for a more comprehensive view of emotional circumstances, may acquire the ability to predict multiple users' feelings at once. A hypergraph is a graph that has more vertices connected by hyperedges than a basic network. The main contribution of proposed method is given below:

1. The present research presents a novel approach to the analysis of visual and emotional sensations in personalized packaging design: Multi-layer ant colony optimization based on Genetic Algorithms.
2. Through the utilization of a substantial dataset, Image-Emotion-Social-Net, which is derived from more than a million photographs on Flickr, the suggested method seeks to improve the precision of emotional detection in user-product interaction.
3. In contrast to conventional packaging layout, that is constrained by the designers' subjective emotional creativity, the new approach offers a more tailored and data-driven way to record unique sensations.
4. The testing results, which show a package design perfection rate of 95.1%, an overall rate of achievement of 98.5%, and a mean square error rate of 1.5%, show that the suggested method works noticeably better than current methods.

The rest of our research article is written as follows: Section 2 discusses the related work on various Distributed Systems Framework, Packaging Design, Visual Perception, Algorithm Optimization. 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. In order to compare the efficacy of Affectiva, Amazon Rekognition, Baidu Studies, Face++, and Microsoft Azure in identifying feelings the author conducted two experiments. First, how well the methods classified images from three different, highly uniform face emotion datasets. The results imply that investigators and programmers may use commercialized face expression detection systems. Using multimodal time-lapse infrared video sequencing, the author [8] presents a three-stage human-computer interaction (3s-HCI) method for identifying emotions and making entertainment recommendations.

Initially detect the palate, students, and nose using the Faster R–CNN structure. The facial ROIs in this thermal clip are monitored by the Multiple Instances Learning (MIL) technique. We discovered, through the use temperatures data and rival categorization systems, that our suggested methodology consistently yielded superior outcomes. One of the main limitations of the study is the limited sample size, which makes it difficult to draw generalizations about feelings among people. In order to identify EEG emotions, the author [26, 25, 16] suggested five classical machine learning (CML) and five ensemble machine learning (EML) methods. The free DREAMER collection contains nine sentiments that can be used for ML-based training of systems.

The food supply chain and the design of brand packaging are becoming more and more dependent on big data and computer-aided design (CAD) technologies. The effectiveness of the food supply chain and the

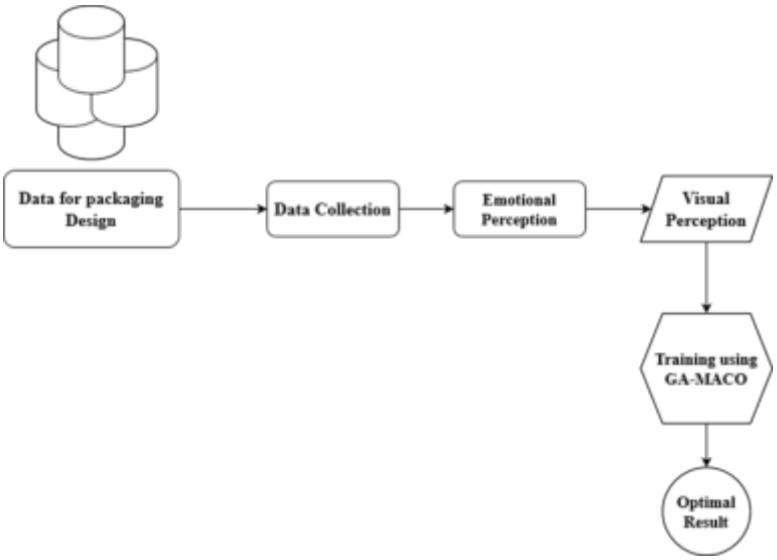


Fig. 3.1: Architecture of proposed method

creativity of brand packaging designers are further enhanced by the use of intelligent packaging technologies. The function of CAD-based and big data-driven autonomous packaging systems in the food supply chain and brand packaging design was investigated by author [10, 5]. The design of brand packaging has additional options thanks to the use of CAD software. CAD software allows architects to create precise models and simulate them, which speeds up the building and optimisation of complicated structures.As a result, packaging design becomes more sophisticated and unique while also increasing design effectiveness and satisfying the wide range of customer concerns about the look and usability of the product [23, 7, 3]. Simulated testing on packing effectiveness, including physical and barrier performance, can also be carried out with CAD technology to guarantee the dependability of packing in real-world applications.

A real-time system called the YOLO (You Only Look Once) method is used to determine broken or faulty goods by analyzing photographs of packing components for external flaws. The device evaluates packaged box quality instantaneously and can be applied in factories and manufacturing lines. The program consists of three basic steps: picture evaluation, item acknowledgment, and robot-assisted faulty item sorting [21,22]. A sorting robot system removes the damaged goods using a surface defect detection technology that combines machine vision with artificial intelligence [13]. The study [28] primarily focuses on how the use of blockchain technology handles the verification of coffee packages and promotes sustainable inclusion of the chain of supply for coffee.

The creation of this technology was sparked by the author's extensive empirical research on emotion recognition and prompt-based sentiment analysis (see [19]), which looks into the prejudices of large-scale pre-trained language model (PLM) technological advances towards emotional computing. The examination also piqued interest in prompt-based classification applications such as emotion recognition [19]. The connection among the layout of the packaging and the probability of consumers shopping online returning to Politeknik IT&B Medan is examined by the author [15]. The sample for this study is made up of 47 students from Politeknik IT&B Medan's Faculty of Business Administration, who were selected throughout both the second and forth quarters of the school year.

3. Proposed Methodology. The proposed methodology for Distributed Systems Framework for Packaging Design Innovation Using Visual and Emotion Perception and Genetic Algorithm based Multi-Layer Ant Colony Optimization. Initially, the packaging design dataset is collected and then the collected data is preprocessed. Next, the Emotion perception computing is evaluated and then the GA-MACO is used for training the dataset. Figure 3.1 shows the architecture of proposed method.

Customized food packaging design is one such trend that has caught the eye of many businesses. Companies are adopting this strategy due to strong effective performance. Instead of aiming for a mass-market approach, businesses can use customized food packaging that includes tailored design, logo or even messaging. The reason for this is that customers are now more concerned with originality, aesthetics, and emotions that the product can potentially spark. Packaging is not purely regarded as a wrapper, but it is an integral part of the product provides value during purchasing, and guarantees.

3.1. Dataset Collection. The suggested technique made use of the Flickr picture collection, whose Flickr30k dataset has become the industry standard for sentence-based image categorization. flickr-image-dataset at kaggle.com/datasets/hsankesara[24]. The present research presents Flickr30k Organizations, an addition to the 158k captions in Flickr30k that includes 276k manually annotated bounding boxes and 244k reference chains connecting image-wide citations of the same objects. They enable us to set a new benchmark for image entity-mention translation.With the integration of picture-text embeddings that frequent object sensors, a color classification algorithm, and a bias for larger objects in object selection, they offer a solid basis for this task. We show that, although our fundamental model is equally precise as more complex state-of-the-art designs, its advantages on image-sentence recall as well as additional problems do not translate to these new methods, emphasizing their limitations and the need for further research.

3.2. Emotion Perception. In order to handle the expectations of consumers and the multifaceted nature of society, the article suggests that the design method of thought be highly regarded.It has been widely employed by corporations and charities to solve social and economic challenges. Over the past ten years, the idea of design thinking has been more and more well-liked among scholars as an approach that can provide novel solutions to issues in exceedingly complex environmental and social situations. Design thinking is well suited for training multidisciplinary teams. Design teams prioritize the relationships among subjects and end customers after doing a psychological assessment first. Recognizing the issue is the first step in the process, which then moves on to solving it, motivation, innovation, and operation, and finally, execution.For every experiment, there were five steps involved.Choosing the item at the store; Opening the package; Preparing and eating the meal; completing another buy.

3.3. Genetic Algorithm. A search heuristic based on Charles Darwin's theory of natural selection is called a genetic algorithm. The method emulates the process of natural selection, whereby the most suited people are selected for procreation in order to generate the offspring of the subsequent generation. The category of evolutionary optimization algorithms includes genetic algorithms (GA). There are three phases to it: reproduction, survival, and fitness.

The genetic algorithm's process is depicted in Figure 3.2. Individual populations are created initially. Fitness value is used to calculate each individual. After individual selection using genetic operators, the fitness calculations are performed on the newly selected individual.

3.3.1. Multi-Layer Ant Colony Optimization algorithm. Setting initial values for variables like α,β, p_g, p_1, q_0, and n ant in the first phase. To determine the initial pheromone, zero, a randomized solution is built. The fitness value is then used to evaluate the population's fitness, and non-dominated solutions or Pareto optimal groupings are discovered. Figure 3.3 displays the workflow chart created by MLACO.

3.3.2. Pareto Set. The MLACO approach preserves a Pareto set. For each solution, two goals (KKM and RC) were established once the strategy for retaining non-dominated responses was developed. The non-dominated systems are found using the goal variables. Every time a new installment yields a non-dominated result, we eliminate dominating alternatives from the group and update the Pareto optimal set. These Pareto groups produce a better response for the populations that will follow, which optimizes the exploring region.

Solutions within the Pareto set don't actually guarantee the maximum NMI or modularity. True net grouping is not the same as optimal modularity[4]. The Pareto set gives the outcome based on NMI and modularity principles, allowing for the maximum possible modularity and NMI numbers. The answer with the highest NMI/modularity values was selected as the global optimum inside the Pareto optimal group.

3.3.3. Initial Pheromone trail. The value τ_0, which has a very tiny number and is constant across all linkages in the graphs, defines the initial quantity of pheromones. The beginning pheromone quantity of τ_0

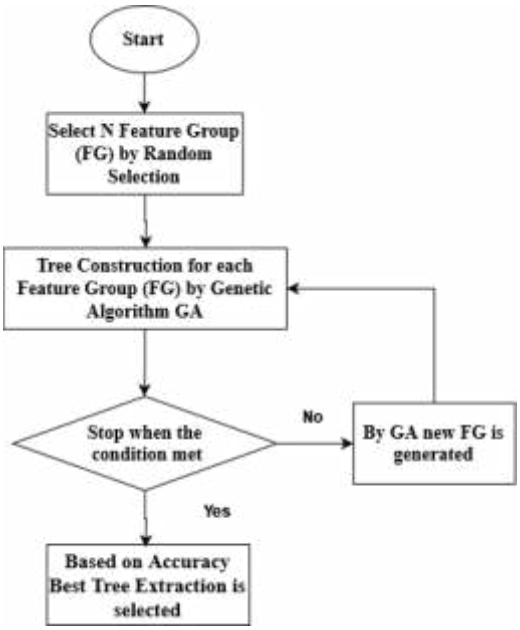


Fig. 3.2: Genetic Algorithm Process

is found using Eq.(3.1).

$$\tau_0 = \frac{1}{n \times [RC\,(S_0) + KKM(S_0)]} \tag{3.1}$$

The terms RC (S0) and KKM(S0) represent the response S0's RC amount and KKM values, respectively, whereas n represents the number of nodes in the network. S0, the initial response, is generated at random. KKM and RC, respectively, are the fitness functions that are determined.

3.3.4. Updating the Pheromone. The pheromone levels are updated with each cycle, therefore this data is utilized to explore a new search region. The usefulness of pheromones will also be diminished by the concept of pheromone evaporation. In MLACO, pheromones will be altered in two ways: locally and internationally. Eq. (3.2) represents the localized pheromone update approach.

$$\tau_{i,j}\,(T) = (1 - p_1)\,\tau_{i,j}\,(T - 1) + p_i\tau_0 \tag{3.2}$$

The localized pheromones evaporating variable, denoted as p1, has a range of 0 to 1. Use the local pheromone updates the rule to lower the number of pheromones if a link is formed between nodes i and j. The globally pheromone updated rules are used to upgrade the pheromone principles while the localized updated rules are utilized to adjust the pheromone levels. Non-Pareto global optimum or Pareto global optimum solutions were among the options for multi-objective function optimization. The global pheromone is altered using Equation (3.3).

$$\tau_{i,j}\,(T) = (1 - p_g)\,\tau_{i,j}\,(T - 1) + \frac{p_g}{RC\,(S) + KKM(s)} \tag{3.3}$$

Here p_g stands for globally pheromones evaporating variable, and its value ranges from 0 to 1.

3.3.5. Process of Multi-layer ACO. In a multi-layer ACO like this, eligible nodes are categorized at different layers. In networks, the number of components equals the size of the network. The total of the anticipated limits for each component equals the number of nodes. For a particular node, the permitted replies

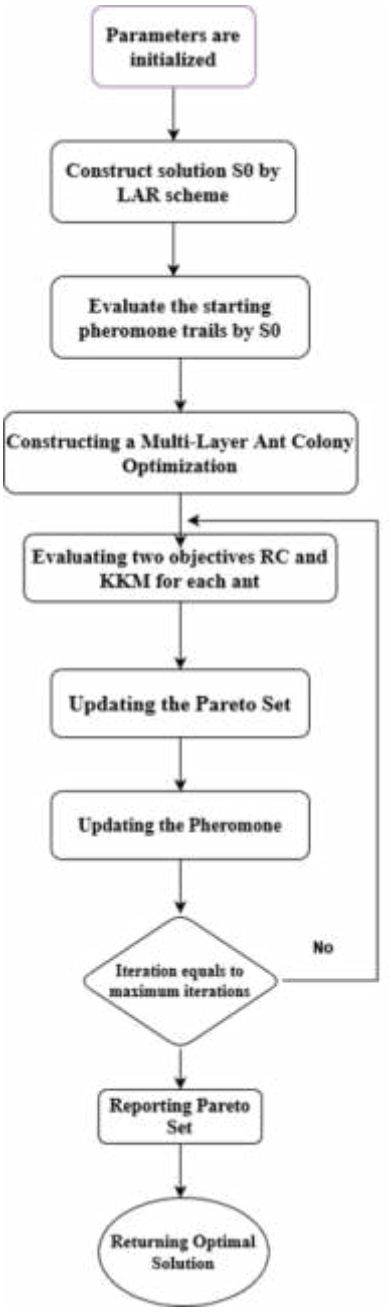


Fig. 3.3: Workflow of MLACO

are in fact the adjacent nodes. Ten levels are required, and each layer's options are determined by the values of its neighboring layers. The allowed nodes in layer 1 are 3 and 7, in layer 2 they were 4 and 5, and so on. I will select acceptable nearby nodes m for an ant k in level l nodes by using Equation (3.4)

$$m = \begin{cases} \arg\max\{[\tau(i,m)]^{\alpha}\left[\eta(i,m)\right]^{\beta}\} & if\ q < q_0 \\ M & otherwise \end{cases} \tag{3.4}$$

Table 4.1: Result of Emotion Rating

Time Step	Emotion Score	Anticipated Emotion
2.5	0.93	Happiness
5	0.52	Relaxation
7.5	0.61	Surprise
10	0.84	Joy
15	0.45	Sad
17.5	0.93	Happiness
20	0.32	Sad

$$P\left(i,m\right)=\frac{\left[\tau\left(i,m\right)\right]^{\alpha}\ \left[\eta(i,m)\right]^{\beta}}{\sum_{\omega\in I_i}\left[\tau\left(i,\omega\right)\right]^{\alpha}\ \left[\eta(i,\omega)\right]^{\beta}} \tag{3.5}$$

An assortment of non-dominated ants is selected to form a Pareto optimal set. A non-dominated mixture is transmitted when an answer is excluded from the Pareto set because it is dominating another response. For this reason, the Pareto set will change following each cycle.

A key feature of this process is the use of a Pareto optimal set to guide the optimization. Non-dominated ants, whose solutions are not outperformed by others in terms of multiple criteria, are selected to form this set. A non-dominated mixture is generated when a solution is excluded from the Pareto set because it dominates another response. As the process progresses, the Pareto set dynamically evolves, ensuring that only the most optimal solutions are retained. This continuous refinement of the Pareto set across iterations allows the multi-layer ACO to effectively converge toward high-quality solutions while maintaining diversity in the search space.

4. Result Analysis. The suggested technique made use of the Flickr picture collection, whose Flickr30k dataset has become the industry standard for sentence-based image categorization. hsankesara/flickr-image-dataset.kaggle.com. The present research presents Flickr30k Entities, an addition to the 158k captions in Flickr30k that includes 276k manually annotated bounding boxes and 244k reference chains connecting image-wide citations of identical objects.

By utilizing the possibility of concurrent multimodal input, GA-MACOcan transform understanding, production, and data interaction. It is more advanced than traditional machine learning. This first section presents an overview of GA-MACOeffective abilities and lays the groundwork for investigating its possible uses and implications. The model's evaluation of the intensity of feelings over a certain time period is shown by the Emotion Rating in Table 4.1, ranging from 0.30 to 1. The categorization that the framework assigns to identify the state of mind of the data that is offered at every stage is known as the Expected Mood. Figure 4.1 shows the result of emotion analysis based on time steps.

