

Scalable Computing: Practice and Experience

Scientific International Journal
for Parallel and Distributed Computing

ISSN: 1895-1767



Volume 21(2)

June 2020

EDITOR-IN-CHIEF

Dana Petcu

Computer Science Department
West University of Timisoara
and Institute e-Austria Timisoara
B-dul Vasile Parvan 4, 300223
Timisoara, Romania
Dana.Petcu@e-uvt.ro

MANAGING AND
TECHNICAL EDITOR

Silviu Panica

Computer Science Department
West University of Timisoara
and Institute e-Austria Timisoara
B-dul Vasile Parvan 4, 300223
Timisoara, Romania
Silviu.Panica@e-uvt.ro

BOOK REVIEW EDITOR

Shahram Rahimi

Department of Computer Science
Southern Illinois University
Mailcode 4511, Carbondale
Illinois 62901-4511
rahimi@cs.siu.edu

SOFTWARE REVIEW EDITOR

Hong Shen

School of Computer Science
The University of Adelaide
Adelaide, SA 5005
Australia
hong@cs.adelaide.edu.au

Domenico Talia

DEIS
University of Calabria
Via P. Bucci 41c
87036 Rende, Italy
talia@deis.unical.it

EDITORIAL BOARD

Peter Arbenz, Swiss Federal Institute of Technology, Zürich,
arbenz@inf.ethz.ch

Dorothy Bollman, University of Puerto Rico,
bollman@cs.uprm.edu

Luigi Brugnano, Università di Firenze,
brugnano@math.unifi.it

Giacomo Cabri, University of Modena and Reggio Emilia,
giacomo.cabri@unimore.it

Bogdan Czejdo, Fayetteville State University,
bczejdo@uncfsu.edu

Frederic Desprez, LIP ENS Lyon, frederic.desprez@inria.fr

Yakov Fet, Novosibirsk Computing Center, fet@ssd.sccc.ru

Giancarlo Fortino, University of Calabria,
g.fortino@unical.it

Andrzej Goscinski, Deakin University, ang@deakin.edu.au

Frederic Loulergue, Northern Arizona University,
Frederic.Loulergue@nau.edu

Thomas Ludwig, German Climate Computing Center and Uni-
versity of Hamburg, t.ludwig@computer.org

Svetozar Margenov, Institute for Parallel Processing and Bul-
garian Academy of Science, margenov@parallel.bas.bg

Viorel Negru, West University of Timisoara,
Viorel.Negru@e-uvt.ro

Moussa Ouedraogo, CRP Henri Tudor Luxembourg,
moussa.ouedraogo@tudor.lu

Marcin Paprzycki, Systems Research Institute of the Polish
Academy of Sciences, marcin.paprzycki@ibspan.waw.pl

Roman Trobec, Jozef Stefan Institute, roman.trobec@ijs.si

Marian Vajtersic, University of Salzburg,
marian@cosy.sbg.ac.at

Lonnie R. Welch, Ohio University, welch@ohio.edu

Janusz Zalewski, Florida Gulf Coast University,
zalewski@fgcu.edu

SUBSCRIPTION INFORMATION: please visit <http://www.scpe.org>

Scalable Computing: Practice and Experience

Volume 21, Number 2, June 2020

TABLE OF CONTENTS

SPECIAL ISSUE ON INTELLIGENCE ON SCALABLE COMPUTING FOR RECENT APPLICATIONS:

Introduction to the Special Issue	157
<i>P. Vijaya, D Binu</i>	
CPU-Memory Aware VM Consolidation for Cloud Data Centers	159
<i>B. Nithiya, R. Eswari</i>	
Bird Swarm Optimization-based Stacked Autoencoder Deep Learning for Umpire Detection and Classification	173
<i>Suvarna Nandyal, Suvarna Laxmikant Kattimani</i>	
Enhanced DBSCAN with Hierarchical tree for Web Rule Mining	189
<i>Neelima Gullipalli, Sireesha Rodda</i>	
A Comprehensive Survey of the Routing Schemes for IoT Applications	203
<i>Dipali K. Shende, Yogesh Angal, S. S. Sonavane</i>	
Chicken-Moth Search Optimization-based Deep Convolutional Neural Network for Image Steganography	217
<i>V.K. Reshma, R.S. Vinod Kumar, D. Shahi, M.B. Shyjith</i>	
An Efficient Dynamic Slot Scheduling Algorithm for WSN MAC: A Distributed Approach	233
<i>Manas Ranjan Lenka, Amulya Ratna Swaini</i>	
Artefacts Removal from ECG Signal: Dragonfly Optimization-based Learning Algorithm for Neural Network-enhanced Adaptive Filtering	247
<i>Talabattula Viswanadham, P Rajesh Kumar</i>	
A Comprehensive Review on State-of-the-Art Image Inpainting Techniques	265
<i>Balasaheb H. Patil, P.M. Patil</i>	
An Efficient Way of Finding Polarity of Roman Urdu Reviews by using Boolean Rules	277
<i>Halima Sadia, Mohib Ullah, Tariq Hussain, Nida Gul, Muhammad Farooq Hussain, Nauman ul Haq, Abu Bakar</i>	
Forecasting the Impact of Social Media Advertising among College Students using Higher Order Statistical Functions	291
<i>Meena Zenith N, Radhika R</i>	

REGULAR PAPERS:

- Novel Metric for Load balance and Congestion Reducing in Network on-Chip** **309**
Abdelkader Aroui, Abou Elhassan Benyamina, Pierre Boulet, Kamel Benhaoua, Amit Kumar Singh
- A Distributed Neural Network Training Method Based on Hybrid Gradient Computing** **323**
Zhen Lu, Meng Lu, Yan Liang
- An Analytical Model of a Corporate Software-Controlled Network Switch** **337**
Valery P. Mochalov, Gennady I. Linets, Natalya Yu. Bratchenko, Svetlana V. Govorova



INTRODUCTION TO THE SPECIAL ISSUE ON INTELLIGENCE ON SCALABLE COMPUTING FOR RECENT APPLICATIONS

P. VIJAYA* AND D BINU†

The special issue has been focussed to overcome the challenges of scalability, which includes size scalability, geographical scalability, administrative scalability, network and synchronous communication limitation, etc. The challenges also emerge with the development of recent applications. Hence this proposal has been planned to handle the scalability issues in recent applications. This special issue invites researchers, engineers, educators, managers, programmers, and users of computers who have particular interests in parallel processing and/or distributed computing and artificial intelligence to submit original research papers and timely review articles on the theory, design, evaluation, and use of artificial intelligence and parallel and/or distributed computing systems for emerging applications. The ten papers in this special issue cover a range of aspects of theoretical and practical research development on scalable computing. The proposal provides an effective forum for communication among researchers and practitioners from various scientific areas working in a wide variety of problem areas, sharing a fundamental common interest in improving the ability of parallel and distributed computer systems, intelligent techniques, and deep learning mechanisms and advanced soft computing techniques. The issue covers wide range of applications, but with scalable problems that to be solved by perfect hybridization of distributed computing and artificial intelligence.

The first paper is “CPU-Memory Aware VM Consolidation for Cloud Data Centers” introduced a CPU-Memory aware VM placement algorithm is proposed for selecting suitable destination host for migration. The Virtual Machines are selected using Fuzzy Soft Set (FSS) method VM selection algorithm. The proposed placement algorithm considers CPU, Memory, and combination of CPU-Memory utilization of VMs on the source host.

In “Bird Swarm Optimization-based stacked autoencoder deep learning for umpire detection and classification”, presented the umpire detection and classification by proposing an optimization algorithm. The overall procedure of the proposed approach involves three steps, like segmentation, feature extraction, and classification. Here, the classification is done using the proposed Bird Swarm Optimization-based stacked autoencoder deep learning classifier (BSO-Stacked Autoencoders), that categories into umpire or others.

In “Enhanced DBSCAN with Hierarchical tree for Web Rule Mining”, proposed an enhanced web mining model based on two contributions. At first, the hierarchical tree is framed, which produces different categories of the searching queries (different web pages). Next, to hierarchical tree model, enhanced Density-Based Spatial Clustering of Applications with Noise (DBSCAN) technique model is developed by modifying the traditional DBSCAN. This technique results in proper session identification from raw data. Moreover, this technique offers the optimal level of clusters necessitated for hierarchical clustering. After hierarchical clustering, the rule mining is adopted. The traditional rule mining technique is generally based on the frequency; however, this paper intends to enhance the traditional rule mining based on utility factor as the second contribution. Hence the proposed model for web rule mining is termed as Enhanced DBSCAN-based Hierarchical Tree (EDBHT).

In “A comprehensive survey of the Routing Schemes for IoT applications”, this review article provides a detailed review of 52 research papers presenting the suggested routing protocols based on the content-based, clustering-based, fuzzy-based, Routing Protocol for Low power (RPL) and Lossy Networks, tree-based and so on. Also, a detailed analysis and discussion are made by concerning the parameters, simulation tool, and year

*Waljat College of Applied Sciences, Rusayl, Muscat, Sultanate of Oman (pvvijaya27@gmail.com)

†Resbee Info Technologies, Tamilnadu, India (ultimatebinu@gmail.com)

of publication, network size, evaluation metrics, and utilized protocols. In “Chicken-Moth Search Optimization-Based Deep Convolutional Neural Network For Image Steganography”, proposed an effective pixel prediction based on image steganography is developed, which employs error dependent Deep Convolutional Neural Network (DCNN) classifier for pixel identification. Here, the best pixels are identified from the medical image based on DCNN classifier using pixel features, like texture, wavelet energy, Gabor, scattering features, and so on. The DCNN is optimally trained using Chicken-Moth search optimization (CMSO). The CMSO is designed by integrating Chicken Swarm Optimization (CSO) and Moth Search Optimization (MSO) algorithm based on limited error.

In “An Efficient Dynamic Slot Scheduling Algorithm for WSN MAC: A Distributed Approach”, an effective TDMA based slot scheduling algorithm needs to be designed. In this paper, we propose a TDMA based algorithm named DYSS that meets both the timeliness and energy efficiency in handling the collision. This algorithm finds an effective way of preparing the initial schedule by using the average two-hop neighbors count. Finally, the remaining un-allotted nodes are dynamically assigned to slots using a novel approach.

In “Artefacts removal from ECG Signal: Dragonfly optimization-based learning algorithm for neural network-enhanced adaptive filtering”, proposed a method utilizes the adaptive filter termed as the (Dragonfly optimization + Levenberg Marquardt learning algorithm) DLM-based Nonlinear Autoregressive with exogenous input (NARX) neural network for the removal of the artefacts from the ECG signals. Once the artefact signal is identified using the adaptive filter, the identified signal is subtracted from the primary signal that is composed of the ECG signal and the artefacts through an adaptive subtraction procedure.