It is very important to use Image Emotional Perception technology in packaging design. We've already talked about the importance of hues and designs in package design as well as the best way to put these elements into practice. To improve the efficiency of information flow, a designer has to incorporate cultural implications into the packaging layout for the product via aesthetic visual symbols or language. Unlike traditional visual symbols, people usually use two-dimensional design images or symbols as graphic symbols. No matter what color scheme or visual design is used, designers have to make sure that the general public can grasp it.The 3s-HCI, AFED-CR, and PLM modeling are used to compare the suggested method's efficacy in raising the caliber of packaging layout. GA-MACO systems are able to recognize and classify the emotions expressed by product designs through the use of photo analytics.GA-MACO can identify a variety of emotions through visual cues like color schemes, gestures, and facial expressions, include happy, passion, confidence, unexpectedly and discontent. GA-MACO enables the quantitative analysis of packaging design perceptions based on feelings. The overall quality of design evaluation is displayed in Fig. 4.2.

Analyzing the expected emotion labels revealed that inaccurate categorization commonly occurred within similar emotional perception. The mean squared error can be computed as the disparity between the anticipated and actual values of the hypergraph. Reducing errors in prediction through improved modeling accuracy,

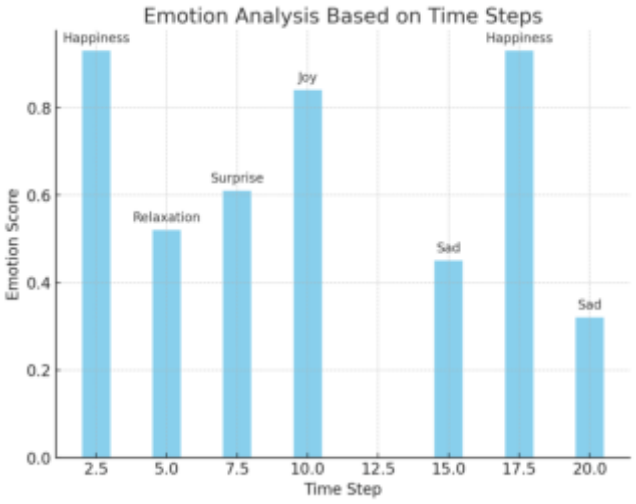


Fig. 4.1: Emotion Analysis using Time Steps

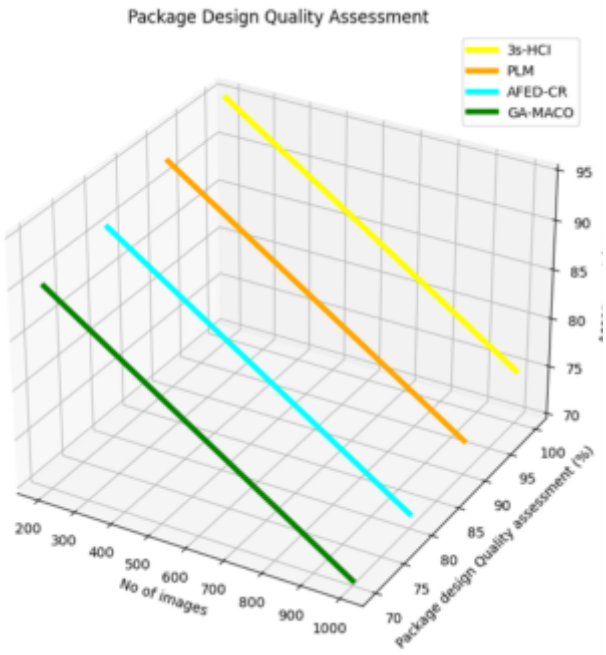


Fig. 4.2: Package Quality Design Assessment

feature evaluation, accuracy of data, and ensembles strategies is the main objective of GA-MACO models during package design. By employing these tactics, GA-MACO could substantially improve the evaluation of packaging design and optimize emotional effect. The mean square error rate can be seen in Fig. 4.3.

5. Conclusion. Human food consumption is becoming increasingly personalized to meet the specific needs of each individual, with a wide variety of product labels easily accessible. Because of this, a lot of businesses are starting to focus on making modern packaging more useful. Over the course of a user-product interaction, emotional reactions and sensory paradigms may change. Conventional packaging layout relies heavily on the

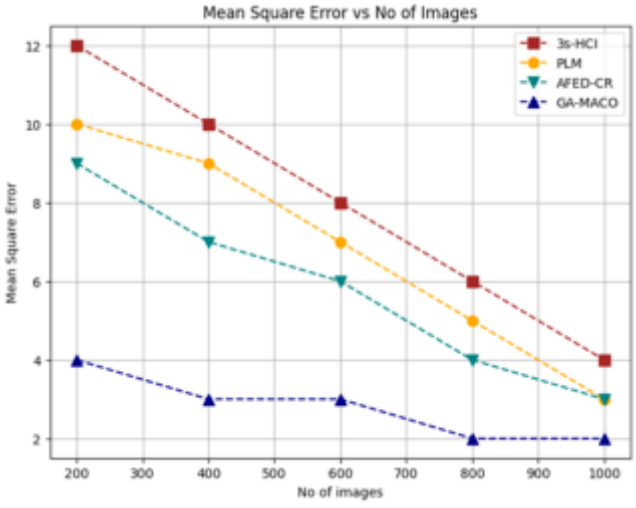


Fig. 4.3: Result of MSE vs number of images

designer's sentimental creativity and past experiences, but it is constrained by unpredictable content and a dearth of professional guidance; previous research on mental image analysis mostly predicted the most common emotions of viewers. Because the feelings evoked by a picture are very personal and vary from viewer to viewer, there are occasions in which the general impression of the image is inadequate for use in reality. The suggested methodology analyzes sight and feelings in order to identify human senses using Multi-Layer Ant Colony Optimization based on Genetic Algorithms. To evaluate classified and multimodal attitude depictions, a large image set known as Image-Emotion-Social-Net is used. More than 9,000 people have contributed more than a million photographs to the Flickr library.The research conducted on this dataset shows that the proposed strategy outperforms various modern approaches in tailored emotional recognition. The research conducted on this dataset shows that the proposed strategy outperforms various modern approaches in tailored emotional recognition. In the case of future studies, they could assess the possible long term effects that such optimization of the packaging design may have on consumers, retentions levels, and sales return.

REFERENCES

[1] H. Cui, *Computer-aided design of hand-drawn art food packaging design based on a deep neural network model*, International Journal for Simulation and Multidisciplinary Design Optimization, 15 (2024), p. 10.

[2] M. Di Capua, A. Ciaramella, and A. De Prisco, *Machine learning and computer vision for the automation of processes in advanced logistics: The integrated logistic platform (ilp) 4.0*, Procedia Computer Science, 217 (2023), pp. 326–338.

[3] N. Ghodsian, K. Benfriha, A. Olabi, V. Gopinath, and A. Arnou, *Mobile manipulators in industry 4.0: A review of developments for industrial applications*, Sensors, 23 (2023), p. 8026.

[4] Y. Guangpeng, H. Yin, and L. Fengze, *Retracted article: Computer vision technology based on image optical processing in visual packaging art design simulation*, Optical and Quantum Electronics, 56 (2024), p. 529.

[5] J. Henriques-Gil, D. Mejia Parra, and B. Simões, *Enhancing decision making in mdo through interactive visual analytics on the web*, in Proceedings of the 28th International ACM Conference on 3D Web Technology, 2023, pp. 1–9.

[6] H.-M. Heyn, E. Knauss, and P. Pelliccione, *A compositional approach to creating architecture frameworks with an application to distributed ai systems*, Journal of Systems and Software, 198 (2023), p. 111604.

[7] J. Jenis, J. Ondriga, S. Hrcek, F. Brumercik, M. Cuchor, and E. Sadovsky, *Engineering applications of artificial intelligence in mechanical design and optimization*, Machines, 11 (2023), p. 577.

[8] J. Jiang, *Computer-aided visual communication design based on image detail enhancement algorithm*, Optical and Quantum Electronics, 56 (2024), p. 652.

[9] S. kui Yang, W. jun Chung, and F. Yang, *Analyzing the packaging design evaluation based on image emotion perception computing*, Heliyon, 10 (2024).

[10] X. Liu, G. Gong, X. Hu, G. Shang, and H. Zhu, *Cognitive enhancement of robot path planning and environmental perception based on gmapping algorithm optimization*, Electronics, 13 (2024), p. 818.

[11] Y. Liu, X. Tao, X. Li, A. W. Colombo, and S. Hu, *Artificial intelligence in smart logistics cyber-physical systems: State-of-the-arts and potential applications*, IEEE Transactions on industrial cyber-physical systems, 1 (2023), pp. 1–20.
[12] S. Maheswaran, *Intelligent cold chain security: Nano power temperature sensors, esp32 and telegram bot integration for temperature assurance and environmental harm prevention*, J. Environ. Nanotechnol, 13 (2024), pp. 17–25.
[13] A. Manavis, K. Kakoulis, and P. Kyratsis, *A brief review of computational product design: a brand identity approach*, Machines, 11 (2023), p. 232.
[14] H. Matyi and P. Tamás, *An innovative framework for quality assurance in logistics packaging*, Logistics, 7 (2023), p. 82.
[15] K. Meduri, G. S. Nadella, H. Gonaygunta, and S. S. Meduri, *Developing a fog computing-based ai framework for real-time traffic management and optimization*, International Journal of Sustainable Development in Computing Science, 5 (2023), pp. 1–24.
[16] S. Ou, Z. H. Ismail, and N. Sariff, *Hybrid genetic algorithms for order assignment and batching in picking system: A systematic literature review*, IEEE Access, (2024).
[17] W. Qi, C.-H. Chen, T. Niu, S. Lyu, and S. Sun, *A multisensory interaction framework for human-cyber–physical system based on graph convolutional networks*, Advanced Engineering Informatics, 61 (2024), p. 102482.
[18] N. Rane, S. Choudhary, and J. Rane, *Enhanced product design and development using artificial intelligence (ai), virtual reality (vr), augmented reality (ar), 4d/5d/6d printing, internet of things (iot), and blockchain: A review*, Virtual Reality (VR), Augmented Reality (AR) D, 4 (2023).
[19] S. Špirková, M. Straka, and A. Saniuk, *Vr simulation and implementation of robotics: A tool for streamlining and optimization*, Applied Sciences, 14 (2024), p. 4434.
[20] B. Vivek, A. Arulmurugan, S. Maheswaran, S. Dhamodharan, A. Dharunash, and N. Gowtham, *Design and implementation of physical unclonable function in field programmable gate array*, in 2023 8th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2023, pp. 152–158.
[21] B. Vivek, S. Maheswaran, P. Keerthana, S. Sathesh, S. Bringeraj, R. A. Sri, and S. A. Sulthana, *Low cost raspberry pi oscilloscope*, in 2018 international conference on intelligent computing and communication for smart World (I2C2SW), IEEE, 2018, pp. 386–390.
[22] B. Vivek, S. Maheswaran, N. Prabhuram, L. Janani, V. Naveen, and S. Kavipriya, *Artificial conversational entity with regional language*, in 2022 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2022, pp. 1–6.
[23] X. Wan and Y. Liang, *Design of intelligent color matching system for cultural and creative products based on data analysis algorithm*, in International Conference on Innovative Computing, Springer, 2023, pp. 667–674.
[24] X. Wang and J. Jiang, *Visual analysis of brand packaging design data based on cad and big data technology*, (2024).
[25] R. Yavich, S. Malev, I. Volinsky, and V. Rotkin, *Configurable intelligent design based on hierarchical imitation models*, Applied Sciences, 13 (2023), p. 7602.
[26] X. Zhang and X. Li, *Optimization of virtual reality in brand identity design and visual recognition. based on image fusion and text assistance*, (2023).
[27] A. Zhu, *Design and research of green concept product packaging based on artificial intelligence technology*, Applied Mathematics and Nonlinear Sciences, (2023).
[28] D. Zhu and G. Liu, *Corporate brand design strategy based on basic graphical language descriptions of visual communication*, (2023).

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: Sep 26, 2024
Accepted: Nov 26, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2275–2285, DOI 10.12694/scpe.v26i5.5025

DISTRIBUTED SYSTEMS FOR EVALUATING AND OPTIMIZING ENVIRONMENTAL ART DESIGN USING IMAGE PROCESSING

HUI WANG*

Abstract. Improving the visual appeal and practicality of areas is a major function of environmental art design. Nevertheless, conventional techniques for assessing and improving such designs are frequently arbitrary and ineffective. In order to assess and improve environmental art design, this study suggests a distributed system that makes use of Generative Adversarial Networks (GANs) and sophisticated image processing methods based on the characteristics of Computer-aided design (CAD). The methodology is based on distributed demand-side management in intelligent energy systems and emphasizes the decentralization of computational resources for increased flexibility and effectiveness.In order to model different design situations and improve based on aesthetic characteristics like color balancing, spatial arrangement, and visual balance, the system uses GANs for creating images and design transmission. This method's implementation in a distributed framework speeds up the assessment procedure and allows for continuous improvement and real-time feedback. A comparative examination of the findings from the experiment highlights the remarkable quality and effectiveness of the strategy presented in this study, which performs better than alternative strategies when it comes to of precision, recall, and F1 score. The results validate the superior performance of the suggested approach in relation to component extraction and recognition in CAD environmental artwork design. It is expected that this will support strong evidence for real-world applications and further research advancements in relevant sectors.

Key words: Distributed Systems, Optimization, Environmental Art Design, Image Processing, Deep Learning

1. Introduction. Computer vision's primary responsibilities include interpreting, evaluating, and comprehending images. An essential duty in the domains of planning for cities, sustainability, and growth of tourism is assessing the overall condition of the environment. Manually observation and scoring are frequently used in conventional landscape evaluation techniques, yet this approach is unreliable and subjective[12]. It is frequently used in evaluations of landscape quality. Utilizing the pixel data of the street view pictures, deep learning-based image segmentation methods are able to divide images into numerous areas and recognize the objects and scenes that belong to each zone[17]. This method offers robust support for evaluating the beauty of landscapes by extracting rich visual data from photos.

With the application of AI technology in visual communication design, the viewer may get involved in the creation process, which not only piques their attention but also helps them connect emotionally with the artist [2]. The singular connection among artistic items and conventional design for visual communication has fundamentally changed due to AI virtual art's communication, opening up a variety of potential sensory experiences.Artificial intelligence has elevated visual perception to a considerable degree. It is an artistic and technological fusion that disrupts the conventional display and opens up new possibilities. People would be treated to a novel creative moment upon its arrival [19].The effective enhancement of art design's job efficiency has emerged as a crucial subject in the discipline of smart art creation.

In recent years, as technological and scientific developments have advanced, pictures have been employed in numerous facets of people's life, including web page image material, face recognition, license plate recognition, and mobile phone imaging. Extensive study on pictures can enhance people's standard of life, alter people's way of life, and improve their living conditions [22, 13]. Numerous disciplines are involved in processing images, such as division, improvement, pairing, reconstruction, change, and categorization. The primary applications of image processing technology include image enhancement and reconstruction, image storage and coding, picture segmentation, image alignment, and picture classification [23]. Among these applications, there are numerous restorations of image, imaging segmentation, and visual distribution-related issues that require optimization.

*School of Art and Design AnHui Business and Technology College,Hefei,230041, China (wangh@ahbvc.edu.cn)

Key design features can be extracted from these photos through processing and analysis using deep learning techniques. In order to give architects more thorough and in-depth design references, the author [31] included geography, water bodies, plant species, and other elements in photographs.Connection and interaction among many factors are critical for ecological design for landscaping.By examining geographical element correlations and comprehending the connections and interactions among elements, deep indication learning technology can offer designers more precise design guidance [18, 9, 6, 25]. One of the most important links to ecological landscape architecture is structural design. Engineers can receive technical assistance, particularly analysis of structures and optimizing, from computer-assisted building and construction [6, 26]. It is possible to more precisely assess the stability and safety of structures by using machine learning construction and facilities which gives architects more dependable design options.

In the modern world where art has gained prominence, creating designs that can not only complement the space but also allow for fluid interactions with the said design and the space it is in, become the need of the hour. On the downside, design assessment approaches to date tend to be subjective, slow, and non-scalable. Irrespective of the impressive features provided by the current approach, it is becoming increasingly clear that modern design challenges require new data-driven solutions. Taking into consideration the increasing potential of artificial intelligence, one can anticipate dramatic changes in this area.