In “A Comprehensive Review on State-of-the-Art Image Inpainting Techniques”, this survey makes a critical analysis of diverse techniques regarding various image inpainting schemes. This paper goes under (i) Analyzing various image inpainting techniques that are contributed in different papers. (ii) Makes the comprehensive study regarding the performance measures and the corresponding maximum achievements in each contribution. (iii) Analytical review concerning the chronological review and various tools exploited in each of the reviewed works.

In “An Efficient Way of Finding Polarity of Roman Urdu Reviews by Using Boolean Rules”, proposed a novel approach by using Boolean rules for the identification of the related and non-related comments. Related reviews are those which show the behavior of a customer about a particular product. Lexicons are built for the identification of noise, positive and negative reviews.

The final paper is “Forecasting the Impact of Social Media Advertising among College Students using Higher Order Statistical Functions”, this research work plans to develop a statistical review that concerns on social media advertising among college students from diverse universities. The review analysis on social media advertising is given under six sections such as: (i) Personal Profile; (ii) Usage; (iii) Assessment; (iv) Higher Order statistics like Community, Connectedness, Openness, Dependence, and Participation; (v) Trustworthiness such as Trust, Perceived value and Perceived risk; and (vi) Towards advertisement which involves attitude towards advertisement, response towards advertisement and purchase intention.



CPU-MEMORY AWARE VM CONSOLIDATION FOR CLOUD DATA CENTERS

B. NITHIYA *AND R. ESWARI †

Abstract. The unbalanced usage of resources in cloud data centers cause an enormous amount of power consumption. The Virtual Machine (VM) consolidation shuts the underutilized hosts and makes the overloaded hosts as normally loaded hosts by selecting appropriate VMs from the hosts and migrates them to other hosts in such a way to reduce the energy consumption and to improve physical resource utilization. Efficient method is needed for VM selection and destination hosts selection (VM placement). In this paper, a CPU-Memory aware VM placement algorithm is proposed for selecting suitable destination host for migration. The VMs are selected using Fuzzy Soft Set (FSS) method VM selection algorithm. The proposed placement algorithm considers both CPU, Memory, and combination of CPU-Memory utilization of VMs on the source host. The proposed method is experimentally compared with several existing selection and placement algorithms and the results show that the proposed consolidation method performs better than existing algorithms in terms of energy efficiency, energy consumption, SLA violation rate, and number of VM migrations.

Key words: cloud computing, VM consolidation, VM Placement, Energy Consumption, Energy Efficiency, SLA Violation Rate.

AMS subject classifications. 68M14

1. Introduction. Due to the increase of storage demands the cloud service providers launched the cloud data centers for satisfying the user needs. Various users' VM requests are complex and required taking into account of multiple resource constraints. So, the physical servers that satisfy the users request will be considered to deploy the selected VMs [1]. Commercial IaaS cloud companies, including Amazon EC2 [2], IBM [3] and Google Compute Engine [4] are offering various types of VM instances with varying types and resource volumes. As a significance, when cloud data center owners are unable to deploy different types of VM requests efficiently, some assets may get overloaded while others remain underutilized. These unbalanced resource usages can eventually lead to unnecessary physical server activation. Thus, a significant concern is needed to balance the load in terms of CPU, memory, storage, and network bandwidth while meeting all requests of VM.

Consolidation is the process of moving running VM from one physical server to another without down time and switch off idle servers to power save mode. There are two types of consolidation: Static where VMs size is fixed; Dynamic where periodical demands in the each VM. Dynamic consolidation is performed in two steps one is migrate VMs from underutilized host and put the host in sleep mode. Another step is to migrate VMs from overloaded host without degrading the performance such as energy consumption, SLA violation rate [5].

For dynamic VM consolidation, there are several VM selection and VM placement algorithms have been proposed for selecting VMs from overloaded host and placing them on to the appropriate destination host. In this paper, dynamic consolidation is considered. In this paper a fuzzy soft set based VM selection algorithm [6] is used for selecting VMs from host which considers all factors such as RAM, CPU, memory, and correlation values. A CPU-Memory aware placement algorithm is proposed which considers both CPU and memory factors for selecting appropriate hosts for deploying selected VMs for migration. The abbreviations are used in this paper are listed in Table 1.1.

The remainder of the paper is structured as follows: Section 2 presents the related works. Section 3 describes the VM consolidation and the proposed CPU-Memory aware VM placement algorithm. In Section 4,

*Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamilnadu, India (nithiyab1988@gmail.com).

†Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamilnadu, India (eswari@nitt.edu).

TABLE 1.1
Abbreviation

Abbreviation	Explanation
VM	Virtual Machine
FSS	Fuzzy Soft Set
MEM	Memory Utilization
RCM	Ratio of CPU utilization to Memory utilization
PCM	Product of CPU utilization to Memory utilization
MCC	Minimum Correlation Coefficient
IQR	Inter-Quartile Range
LR	Local Regression
LRR	Local Robust Regression
MAD	Median Absolute Deviation
THR	Threshold
RS	Random Solution
MU	Minimum Utilization
MC	Maximum Correlation
MP	Meet Performance
AFT-FS	Adaptive Four Threshold based Fuzzy VM Selection
PABFD	Power-Aware Best-Fit Decreasing
EC	Energy Consumption
SLA violation rate	Service Level Agreement violation rate
SVTAH	SLA Violation Time per Active Host
PDCVM	Performance Degradation Caused by VM Migration
ESV	Product of Energy and SLA Violation
EE	Energy Efficiency

the experimental setup and evaluation metrics are discussed. In Section 5, the experimental results of existing and proposed methods are discussed. Finally, Section 6 concludes the work with future extension.

2. Related Work. The dynamic VM consolidation is done by 3 levels: Host classification, VM Selection, and VM Placement. Several Selection, and Placement algorithms have been proposed by researchers.

2.1. Host Classification. For host allocation the following algorithms are used by researchers [5]: Inter Quartile Range (IQR), Local Regression (LR), Local Robust Regression (LRR), Median Absolute Deviation (MAD), and Static Threshold (THR). The authors [5] found that THR is the best VM allocation algorithm than others.

An adaptive four threshold method [7] is used to classify the hosts. The thresholds are determined using K-means clustering midrange inter quartile range algorithm [8].

2.2. VM Selection. The Minimum Migration Time (MMT) algorithm [5] migrates a VM that needs minimum migration time to complete the migration from overloaded host to less loaded host. The migration time could be calculated as the ratio of amount of RAM utilized by VM to the network bandwidth available for the host.

Higher the usage of resources in the server by the applications, greater the chance of the server getting overloaded. So, the authors in [5] proposed maximum correlation algorithm to select the VMs that have higher correlation of the CPU utilization compared with other VMs for migration. The multiple correlation coefficient is applied to evaluate the correlation between CPU utilization of VMs.

In Minimum Utilization (MU) algorithm [5], the VM that uses high CPU will not be considered for migration since its migration increases downtime (the period during which the service is unavailable due to there being no currently executing instance of that VM). So, the VMs that have minimum CPU utilization is selected for migration. The Random Selection (RS) algorithm [5] selects VMs randomly without any rules.

Meet Performance (MP) selection algorithm was proposed by [10] virtual which may vary from the other

selection algorithms. Their algorithm compares host's utilization deviation with upper threshold and with CPU utilization of VMs in the host. The selection of VMs is based on the comparison results. The VM that has the lowest resource satisfaction will get higher priority to be migrated.

An adaptive fuzzy based VM selection algorithm (AFT_FS) was proposed by us [7] which uses four threshold values to detect overloaded hosts and a fuzzy-based approach to select VMs for migration.

All of the above mentioned selection algorithms have considered any one of the selection factors such as either RAM or CPU or Memory or Correlation values for selecting VMs during migration.

A Fuzzy Soft Set (FSS) based VM Selection algorithm was proposed by us [9] to achieve the optimal selection of VMs for migration. The algorithm considers all four factors at a time, and accurately finds which VM has to migrate from the overloaded hosts in the cloud data center.

2.3. VM Placement. Virtual Machine Placement is a crucial issue in cloud computing. It is a method where most appropriate physical machine (PM) will be selected to place VMs. The VM placement also plays important role during dynamic VM consolidation. Most of the researchers concentrate on initial VM placement. Many heuristic and meta-heuristic algorithms have been proposed for initial VM placement such as Ant Colony System (ACS) embedded with First Fit Decreasing (FFD) [11], Grey Wolf Optimization (GWO) [12], Genetic Algorithm [13] etc. But our research work mainly focuses on the VM placement during VM consolidation. Since the meta-heuristic algorithms consume more time to take the decision for PM-VM pair during consolidation, our research work considers only heuristic algorithms to select appropriate destination host.

PABFD algorithm [5] is most commonly used to place VMs onto the destination hosts. It sorts all the VMs in the decreasing order of their CPU utilization values and allocates each VM to a host that provides the minimum raise of energy consumption due to this allocation.

Most of the VM consolidation methods use this algorithm to select the hosts to which the VMs can be placed. This placement algorithm considers only the CPU utilization of hosts during host selection.

The Minimum Correlation Coefficient (MCC) algorithm was proposed by [10]. It finds the correlation coefficient between chosen VM and target host based on the utilization of CPU alone. The authors place the chosen VM to the host that has the minimum correlation coefficient with it.

This research work tries to investigate the impact of memory utilization, CPU utilization, and the combination of both for selecting the target host. The proposed algorithm finds the minimum correlation coefficient between chosen VM and target host based on memory, CPU, Ratio of memory to CPU, and Product of memory and CPU.

3. Proposed Work. In this paper, the CPU-Memory aware VM placement algorithm is proposed for cloud data centers. The algorithm considers three different utilization matrices based on CPU and Memory utilization of VMs on targeted hosts during p time slices. The association (correlation) between VM and host will be separately calculated for each resource (utilization matrix). The host that gives minimum correlation will be selected for placement of VM. The nomenclature used in this paper are listed in Table 1. and the overall flow chart of VM Consolidation is shown in Fig 3.1.

3.1. VM Consolidation. VM Consolidation is used to maintain the balance between energy and QoS. An efficient VM consolidation method should minimize energy consumption, SLA violation rate, and maximize energy efficiency. It should also have efficient VM migration, and minimum number of active hosts at a given time. Fig 3.2 shown that the host classification of datacenter.

It considers the following steps as given below.

- All hosts in the data centers are clustered into 5 groups using K-Means Inter Quartile Range clustering algorithm [7]: overloaded hosts, normally loaded hosts, little loaded hosts, less loaded hosts, and idle hosts.
- Migrate VMs if any on idle hosts to less loaded hosts and move hosts to power save mode.
- Migrate all VMs from little loaded hosts to normally loaded host. Then move all little loaded hosts to power save mode.
- Select VMs from the overloaded hosts using FSS algorithm and migrate them to less loaded hosts.

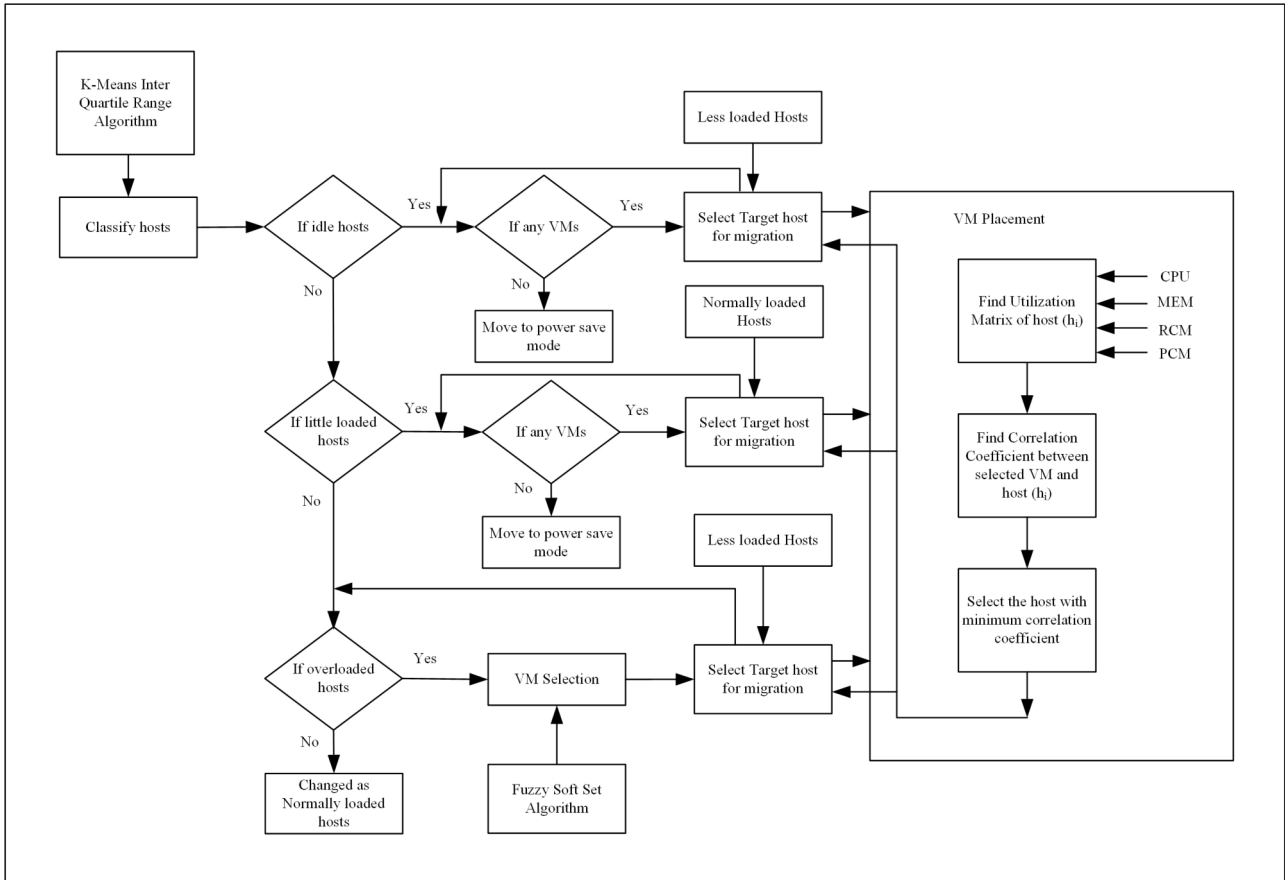


FIG. 3.1. Flow Chart of VM Consolidation

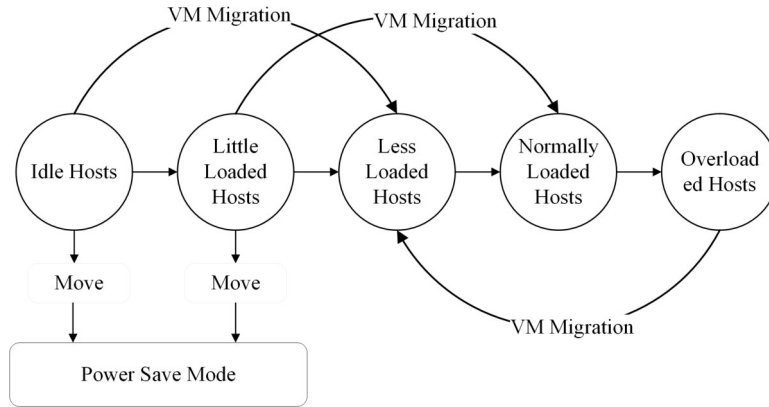


FIG. 3.2. Host Classification

3.2. CPU-Memory aware VM Placement Algorithm. Once a VM is selected using FSS for migration, a VM placement algorithm will be used to select an appropriate host for migration. The placement algorithm must minimize power consumption, and SLA violation rate.

3.2.1. VM Placement. The proposed algorithm consolidates VMs using three different ways of placement algorithm. The first one is based on memory utilization of selected VMs and second and third are based on

Algorithm 1 VM_Consolidation**Require:** Hosts in datacenter, VM_list**Ensure:** Select the Target_host

```

1: Cluster all hosts into five: Overloaded_hosts, Normallyloaded_hosts, Lessloaded_hosts, Littleloaded_hosts, Idle_hosts based on threshold values
2: Placement_Policy ← Get Placement_Policy
3: if (Idle_hosts) then
4:   for each host  $h_i$  in Idle_hosts do
5:     for each VM  $v$  in host  $h_i$  do
6:       Target_host ← Select_host ( $v$ , Lessloaded_hosts, Placement_Policy)
7:       Migrate  $v$  to Target_host
8:     Move host  $h_i$  to power save mode
9: else if (Littleloaded_hosts) then
10:  for each host  $h_i$  in Littleloaded_hosts do
11:   for each VM  $v$  in host  $h_i$  do
12:     Target_host ← Select_host ( $v$ , Normallyloaded_hosts, Placement_Policy)
13:     Migrate  $v$  to Target_host
14:   Move host  $h_i$  to power save mode
15: else(Overloaded_hosts)
16:  for each host  $h_i$  in Overloaded_hosts do
17:   repeat
18:     Select VM_Migrate list in  $h_i$  using fuzzy soft set
19:     for each VM  $v$  in VM_Migrate list do
20:       Target_host ← Select_host ( $v$ , Lessloaded_hosts, Placement_Policy)
21:       Migrate  $v$  to Target_host
22:   until host  $h_i$  becomes Normallyloaded

```

CPU and memory utilization of selected VMs for every p time slice.

VM Consolidation is given in algorithm 1. Step 1 clusters the hosts into five: Overloaded_hosts, Normallyloaded_hosts, Lessloaded_hosts, Littleloaded_hosts, and Idle_hosts based on their threshold values. Step 2 gets the Placement_Policy. Steps 3 to 8: Select target hosts from Lessloaded_hosts to place all VMs from idle host. This host selection will be repeated for all idle hosts. Now idle hosts are moved to power save mode. Steps 9 to 14: Select target hosts from Normallyloaded_hosts to place all VMs from Littleloaded_hosts. This host selection will be repeated for all little loaded hosts. Now Littleloaded_hosts are moved to power save mode. Steps 15 to 22: Select target host from Lessloaded_hosts for every VM in the overloaded hosts for migration. This host selection will be repeated for all overloaded hosts.

3.2.2. Target Host Selection. Algorithm 2 selects the host to which the chosen VM to be migrated. It uses the Minimum Correlation Coefficient (MCC) [10] to represent the association between chosen VM and target host.

The target host selection algorithm is given in algorithm 2. The algorithm receives the type of hosts from algorithm1. Steps 1 to 3: find the pool of hosts which satisfy the VM v 's demand and the total CPU usage of host and VM less than threshold values. Steps 4 to 10: find utilization matrix and squares of correlation coefficient between chosen VM v and all hosts in host pool list. Step 11 selects the target host that has the minimum squared correlation coefficient.

3.2.3. Finding Utilization Matrix. The VM placement is based on three different utilization matrices.

1. Memory Utilization (MEM) Memory utilization method consider only memory resource. For a host h_i with m VMs, the memory utilization of m VMs are collected during p time slices. Now the utilization

Algorithm 2 Select_host (v , Loaded_hosts, Placement_Policy)**Ensure:** Select the Target_host

- 1: **for** each host h_i in Loaded_hosts **do**
- 2: **if** ($(h_i$ satisfies v 's demand) && ($(h_i$'s usage v 's demand) < thr)) **then**
- 3: Host_pool_list $\leftarrow h_i$
- 4: **for** each host h_i in Host_pool_list **do**
- 5: Umatrix [m] [p] \leftarrow Utilization_Matrix (Placement_Policy, h_i)
- 6: **for** each time slice k in p **do**
- 7: Find the resource utilization of h_i

$$sum_util_i[k] = \sum_{j=1}^m Umatrix_util_i[j][k] \quad (3.1)$$

- 8: Calculate correlation coefficient between v and h_i

$$\rho_i = \frac{EC[(res_VM' \frac{1}{p} \sum_{(k=1)}^p res_VM_util_k)(res_Host' \frac{1}{p} \sum_{(k=1)}^p sum_util_i[k])]}{\sqrt{(Var(res_VM))} \sqrt{(Var(res_Host))}} \quad (3.2)$$

where resource_VM and resource_Host refer to the current resource utilization of the chosen j^{th} VM and the host h_i . $\frac{1}{p} \sum_{(k=1)}^p res_VM_util_k$ and $\frac{1}{p} \sum_{(k=1)}^p sum_util_i[k]$ refer to the total resource utilization of the chosen j^{th} VM and the host h_i during p time slices. The variance of the resource utilization of the chosen j^{th} VM and host h_i are calculated as

$$Var(res_VM) = E[(res_VM - \frac{1}{p} \sum_{(k=1)}^p res_VM_util_k)^2] \quad (3.3)$$

$$Var(res_Host) = E[(res_Host - \frac{1}{p} \sum_{(k=1)}^p sum_util_i[k])^2] \quad (3.4)$$

- 9: Compute squares of the correlation coefficient

$$\rho = \rho_1^2, \rho_2^2, \dots, \rho_n^2 \quad (3.5)$$

- 10: Target_host \leftarrow host that has the minimum ρ
- 11: return Target_host

Algorithm 3 Utilization_Matrix (Placement_Policy, h_i)**Ensure:** Utilization_Matrix

- 1: **if** (Placement_policy = 'Mem') **then**
- 2: Compute memory utilization matrix using Eqn. 3.6.
- 3: **else if** (Placement_policy = 'RCM') **then**
- 4: Compute RCM utilization matrix using Eqn. 3.7.
- 5: Find RCM values using Eqn. 3.8.
- 6: **else** (Placement_policy = 'PCM')
- 7: Compute PCM utilization matrix using Eqn. 3.9.
- 8: Find PCM values using Eqn. 3.10.

matrix of h_i is calculated as the memory utilization of m VMs on h_i at each time slice.