Design evaluation and generation are the fields where Generative Adversarial Networks can demonstrate the best results in addressing these challenges. GANs are primed to provide a comprehensive set of variants for each design and then CAD offers a great deal of organized approaches needed for the integration. The goal of the research is to make use of these technologies within the distributed systems framework allowing for the outsourcing of computational power to allow flexibility and constant scaling along with near-instantaneous feedback. Extending the applications of demand-side management in intelligent energy systems, aid is provided in the form of efficient resource allocation and redesigning of environmental arts.

The main contribution of proposed method is given below:

1. The distributed system architecture presented in this work decentralizes computing resources for environmental art design evaluation and optimization. Designing possibilities are created, modeled, and transferred using Generative Adversarial Networks (GANs).
2. This allows the framework to improve visual harmony, color balance, and spatial layout—all important aspects of natural art creation. GANs retain computing efficiency while improving the overall quality of design variants.
3. The study makes use of advanced image processing techniques designed for CAD-based design of environmental components.
4. A thorough experimental assessment is carried out, which shows that in regards to precision, recall, and F1 score, the suggested system performs better than conventional image processing and design assessment methods. This confirms the program's exceptional ability to gather and identify design components from environmental artwork created using CAD.

The rest of our research article is written as follows: Section 2 discusses the related work on various Distributed Systems, Optimization, Environmental Art Design and Image Processing, Deep Learning. 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 conventional techniques of designing landscapes typically lack accuracy and impartiality because they rely too much on the expertise and intuition of the designers. More thorough and in-depth design guides are made available to designers with the progressive implementation of landscape design. Within the field of landscape design, the author [20, 28, 24] investigated the use of neural network-assisted intelligence and CAD vision technologies. Environmental design features can be precisely measured and modeled using CAD visual technological advances, and they can be intelligently analyzed and optimized with the use of neural networks and intelligent technology [21, 5, 10, 16]. Various components must be recognized and categorized in designing landscapes. Automatic detection and categorization of these components is possible using CAD visual technological advances, increasing design efficiency and accuracy.

An all-encompassing design approach that blends design, creativity, and environment is called environmen-

tal art development [14]. Designers of environmental art must separate out from the crowd and incorporate into their designs the components that have aesthetic value. Conventional approaches, on the other hand, lack objectivity and precision because they frequently rely on the experience and intuition of the creator to determine artistic and creative components [11]. The technique of identifying environment design and artistic components was examined by the author [3]. It explains how to create and apply an algorithm and how to use it to recognize design features and works of environmental art. Large-scale datasets can be used to train and deduce the method despite the need to formally describe the geographical distribution of the information.

The field of visual communication is now at the forefront of modern art design due to a shift in people's aesthetic sensibility, and a lot of pieces are constantly in the public eye. Over time, interest has also grown in the creative visual communication design. In [4] the author provided an example of how digital equipment is used in the production of art for digital media as well as how it might improve artistic practice in connection to digital media. By examining the features of ornamental paintings instruction in every aspect of graphic layout, the author talked about the ways to update the curriculum, instructional strategies, and instructional ideas[15, 8, 30]. A significant part of the human past and cultural legacy is the cultural environment, which includes both natural and man-made environments.

Modern technological methods are required for recording and presenting this priceless historical treasure so that it can be better preserved and passed down to future generations [1]. A reliable and accurate method for measuring and modeling that can be utilized to capture comprehensive data on cultural landscapes is point cloud 3D modeling. Using a case study of a garden, the author [29] presented the technique of using point cloud three-dimensional models to document cultural environments and illustrated the use of spatial integration technologies in landscape architecture. The dimensions, form, substance, and other details associated with cultural settings can be adequately represented by this representation, offering solid backing for later planning and preservation [27, 7]. An essential source for guidance for landscape designers is geographic data. Geospatial data and 3D models can be integrated via geographic integration technology to give designers deeper and more thorough design guidelines [18]. Geographic integrating the internet, for instance, may be employed to gather data on elevation, plant transportation, environment, etc.

In spite of the desire to create eye-catching artistic designs, the evaluation and optimization of those designs remain a considerable challenge. Conventional methods of design and detail orientation are predominantly subjective and rely on guesswork which often leads to contradictions and ineffective scenes. Archcad drawings often disregard the changing nature of such aesthetic aspects as their arrangement and scale. Although some progress has been made, for instance, CAD tools enable accurate design, many of them still struggle to encapsulate a wide range of design aspects like integration of color and balance in space. The use of computing technologies for image processing in environmental art design is still largely undeveloped particularly the use of jot GANs of these methodologies in other fields has proven to be very effective in creating a wide range of designs.

3. Proposed Methodology. The proposed methodology for Evaluating and Optimizing Environmental Art Design Using Image Processing in Distributed Systems based onGenerative Adversarial Network (GAN) and the characteristics of CAD. In this work, initially the data is collected and stored in distributed networks. The collected data is used for pre-processing and then the feature is extracted. Finally, the GAN is used for training and optimizing the dataset. In figure 3.1 shows the architecture of proposed method.

The recognition and extraction of components is essential in enhancing the accuracy and productivity of CAD in the rendering of environmental designs. These methods consist of the analysis of the overall design by breaking it down into its components and features such as shape, pattern, texture, and even the structural components, for further detail, modification, and enhancement. Thus, if certain features are extracted, the designers of the project can devote their energy to circling single parts without changing the complete structure. For example, it is possible to detail a texture or a geometric pattern that has been removed from the general design for modifications. Components in CAD systems should be presented as independent elements of design to be further examined for features such as materials, structural, or decorating purposes. This isolation of elements makes it relatively easy to assess the design as a whole and within elements of design so that there is a fast modification of the design in consideration of the less time taken to evaluate an element within a design.

3.1. Data Pre-processing. The recognition and extraction method is presented automatically in this piece. This technique makes extensive use of GAN, pattern recognition, image processing, and other techniques

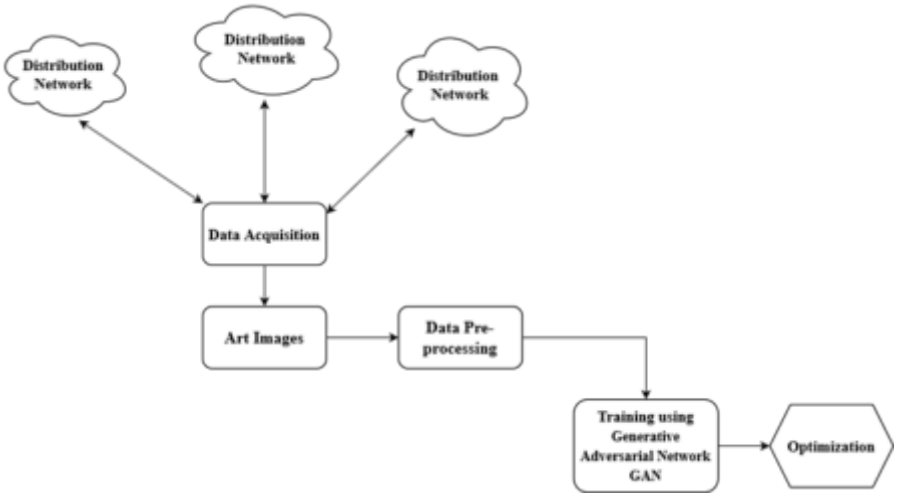


Fig. 3.1: Architecture of Proposed Method

to extract and identify design aspects automatically, hence increasing the effectiveness and caliber of the creation procedure.For recognition and extraction techniques to be accurate and effective, preparing the data is essential. The term complexity, which variation, and huge dimensions are common features of original data in CAD environment designs. In order to guarantee that the data supplied to GAN is precise and consistent, information cleaning—the main preprocessing task—aims to eliminate any superfluous, redundant, or incorrect components from the initial information. Duplicate drawings of designs or features may arise throughout the data collection or organization procedure.

For this reason, duplicate data is found and eliminated using the appropriate methods or tools in this article. Furthermore, certain information may lack specific design features or characteristics, which will result in data that is insufficient.In this situation, it can be handled by interpolation, deletion, or estimation using associated information. Furthermore, faults or other factors related to the acquisition procedure could introduce noise or anomalous values into the data. These parameters will negatively affect the model's learning. It is possible to effectively remove outliers as well as noise from these data by employing filters, statistical approaches, or other noise removal strategies.

3.2. Training and Optimization using GAN. The creation of realistic data has allowed Generative Adversarial Networks (GANs) to transform a number of machine learning disciplines. In this work, GANs are used in conjunction with image processing methods to maximize environmental art design. Two neural networks, the Generator (G) and the Discriminator (D), which cooperate within a min-max gaming structure, make up the GAN design.

The Generator is in charge of creating artificial visuals that mimic aspects of real-world layout, like aesthetic harmony, color balance, and spatial organization. In order to provide fresh design examples, it learns how to map points from a latent space (random noise). At the other side, the Discriminator's job is to differentiate among actual design examples and those generated by the Generator. The Generator gains the capacity to create images that are more realistic as it trains, while the Discriminator keeps getting better at telling actual photos from produced ones.

A min-max game among the two networks can be used to illustrate the main objective of the GAN development procedure:

$$\min_{G}\max_{D} V\left(D,G\right)=\mathbb{E}_{x\sim P_{d}(x)}\left[logD\left(x\right)\right]+\mathbb{E}_{rnv\sim P_{rnv}(rnv)}[\log?(1-D(G(rnv)))] \tag{3.1}$$

In this formula, the Discriminator optimizes its incentive by successfully detecting genuine samples from the dataset and differentiating them from produced samples, while the Generator aims to limit the Discriminator's

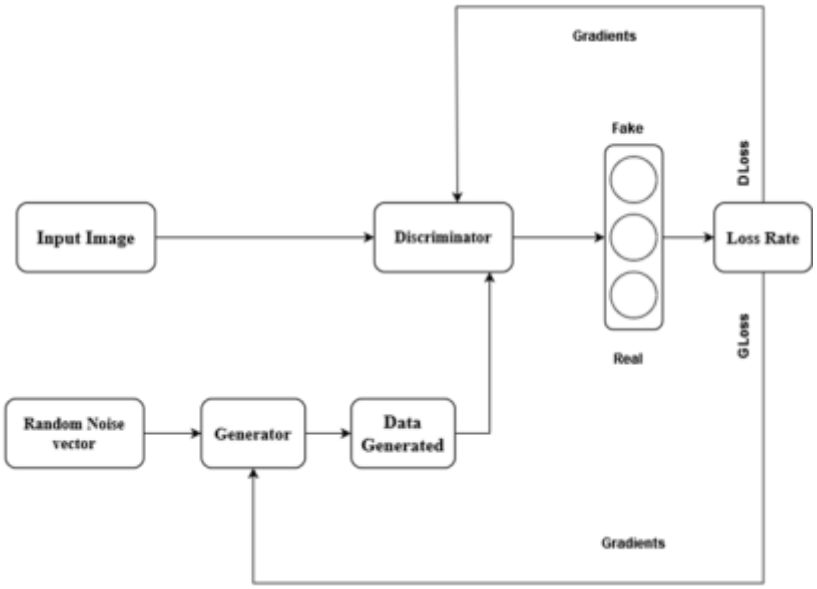


Fig. 3.2: Structure of GAN

capacity to do so. The Generator generates synthetic samples G(z) using the random noise variable $P_z(z)$, and the Discriminator assesses the probability that each sample is created $D(G(z))$ or real $D(x)$.

The Generator is tuned to generate incredibly lifelike artistic creations that imitate actual training instances as the system goes through its iterative learning procedure. As this is going on, the Discriminator gets better at differentiating real photographs from fake ones, making it harder and harder to tell both of them apart.

The Generator makes use of a deep neural network with numerous convolution layers to produce high-quality images that capture the creative elements that comprise environmental layouts. The input photographs are edited to match the design of the Generator after being shrunk to $128 \times 128 \times 3$ pixels. Four convolution layers combined with ReLU activations produce aesthetically pleasing and incredibly intricately designed pictures.Through information sharing among the encoder and the discriminator, the discriminator integrates an encoder to improve feature extraction. Although the encoder makes sure the created images retain the visual integrity needed for environmental artwork, the discriminator focuses on differentiating between actual and produced pictures.

The following is a representation of the loss function for reconstruction for this process:

$$\mathcal{L}_{recons}^{pix} = \mathbb{E}_{q \sim D_{encoder}(x), x \sim I_{real}} [|\,|\kappa\,(q) - \tau\,(x)|\,|] \tag{3.2}$$

In this case, $\tau(x)$ indicates the discriminator's feature map, while $\kappa(q)$reflects the decoder's operations on real image characteristics IrealIreal. In order to guarantee that the created designs closely resemble the visual traits of actual environmental art, the aim is to reduce this reconstruction loss.This method guarantees that the Generator generates high-quality synthetic design pictures that match the visual and spatial qualities pertinent to ecological art design, in addition to deceiving the Discriminator.In figure 3.2 shows the structure of GAN architecture.

The first challenge is the deployment of the system requires particular strategies for maintaining the effectiveness and efficiency of the system while managing resource allocation, latency, data synchronization, and scalability across different environments. The allocation of resources to nodes in the distributed system was one of the major challenges. Working with computational tasks which include running Generative Adversarial networks (GANs) or processing digital models based on complex CAD data, achieving balanced workloads was vital. This was handled by demand-side management approaches where resources were dynamically reallocated

Table 4.1: Dataset Description [12]

Dataset	**Design element category**	**Number training samples of set**	**Number verification samples of set**	**Number test samples of set**
Dataset A	Material selection, color coordination, and space layout	5000	1000	2000
Dataset B	Space organization, color coordination, material choice, and lighting	8000	2000	3000
Dataset C	Space planning, color coordination, material choice, lighting configuration, furnishings, and decoration	10000	3000	4000

as per task priority and node constraints. Advanced scheduling algorithms were employed to evenly distribute computational loads and avert bottlenecks. Another difficulty involved the restriction of latency in the real time feedbacks and the iterative processes of continuous improvements. It is fair to assume that high latency could cause a disturbance to the iterative approach of design evaluation. To solve such a problem, parallel processing and predictive caching techniques were used to retrieve frequently accessed data and allow for the execution of multiple tasks at the same time. Furthermore, the implementation of edge computing was used to localise the tasks and thus reduce the dependency on the central nodes and improve the response time. Also, data synchronization between geographically distributed nodes was also a challenge.

4. Result Analysis. Large multi-core CPUs, GPUs, or TPUs that have the capability of running the GANs or performing demanding image processing tasks are the main computational resources for the suggested distributed framework. These resources should be able to provide efficient handling of large databases as well as allow real time interaction within the system. However, due to the large amount of data that is created when evaluating and optimizing designs of the environmental art, distributed storage systems are necessary for data management. Furthermore, a network infrastructure that is bandwidth-intensive is also paramount so as to enable effective inter-node communication within the distributed system. To overcome the problem of scalability, cloud resources and distributed computing concepts are utilized which makes it possible for such workload and or dataset size to increase without limits. The system applies adverse management techniques and therefore achieves better resource efficiency by ensuring that all tasks with computational requirements are distributed across available nodes in a dynamic manner. As a result, the latencies are somewhat controlled and bottlenecks are avoided, which makes the systems useful in real life large scale applications like urban planning for architecture.

Numerous simulation tests have been carried out as part of this inquiry in order to validate the effectiveness of the suggested method for the recognition and extraction of design elements. Following this, a thorough study and explanation of the results have been conducted. Here is the setup for the test: This part uses TensorFlow, a Python-based GAN structure, to conduct the study. Three public CAD environment creative design information sets—Dataset A, Dataset B, and Dataset C—are used in this study's training and evaluation of the model. The labeling information for these three sets of data covers a variety of design features, including choosing materials, matching of colors, and spatial design. Learning is done on the training set, and optimisation is done using the aforementioned techniques. In order to monitor the model's training environment, its precision and loss are documented throughout the training phase. The model's variables are saved and evaluated on the test set once the model reaches its maximum efficiency on the validation set. Table 4.1 shows the dataset description.