$$Mem_util_i[m][p] = \begin{bmatrix} Mem_{11} & Mem_{12} & Mem_{13} & \cdots & Mem_{1p} \\ Mem_{21} & Mem_{22} & Mem_{23} & \cdots & Mem_{2p} \\ \vdots & \vdots & Mem_{jk} & \ddots & \vdots \\ Mem_{m1} & Mem_{m2} & Mem_{m3} & \cdots & Mem_{mp} \end{bmatrix} \quad (3.6)$$

where $[RCM]_{jk}$ refers to the ratio of CPU utilization to memory utilization of VM j on host h_i during time slice k

2. Ratio of CPU utilization to Memory utilization Algorithm (RCM) RCM method considers both CPU resource and memory resource. Consider a host h_i and assume that it has m VMs. The CPU utilization and memory utilization of m VMs are collected during p time slices. Now the utilization matrix of h_i is calculated as the ratio of CPU utilization to memory utilization of m VMs on h_i at each time slice.

$$RCM_util_i[m][p] = \begin{bmatrix} RCM_{11} & RCM_{12} & RCM_{13} & \cdots & RCM_{1p} \\ RCM_{21} & RCM_{22} & RCM_{23} & \cdots & RCM_{2p} \\ \vdots & \vdots & RCM_{jk} & \ddots & \vdots \\ RCM_{m1} & RCM_{m2} & RCM_{m3} & \cdots & RCM_{mp} \end{bmatrix} \quad (3.7)$$

$$RCM_{jk} = \frac{CPU_util_{jk}}{Mem_util_{jk}} \quad (3.8)$$

where $[RCM]_{jk}$ refers to the ratio of CPU utilization to memory utilization of VM j on host h_i during time slice k

3. Product of CPU utilization to Memory utilization Algorithm (PCM) PCM method considers both CPU resource and memory resource. For host h_i with m VMs. The CPU utilization and memory utilization of m VMs are collected during p time slices. Now the utilization matrix of h_i is calculated as the product of CPU utilization to memory utilization of m VMs on h_i at each time slice.

$$PCM_util_i[m][p] = \begin{bmatrix} PCM_{11} & PCM_{12} & PCM_{13} & \cdots & PCM_{1p} \\ PCM_{21} & PCM_{22} & PCM_{23} & \cdots & PCM_{2p} \\ \vdots & \vdots & PCM_{jk} & \ddots & \vdots \\ PCM_{m1} & PCM_{m2} & PCM_{m3} & \cdots & PCM_{mp} \end{bmatrix} \quad (3.9)$$

$$PCM_{jk} = CPU_util_{jk} \times Mem_util_{jk} \quad (3.10)$$

4. Experimental Setup. There are many difficulties that faces during testing and experimentation of Cloud Computing like demand for energy-efficient for IT technologies, demand time saving, and controlling the evaluation of algorithms, applications, and policies before real cloud products. One of the suitable approaches to make all these difficulties as easy is the simulations tools. The objective of this simulation tool is to offer an extensible framework that enables simulation, modeling, experimentation of Cloud computing infrastructures and application services. For this reason, simulation has been chosen to evaluate the performance of the algorithms.

4.1. Cloudsim Toolkit. The implementation was done using Cloudsim Toolkit [14]. The toolkit has been developed by the Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne. CloudSim is completely written in Java. Netbeans or Eclipse IDE is used to run Cloudsim Toolkit. The simulation platform supports for modeling and simulation of large-scale Cloud computing data centers, virtualized server hosts, with customizable policies for provisioning host resources to virtual machines, energy-aware computational resources. And support for user-defined policies for allocation of hosts to virtual machines and policies for allocation of host resources to virtual machines. The data center is set up with 800 heterogeneous

hosts (physical nodes), half of which consists of HP Proliant ML 110 G4 and another half of which consists of HP Proliant ML 110 G5. The virtual machine's characteristics are corresponding to Amazon EC2. There are four types of virtual machines are considered such as High-CPU Medium instance, Extra-Large instance, Small instance, and Micro instance. The real-time workload data traces are used for this experiment as a part of the CoMon project Monitoring infrastructure for PlanetLab. The workload data used in the CPU utilization were taken from more than 500 places around the world [15]. In this experiment, 1516 VMs are chosen from '22/March/2011' dataset in workload traces.

4.2. Energy Consumption Model. A non-profit corporation called Standard Performance Evaluation Corporation (SPEC) is formed to establish, maintain and endorse standardized benchmarks and tools to evaluate performance and energy efficiency for the newest generation of computing systems [16] and [17]. All real data of energy consumption are derived from SPEC power benchmark. The different workload levels of energy consumption in the hosts are shown in Table 4.1.

TABLE 4.1
Host Energy Consumption in Different WorkLoad Levels

Server	HP G4	HP G5
0%	86	93.7
10%	89.4	97
20%	92.6	101
30%	96	105
40%	99.5	110
50%	102	116
60%	106	121
70%	108	125
80%	112	129
90%	114	133
100%	117	135

4.3. Evaluation Metrics. The following evaluation metrics are considered to compare the efficiency of proposed and existing methods.

4.3.1. Energy Consumption. The energy consumption by physical nodes in data center is mostly measured by CPU, memory, and network interfaces. Compared to other computing resources, the CPU consumes more energy. Also, most of these analyses have shown that an idle server consumes almost 70% of energy. This evidence legitimizes the method of converting the idle server to the power saver mode to reduce energy consumption. The energy model [18] is defined as, where E_{max} is the maximum energy consumed by a fully utilized server; k is the fraction of energy consumed by idle hosts (i.e., 70%), and CPU_Host is the host's CPU utilization. In our experiment, E_{max} value is set as 250W which is a constant value for modernized servers.

Due to the workload uncertainty, the CPU utilization may change over time and is defined as CPU_Host(t). Thus, the energy consumption by a host EC can be illustrated as an integral of power consumption over a while.

4.3.2. SLA Violation Rate. The SLA violation rate is one of the factors of QoS. It occurs while migrating VMs from overloaded host. It is based on two metrics [5] such as SVTAH and PDCVM.

1. SLA Violation Time per Active Host (SVTAH) SVTAH decides which active host has reached the 100% CPU utilization during the time. It is given in Eqn. 4.1,

$$SVTAH = \frac{1}{M} \sum_{i=1}^M \frac{T_{pi}}{T_{qi}} \quad (4.1)$$

where M is the number of hosts in the data center; T_{pi} is total time during which the i^{th} host reaches 100% CPU utilization; T_{qi} is total time of host i being in an active state.

2. Performance Degradation Caused by VM Migration (PDCVM) The overall performance will be degraded due to the VM migrations most of the time. It is formulated as

$$PDCVM = \frac{1}{N} \sum_{j=1}^N \frac{P_{dj}}{P_{ri}} \quad (4.2)$$

where N denotes the number of VMs; P_{dj} shows the estimation of performance degradation caused by migration of j^{th} VM; P_{ri} shows the lifetime of the total CPU capacity requested by j^{th} VM.

Both the SLA metrics are equally used to measure the SLA violation independently. It is obtained by multiplying SVTAH and PDCVM which is given Eqn. 4.3.

$$SLAV = SVTAH \times PDCVM \quad (4.3)$$

3. Energy SLA Violation (ESV) ESV is a combined metric that captures both energy consumption and SLA Violation rate, which are denoted as Energy Consumption (EC) and SLA Violation rate (SLAV).

$$ESV = EC \times SLAV \quad (4.4)$$

4. Energy Efficiency (EE) Energy Efficiency (EE) can be incorporated into forms of Energy Consumption and SLA Violation rate. Where EC is energy consumption. It is formed as

$$EE = \frac{1}{P_c \times SLA} \quad (4.5)$$

5. Improvement Rate The percentage improvement of the proposed algorithm is computed using the following Eqn. 4.6.

$$\varphi = \left(1 - \frac{\text{Proposed Method}}{\text{Existing Method}}\right) \times 100 \quad (4.6)$$

5. Results and Discussions. The proposed CPU-Memory aware placement algorithms are compared with following existing algorithms: PABFD [5], and MCC[10]. The parameter d varies from 0.6 to 1.0 by increase of 0.1 [5]. For $d < 0$ there is no CPU utilization of VMs and $d > 1$ there is no variation in these objectives. Hence the value of d is considered between 0.6 to 1.0 for all the algorithms. The impact of the proposed algorithm is experimentally tested with various selection algorithms and the obtained results are tabulated (Table 5.1).

The maximum efficiency is obtained when the correlation is based on memory utilization of hosts and VMs and it also takes less number of VM migrations for consolidation. The performance of the proposed method is discussed with respect to each metric as given below:

5.1. Energy Consumption. The main objective of this paper is to design a VM placement algorithm so that the energy consumption is reduced. Energy consumption is calculated by taking into all hosts throughout the simulation by mapping of CPU and different workload levels from Table 4.1. In every iteration the CPU utilization is measured and energy consumption is calculated from Table 4.1. The results obtained for existing and proposed methods are shown in Fig 5.1 and 5.2. From the figures it is observed that THR_FSS_MEM gives the minimum energy consumption than others. The proposed memory utilization based placement algorithm consumes less energy than other algorithms. It reduces energy consumption by 9.52 % than MCC and 1 % than PABFD. The minimum amount of energy consumed by the proposed algorithm is 12.16 KWh whereas the minimum energy consumption among the existing algorithms 12.28 KWh. Moreover the minimum energy consumption obtained by memory is based on the FSS selection algorithm as shown in Fig 3. The existing selection algorithms generate competitive results with FSS when they are combined with any of the proposed algorithms.

TABLE 5.1
Comparison of various Placement Algorithms

Algorithms	Energy Efficiency	Energy Consumption (KWh)	SLA Violation Rate ($\times 10^{-4}$)	Number of VM migrations
THR_RS_PABFD_0.6	180.48	18.3	3.03	2010
THR_MC_PABFD_0.7	124.29	27	2.98	3644
THR_MMT_PABFD_0.6	254.18	35.19	1.12	4899
THR_MU_PABFD_0.9	120.99	32.93	2.51	4597
THR_MP_MCC_0.6	240.26	38.29	1.09	2857
THR_RS_MCC_0.6	213.75	14.1	3.32	1774
THR_MC_MCC_0.7	236.69	15.7	2.69	1856
THR_MMT_MCC_0.6	164.13	14.1	4.32	1457
THR_MU_MCC_0.6	168.95	14.89	3.98	1588
THR_RS_MEM_0.7	161.77	13.55	4.56	1649
THR_RS_RCM_1.0	490.97	16	1.27	1837
THR_RS_PCM_0.9	166.11	17.8	3.38	2077
THR_MC_MEM_0.6	179.43	14.39	3.87	1666
THR_MC_RCM_0.7	239.8	17.7	2.36	1760
THR_MC_PCM_0.9	197.02	18.39	2.76	1567
THR_MMT_MEM_0.6	239.93	16.37	2.55	1631
THR_MMT_RCM_0.7	183.57	13.68	3.98	1418
THR_MMT_PCM_1.0	230.84	14.77	2.93	1735
THR_MU_MEM_0.9	130.26	17.09	4.49	1782
THR_MU_RCM_0.6	185.67	12.83	4.20	1670
THR_MU_PCM_1.0	122.38	12.16	6.72	1374
THR_FSS_MCC_0.7	680.12	13.44	1.09	1564
THR_FSS_PABFD_0.7	318.6	12.28	2.56	1670
THR_FSS_MEM_0.6	704.79	12.49	1.14	1257
THR_FSS_RCM_0.7	594.35	15.55	1.08	1625
THR_FSS_PCM_0.7	485.84	18.56	1.11	1629

5.2. SLA Violation Rate. SLA violation is one of the key factors of Quality of Service (QoS). It is calculated by taking into two scenarios. First thing is to find overloaded host detection and the next is incurred for migration. K-Means Inter Quartile Range clustering algorithm [7] efficiently finds out the overloaded host. If host overload is predicted efficiently then there will be fewer migrations which will reduce the SLA violation rate. The SLA violation rate is also based on the number of VM migrations. If number of violations are less then SLA violation rate will be less. The proposed RCM (FSS_RCM_0.7) based placement algorithm obtains minimum SLA violation rate as it gets less number of VM migrations. The obtained experimental results are shown in Fig 5.3 and 5.4. It reduces SLA violation by 1 % than MCC and 3.4 % than PABFD. Moreover FSS selection algorithm outperforms other algorithms in terms of generating minimum SLA violation rate. Fig 5.4. shows that the existing selection algorithms reduce the SLA violation rate when they are combined with the proposed placement algorithms.

5.3. Energy Efficiency. The energy efficiency is inversely proportional to Energy Consumption and SLA violation rate. The algorithm which gets minimum EC and SLA violation rate will get minimum energy efficiency. Fig 5.5. shows the energy efficiency obtained by various placement algorithms. It is observed that the FSS based memory aware placement algorithm maximizes the energy efficiency. It outperforms MCC by 3.62 % and PABFD by 121.21 % in terms of improving energy efficiency. The maximum energy efficiency obtained by FSS_MEM_0.6 is 704.79 and the maximum among other algorithms is 680.12. Hence, the FSS_MEM VM placement algorithm is the most energy efficient for VM placement.

5.4. Number of VM Migrations. Less number of VM migrations means efficient VM consolidation and minimum SLA violation rate. Since the proposed placement algorithm uses five thresholds to classify the data

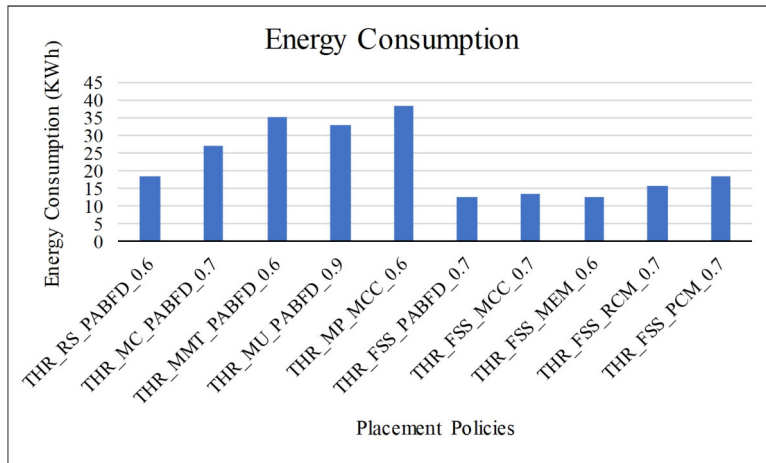


FIG. 5.1. Comparison of Energy Consumption using various VM placement Algorithms

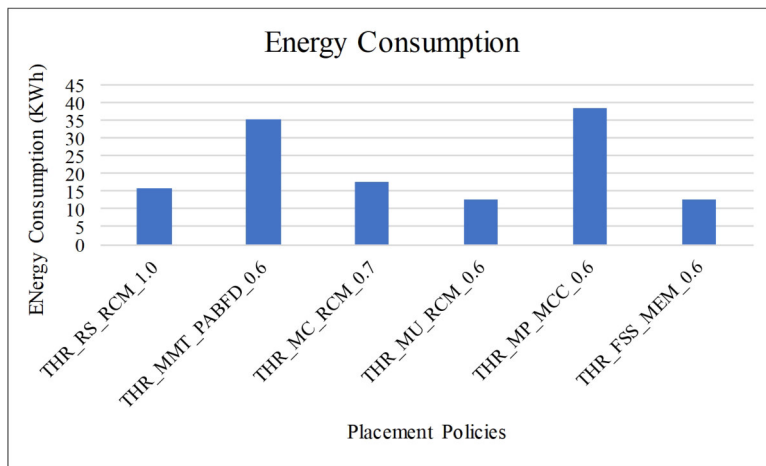


FIG. 5.2. Energy Consumption using VM placement Algorithms

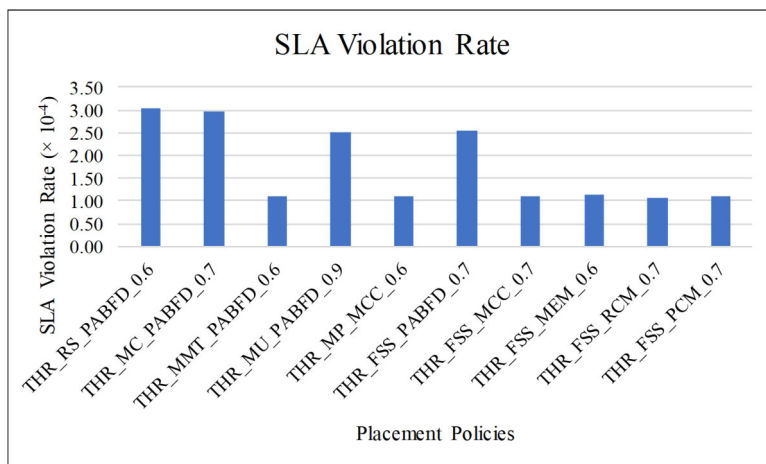


FIG. 5.3. Comparison of SLA Violation Rate using various VM placement Algorithms

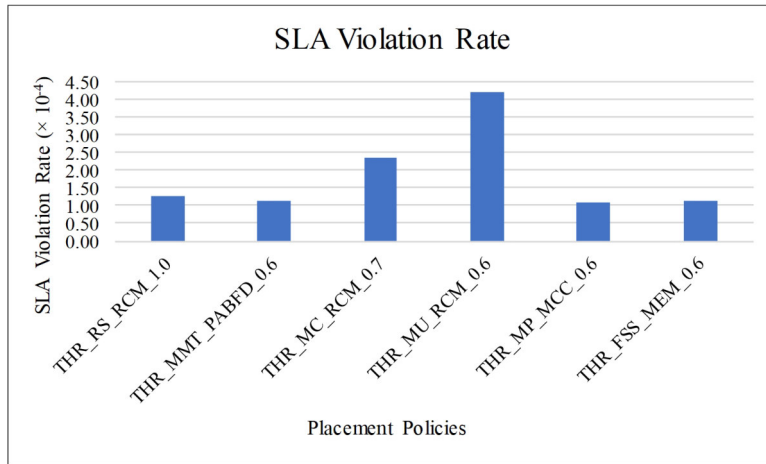


FIG. 5.4. SLA Violation Rate using various VM placement Algorithms

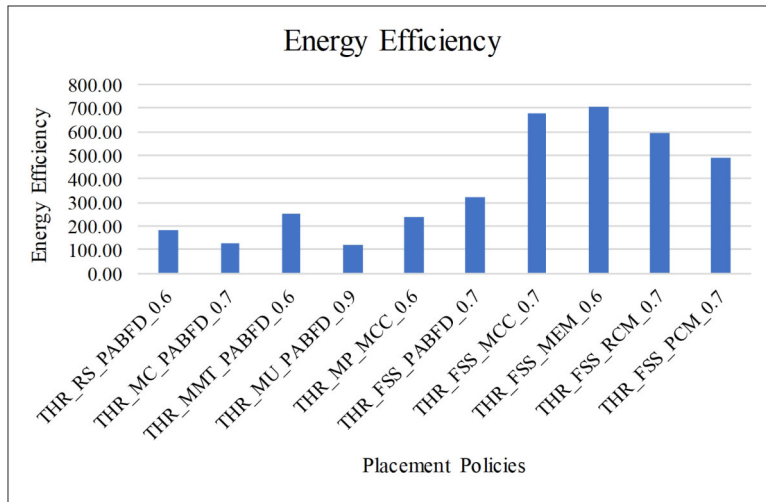


FIG. 5.5. Comparison of Energy Efficiency using Various VM Placement Algorithms

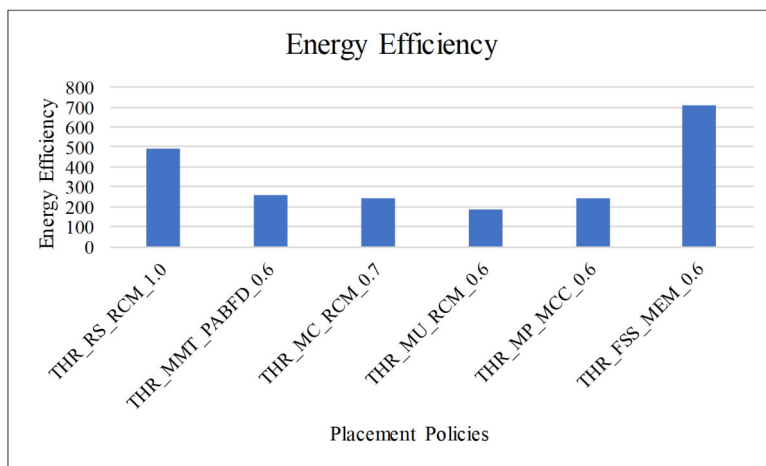


FIG. 5.6. Energy Efficiency using various placement Algorithms

center hosts into five clusters of hosts and applies fuzzy soft set for VM selection it obtains minimum number of VM migrations. The number of VM migration caused by FSS_MEM 0.6 is 1257 whereas the minimum caused by other algorithms is 1457. FSS_MEM algorithm resulted in 13.73 % reduction in migration than the other existing VM placement algorithms.

5.5. Observations. From the simulation results the following observations are made:

1. The existing selection algorithms (RS, MC, MMT, MU, MP) used two thresholds while algorithm FSS used four thresholds for host classification. In our previous work [7], it is identified that the adaptive four threshold algorithm is more effective than two and three threshold algorithms due to better prediction of overloaded hosts and underutilized hosts.
2. The existing VM selection (RS, MC, MMT, MU, MP) algorithms have considered any one of the factors such as either RAM or CPU or Memory or Correlation values for selecting VMs for migration. But are FSS based VM selection algorithm takes into consideration all the four factors at the same time. From the results of [9] it is observed that the FSS algorithm accurately finds which VM has to be migrated from the overloaded hosts.
3. The existing VM placement algorithms (PADFD, MCC) are compared with proposed VM placement algorithms (MEM, RCM, PCM). During the VM placement, the algorithm FSS_MEM achieves maximum energy efficiency. It is experimentally proved that the latter has better performance than the formers.
4. The proposed FSS based VM placement algorithms outperform other selection and placement algorithms in terms of maximizing energy efficiency, minimizing energy consumption, minimizing SLA violation rate, and minimizing the number of VM migrations.