It appears that the loss numbers for the various tests fluctuate, suggesting that the model's performance changes with each iteration.It's possible that there is inconsistency in the optimization process or that the algorithms are still learning and altering parameters because none of the tests appear to show a continuously decreasing loss.Although the behavior of each test varies, overall, throughout the course of the 10 repetitions, the loss values lie between 0.1 and 0.7. The four tests are distinguished by their corresponding colors in the description on the left. Every test seems like a study or a model that is being optimized, and the legend makes

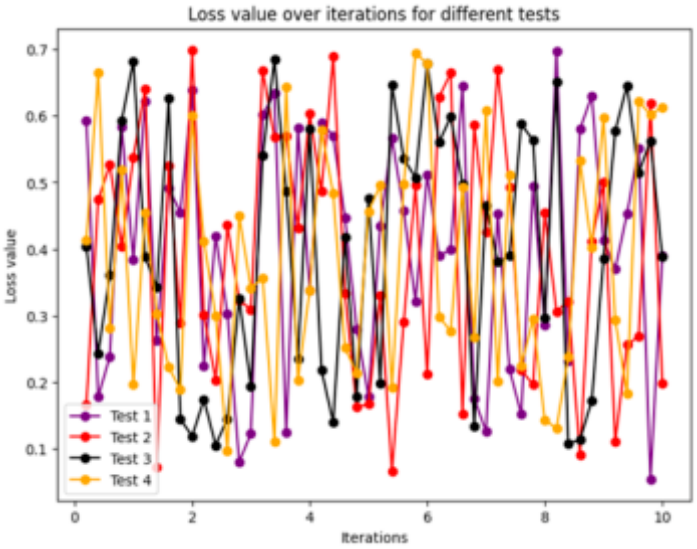


Fig. 4.1: Loss Value over Iterations

Table 4.2: Experimental Result of Accuracy, Recall, and F1-score

Dataset	**Design element category**	**Accuracy rate**	**Recall rate**	**F1 value**
Dataset A	Space Layout	93.3%	90.6%	92.7%
	Colour matching	88.5%	86.3%	87.2%
	Material selection	91.2%	88.4%	89.8%
Dataset B	Space Layout	93.8%	91.1%	92.5%
	Colour matching	90.3%	93.7%	92.7%
	Material selection	94.9%	89.7%	93.5%
	Lighting setting	93.3%	91.2%	93.7%
Dataset C	Space Layout	95.7%	92.1%	95.7%
	Colour matching	89.5%	94.7%	90.9%
	Material selection	90.4%	90.6%	90.2%
	Lighting setting	92.7%	90.4%	92.1%
	Decoration and furnishings	93.8%	92.6%	93.1%

it easier to figure out which line goes with which test. Figure 4.1 shows the result of Loss Value.

This section presents the detection and extraction outcomes of the suggested method on several data sets, following a series of trials. Table 4.2 displays particular outcomes indicators.

The most inconsistent and lowest accuracy rates are displayed by the Traditional Image Processing Methods, suggesting that this approach may not be the best fit for the given task.Although the RNN technique outperforms the old technique, it still exhibits significant accuracy fluctuations, suggesting that further tuning or stability in the learning process may be necessary for the model.The highest performance is shown by the Proposed Method , which maintains a high degree of accuracy with comparatively less volatility, indicating that it is more robust or effective for the task at hand.

This graphic 4.2 shows that the accuracy of the Proposed Method constantly beats that of the other two methods (conventional and RNN) over the course of the iterations. This implies that the suggested approach has a more successful model architecture or is more adept at learning from the data.Significant instability is shown by the traditional image methods of processing, which may indicate that these techniques are out of date or not task-specifically tuned.The RNN Method performs moderately, indicating that it has promise but

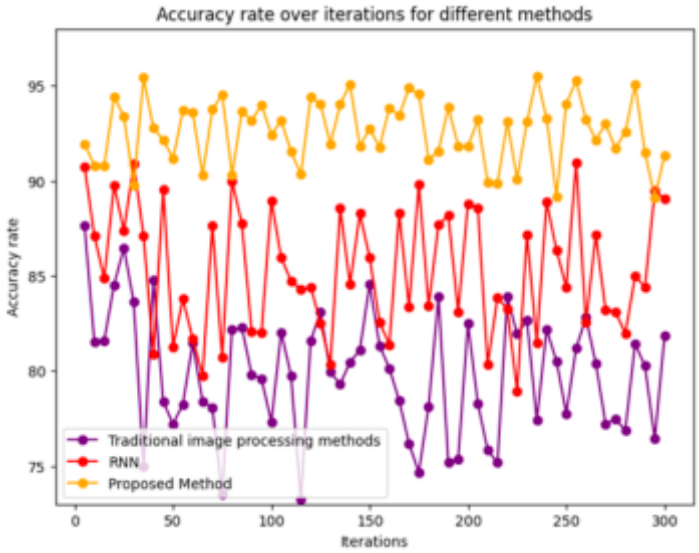


Fig. 4.2: Accuracy

might require more improvement. The Proposed Method gets the best and most stable accuracy, following by the RNN, with the Traditional Image Processing Methods trailing behind. The plot, which compares all three approaches over numerous iterations, illustrates this.

Traditional image processing techniques: This technique performs the worst in terms of recall rate, ranging from about 75 to 85. The line has several changes and is rather unpredictable.RNN: This approach outperforms the conventional techniques, with recall rates ranging from 80 to 90. While variable, it continues to be higher than the conventional approach.Proposed strategy: With a recall rate that hovers around 90 to 95, this strategy is the most effective on a regular basis. More stability is indicated by the significantly smaller volatility.Recall rate is a continuous advantage of the Proposed Method over both the RNN and the Traditional Image Processing Methods.The two alternative methods perform better in recall, but the traditional image processing methods show a lot of instability.Though it is still fewer reliable and successful than the suggested method, the RNN outperforms more conventional approaches. In figure 4.3 result of Recall.

In terms of F1 value, the Proposed Method reliably performs better than the RNN and Traditional Image Processing Methods respectively.The F1 values of the Traditional Methods drop noticeably after numerous iterations, resulting in poor performance and instability.Although RNN outperforms conventional techniques, its performance is inconsistent and frequently varies greatly.With the highest F1 values and the most stability, the Proposed Method is probably the best solution for the given situation because it effectively balances precision and recall. In figure 4.4 shows the result of F1-score.

5. Conclusion. Enhancing spaces' aesthetic appeal and usefulness is one of environmental art design's main goals. However, traditional methods for evaluating and enhancing such designs are often arbitrary and unproductive. This study proposes a distributed system that utilizes advanced image processing techniques based on computer-aided design (CAD) features and Generative Adversarial Networks (GANs) to evaluate and enhance environmental art design. The methodology stresses the decentralization of computational resources for enhanced flexibility and efficacy. It is based on distributed demand-side management in intelligent energy systems. The system uses GANs for image creation and design transmission in order to simulate various design scenarios and improve depending on aesthetic qualities including color balancing, spatial arrangement, and visual balance. The use of this technique in a distributed system facilitates real-time feedback, continuous improvement, and expedites the evaluation process. An analysis of the experiment's results in comparison reveals the exceptional quality and efficacy of the method described in this study, outperforming other techniques

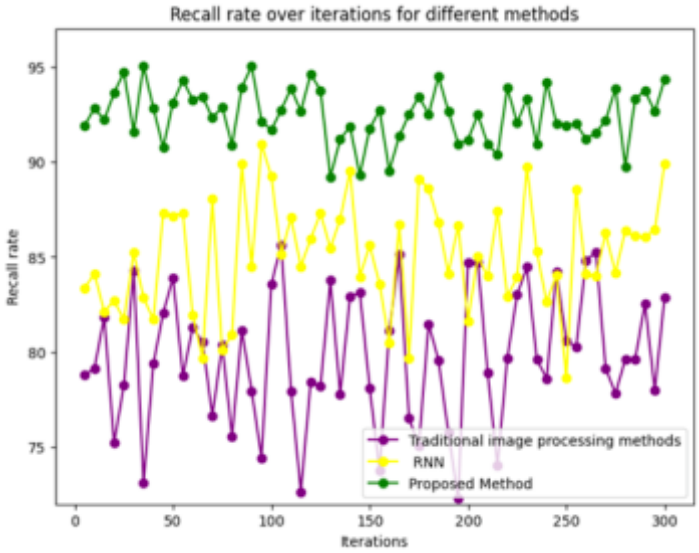


Fig. 4.3: Recall Rate

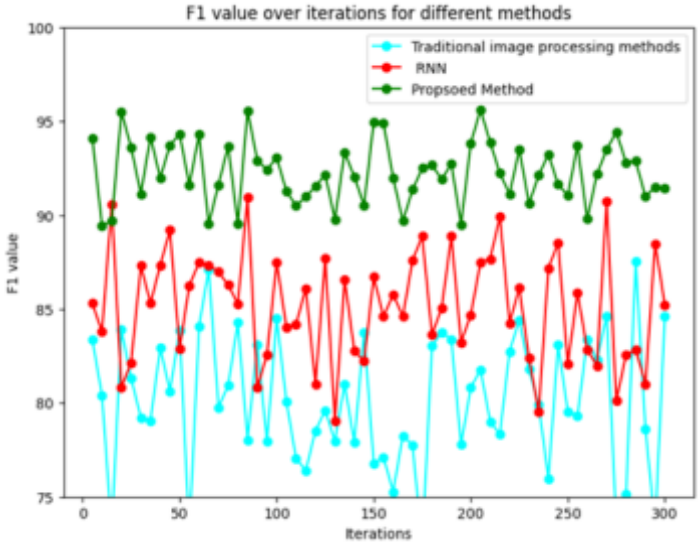


Fig. 4.4: F1-Score

in terms of precision, recall, and F1 score. The outcomes confirm that the recommended method performs better when it comes to component extraction and recognition in CAD environmental artwork design.

Acknowledgement. Key Project of Humanities and Social Sciences Research in Anhui Province in 2021, Campus Landscape Sorting and Comparative Study of Anhui Universities, Project Number: SK2021A1089.

REFERENCES

[1] D. Ali, A. A. Salih, O. M. Ahmed, A. A. Yazdeen, R. Majeed, and T. M. G. S. Abdullah, *A comparative review of*

sustainable enterprise system frameworks: Integrating web technology, clouding, ai, iot, and security for green business transformation, Journal of Information Technology and Informatics, 3 (2024).

[2] S. Aminizadeh, A. Heidari, M. Dehghan, S. Toumaj, M. Rezaei, N. J. Navimipour, F. Stroppa, and M. Unal, *Opportunities and challenges of artificial intelligence and distributed systems to improve the quality of healthcare service*, Artificial Intelligence in Medicine, 149 (2024), p. 102779.

[3] R. Anand, D. Pandey, D. N. Gupta, M. Dharani, N. Sindhwani, and J. Ramesh, *Wireless sensor-based iot system with distributed optimization for healthcare*, Meta Heuristic Algorithms for Advanced Distributed Systems, (2024), pp. 261–288.

[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] F. Cuares and J. I. Teleron, *Harmony in nodes: Exploring efficiency and resilience in distributed systems*, (2024).

[6] P. K. Devulapalli, S. K. GAE, S. B. Maganti, and S. Rachapogula, *Image transmission in mobile wireless multimedia sensor networks using cat swarm optimization*, Multimedia Tools and Applications, 83 (2024), pp. 17557–17578.

[7] Z. Han, R. Zhou, C. Xu, Y. Zeng, and R. Zhang, *Inss: An intelligent scheduling orchestrator for multi-gpu inference with spatio-temporal sharing*, IEEE Transactions on Parallel and Distributed Systems, (2024).

[8] I. A. Hashem, A. Siddiqa, F. A. Alaba, M. Bilal, and S. M. Alhashmi, *Distributed intelligence for iot-based smart cities: a survey*, Neural Computing and Applications, (2024), pp. 1–36.

[9] J. Issa, L. M. Abdulrahman, R. M. Abdullah, T. M. G. Sami, and B. Wasfi, *Ai-powered sustainability management in enterprise systems based on cloud and web technology: Integrating iot data for environmental impact reduction*, Journal of Information Technology and Informatics, 3 (2024), p. 154.

[10] Z. Jalali Khalil Abadi and N. Mansouri, *A comprehensive survey on scheduling algorithms using fuzzy systems in distributed environments*, Artificial Intelligence Review, 57 (2024), p. 4.

[11] J. Jeyaraman, S. V. Bayani, and J. N. A. Malaiyappan, *Optimizing resource allocation in cloud computing using machine learning*, European Journal of Technology, 8 (2024), pp. 12–22.

[12] P. Jiao and W. Cao, *Identification and extraction of environmental art design elements based on computer vision and neural network*, (2024).

[13] S. I. Kaitouni, I. A. Abdelmoula, N. Es-sakali, M. O. Mghazli, H. Er-retby, Z. Zoubir, F. El Mansouri, M. Ahachad, and J. Brigui, *Implementing a digital twin-based fault detection and diagnosis approach for optimal operation and maintenance of urban distributed solar photovoltaics*, Renewable Energy Focus, 48 (2024), p. 100530.

[14] V. S. K. Koppisetti, *Machine learning at scale: Powering insights and innovations*, ESP International Journal of Advancements in Computational Technology (ESP-IJACT) Volume, 2 (2024), pp. 56–61.

[15] Y. Liu, Y. Xu, and R. Song, *Transforming user experience (ux) through artificial intelligence (ai) in interactive media design*, (2024).

[16] S. Maheswaran, *Intelligent cold chain security: Nano power temperature sensors, esp32 and telegram bot integration for temperature assurance and environmental harm prevention*, J. Environ. Nanotechnol, 13 (2024), pp. 17–25.

[17] G. Maier, A. Albanese, M. Ciavotta, N. Ciulli, S. Giordano, E. Giusti, A. Salvatore, and G. Schembra, *Watchedge: Smart networking for distributed ai-based environmental control*, Computer Networks, 243 (2024), p. 110248.

[18] A. Mallik, J. Xie, and Z. Ran, *A performance analysis modeling framework for extended reality applications in edge-assisted wireless networks*, in 2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS), IEEE, 2024, pp. 726–737.

[19] A. Matei and M. Cocoșatu, *Artificial internet of things, sensor-based digital twin urban computing vision algorithms, and blockchain cloud networks in sustainable smart city administration*, Sustainability, 16 (2024), p. 6749.

[20] Y. Y. F. Panduman, N. Funabiki, E. D. Fajrianti, S. Fang, and S. Sukaridhoto, *A survey of ai techniques in iot applications with use case investigations in the smart environmental monitoring and analytics in real-time iot platform*, Information, 15 (2024), p. 153.

[21] E. Ramos-Sorroche, J. Rubio-Aparicio, J. Santa, C. Guardiola, and E. Egea-Lopez, *In-cabin and outdoor environmental monitoring in vehicular scenarios with distributed computing*, Internet of Things, 25 (2024), p. 101009.

[22] A. K. Srivastava, S. Kumar, and M. Zareapoor, *Self-organized design of virtual reality simulator for identification and optimization of healthcare software components*, Journal of Ambient Intelligence and Humanized Computing, (2024), pp. 1–15.

[23] K. Thangavel, R. P. Perumal, K. F. Hussain, A. Gardi, and R. Sabatini, *Multidisciplinary design and optimization of intelligent distributed satellite systems for earth observation*, Acta Astronautica, 216 (2024), pp. 411–427.

[24] B. Vivek, A. Arulmurugan, S. Maheswaran, S. Dhamodharan, A. Dharunash, and N. Gowtham, *Design and implementation of physical unclonable function in field programmable gate array*, in 2023 8th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2023, pp. 152–158.

[25] B. Vivek, S. Maheswaran, P. Keerthana, S. Sathesh, S. Bringeraj, R. A. Sri, and S. A. Sulthana, *Low cost raspberry pi oscilloscope*, in 2018 international conference on intelligent computing and communication for smart World (I2C2SW), IEEE, 2018, pp. 386–390.

[26] B. Vivek, S. Maheswaran, N. Prabhuram, L. Janani, V. Naveen, and S. Kavipriya, *Artificial conversational entity with regional language*, in 2022 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2022, pp. 1–6.

[27] Z. Wang, M. Goudarzi, M. Gong, and R. Buyya, *Deep reinforcement learning-based scheduling for optimizing system load and response time in edge and fog computing environments*, Future Generation Computer Systems, 152 (2024), pp. 55–69.