6. Conclusion. In this paper, the VM placement problem is addressed for improving the resource utilization across multiple dimensions with the goal of maximizing energy efficiency and minimizing SLA violation rate. Multiple resource-constraint factors, such as CPU utilization and Memory utilization are used to migrate VMs onto the appropriate hosts in cloud data centers. CPU-Memory aware VM placement algorithm is proposed which considers three variations of resource utilizations: Memory, Ratio of CPU to memory utilization (RCM), and Product of CPU and memory (PCM) utilization. The proposed algorithm is implemented for real-world dataset and the experimental results are compared with existing selection and placement algorithms for various metrics. The results show that THR_FSS_MEM outperforms energy efficiency by (3.62 %) than MCC and the (121.21%) than MCC. THR_MU_PCM outperforms energy consumption (9.52 %) than MCC and (1%) than PABFD. THR_FSS_RCM outperforms SLA violation rate (3.4 %) than PABFD and (1%) than MCC. THR_MEM_0.6 outperforms number of VM migration by (13.73 %) than PABFD and MCC.

Currently the serial processing with multiple iteration is used to process the VM placement methods. All the four methods give the best result. But it consumes more time during the implementation. The parallel processing of all the 4 VM placement methods like CPU, MEMORY, RCM, and PCM method together was a challenge. The above challenges will be overcome in future using GPU systems.

From the experiments and detailed analysis, the VM placements can be done using either MEM or PCM or RCM strategies with fuzzy soft set VM selection policy. They are giving competitive results in terms of generating quality of service during VM consolidation. In the future work, the machine learning or deep learning based prediction method will be applied to dynamically predict VM placement method.

REFERENCES

- [1] HAO JIN, DENG PAN, JING XU, AND NIKI PISSINOU, *Efficient vm placement with multiple deterministic and stochastic resources in data centers*, *IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2012, pp. 2505–2510.
- [2] DONALD J. DALY AND DONALD J. DALY [https://aws.amazon.com/ec2/Economics 2: Ec2](https://aws.amazon.com/ec2/Economics%20Ec2), 1987.
- [3] CODE AND RESPONSE <https://www.ibm.com/us-en/>
- [4] GOOGLE CLOUD <https://cloud.google.com/products/>
- [5] ANTON BELOGLAZOV AND RAJKUMAR BUYYA *Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers*, *Concurrency and Computation: Practice and Experience*, Wiley Online Library, 2012, pp. 1397–1420.
- [6] NITHIYA BASKARAN AND R ESWARI *An Efficient Threshold-Fuzzy Based Algorithm for VM Consolidation in Cloud Datacenter*, 2020, *International Journal of Grid and High Performance Computing*, IGI, (In press).

- [7] NITHIYA BASKARAN AND R ESWARI *Adaptive threshold-based algorithm for multi-objective vm placement in cloud data centers*, In International Conference on Frontier Computing, Springer, 2018, pp. 118–129.
- [8] ZHOU ZHOU, JEMAL ABAWAJY, MORSHED CHOWDHURY, ZHIGANG HU, KEQIN LI, HONGBING CHENG, ABDULHAMEED A ALE-LAIWI, AND FANGMIN LI *Minimizing sla violation and power consumption in cloud data centers using adaptive energy-aware algorithms*, Future Generation Computer Systems, Elsevier, 2018, pp.836–850.
- [9] NITHIYA BASKARAN AND R ESWARI *Fuzzy Softset Based VM Selection in Cloud Datacenter*, International Conference on Intelligent Computing and Control Systems [ICICCS 2019], IEEE Xplore Digital Library, IEEE, (In Press).
- [10] XIONG FU AND CHEN ZHOU *Virtual machine selection and placement for dynamic consolidation in cloud computing environment*, Frontiers of Computer Science, Springer, pp.322–330, 2015.
- [11] ALHARBI, FARES AND TIAN, YU-CHU AND TANG, MAOLIN AND ZHANG, WEI-ZHE AND PENG, CHEN AND FEI, MINRUI *An ant colony system for energy-efficient dynamic virtual machine placement in data centers*, Expert Systems with Applications, Elsevier, 2019, pp.228–238.
- [12] AL-MOALMI, AMMAR AND LUO, JUAN AND SALAH, AHMAD AND LI, KENLI *Optimal virtual machine placement based on grey wolf optimization*, Electronics, Multidisciplinary Digital Publishing Institute, 2019, pp.283.
- [13] PARVIZI, ELNAZ AND REZVANI, MOHAMMAD HOSSEIN *Utilization-aware energy-efficient virtual machine placement in cloud networks using NSGA-III meta-heuristic approach*, Cluster Computing, Springer, 2020, pp. 1–23.
- [14] RODRIGO N CALHEIROS, RAJIV RANJAN, ANTON BELOGLAZOV, CÉSAR AF DEROSE, AND RAJKUMAR BUYYA, *Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms*, Software: Practice and experience, Wiley Online Library, 2011, pp.23–50.
- [15] KYOUNGSOO PARK AND VIVEK S PAI, *Comon: a mostly-scalable monitoring system for planetlab*, ACM SIGOPS Operating Systems Review, ACM, 2006, pp.65–74.
- [16] XIAOBO FAN, WOLF-DIETRICH WEBER, AND LUIZ ANDRE BARROSO, *Power provisioning for a warehouse-sized computer*, ACM SIGARCH computer architecture news, volume 35, ACM, 2007, pp. 13–23.
- [17] DARA KUSIC, JEFFREY O KEPHART, JAMES E HANSON, NAGARAJAN KANDASAMY, AND GUOFEI JIANG, *Power and performance management of virtualized computing environments via lookahead control*, Cluster computing, Springer, 2009, pp.1–15.
- [18] ANTON BELOGLAZOV, JEMAL ABAWAJY, AND RAJKUMAR BUYYA, *Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing*, Future generation computer systems, Elsevier, 2012, pp.755–768.

Edited by: P. Vijaya

Received: Dec 10, 2019

Accepted: Jun 10, 2020



BIRD SWARM OPTIMIZATION-BASED STACKED AUTOENCODER DEEP LEARNING FOR UMPIRE DETECTION AND CLASSIFICATION

SUVARNA NANDYAL* AND SURVANA LAXMIKANT KATTIMANI†

Abstract. One of the most-watched and a played sport is cricket, especially in South Asian countries. In cricket, the umpire has the power to make significant decisions about events in the field. With the growing increase in the utilization of technology in sports, this paper presents the umpire detection and classification by proposing an optimization algorithm. The overall procedure of the proposed approach involves three steps, like segmentation, feature extraction, and classification. At first, the video frames are extracted from the input cricket video, and the segmentation is performed based on the Viola-Jones algorithm. Once the segmentation is done, the feature extraction is carried out using Histogram of Oriented Gradients (HOG), and Fuzzy Local Gradient Patterns (Fuzzy LGP). Finally, the extracted features are given to the classification step. Here, the classification is done using the proposed Bird Swarm Optimization-based stacked autoencoder deep learning classifier (BSO-Stacked Autoencoders), that categories into umpire or others. The performance of the umpire detection and classification based on BSO-Stacked Autoencoders is evaluated based on sensitivity, specificity, and accuracy. The proposed BSO-Stacked Autoencoder method achieves the maximal accuracy of 96.562%, the maximal sensitivity of 91.884%, and the maximal specificity of 99%, which indicates its superiority.

Key words: Umpire classification, Viola-Jones algorithm, Bird Swarm Optimization, Stacked autoencoders deep learning, Histogram of Oriented Gradients, Fuzzy Local Gradient Patterns

AMS subject classifications. 68T05

1. Introduction. Cricket is one of the popular games after soccer in the worldwide. Nowadays, matches are viewed and shared internationally through live satellite broadcasting with the highest viewership rating [1, 9]. Some of the cricket playing nations are Australia, New Zealand, England, India, South Africa, Pakistan, Zimbabwe, Sri Lanka, and Bangladesh. Television broadcasters, such as star sports, and ESPN consists of huge repositories of cricket videos. The analysis of cricket video has been gained more attention in digital video processing. The cricket video analysis is a challenging task, due to its complexities [10, 11]. Existing approaches of cricket videos are classified into genre-independent or genre-specific [17]. Cricket video is chosen as the initial application for entertainment purposes. In the cricket videos, the video contents are edited or recorded using various style formats [12, 14]. Some works have been done in the field that specially targets on cricket videos. One of the flourishing attempts for extracting semantic events from cricket video was based on a multi-level hierarchical framework [10, 16]. This framework employed audio-visual features for categorizing video segments [7]. On the other hand, it has been revealed that camera motion parameters are utilized for classifying limited events in cricket video [13].

Several object categories, such as objects of uninterested, and interest regions present in the video are to be segmented. The video object that is separated from its context is incomplete. These systems must be enforced with the context information. In cricket, the umpire is one of the people with high power for making decisions about events in the field. The umpire used gestures, hand signals, and poses [1]. Object extraction is one of the most crucial components in the framework, since the objects are used as the input for the event extraction process. Object detection from a frame or a video sequence has attracted the attention of many engineers working in the field. In current time object detection technology is well recognized [6, 8]. The motion of the ball and the players is important for understanding any game. The player, as well as the object identification,

*Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Kalaburgi, Karnataka 585102, India (suvarna_nandyal@yahoo.co.in).

†Department of Computer Science and Engineering, B.L.D.E.A's V.P.Dr.P.G.Halakatti College of Engineering and Technology (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Vijayapur, Karnataka 586103, India (suvarnaky1977@gmail.com).

is performed based on contextual color-based segmentation. Then, the location or tracking information is possible by simply finding a similar object in explicit tracking or successive frames [19]. Various algorithms are implemented for tracking and segmenting the video objects, like Continuously Adaptive Mean Shift algorithm (CAMSHIFT), partial list square analysis, Threshold decision, and diffusion distance, etc. [20]. From that CAMSHIFT [2] is a popular method for visual tracking with a minimal computational cost.

There are many techniques utilized to the task for umpire detection, from the classic methods of Hidden Markov Models (HMMs), and Gaussian Mixture Models (GMMs) to machine learning methods [35] of multilayer perceptron (MLP) and static var compensator (SVC), then further moves to the deep neural learning methods [31] of Convolutional neural network (CNN) and LSTM [12]. The k-nearest neighbor algorithm (KNN) is also employed for classification. Motion is an important feature for representing video sequences. Motion texture is the feature, which is derived from the motion field between video frames, like motion vector field or optical flow field [7]. These features are used in conjunction for devising a set of multicategory classifiers with support vector machines (SVMs) [18]. The CNN [1] is also outperformed in object detection and image classification [34], which is the integration of classifier and feature extractor. The convolutional layers of CNN are the feature extractors. They learn the representations from the input data automatically. The previous layers in CNN learn more generic features, like colour blobs, edges, and shapes. Deep learning methods are used in multidimensional applications [28], and online applications [29, 30]. The final fully connected layers of CNN employed these features and divided the data into one of the classes [18].