[28] Y. Wu, *Reference image aided color matching design based on interactive genetic algorithm*, Journal of Electrical Systems,

20 (2024), pp. 400–410.
[29] A. H. Younus, A. A. Salih, O. M. Ahmed, A. A. Yazdeen, R. M. Abdullah, and T. M. G. Sami, *Web-based and cloud-computing influences on resource utilization optimization for sustainable enterprise systems with ai, iot, and security*, management, 12 (2024), p. 13.
[30] M. Zhang, B. Yuan, H. Li, and K. Xu, *Llm-cloud complete: Leveraging cloud computing for efficient large language model-based code completion*, Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 5 (2024), pp. 295–326.
[31] X. Zhang, J. Wang, Y. Zhou, H. Wang, N. Xie, and D. Chen, *A multi-objective optimization method for enclosed-space lighting design based on mopso*, Building and Environment, 250 (2024), p. 111185.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: Sep 26, 2024
Accepted: Nov 26, 2024


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2286–2295, DOI 10.12694/scpe.v26i5.5037

NETWORK TRAFFIC ANOMALY DETECTION ALGORITHMS ON DISTRIBUTED SYSTEMS USING COGNITIVE INTELLIGENCE

NING PAN*

Abstract. Network traffic monitoring is one of the important roles to maintain the security and confidentiality between distributed systems particularly in detecting early cyber threats. Distributed system is a large network interconnected device, which are connected with one another. So, detecting anomalies is a major challenge in these systems. Traditional systems fail to detect anomalies in early stages, because threats are too advanced which are not handled by the traditional capacities. To address this issue the present study proposed and improved version network traffic monitoring system called (Net-IV). This approach combines the advantages of 1D-CNN, Long Term-Short Term Memory (LSTM) and GRU (Gated Recurrent Units). Accoridng to this, 1D-CNN which is well known for its feature extraction ability, whereas LSTM helps to analyse the temporal dependencies, finally GRU refine the overall performance and helps to detect anomalies with greater precision. The model was evaluated using CIC-1DS-2017 dataset, a complete benchmark dataset for intrusion detection system. Through the simulation, we observed that the suggested Net-IV achieves a remarkable accuracy rate of 99.78% and F1-Score of 99.56% which is 0.05% higher than the existing DCGCANet model. Thus, the results suggested that the proposed Net-IV system could be effectively installed in real-time, to protect the distributed system confidentialities from various forms of cyber-attacks.

Key words: Network traffic anomaly detection, CNN, LSTM, GRU, distributed system, cognitive intelligence

1. Background.

1.1. Distributed Systems Advantages and the Impacts of Cyber Threat. Distributed system is one of the advanced technologies which interconnected with multiple systems to achieve a specific goal. This involves the advantages of resource sharing, fault tolerance, scalability and parallel processing which is very helpful for large scale applications like cloud computing, data processing, and network services [13, 1]. One of the main advantages of the distributed system is their ability to operate efficiently and continuously, even in the situations of individual system failure. But however, the decentralized nature of distributed systems also introduces risks in the form of cyber-attacks, which targets any node or system can disturb the entire process and the confidentiality of the privacy data. Distributed Deniel of Service (DDoS) attacks, man in the middle attacks, and malware propagation are some of the common risks which impacts the data privacy availability and confidentiality [12, 4]. Since distributed systems depends on network communication for, by using this advantage these attacks are spread quickly and leads to the sever impacts like network corruptions, data breaches, leakages in privacy data. So, there is an increasing demand of advanced anomaly detection algorithms, cognitive intelligence solutions to continuously monitor the network traffic, detecting anomalies in real-time and prevent possible security risks [9]. This helps to anomaly free transformations between the systems and helps to improve the security of distributed systems also to guarantees the safety of confidential data.

1.2. Previous Techniques and its Drawbacks. Most effective techniques are proposed in the existing articles to address the anomaly detection issues, but however these techniques also face some common limitations which is discussed below.

This study [3] used the benefit of unstructured log analysis technique with Finite State Automaton (FSA) for anomaly detection. Thos approach mainly depends on log keys; these keys are not able to capture all the system behaviors effectively. This paper [16] identifies the anomalies by using the triangular approach, this model also fails in accurately identifies the anomalies due to the limitations of big dimensionality. VeLog a VAE variational autoencoder method is used to detect anomalies in distributed systems is the main of this study [14],

*Hubei University Big Data Center,Wuhan,430062, China, email: ningpancampupeas@rediffmail.com

but however we find the difficulties regarding fine tuning the models in real-time large-scale environments. In this study [8] Fuzzy base clustering approach is used to handle the network anomalies, but there is an increase in clustering it leads to computational overload and make the possibility in network complexities. This study [5] used a deep learning based BiLSTMtransformer-based model for anomaly detection. This approach also highlights the drawback regarding extensive computational resource.

1.3. Suggested Net-IV and its advantages. By thoroughly analysing the existing studies advantages and disadvantages, the present study focuses to present the hybrid model called Net-IV, which combines the benefits of powerful deep learning techniques: 1D-CNN, LSTM, and GRU, which are combined to create the effective structure to address the risk of threats, also detects the anomalies early to avoid the severe impacts.

The main aim of the study is to construct a new network traffic anomaly detection system called Net-IV and examine its ability to learn and operate on distributed cognitive intelligence systems. we seek to improve the performance and speed of anomaly detection by integrating 1D-CNN for feature embedding, LSTM to capture temporal dependencies, and GRU for performance tuning and precision improvement. The study focuses on offering a comprehensive method to secure distributed systems in real time as an alternative to existing solutions that are incapable of addressing the problem of recognizing low and high level cyber attacks including advanced persistent threats.

The main contributions of the paper as follows

1. In Net-IV, ID-CNN is responsible for extracting the essential features from raw network traffic data by identifying abnormal patterns and provide the strong foundation for further analysis.
2. LSTM and GRU are used to capture temporal dependencies in the data and allows the system to understand the sequence and flow of traffic over time.
3. Finally, Net-IV model is simulated using CIC-IDS-2017 network traffic dataset

1.4. Literature Discussions. According to the proof of 2016 Mirai-injected IP camera attack, risk will take place due to the, unsecured cheap devices, some of the attacks happening is denial-of-service (DDoS) attacks in IoT environments. By analyzingthis, the study [6], suggested D-PACK mechanismwhich combines a CNN with an unsupervised autoencoderdetects traffic anomalies with near-perfect accuracy and a low false-positive rate. Existing machine learning based anomaly detection facility commonly struggles due to the adaptability and handling the multiple data. To address this, the study [17] proposed LSTM based approach to classify the abnormal traffic data with high precision. Article [15] also suggests the LSTM based period wise detection approach to achieve the better performance in real time anomaly detection. Study [11] investigates the anomaly detection by using the proposed model as conditional variational autoencoder (CVAE) and random forest (RF) classifier to effectively improve the detection ability of anomalies. Traditional deep learning drawbacks in anomaly detection and how it is addressed using the advanced deep learning solution was discussed in the study [10]. This study achieves encouraging scores demonstrating how deep learning can improve anomaly detection by processing raw network traffic data without requiring domain expertise for feature selection.

The task of identifying anomalies in distributed systems is an overwhelming task that comes with many complications due to the size, complexity, and changing nature of these ever expanding networks. Distributed systems often consist of various interconnected devices that produce large amounts of diverse data making it hard to detect thinly layered or hidden anomalies related to cyber threats. Conventional means of detecting anomalies, such as static rules or baselines, are ill-equipped to respond to changing attack patterns by unforeseen or sophisticated threats. For example, stealthy types of Advanced Persistent Threats (APTs) target other forces over a long period to penetrate the systems and remain undetected by conventional methods. To addition, zero day attacks focus on unknown vulnerabilities making them able to bypass signature based detection systems. Distributed Denial of Service (DDoS) attacks propagate vast amounts of legitimate traffic never been witnessed before making conventional systems unable to determine whether an attack is in progress or whether there is a peaceful activity. Encrypted traffic and polymorphic malware that constantly mutates even more contribute to the problems of detection. Traditional approaches fail in both handling that amount of data in a reasonable time frame or analyzing the current scene of traffic therefore making comprehensive distrust of such distributed systems possible. These challenges underline the need for advanced anomaly detection techniques, cognitive intelligence models for instance, which are fitted for the needs of such complicated environments.

The traditional anomaly detection approaches are limited in many ways to effectively identify early-stage

and more complex threats. Even though these systems can detect anomalies, advanced cognitive and predictive threat analytics would assess predefined signatures, static threshold limits, or statistical baselines. More advanced threats, such as Advanced Persistent Threats (APTs) and zero-day exploits, have been purposely developed to defeat these types of detection mechanisms–they target becoming undetectable or take advantage of never before seen attack vectors. For instance, polymorphic malware makes frequent alterations to its code structure hence bypassing signature-based systems. Encrypted traffic, on the other hand, harbors vicious activity as traditional approaches cannot analyze encrypted patterns.

Similarly, they cannot process real-time large volumes of high-dimensional data streams produced by distributed environments. This leads to slow times in detection and response times, especially in the presence of compounds in the traffic patterns. The temporal dependencies of sequential data are also not confined by traditional approaches; for example, an attack could be started by minor changes in user behaviors or the flow of a network before it becomes fully aggressive. This absence of context compounds the issues of very high false positive rates and undermines the confidence in these systems.

2. Methodology.

2.1. Net-IV Structure Outline. The proposed methodology consists of 3 layers: Initial layer involves, 1D-CNN based feature extraction, LSTM for further analysing the extracted patterns and finally GRU helps to detect anomalies, LSTM and GRU are the two main Recurrent Neural Network layers helps to thoroughly analysing the traffic patterns and detect the anomalies with high precision. Here, ID-CNN processes raw network traffic data to extract meaningful spatial features. This componenthelps to identify the relevant patterns by reducing the dimensionality and fed into the layer of LSTM and GRU for further analysis. Together with these layers (LSTM and GRU), the model thoroughly analyze the normal behavior of traffic sequence and identifying abnormalities with high accuracy. The final component of the architecture is fully connected layer that combines the outputs of both LSTM and GRU followed by the softmax layer for classification. This allows the model to identify the final predictions about the network traffic is normal or abnormal. The whole framework is designed to handle the large-scale network traffic data in real-time. Finally, the model is assessed using the real-world comprehensive dataset and highlight the effective output, and acts as an efficient contribution for monitoring traffic in distributed systems. Figure 2.1 presents the visual illustration of proposed architecture.

The Net-IV model brings together three cutting-edge branches of neural networks; namely, 1D-CNN, LSTM, and GRU into the network traffic anomaly detection process to enhance accuracy and precision. The 1D-CNN (One-Dimensional Convolutional Neural Network) deals with feature extraction, allowing for the identification of important spatial features in the traffic flow data including packet sizes and flow attributes. Its convolutional layers go through the raw inputs of the data and capture high-order features about the most relevant attributes for anomaly detection while noise and redundancy are minimized.

2.2. 1D-CNN based Feature Extraction. The initial layer of 1D-CNN performs the initial feature extraction in the suggested Net-IV architecturewhich helps to handling the unprocessed network traffic data. The input layer, convolutional layer, pooling layer, fully connected layer, and output layer are the main layers of conventional 1D-CNN architecture. The input layer receives the time-series, raw network traffic data and prepares it for convolutional processes, which will further convert it. Applying convolutional kernels to the input data is the main job of the next convolutional layer in Net-IV. This allows for the extraction of important spatial information from the network traffic data, about anomaly patterns.

The convolutional kernel in a 1D-CNN refers one-dimensional array. The convolution procedure significantly reduces the number of parameters in a 1D-CNN by sharing weights among neurons. Then output of convolutional layer's expressed as

$$ax_n^j = g\left[\sum_{i\in N}\left(ax_{n-1}^i kl_n^{ij}\right) + bi_n^j\right] \tag{2.1}$$

Here, ax_n^j denotes the output of the $j-th$ neuron in the $n-th$ layer, ax_{n-1}^i is the previous layer input and kl_n^{ij} denotes the kernel connecting the $i-th$ neuron pf the previous layer to the $j-th$ neuron in the current layer, bi_n^j is the bias terms and g is the activation function. After convolution the pooling operation is applied

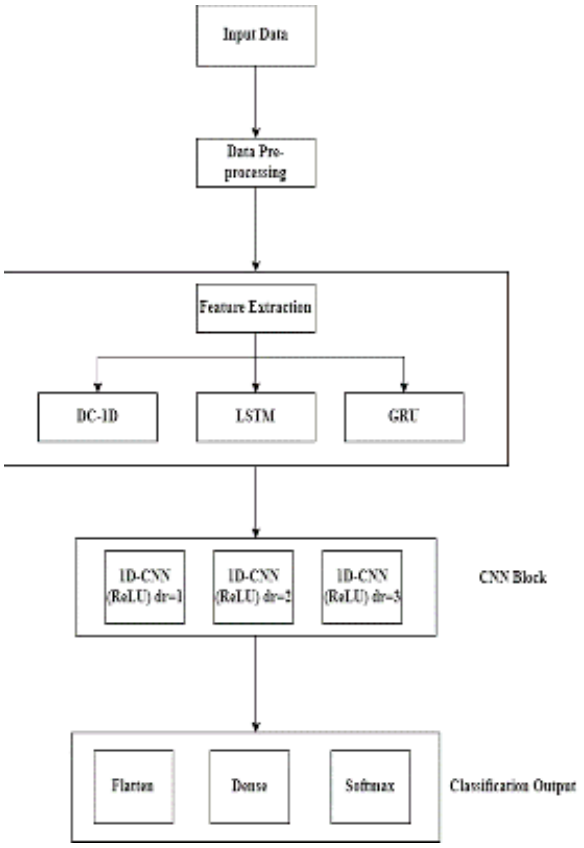


Fig. 2.1: Suggested Framework

downsample the data and reduce the complexity. Pooling can perform using either max pooling or average pooling. Max pooling selects the maximum value within the specific window and average pooling calculates the average value in the same window. These can be defined as

$$pz_n^i = maxpool(ax_{n-1}^i, s1, s2) \tag{2.2}$$

$$pz_n^i = avepool(ax_{n-1}^i, s1, s2) \tag{2.3}$$

Here, pz_n^i denotes the output of the pooling operation in the i-th neuron in the n-th layer and s1, s2 is the pooling scale and step size respectively. This pooling helps to reduce the size of feature map and safeguard the most important data when the process of eliminating redundant data takes place.

2.3. Integration of LSTM-GRU layers. The integration of GRU and LSTM networks in the Net-IV architecture provideseffective way for managing the temporal dependencies present in network traffic data. The sequential data processing abilities of LSTM and GRU helps to perfectly identifying abnormalities in real-time network traffic data.

Using its cell state, input gate, forget gate, and output gate processes, LSTM is used in Net-IV to maintain long-term dependencies in the network data. The LSTM network changes its hidden state for a given input sequence $\{ax\}_{t=1}^{N}$?, where ax_thighlights a sequence of input vectors and N denotes the total number of occurrences. This can be expressed as

$$pz_t^i = \sigma(w\ ax_t^i + u\ pz_{t-1}^i) \tag{2.4}$$

where w and u are the weight matrices and σ is the activation function. The hidden state pz_t^i at each time step is used to capture long term dependencies in the input data. The LSTM output is then, processed through the fully connected layer (FCN) and the softmax function for classification, this process is expressed as

$$p\left(y-i \mid ax\right)=softmax\ \ (wax+bi)=\frac{e^{w_i ax+bi_i}}{\sum_j e^{w_j ax+bi_j}} \tag{2.5}$$

The LSTM process helps to learn the temporal dependencies over long sequence, but it highlights the limitations of computationally expensive performance. To reduce the complexity in managing the performance Net-IV involves GRU layer on the other side of LSTM. GRU are simply faster using few parameter functions. It involves the update and reset gate to process the flow of information. These are expressed as the following

$$rt_t = sig(w_{rt} \cdot [h_{t-1}, ax_t]) \tag{2.6}$$

$$pz_t = sig(w_{pz} \cdot [h_{t-1}, ax_t]) \tag{2.7}$$

$$\tilde{h}_t = tanh(w_h \cdot [rt_t \times h_{t-1}, ax_t]) \tag{2.8}$$

$$h_t = (1 - pz_t) \cdot h_{t-1} + pz_t \cdot \tilde{h}_t \tag{2.9}$$

Here rt_t controls the reset gate, pz_t is the update gate and h_t is the hidden state at time step t. When compares to LSTM GRU simplify the process of the model and effectively capturing the dependencies.

The Net-IV architecture makes use of both the efficiency of GRU and LSTM by combining the outputs of the two. With the least amount of training time and computational resources, our hybrid technique guarantees that the model can accurately identify complex anomalies in network data. To train both models, the back-propagation algorithm is used, this helps to minimizes the cost function CL, which is expressed as

$$CL = \sum_{i=0}^{|D|} \log(p\left(ay\right) = (ay_i \mid ax_i, w, bi)) \tag{2.10}$$

This is the final optimization which ensures the model identifies the best parameters for detecting anomalies in large-scale real time network environments.

3. Results and Experiments.

3.1. Dataset Description. Developed in 2017, the CIC IDS 2017 dataset was created by the Faculty of Computer Science from the University of New Brunswick, aiming to improve the ISCX 2012 dataset. Advancing the earlier work, this dataset defines a realistic depiction of network usage and is intended to remedy the drawbacks present in prior intrusion detection system (IDS) datasets. The researchers claim that the CIC IDS 2017 dataset complies with 11 requirements regarding the design of IDS datasets, such as full network configuration, labeled set, different types of attacks, and extensive meta-data. These criteria allow us to say that the dataset is representative and of adequate size to assess the effectiveness of the IDS.