In this paper, an umpire detection and classification is developed based on BSO-Stacked Autoencoders. The overall procedure of the proposed method involves segmentation, feature extraction, and umpire classification. At first, the video frames are extracted from the input cricket video, and then, the segmentation is done based on the Viola-Jones algorithm. After that, the feature extraction is performed using HOG, and Fuzzy LGP. Once the feature extraction is done, the classification is performed using the proposed BSO-Stacked Autoencoders, which categories into umpire or others.

The main contribution of this paper: Developing umpire detection and classification approach using the proposed BSO-Stacked Autoencoders, in which the Stacked Autoencoder is trained using BSO for effective classification.

The rest of the paper is organized as follows: Section 2 describes the literature survey of eight existing techniques. Section 3 discusses with umpire detection and classification using BSO-Stacked Autoencoders, and section 4 discusses the results of the proposed BSO-StackedAutoencoders. Finally, section 5 concludes the paper.

2. Literature Survey. This section reveals the literature review of several methods related to umpire pose detection, pitch frame classification, batting shots recognition, event, and activity detection are described, and analyzed as follows: Aravind Ravi et al.[1] developed an approach for umpire pose detection using transfer learning to generate cricket highlights. The features were extracted from pre-trained networks. Then, the linear SVM is employed to detect the pose of the umpire. This method has improved training overhead, but did not consider other classification methods for better performance. A. Sasithradevi et al.[2] presented an approach for content-based video retrieval. Histogram of Fourier Coefficients (HFC) was introduced for indicating the objects into video frames. Additionally, the performance of a video retrieval system is validated using Extreme Learning Machine, and Random Forest. The method did not include the temporal information and trajectory information for representing objects in videos. M. Ravinder and T. Venugopal [3] developed the Bag-of-visual-words method to detect, and classify pitch frames in the cricket video. Local Binary Patterns (LBP), Color+Texture+Edge (CTE), and Scale-Invariant Feature Transform (SIFT) are the three features was employed for classification. The main drawback of this method is required more iteration. Muhammad Zeeshan Khan et al.[4] employed a deep convolutional neural network (Deep CNN) for cricket batting shots recognition. For batting shot recognition, several actions, like bowling styles, number of scores, locations of players in the field, and batting shots are analyzed. The method does not consider video to automatic textual commentary generation, information retrieval, and decision making. Sujoy Paul et al. [5] developed weakly-supervised activity localization and classification (W-TALC). Initially, a weakly-supervised module, and Two Stream-based feature extractor network were employed for extracting features. After that, Multiple Instance Learning Loss (MILL) and Co-Activity Similarity Loss (CASL) were established to learn the weights of the

network only the video-level labels of training videos. The method failed to consider CASL to other issues in computer vision.

Mohammad Ashraf Russo et al.[6] employed the combination of CNN, and recurrent neural network for classifying sports videos. The input to the network is a sequence of RGB color frames. Then, the frames were subjected to separate convolutional layers for classification. The method failed to consider other deep learning-based techniques to adopt more weight in the specific dataset. Mahesh M. Goyani et al.[7] developed the keyframe detection-based method to minimize computation time. This framework performed top-down event classification and detection based on the hierarchical tree. At first, the keyframes were extracted using Hue Histogram difference for indexing, and then, the logo transitions were detected. After that, crowd frames were detected using edge detection. Finally, the frame categorized into the player of team A, and team B, and umpire based on skin colour. The smaller dataset is not considered in this method. Sandesh Bananki Jayanth and Gowri Srinivasa [8] developed visual content-based techniques for extracting cricket pitch frames automatically. At first, the pre-processing was performed for selecting the subset of frames. Then, the filtering was done to reduce the search space by removing the frames. Then, the subset of frames was given to the Statistical Modelling of the grayscale (SMoG). After that, the Component Quantization-based region of interest extraction (CQRE) was introduced for pitch frames extraction. Other classifiers were not considered to recognize players and detect events (such as a goal). Hari R and Wilsy M [32] developed a method for detect the presence of the Umpire by the thresholding based color segmentation algorithm. This method produced good output for stored cricket video and also be used in real time with a dedicated hardware support. Anyhow, complex object identification, and object tracking methods were the difficult tasks in this method. Mahesh M. Goyani et al. [33] developed a keyframe detection based approach, which was highly accurate with less computational time. Anyhow, it did not overcome the illumination related problems.

2.1. Challenges.

- The analysis of cricket video is a more challenging task due to the complexities of the game in itself, like various formats of matches, day and night matches (cause illumination related problem), dynamic playing conditions (pitches, field area and so on), and duration [15]. In the proposed system, Fuzzy LGP is used to enhance the distinguishing capability, and to reduce the complexities of the game.
- In [5], the W-TALC framework is developed for temporal activity localization and classification. Here, the accuracy was found better, but the W-TALC framework never been examined on smaller datasets for umpire classification. In this work, the classification is done by the proposed BSA-based stack autoencoder, which is applicable for any size of dataset.

3. Proposed Umpire Detection and Classification Using BSO-Based Stacked Auto Encoder.

This section presents the proposed umpire classification approach using a BSO-based stacked autoencoder. Figure 3.1 deliberates the schematic diagram of the proposed BSO-based stacked autoencoder for umpire detection and classification. At first, the input cricket video is converted into video frames. Subsequently, the segmentation is performed using the Viola Jones algorithm. Once the segmentation is done, the feature extraction is carried out using HOG [16], and Fuzzy LGP [24]. At last, the classification is performed based on extracted features using the proposed BSO [22]-based stacked autoencoder deep learning classifier [21] (BSO-based stacked autoencoder) that categories into umpire or others.

Let us consider G be the input video sequence and it comprises of M number of frames and is denoted as $G = \{L_a; 1 \leq a \leq M\}$. Thus, the frames extracted from is expressed as, $L_a = \{L_1, L_2, \dots, L_M\}$.

3.1. Segmentation of humans in each frame using Viola Jones. This section presents the segmentation of face region from the extracted video frames. Different types of algorithms are used for face detection in previous works. The traditional approaches used for face detection are binary classification techniques, posing high computational complexity issues. The Viola-jones face detection algorithm is utilized for detecting face region from the extracted frames. Compared to other face detection algorithm, the computational complexity of the Viola-Jones [23] is limited that makes it suitable for various applications, like image databases, user interfaces, teleconferencing and so on. The steps involved in the Viola-Jones face detection algorithm are as follows: the creation of an integral image, feature selection by Haar-like a feature, cascading classifiers, detection of the face region, and Ada-boosting classifier based feature selection. After classification and Ada-boost

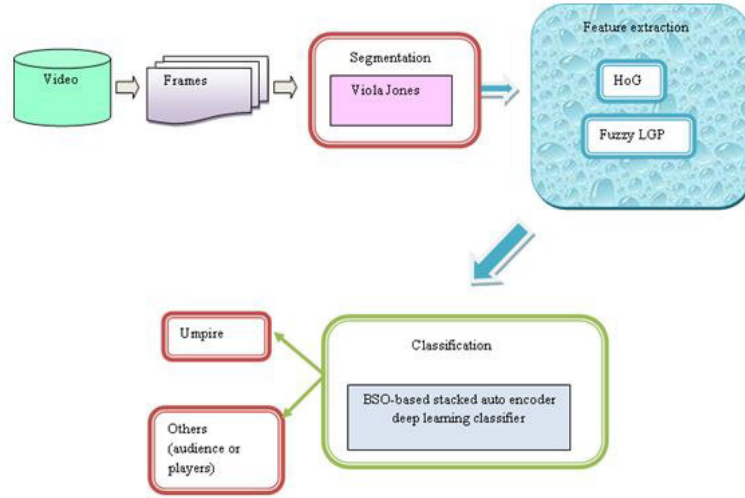


FIG. 3.1. Block diagram of the proposed approach for umpire classification

classifier based feature selection task is performed, the face regions are detected by the Viola-Jones algorithm and is expressed as,

$$(3.1) \quad J = VJ(L_a)$$

VJ indicates the function representing the Viola-Jones algorithm for face detection. The segmented region is expressed as J .

3.2. Feature Extraction. After segmentation, feature extraction is performed using fuzzy HoG, and Fuzzy-LGP. The feature extraction step carried out in this paper is explained as follows:

a) Fuzzy LGP: Fuzzy LGP [24] is the integration of LGP with fuzzy logic, which is used to enhance the distinguishing capability, and to reduce the noise effects. The crisp form of LGP utilizes the neighbourhood pixel properties to explain each pixel. It is more resistant, efficient, and simple to make changes in gray-level using lighting variations. The fine texture properties are effectively captured by the LBP patterns. However, LBP utilizes hard thresholding to compute the code, and has minimum discrimination power, and is sensitive against noise. The Fuzzy-LGP descriptor is an advanced form of LGP that carries highly discriminative features. In Fuzzy-LGP, every pixel is regarded as any number of LBP codes, which contributes to the Fuzzy-LGP bin histogram. The membership function is calculated as,

$$(3.2) \quad h_0(d) = \left\{ \begin{array}{ll} 0; & g_d \geq g_{center} + E \\ \frac{E - g_d + g_{center}}{2 \cdot E} & g_{center} - E < g_d < g_{center} + E \\ 1; & g_d \geq g_{center} - E \end{array} \right\}$$

$$(3.3) \quad h_1(d) = 1 - h_0(d)$$

where, $h_1()$ and $h_0()$ represents the membership functions, E denotes the threshold parameter, which is used to control the fuzziness degree, g_d denotes the neighboring pixel, and d is denoted as the total number of pixels. The membership function determines the contribution of LBP code into the FLBP histogram. The LGP code contribution is defined as

$$(3.4) \quad F(X) = \prod_{d=0}^8 h_{c_d}(d)$$

where, $c_d \in \{0, 1\}$. The sum of neighbourhood contribution is equal to unity, and is expressed as,

$$(3.5) \quad \sum_X^{255} F(X) = 1$$

Crisp LGP histogram has zero value bins, whereas FLGP histogram has non-zero value bins. Hence FLGP is more informative than crisp LGP.

b) HOG: HOG [16] is a type of feature descriptor for extracting umpires. This technique utilizes gradients for localizing the image. The HOG features are extracted using the magnitude and orientation. Gradients (H_u, H_v) are calculated in both vertical and horizontal directions for all pixels in the frame. Gradient magnitudes and directions are estimated using the below equations,

$$(3.6) \quad h = \sqrt{H_u^2 + H_v^2}$$

$$(3.7) \quad \theta = \tan^{-1}\left(\frac{H_v}{H_u}\right)$$

After the extraction of features from every face region, the concatenated feature is expressed as,

$$(3.8) \quad J^D = \{F || H\}$$

The extracted features are denoted as J^D with dimension $1X(M \times S)$. where, M signifies the total number of bins, S refers to the total number of extracted dimensions of the HOG features.