The dataset encompasses daily traffic and attack data for five days, yielding more than 225,745 network packets, which may have up to 80 different features. It incorporates not only normal network activity but also a range of intrusion activities, collected within a week. The simulated attacks in this dataset are divided into 7 types, which are Brute Force Attack, Heartbleed attacks, Botnet, DoS attacks, DDoS attacks, Web Attacks, and Infiltration Attacks.

The present research takes into consideration the DDoS (Distributed Denial of Service) attacks where a target system or network is assaulted with a flood of traffic mostly from a botnet to bring the system down. Please note some of the key traits associated with DDoS attacks are: inter-arrival times (IAT) of flows (minimum, mean, maximum), flow bandwidth measures, and flow duration. The higher these attributes' values are, the

Table 3.1: Dataset Features

Category	**Details**
Dataset	CIC-IDS-2017
Training Data Period	July 3-July 7, 2017
Types of Traffic	Normal (BENIGN) + 9 Attack Types
Training and Testing Ratio	7:3
Attacks Considered	(DDoS, Brute Force, Infiltration, etc.) - 9
Framework	TensorFlow
Operating System	Windows 11 64-bit
CPU	Intel i7-12700H
GPU	NVIDIA GeForce RTX 3070 Ti
Programming Language	Python 3.6
IDE	PyCharm 2020.1

greater the likelihood that DDoS has been employed. In the case of abuse network behavior, the attributes such as B. Len Min sub server byte length, the number of subflows, total packet size and mean packet size are of more relevance.

The proposed study evaluated using the CIC-IDS-2017 dataset inspired from[2], according to the details we extract the necessary features to evaluate the proposed Net-IV. Table 3.1 provides the clear illustration about the dataset.

3.2. Pre-processing CIC-IDS-2017 Dataset. Two important processes are involved in preprocessing the CIC-IDS-2017 dataset: Feature normalization and one-hot encoding. Non-numeric features, like category attributes, are converted into binary vectors using one-hot encoding. Then it is used as input in the deep learning model. The model was able to process features such as 'protocol_type' efficiently because they were converted into binary representations. Features were also normalized to address the significant differences in feature scales. The performance of the model is affectedby some features, such as "Flow IAT Max" and "Total Length of Fwd Packets". To address this, all features were scaled to a range of [0, 1] using mean-variance normalization.

According to parameter setting, inspired from the study[7], according to the procedures we perform the evaluation of proposed framework.

3.3. Evaluation Criteria. The proposed model is evaluated using the common performance metrics of accuracy, recall, precision and F1-Score. The process of evaluation is conducted according to the attack types. Figure 3.1 highlights the model's efficacy in different attack types. According the epoch of 10 the accuracy of the proposed model reached up to 99.78%. Overall, the Bot and Portscan highlights the slight decrease. But when compared with the accuracy, the proposed obtains an effective score in all the attack types.

Figure 3.2 presents the efficacy of models in terms of accuracy, precision, recall and F1-score, according to this the proposed Net-IV model is compared with the existing KNN, ID3, MLP, RF, CNN, AFM-1CNN-1D, CNN-GRU, DCGCANeT. From the figure we observed that the proposed model significantly out performs all the models with its remarkable accuracy. Also, it shows that the proposed Net-IV achieves the accuracy about 99.78%, 99.82% and 99.72% of precision and recall respectively. And finally, the F1-score is about 99.58%, when compared with the existing DCGCANet the proposed model significantly outperforms the DCGCANet model due to its additional involvement of GRU

Similarly Figure 3.3 presents the efficacy of models in terms of computation time. The computation time of the suggested model 250.12 seconds, when compared with the tradition individual techniques of KNN, RF, CNN, ID3 the fusion model utilizes the additional timing to process the data effectively, as a results, it slightly higher than the traditional individual models. When compared with the existing fusion methods the proposed model really beats the other models with its less inference time.

Figure 3.4 presents the results of generalization ability of the models. According to this we involve 4 subsets called SP1, SP2, SP3 and SP4 to test the entire CIC-IDS-2017 dataset was inspired from[7]. In this section

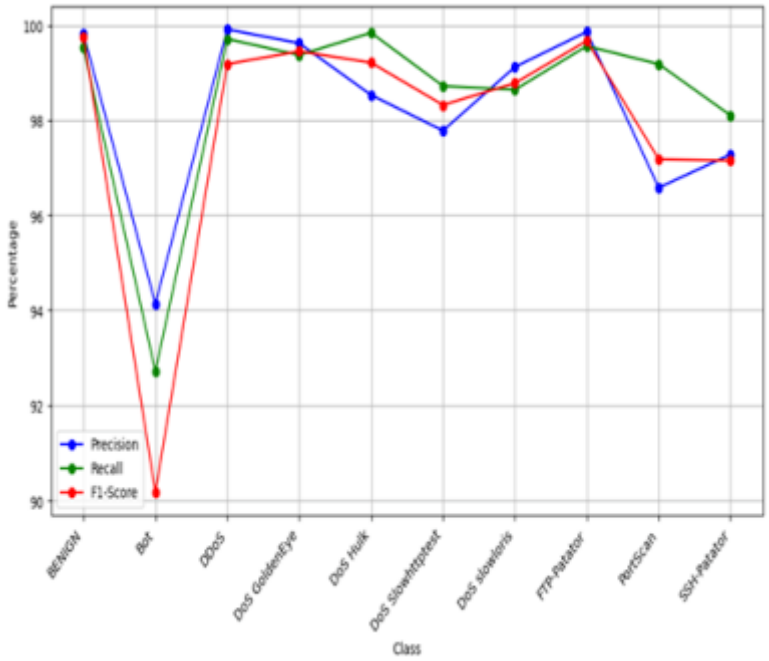


Fig. 3.1: Performance Evaluation under CIC-IDS-2017 Based Attack Types

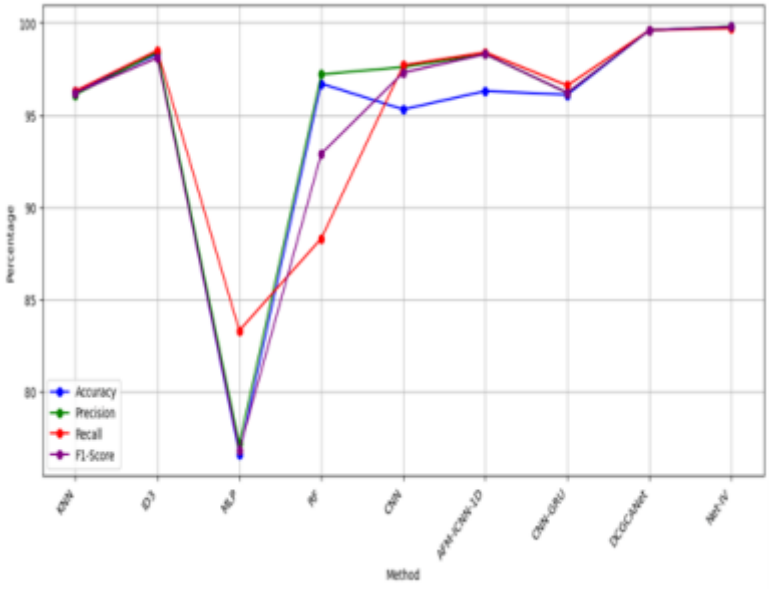


Fig. 3.2: Overall Performance Comparison of Models

the proposed Net-IV model is compared with the existing effective DCGCANet model. This model is a fine competitor of proposed Net-IV model. But the result highlights the proposed Net-IV model outperforms the DCGCANet with its effective outcomes.

The use of 1D-CNN, LSTM, and GRU for network related activity monitoring thesis can be explained

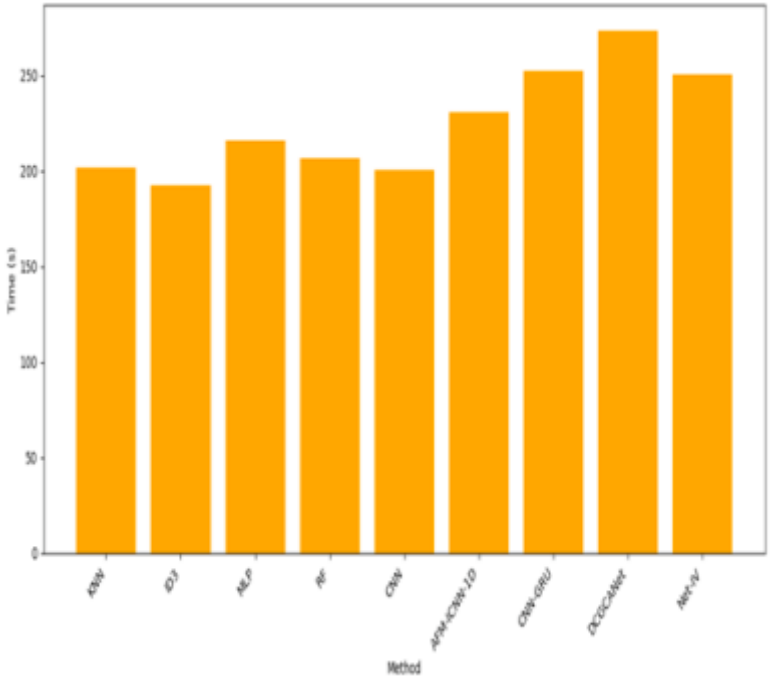


Fig. 3.3: Comparison of Computational Time

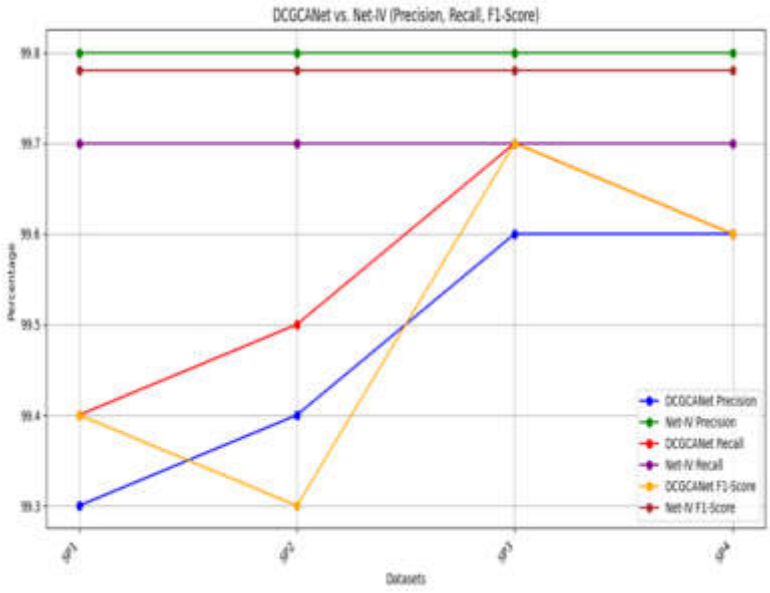


Fig. 3.4: Subset Based Efficiency Analysis

by the fact that these networks show a combination of complementing strengths in capturing sequential and high dimensional data respectively. The one dimensional convolution network (1D-CNN) was picked mainly because it is more efficient at extracting particular traffic data spatial features such as packet size and flow

patterns at less computational cost than 2-D CNN. The convolutional layers within it capture local patterns quite effectively which are very useful in detecting any abnormalities within network traffic. However, a Long Short-term Memory (LSTM) LSTM was added to solve the requirement of the network traffic's time related features in sequential data analysis. Long short term memory (LSTM) networks are very good at remembering information for long periods, using its memory cells to assist in detecting time dependent, minute anomalies quite easily, such as advanced persistent threats (APTs) and traffic variability over time.

LSTMs and GRUs (Gated Recurrent Units) governance or structures are thought to be integrated due to their hierarchical efficient structure and their potential to cause problems often referred to as the vanishing gradient in recurrent neural networks (RNNs). Because GRUs have fewer parameters than LSTMs, GRUs are quicker to train and validate while maintaining precision. The model's overall effect is improved since maximum memory and processing efficiency are used.

4. Conclusion. The present study introduced the effective Net-IV model to analyse and identify the network traffic anomalies. The proposed model effectively combines the advantages of ID-CNN, LSTM and GRU to finely detect the abnormalities in network raw data. The suggested model is evaluated with CIC-IDS-2017 dataset, a comprehensive benchmark dataset used to evaluate in the intrusion detection scenarios. By compare the results of our proposed model with the existing models particularly, with the existing DCGCANet model, Net-IV proves it efficacy with the accuracy score of 99.78% and 99.56% of F1-score which is 0.05 % improvement when compared with the existing DCGCANet model. The future search is needed to effectively reduce the computational complexities was raised due to the fusion of the models. The future search is needed to effectively reduce the computational complexities was raised due to the fusion of the models. When incorporating continuous learning frameworks, the model will be able to smoothly transition towards new or newer threat behaviours without the need for total retraining, and therefore maintain their reliability and relevance over time.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, *Anomaly detection: A survey*, ACM computing surveys (CSUR), 41 (2009), pp. 1–58.

[2] K. Fotiadou, T.-H. Velivassaki, A. Voulkidis, D. Skias, S. Tsekeridou, and T. Zahariadis, *Network traffic anomaly detection via deep learning*, Information, 12 (2021), p. 215.

[3] Q. Fu, J.-G. Lou, Y. Wang, and J. Li, *Execution anomaly detection in distributed systems through unstructured log analysis*, in 2009 ninth IEEE international conference on data mining, IEEE, 2009, pp. 149–158.

[4] R. A. A. Habeeb, F. Nasaruddin, A. Gani, I. A. T. Hashem, E. Ahmed, and M. Imran, *Real-time big data processing for anomaly detection: A survey*, International Journal of Information Management, 45 (2019), pp. 289–307.

[5] P. Han, H. Li, G. Xue, and C. Zhang, *Distributed system anomaly detection using deep learning-based log analysis*, Computational Intelligence, 39 (2023), pp. 433–455.

[6] R.-H. Hwang, M.-C. Peng, C.-W. Huang, P.-C. Lin, and V.-L. Nguyen, *An unsupervised deep learning model for early network traffic anomaly detection*, IEEE Access, 8 (2020), pp. 30387–30399.

[7] C. Ji, H. Yu, and W. Dai, *Network traffic anomaly detection based on spatiotemporal feature extraction and channel attention*, Processes, 12 (2024), p. 1418.

[8] H. Kumarage, I. Khalil, Z. Tari, and A. Zomaya, *Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling*, Journal of Parallel and Distributed Computing, 73 (2013), pp. 790–806.

[9] H. Li and Y. Li, *Logspy: System log anomaly detection for distributed systems*, in 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), IEEE, 2020, pp. 347–352.

[10] G. Marín, P. Casas, and G. Capdehourat, *Rawpower: Deep learning based anomaly detection from raw network traffic measurements*, in Proceedings of the ACM SIGCOMM 2018 Conference on Posters and Demos, 2018, pp. 75–77.

[11] M. Monshizadeh, V. Khatri, M. Gamdou, R. Kantola, and Z. Yan, *Improving data generalization with variational autoencoders for network traffic anomaly detection*, IEEE Access, 9 (2021), pp. 56893–56907.

[12] A. Pamukchiev, S. Jouet, and D. P. Pezaros, *Distributed network anomaly detection on an event processing framework*, in 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), IEEE, 2017, pp. 659–664.

[13] A. D. Pazho, G. A. Noghre, A. A. Purkayastha, J. Vempati, O. Martin, and H. Tabkhi, *A survey of graph-based deep learning for anomaly detection in distributed systems*, IEEE Transactions on Knowledge and Data Engineering, 36 (2023), pp. 1–20.

[14] Y. Qian, S. Ying, and B. Wang, *Anomaly detection in distributed systems via variational autoencoders*, in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2020, pp. 2822–2829.

[15] Z. Shi, J. Li, C. Wu, and J. Li, *Deepwindow: An efficient method for online network traffic anomaly detection*, in 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International

Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE, 2019, pp. 2403–2408.

[16] S. Thudumu, P. Branch, J. Jin, and J. Singh, *A comprehensive survey of anomaly detection techniques for high dimensional big data*, Journal of Big Data, 7 (2020), pp. 1–30.

[17] Y. Zhong, W. Chen, Z. Wang, Y. Chen, K. Wang, Y. Li, X. Yin, X. Shi, J. Yang, and K. Li, *Helad: A novel network anomaly detection model based on heterogeneous ensemble learning*, Computer Networks, 169 (2020), p. 107049.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: Sep 27, 2024
Accepted: Nov 26, 2024

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 5, pp. 2296–2305, DOI 10.12694/scpe.v26i5.5039

DISTRIBUTED SYSTEMS FOR SIMULATION ANALYSIS OF MOTOR DRIVE SYSTEMS USING ADAPTIVE ALGORITHMS

LILI KONG* AND CHUNQIU YI†

Abstract. High processing needs and latency make it difficult to simulate motor drive systems in distributed environments, which affect accuracy and real-time performance. Optimizing motor control and cutting energy use require effective modeling. Conventional simulation techniques have trouble scaling up and down, frequently demanding large amounts of resources and being unable to adjust to changing load circumstances, which leads to sluggish or imprecise simulations. This method improves scalability and lowers latency by dynamically adjusting computing loads in a distributed system through the use of adaptive algorithms. It improves the accuracy and efficiency of simulation by utilizing real-time adaption and parallel processing. The purpose of this work was to suggest distributed systems for motor drive system simulation analysis utilizing adaptive algorithms. Initially, the dataset was gathered from a test bench-mounted momentum permanent magnet synchronous motor (PMSM) in a three-phase system motor vehicle. The exponentially weighted moving standard deviation (EWMS) utilized in standardized data process representations for training. We proposed the Adaptive Controller with dynamic fuzzy system ensemble (AC-DMFSE) for distributed systems for simulation analysis of motor drive systems. To optimize motor performance in dynamic situations, adaptive techniques are used, such as fuzzy logic-based optimization and model predictive control. Our test findings show that the suggested distributed technique reduced simulation times and MSE while enhancing the accuracy of system performance evaluation. The foundation for scalable and effective motor drive system simulations is laid by this work, which also offers insightful information for improving the systems' performance in practical applications.

Key words: Simulation Analysis, Motor Drive Systems, permanent magnet synchronous motor (PMSM), exponentially weighted moving standard deviation (EWMS), Adaptive Controller with dynamic fuzzy system ensemble (AC-DMFSE)

1. Introduction. To create a distributed system that uses adaptive algorithms to simulate and analyze motor drive systems. In this regard, motor drive systems are essential parts of many industrial applications, such as automation systems, manufacturing machinery, and electric cars. Conventional simulation techniques for examining complex systems frequently result in inefficiencies since they need large amounts of computing power and centralized handling. Many processing nodes operate in parallel by utilizing a distributed system technique, greatly cutting down on the amount of time needed for simulations. The system incorporates adaptive algorithms to enhance simulation speed and accuracy by dynamically modifying system parameters in response to demands and conditions in real time [4].

Traction motors for new energy vehicles (NEVs) need be modified for challenging operating conditions in contrast to industry motors. They alternate between motoring and generating on a regular basis. The automobile industry demands a lot from its vehicles: frequent starting and stopping, rapid acceleration and deceleration, high torque at low speeds and high power during high-speed climbing, high power density, large, highly efficient operating area, low vibration and noise, high reliability, and a high performance-to-price ratio. The essential components for transforming the electromechanical energy in NEVs are traction motors and motor power electronic controllers [2].

Either an end-to-end controller or a structured controller can serve as the foundation for the AC system. The end-to-end controller is a single core controller that processes input from sensors to provide the required torque [1]. Conversely, the organizational framework is based on an upper and a lower controller. In the AC command hierarchy, every controller has a major function. While the higher control functions as a decision layer, the lower controller is a control layer. The car would maintain a speed that the driver had initially pre-set;

*College of Modern Science and Technology China Jiliang University Yiwu, China 322000 (lilikongintelligent@rediffmail.com)
†Jiangsu Xuzhou Construction Machinery Research Institute, Xuzhou, Jiangsu, 221005, China

this is known as a speed control mode, provided there were no leading vehicles ahead of it and the road was clear [14].

Key advantages of this system include increased precision, scalability, and shorter simulation times all of which are critical for enhancing motor drive designs, boosting energy efficiency, and cutting development costs in many industries [20]. Real-world implementation, however, is not without its difficulties. These include managing the complexity of integrating adaptive algorithms into various hardware and software ecosystems, minimizing network latency, and guaranteeing synchronization across dispersed nodes. In order to fully utilize distributed systems in motor drive simulations, several issues need to be resolved. As a result, we suggested using distributed systems and adaptive algorithms to simulate and analyze motor drive systems.

The design and deployment of sophisticated motor drive systems in modern vehicles as well as in industrial applications demands innovative simulation methods that are accurate and efficient. Distributed environments offer an interesting opportunity for overcoming the challenges of simulating large systems, but current approaches do not perform well in integrating optimization techniques with dynamically changing scenarios. This research is driven by the necessity to create a framework that is adaptive and scalable while integrating advanced tools such as fuzzy logic, model predictive control and real time adaptation. By altering the amount of computing workload and decreasing the amount of latency, the proposed Adaptive Controller with Dynamic Fuzzy System Ensemble (AC-DMFSE) aims at transforming the way motor drive system simulations are performed. As such, this study aims at attaining the desired theoretical level and implementing sophisticated methods for the improvement of the motor control, energy consumption and performance of distributed systems.

1.1. Research contributions. The study contribution concerned networked systems for adaptive algorithm-based simulation analysis of motor drive systems.

1. The dataset was originally collected from a three-phase system motor vehicle with a momentum permanent magnet synchronous motor (PMSM) installed on a test bench.
2. The training representations for standardized data processes use the exponentially weighted moving standard deviation (EWMS).
3. We presented the Adaptive Controller with Dynamic Fuzzy System Ensemble (AC-DMFSE) for distributed systems purpose of simulating motor drive systems,. Adaptive methods, like fuzzy logic-based optimization and model predictive control, are used to maximize motor performance in dynamic conditions.

1.2. Organization of the research, . The remainder of the document is structured as follows: In Section II, pertinent material is reviewed. We discuss the concept, the methodology, and the importance of the intended effort in Section III. The procedures and results of the experiment are described in Section IV. Section V summarizes the analysis and offers some recommendations for further study.

2. Related works. A summary of the most current particular investigations is provided below, along with the accompanying solutions. The research [5] improved the intelligence, adaptability, and resilience of smart production-logistics systems, a self-adaptive collaborative control (SCC) mode is suggested. Cyber-physical systems (CPSs) and the industrial Internet of things (IIoT) are utilized to gather and interpret real-time status data for the purpose of optimization and decision making. This work [16] offered an adaptive Fractional Order PID (FOPID) controller that uses the Artificial Bee Colony (ABC) method to enhance the performance of a Brushless DC (BLDC) motor. This work [15] examined a class of nonlinear systems that have sensor failures and an uncertainty problem. The architecture of the suggested control strategy is predicated on the backstepping process. In order to demonstrate the viability of the suggested system, a simulation of a single-link robot arm example is performed. The study [13] covered the history of electric vehicle (EV) motors, different types of EV motors, mathematical modeling of EV motor drives, and EV motor design process. The hardware outcomes have also been contrasted using various BLDC hub and SRM control strategies. This work [12] demonstrated its superiority over all existing signature identification techniques by outperforming them all. Afterwards, the indices created with the suggested technique were fed into a fuzzy decision box. The ultimate outcomes, which combined suggested methods with fuzzy decisions, were discovered to be 100% correct and effective. The research [10] suggested a unique control scheme for a commuter pull-in hybrid vehicle's adaptive real-time energy management. As long as the deterministic driving condition is met, the simulation results demonstrate

that the suggested technique can achieve an optimal energy distribution on a nearly global optimal level, or near the dynamic programming (DP) level. In order to improve cloud computing data access storage techniques, HDFS was established in the study [17]. Next, the effect algorithm is used to optimize both the data block size and the Internet of Things' topology. At last, the file storage design has been optimized. It is demonstrated through simulated experiments that the optimized cloud storage approach provides clear performance benefits in terms of memory utilization and file read and write speeds. This work [3] introduced the frequency control for the engine generators that manage an islanded MMG's frequency. By taking operating point fluctuations and load uncertainty into account, the suggested method minimizes the rejection of uncertain disturbances and lessens the effects of nonminimum phase dynamics brought on by engine delay. A virtual adaptive inertia control (VAIC) approach is presented in this research [18]. The dual extended Kalman filter is used to estimate the states of energy storage battery packs (ESBPs) in real time. Throughout the microgrid's whole operation, it can reduce voltage fluctuations, enhance system stability, and accomplish decentralized and coordinated control. This work [11] described a DTC of the Double Fed Induction Motor (DFIM) powered by two voltage inverters. Based on a Genetic Algorithm (GA), which has been proposed for optimizing the PID controller's parameters through a weighted combination of objective functions, the DFIM's regulation speed is regulated using a PID controller. The research [6] suggested a optimizing the energy distribution of hybrid energy storage systems (HESS) and its enhanced semi-active architecture in order to further minimize energy loss and degradation of battery capacity. The simulation's findings demonstrate that, under various driving scenarios, battery capacity degradation and energy loss are reduced when compared to the conventional MMC and semi-active topologies. This work [9] presented a controller tuning algorithm called the multi-role exploration strategy distributed deep deterministic policy gradient (MESD-DDPG). It expands on the deep deterministic policy gradient (DDPG) by using a multi-role exploration strategy among other tricks. The simulation results show that the MESD-DDPG adaptive PI controller performs exceptionally well and is highly adaptive.

2.1. Research Gap. The motor drive system dynamic modeling in a distributed environment is extremely challenging mainly due to the high computational requirements, time lag, and lack of scalability. Traditional simulation methods are primarily impeded by harsh static resource provisioning and adjustment to the workload, thus being inefficient and inaccurate in real-time performance. These conventional approaches have also reported the use of energy optimization without considering the dynamic nature of the conditions under which the control systems are implemented. Furthermore, very few studies have reported the fusion of adaptable algorithms such as fuzzy logic based optimization and model predictive control into distributed systems for the case of motor simulation. This gap calls for appropriate mechanisms which are urgently needed to deal with the timing characteristics, precision, and the conformativeness while applying the proposed models in real-life scenarios.

2.2. Problem statement. The difficulty of precisely modeling and analyzing motor drive systems in real-time across distributed environments is the root cause of the issue in Distributed Systems for Simulation Analysis of Motor Drive Systems using Adaptive Algorithms. As motor drive systems become increasingly complex, traditional simulation techniques find it difficult to meet the needs of large computational loads, real-time performance, and adaptability. Adaptive algorithms provide an answer by optimizing performance through dynamic parameter adjustments. To ensure simulation accuracy and efficiency, however, integrating these methods into a distributed system for real-time simulation poses a number of issues, including synchronization, communication latency management, and distributing the computing load among nodes. The problem is even more difficult in real-time settings where system design, control validation, and optimization depend on precise and timely simulation findings. Additionally, conventional methods are constrained by their incapacity to scale effectively, adjust to varying system characteristics, or handle the massive volumes of data produced during the simulations, leading to a slower analysis and inferior overall performance.

3. Proposed system model. The lower computer measurement implements random position control in the distributed motor control system with random position control. The random number algorithm generates the sequence, and then control debugging occurs. The distributed motor measurement and control system's adaptive analysis extraction principles should be compatible with random position control technology, high precision control, precise signal collection, stability and reliability, and model training and generalization capacity. For verification and analysis, the research chooses the random position control and the random number

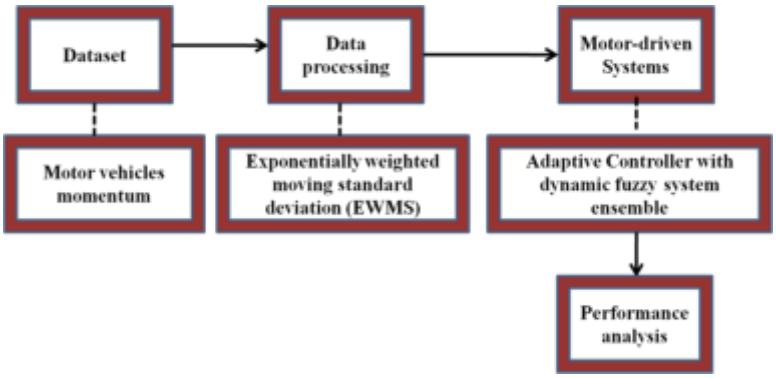


Fig. 3.1: Overall Proposed System

sequence performance. Fig 3.1 depicts the overall proposed architecture.

In the case when Fuzzy Logic is used for control purposes, it fakes human Qualitative Judgment in order to modify motor control parameters. It is a rule based method wherein input variables which include, motor speed, load torque, temperature etc. are controlled through fuzzy membership functions.
Input Variables: Motor speed interruptions caused by deviations, and variations in load torque.
Fuzzy Rules: If the speed deviation is high and the load torque is low, increase the motor current.

These rules allow motor parameters such as current, voltage, etc. to be constantly adjusted in order to improve energy consumption efficiencies whilst ensuring operational stability. Fuzzy Logic achieves this by having the motor performance monitored automatically so as to adjust the motor performance and reduce energy wastage according to changes in load behavior.

Fuzzy control in addition helps oneself to application development by forecasting the future state of the system including a reasonable state error. It can also be defined as PFC where P is predictive and F is Fuzzy control. It is one of the methods of control operation and predictions that use the physical model of the motor system to establish the probable trajectories in the time frame that has been set earlier. Some of the steps undertaken while working with MPC include:

1. State Estimation: The current state of the system is defined as the last recorded state utilization information such as the motor voltage and current.
2. Optimization: The main function of the MPC observational algorithm is to establish the probable movements of the motor in the near future. It achieves this as it has a pre-defined cost function most of the time for the speed deviation minimization or the power usage minimization.
3. Implementation: The control action must first be selected from the sequence and this action repeats itself as new information is collected about the system. The proposed architecture also utilizes fuzzy logic and Adaptive Controller with Dynamic Model Predictive Fifth Element which integrates model predictive capability in structure control with fuzzy adaptable rule base decision logic integration. Fuzzy helps out in reaction orientation to current active participation, whereas MPC assures maintaining and controlling the robustness in a more strategic sense over a long period.

3.1. Dataset. The dataset comprises 140 hours of multivariate indicators taken at a frequency of 2 Hz from an a 3-phase system motor vehicles momentum permanent magnet synchronous motor (PMSM) (52 kW) that is installed on a test bench[8]. The motor speed of the PMSM is controlled by torque, and it is fed by a two-level IGBT inverter (Semikron: 3xSKiiP 1242GB120-4DW). In addition, a test bench is equipped with a dSPACE DS1006MC rapid-control-prototyping system. dSPACE analog-digital converters, which have been synchronized with the control task, recorded all measurements.

3.2. Data processing. Standardized data representations must be used for training purposes. Each feature is then scaled further to show unit variance in the training set by subtracting the mean from every observer. All input signals x have their exponentially weighted moving average (EWMA) and exponentially

weighted standard deviation (EWMS) conformed to the original input space. Thus, as extra regressors, for each signal z_s at time step s, the following two values are given.

$$\mu_s = \frac{\sum_{k=0}^{s} y_k z_{s-k}}{\sum_{k=0}^{s} y_k} \tag{3.1}$$

$$\sigma_s = \frac{\sum_{k=0}^{s} y_k (z_s - \mu_s)}{\sum_{k=0}^{s} y_k} \tag{3.2}$$

Here $y_k = (1-\alpha)_k$ and s is an arbitrary span and $\alpha = 2/s+1$. Strong linear regressors are indicated by different smoothed versions of all time series and their standard deviation that result from multiple, differently spanned EWMAs and EWMSs. To limit memory requirements, only four alternative values for the span in the EWMA and EWMS computation are simultaneously examined. In light of this, the total number of input quantities, which is constant across all tests, indicates 108 characteristics.

Combined with the linear regression the EWMA shifts the focus from the noise in the short term fluctuations in the data. It helps trends or patterns to be more identifiable. This is handy in dealing with inputs with lots of noise, such as in sensor readings or in motor control datasets. With EWMA, more importance is placed on more current observations, allowing for more changes to be introduced to the entire dataset which is ideal for monitoring ever-changing conditions. EWMS takes into account variability in the data on a moving basis, with the most emphasis given to the most current and recent ones, thereby assisting in the interpretation of the changes in data dispersion or volatility most up to the moment. This model stated that the cubic weighting function can effectively suppress the influence of outliers and past observations' noise while still allowing for standard estimates of variability since skewed distributions are known.

Shift of variance in processes and systems is easy to be captured by EWMS however such change can indicate the existence of problem or instability in the system which may lead to failure. This is vital in many industrial settings especially in motor drive systems.

3.3. Adaptive Controller with dynamic fuzzy system ensemble for motor-driven Systems. The motor's performance is tested during the entire process of design, development, and manufacture. The window motor in use has a start-stop location and start-stop interval that follow a normal distribution law rather than being uniformly distributed with equal probability. This must adhere to the specifications for the motor test, accurately depict the motor's operating circumstances, and simulate passenger handling behaviors during the test. In order to increase test efficiency and achieve adaptability, the research incorporates random position control into the motor control system. A distributed motor control system with one primary control center managing three motors for testing is the need for the system test.

There are n-order subsystems in the multivariable PMSM servo system breakdown. For each stage in the design process, the corresponding state variables' control law and parameter adaptive law are solved to enable the step's state variables to reach the necessary asymptotic properties. Ensure that the goal angle z_c is tracked by the motor angle $\dot{z}_1$.

Step 1: In accordance with (3.3) first, insert two errors.

$$\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_2 = b_1 z_{31} + b_1 z_2 + e + c \\ \dot{z}_{31} = b_2 z_{31} + b_2 z_2 z_{32} + b_3 z_2 + b_3 v_p \\ \dot{z}_{32} = b_2 z_{32} + b_2 z_2 z_{31} + b_3 v_c \end{cases} \tag{3.3}$$

$$w_1 = z_1 - z_c \tag{3.4}$$

$$w_2 = z_2 - \alpha_1 \tag{3.5}$$

Step 2: Next, as per (3.3), introduce the third fault. Since there is no field weakening control with this control approach $\alpha_{21}(\text{the d} - \text{axis current}) = 0$.

$$w_{31} = z_{31} - \alpha_{21} \tag{3.6}$$

Step 3: The third and fourth differential equations in equation (3.3) are addressed at this point. The final phase involves simultaneously solving two control laws because this system has multiple inputs. After determining in Step 2 that α_{21} is a function of z_{31}, z_{32} and $\alpha_{31} = 0$, errors (3.7) and (3.8) from the previous step are built based on this information to determine the system's final input.

$$w_{31} = z_{31} - \alpha_{31} \tag{3.7}$$

$$w_{32} = z_{32} \tag{3.8}$$

The real PMSM servo steering system's unknown properties are taken into consideration in the control technique design. If certain factors can be precisely determined beforehand using sophisticated measurement techniques, create the FSE combination model for forecasting. Using each fuzzy logic classifier, this method sorts the dataset and generates a probability outcome. They are mixed probability outputs to produce an ensemble output. The formula for the FSE combination prediction model is provided by equation (3.9):

$$\hat{z}_s = \sum_{k=s}^{m} x_k \hat{z}_s^k \tag{3.9}$$

In equation (3.13), $\hat{z}_s^k$ denotes the predicted value of the kth predictive model for the given year t, x_k stands for the kth model's weight coefficient, $\hat{z}_s$ for the combined model for the same year s, and m is the number of cities. Equation (3.10) shows how the weights are calculated using the DMFSE approach, which uses mean square error:

$$x_k = \overline{\left[\left[\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right] * \left\{\sum_{k=1}^{m} \left|\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right|^{-1}\right\}\right]} \tag{3.10}$$

The discount factor is denoted by γ, while the data length utilized for weight calculation is represented by t. The FSE model can be obtained as illustrated in equation (3.11) by replacing equation (3.11) into Eq. (3.10).

$$\hat{z}_s = \sum_{k=s}^{m} x_k \hat{z}_s^k = \sum_{k=s}^{m} \overline{\left[\left[\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right] * \left\{\sum_{k=1}^{m} \left|\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right|^{-1}\right\}\right]} \tag{3.11}$$

We implement the FSE mixed-model predictive model that has been tuned with the controller method. The goal of this study is to convert the discount factor γ used in weight calculations into a matrix so $\gamma^{\,t-s}_{\,+1}$ that the difference between expected and actual values at various time periods can be eliminated. Equations (3.12) and (3.13) represent the weight equation and the improved predictive model, respectively.

$$x_k = \frac{1}{\left[\left[\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right] * \left\{\sum_{k=1}^{m} \left|\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right|^{-1}\right\}\right]} \tag{3.12}$$

$$\hat{z}_s = \sum_{k=s}^{m} x_k \hat{z}_s^k = \sum_{k=s}^{m} \frac{\mathbb{C}_s^k}{\left[\left[\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right] * \left\{\sum_{k=1}^{m} \left|\sum_{s=1}^{t} \gamma^{\,t-s}_{\,+1} (z_s - \hat{z}_s^s)^2\right|^{-1}\right\}\right]} \tag{3.13}$$

Table 4.1: Evaluation of Current and Proposed Methods' Performance

Methods	**Performance Metric of Proposed and Existing Methods**		
	MSE (%)	**Simulation Time (Sec)**	**Motor Frequency (%)**
IPSO-GA[19]	56	90	82
LFDNM[7]	45	77	90
AC-DFSE [Proposed]	37	65	97

Find the best matrix of discount factors. The selection of $\mathbb{C}$ should not be done randomly, since it could result in poor prediction performance. Hence, the optimization problem of the objective function is used to determine the discount factor matrix γ^{t-s+1}. Using a variety of forecast time points and models, the controller seeks out the best discount factors. Get the best potential predictions by plugging in the best possible discount factors into the model.

The effectiveness and dependability of the entire control system can be raised with this tactic. These findings show that the measuring and control system performs exceptionally well in terms of voltage stability. In the meantime, it is possible to retain the least amount of voltage variations under various working situations, guaranteeing the motor's excellent stability and dependability. As a result, the data collection system can record the electrical current and voltage level at the motor's two ends in real time. The accuracy test requirements are satisfied because the relative error is within 0.8.

The AC-DMFSE model is incorporated in a multi-variable Permanent Magnet Synchronous Motor (PMSM) servo system which is subdivided into several subsystems. Each subsystem is controlled with adaptive laws to accomplish the required performance by managing state variables. laws are responsible for attempting to alter parameters of the motor, including Zc (the goal angle). They work by controlling torque, speed, and voltage amongst many other variables in real time.

4. Experimental analysis. The numerical evaluation of AC-DFSE for distributed systems for simulation analysis of motor drive systems is presented in this part. This research proves that the offered strategy works wonders. In this study, we evaluate our suggested approach in comparison to current methods, including improved particle swarm optimization–genetic algorithm (IPSO-GA) [19], and Lévy flight distribution Nelder–Mead (LFDNM) algorithm [7]. The following indicators are essential for assessing how well distributed systems for simulation analysis of motor drive systems. This study examined a number of parameters, including motor frequency, MSE and stimulation time. Table 4.1 compares the performance of the existing approaches with the recommendations.

4.1. Motor Frequency. The term "motor frequency" describes the number of cycles per hour at which the electrical system of the motor runs. It is equivalent to the supply frequency of the electrical current that powers alternating current (AC) motors. Usually, the electrical supply's frequency and the motor's speed are proportionate.

Fig. 4.1 shows how effective the motor frequency is. With an efficacy of at least 97%, the recommended AC-DFSE is far less effective than cutting-edge methods like 82% IPSO-GA and 90% LFDNM. Since motor frequency directly impacts the motor's speed and performance, it is a crucial component in motor drive system simulations. The dynamic behavior of the motor may change if its frequency is changed, which could jeopardize the accuracy of the simulation analysis.

4.2. MSE (Mean Squared Error). A statistical metric called mean squared error is used to express how much the actual and anticipated values differ from one another. The average of the squared differences between these two sets of values is determined.

$$MSE = \frac{1}{q}\sum_{l=1}^{q} |a_m - \hat{a}_m| \tag{4.1}$$

Here,

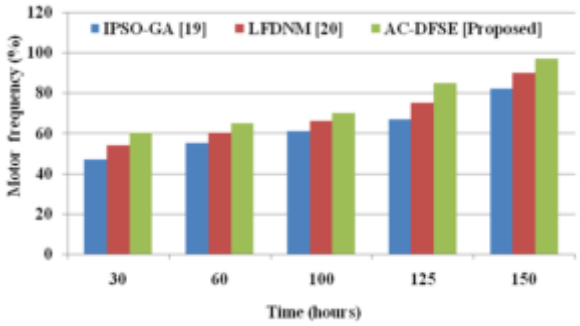


Fig. 4.1: Motor Frequency

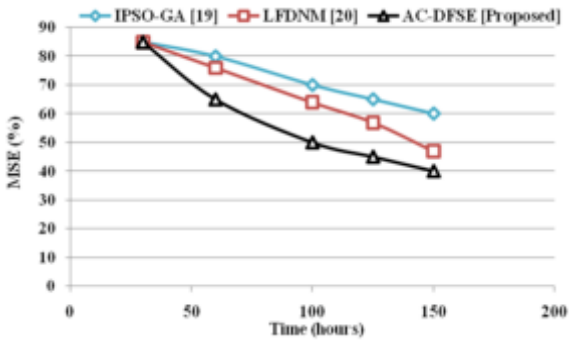


Fig. 4.2: MSE Analysis

a_m is the observed value,
$\hat{a}_m$is the predicted value,
$\overline{a}_m$ is the mean of the observed values,

The effectiveness of the current and recommended strategies is shown in Fig. 4.2. At least 40% efficacy is produced by the suggested AC-DFSE, a far cry from state-of-the-art methods like 60% IPSO-GA and 47% LFDNM. MSE is frequently used in motor drive simulations to assess an adaptive algorithm's effectiveness by gauging how effectively the algorithm can forecast the behavior of the system over time. The predictions are shown by lower mean square error (MSE), which is important for optimizing the algorithm's parameters.

4.3. Simulation Time. The term "simulation time" describes the amount of time that a motor or system is exposed to a specific input or excitation, like torque, voltage, or current. It is the period of time that the motor's reaction to this input is monitored and examined in the context of simulation. The motor's dynamic response analysis is impacted by the duration of stimulus. Engineers can predict transient behaviors and modify the control algorithm by observing how quickly the motor responds to changes in input by adjusting the stimulation time.

Figure 4.3 illustrates the effectiveness of the existing and recommended solutions. Compared to state-of-the-art methods like 81% IPSO-GA and 67% LFDNM, the suggested alternatives can achieve up to 55% maximum efficiency. IPSO-GA and LFDNM tests show that the suggested method outperforms the existing one in terms of Simulation time for the motor drive system.

5. Conclusion. This work proposed AC-DMFSE for distributed systems of simulating motor drive systems. The dataset was originally collected from a three-phase system motor vehicle with a momentum PMSM installed on a test bench. The training representations for standardized data processes use the EWMS. Adaptive

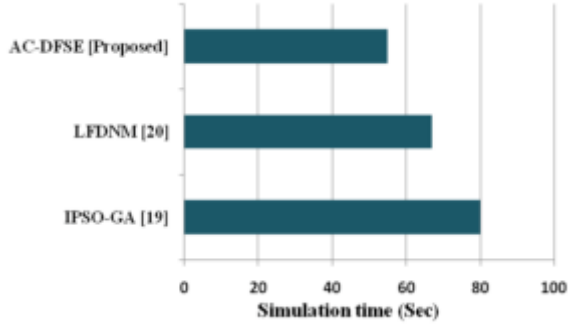


Fig. 4.3: Simulation Time

methods, like fuzzy logic-based optimization and model predictive control, are used to maximize motor performance in dynamic conditions. Simulations and experimental validation show that the robot performs better than existing methods in improving accuracy, simulation time and MSE. According to the simulation results, the suggested AC-DMFSE performed better than the others in terms of motor frequency (97%), simulation time (55sec), and MSE (40%). Challenges such as network latency, synchronization problems between distributed nodes, and the difficulty of implementing adaptive algorithms across heterogeneous hardware platforms are encountered by the distributed system for simulation study of motor drive systems. Furthermore, communication network performance may limit scalability, and sophisticated data handling methods may be needed to manage big information in real time. Future developments might concentrate on enhancing scalability through cloud computing, integrating machine learning for more adaptive algorithms, and optimizing network infrastructure for reduced latency. In future, models that would merge this technique with neural networks or other deep models would enhance their capacity to perform on tasks that have a non-linear trend and interdependencies.

Acknowledgement. Zhejiang Province Department of Education Research Funding Project (Y202147315).

REFERENCES

[1] B. Almadani, N. Alshammari, and A. Al-Roubaiey, *Adaptive cruise control based on real-time dds middleware*, IEEE Access, 11 (2023), pp. 75407–75423.
[2] W. Cai, X. Wu, M. Zhou, Y. Liang, and Y. Wang, *Review and development of electric motor systems and electric powertrains for new energy vehicles*, Automotive Innovation, 4 (2021), pp. 3–22.
[3] M. Davari, W. Gao, Z.-P. Jiang, and F. L. Lewis, *An optimal primary frequency control based on adaptive dynamic programming for islanded modernized microgrids*, IEEE Transactions on Automation Science and Engineering, 18 (2020), pp. 1109–1121.
[4] O. Gheibi, D. Weyns, and F. Quin, *Applying machine learning in self-adaptive systems: A systematic literature review*, ACM Transactions on Autonomous and Adaptive Systems (TAAS), 15 (2021), pp. 1–37.
[5] Z. Guo, Y. Zhang, X. Zhao, and X. Song, *Cps-based self-adaptive collaborative control for smart production-logistics systems*, IEEE transactions on cybernetics, 51 (2020), pp. 188–198.
[6] L. Hu, Q. Tian, C. Zou, J. Huang, Y. Ye, and X. Wu, *A study on energy distribution strategy of electric vehicle hybrid energy storage system considering driving style based on real urban driving data*, Renewable and Sustainable Energy Reviews, 162 (2022), p. 112416.
[7] D. Izci, *Design and application of an optimally tuned pid controller for dc motor speed regulation via a novel hybrid lévy flight distribution and nelder–mead algorithm*, Transactions of the Institute of Measurement and Control, 43 (2021), pp. 3195–3211.
[8] W. Kirchgässner, O. Wallscheid, and J. Böcker, *Estimating electric motor temperatures with deep residual machine learning*, IEEE Transactions on Power Electronics, 36 (2020), pp. 7480–7488.
[9] J. Li and T. Yu, *A new adaptive controller based on distributed deep reinforcement learning for pemfc air supply system*, Energy Reports, 7 (2021), pp. 1267–1279.
[10] P. Li, X. Jiao, and Y. Li, *Adaptive real-time energy management control strategy based on fuzzy inference system for plug-in hybrid electric vehicles*, Control Engineering Practice, 107 (2021), p. 104703.
[11] S. Mahfoud, A. Derouich, N. El Ouanjli, M. El Mahfoud, and M. Taoussi, *A new strategy-based pid controller optimized by genetic algorithm for dtc of the doubly fed induction motor*, Systems, 9 (2021), p. 37.

[12] H. Malik, A. Iqbal, and A. K. Yadav, *Soft computing in condition monitoring and diagnostics of electrical and mechanical systems*, vol. 1096, Springer, 2020.
[13] D. Mohanraj, J. Gopalakrishnan, B. Chokkalingam, and L. Mihet-Popa, *Critical aspects of electric motor drive controllers and mitigation of torque ripple*, IEEE Access, 10 (2022), pp. 73635–73674.
[14] Z. Nie and H. Farzaneh, *Adaptive cruise control for eco-driving based on model predictive control algorithm*, Applied Sciences, 10 (2020), p. 5271.
[15] Y. Sun, C. Gao, Y.-H. Yang, and L.-B. Wu, *Multiple filters-based output feedback fault-tolerant control for a class of uncertain nonlinear systems with event-triggered mechanism*, International Journal of Control, Automation and Systems, 22 (2024), pp. 151–162.
[16] K. Vanchinathan and N. Selvaganesan, *Adaptive fractional order pid controller tuning for brushless dc motor using artificial bee colony algorithm*, Results in Control and Optimization, 4 (2021), p. 100032.
[17] M. Wang and Q. Zhang, *Optimized data storage algorithm of iot based on cloud computing in distributed system*, Computer Communications, 157 (2020), pp. 124–131.
[18] W. Xing, H. Wang, L. Lu, X. Han, K. Sun, and M. Ouyang, *An adaptive virtual inertia control strategy for distributed battery energy storage system in microgrids*, Energy, 233 (2021), p. 121155.
[19] Z. Yang, C. Lu, X. Sun, J. Ji, and Q. Ding, *Study on active disturbance rejection control of a bearingless induction motor based on an improved particle swarm optimization–genetic algorithm*, IEEE Transactions on Transportation Electrification, 7 (2020), pp. 694–705.
[20] Z. Zhao, H. Taghavifar, H. Du, Y. Qin, M. Dong, and L. Gu, *In-wheel motor vibration control for distributed-driven electric vehicles: A review*, IEEE Transactions on Transportation Electrification, 7 (2021), pp. 2864–2880.

Edited by: Rajkumar Rajavel
Special issue on: Cognitive Computing for Distributed Data Processing and Decision-Making in Large-Scale Environments
Received: Sep 27, 2024
Accepted: Nov 26, 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.