3.3. Umpire classification using BSO-based stack autoencoder. Once the features are extracted using fuzzy LGP, and HOG, the extracted features are given to the proposed BSA-based stack autoencoder for umpire classification. For the effective classification, stack autoencoder deep learning classifier is used, and the weights-biases are determined using the proposed algorithm optimally.

i) Architecture of stacked autoencoder deep learning. One of the building blocks of Deep Neural Network (DNN) is autoencoder. Stacked autoencoder [21] is the most commonly employed deep learning techniques. Figure 3.2 shows the architecture of the stacked autoencoder. Here, the autoencoders are stacked hierarchically. The architecture of autoencoder possesses three units, input visible units, output visible units, and hidden units, denoted as S , K and N . An autoencoder is utilized for encoding the input vector into an advanced stage hidden representation. In the encoder phase, the deterministic mapping R_ϕ , transforms the input vector into the hidden vector, and is expressed as,

$$(3.9) \quad B = fun(Q_1 P + A_1 P)$$

where, Q_1 indicate the weight matrix, and A_1 refers to the bias vector. The term P refers to reconstruction. Then the hidden representation is decoded and back to reconstruction factor is given as follows,

$$(3.10) \quad P = fun(Q_2 K + A_2 K)$$

where, Q_2 denote weight matrix, A_2 represent the bias vectors, and K hidden units.

ii) Training phase. a) Fitness Evaluation: The fitness function is computed for the individual solution for better results. The output has the limited value of the error, which considered as optimal output. The best solution is determined at the previous iteration as each solution craves to obtain a better position. To mitigate the cost function, and the squared reconstruction error, the autoencoder is trained using backpropagation algorithm expressed by

$$(3.11) \quad Jac(Q, A) = \frac{1}{2Y} \sum_{j=1}^Y ||P_j - P_{target}||^2$$

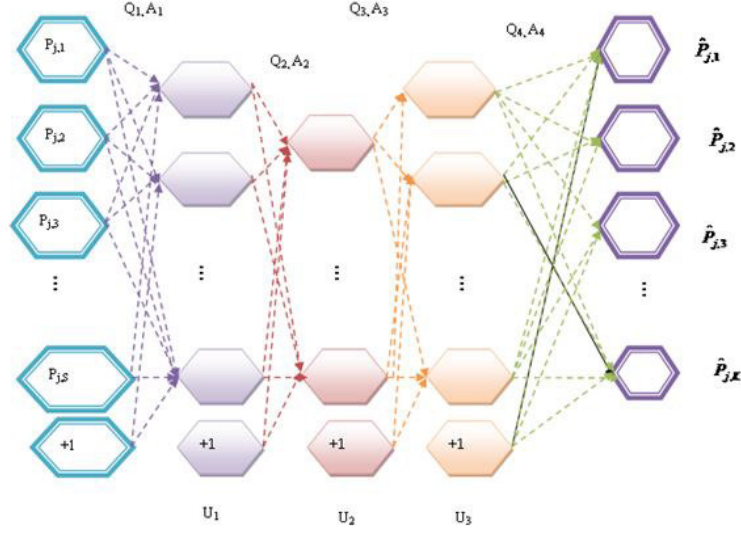


FIG. 3.2. Architecture of stacked autoencoder classifier

where P_j represents the estimated output, and the target output is denoted as P_{target} . The term Y denotes the total number of training samples.

b) Algorithmic description of the proposed BSO-Stacked autoencoder BSO [22] is duly based on the social behaviors of birds that follow some idealistic rules that follow: The individual bird switches between the vigilance behavior and foraging behaviour of birds. When foraging is in progress, individual bird records and updates the previous experience and their position, and also, the best experience of the swarms is updated, which is regarding the location of food. In case of vigilance behaviour, the individual birds move towards the center of swarms. The vigilance behaviour gets affected when there is any possibility of interference. At the same time, the bird switches between scrounging and producing when birds are trying to fly from one site to another. On the other hand, producers engage in active search for food. In addition, there is a perfect balance between exploration and exploitation in BSA. The Algorithmic steps of the proposed BSO-Stacked Autoencoders are illustrated below:

a) Initialization: In the first step, the parameters of optimization including the population are initialized, which includes: $\{X_{i,j}; 1 \leq i \leq y; 1 \leq j \leq z$ where, the population size is denoted as y , τ_{max} denotes the maximal iteration, Pro represents the probability of foraging food, and the frequency of flight behaviour of birds is denoted as fr . Here, $X \in Q_1, Q_2, A_1, A_2$.

b) Evaluation of objective function: The selection of the optimal location of the bird is performed based on the minimization problem. The minimal value of the objective function describes the better solution, and therefore, the solution with the limited value of error is selected as the optimal solution. The error is estimated using equation (11).

c) Location update of the birds: For updating their positions, the birds have three stages, which are decided based on the probability. Whenever the random number, then the update is based on the foraging behaviour or else, the vigilance behaviour commences. On the other hand, swarm splits as scroungers and producers, which is modelled as flight behaviors. Finally, the feasibility of the solutions is verified and the best solution is retrieved.

d) Foraging behavior of birds: The individual bird searches the food based on their own experience, and the behaviour of the swarm, which is given below. The standard equation of the foraging behavior of birds is given as follows,

$$(3.12) \quad X_{i,j}^{\tau+1} = X_{i,j}^{\tau} + (P_{i,j} - X_{i,j}^{\tau}) \times G \times random(0,1) + [C_j - X_{i,j}^{\tau}] \times random(0,1)$$

where $X_{i,j}^{\tau+1}$ and $X_{i,j}^{\tau}$ denotes the location of i^{th} bird in j^{th} dimension at $\tau + 1$ and τ . $P_{i,j}$ refers to the previous best position of the i^{th} bird, $random(0, 1)$ is the independent uniformly distributed numbers, and C_j indicates the best previous location shared by the bird swarm. The positive numbers are denoted as G and V , and $P_{i,j}$ denotes the personal best solution. C_j represents the global best solution.

e) Vigilance Behaviour of Birds: The birds move towards the center during which the birds compete with each other and the vigilance behaviour of birds is modelled as,

$$(3.13) \quad X_{i,j}^{\tau+1} = X_{i,j}^{\tau} + B_1(\mu_j - X_{i,j}^{\tau}) \times random(0, 1) + B_2[M_{kj} - X_{i,j}^{\tau}] \times random(-1, 1)$$

$$(3.14) \quad B_1 = w_1 \times exp\left(\frac{-Ff(M)_i}{\sum Ff + \delta} \times y\right)$$

$$(3.15) \quad B_1 = w_1 \times exp\left[\left(\frac{Ff(M)_i - Ff(M)_\tau}{|Ff(M)_\tau - Ff(M)_i| + \delta}\right) \frac{y \times Ff(m)_\tau}{\sum Ff + \delta} \times y\right]$$

where y represents the number of birds, w_1 and w_2 are the positive constants lying in the range of $[0, 2]$, $Ff(M)_i$ represents the optimal fitness value of i^{th} bird, and $\sum Ff$ corresponds to the addition of the best fitness values of the swarm. δ be the constant that keeps optimization away from zero-division error. T refers to the positive integer ($j \neq i$). Whenever the bird approaches the center of the swarm, there is a tendency to compete with each other. The average fitness $\sum Ff$ value of the swarm corresponds to the indirect effect caused by surroundings when the swarm approaches the surroundings. The mean μ_h refers to the h^{th} element of the average position of the swarm.

f) Flight behaviour: This behaviour of the bird's progress when the birds fly to another site in case of any threatening events and foraging mechanisms. When the birds reach a new site, they search for food. Few birds in the group try acting as producers and few as scroungers. The behaviour is modelled as

$$(3.16) \quad X_{i,j}^{\tau+1} = X_{i,j}^{\tau} + random\ r(0.1) \times X_{i,j}^{\tau}$$

$$(3.17) \quad X_{i,j}^{\tau+1} = X_{i,j}^{\tau} + (X_{\gamma,j}^{\tau} - X_{i,j}^{\tau}) \times fl \times Rr(0, 1)$$

where $Rr(0, 1)$ refer to the Gaussian distributed random number with zero mean and standard deviation, $X_{i,j}^{\tau+1}$ and $X_{i,j}^{\tau}$ denotes the location of i^{th} bird in j^{th} dimension at $(\tau + 1)$ and τ .

g) Checking the feasibility of solution: The feasibility of the solution is computed based on the objective function. If the newly generated solution is best than the previous one, then it is changed by the new solution.

h) Termination: Repeat the steps for the maximal iterations until the global optimal best solutions are determined. Thus, the optimization algorithm discussed in this section aims at determining the optimal multipath for enabling secure communication in the network. Algorithm 1 depicts the pseudo-code of the proposed BSO-Stacked Autoencoders, which demonstrates the step-wise description of the algorithm.

The flowchart for the proposed BSO-Stacked Autoencoders is given in Figure 3.3.

4. Discussion of Results. The results obtained by the proposed BSO-Stacked Autoencoders are described in this section. The proposed BSO-Stacked Autoencoders is analyzed based on the performance using three measures, which include accuracy, sensitivity, and specificity.

4.1. Experimental setup. The experimentation of the developed approach is performed using manually collected database, and the implementation is done in the MATLAB tool. Here, the 20 videos are collected from YouTube.

Algorithm 1: Pseudocode for the proposed BSO-Stacked Autoencoders

```

Data: Bird swarm population  $\{X_i, j; 1 \leq i \leq y; 1 \leq j \leq z;\}$ 
Result: Best solution
1 Procedure:
2 Begin
3 Population initiation:  $\{X_i, j; 1 \leq i \leq y; 1 \leq j \leq z;\}$ 
4 Read the parameters  $y$  -population size;  $\tau_{max}$  maximal iteration,  $Pro$  -probability of foraging food,  $fr$  -frequency of
  flight behaviour of birds
5 Determine the fitness of the solutions
6 while  $\tau < \tau_{max}$  do
7   for  $g = 1 : y$  do
8     if  $R(0, 1) < Pro$  then
9       | Foraging behaviour using equation (3.12)
10    else
11     | Vigilance behaviour using equation (3.15)
12   Split the swarm as scroungers and producers
13   for  $g = 1 : y$  do
14     if  $g$  is a producer then
15       | Update using equation (3.16)
16     else
17       | Update using equation (3.17)
18   Check the feasibility of the solutions
19   Return the best solution
20    $\tau = \tau + 1$ 
21 Optimal solution is obtained
22 End

```

4.2. Performance metrics. The performance of the developed BSO-Stacked Autoencoders is evaluated using the metrics, namely accuracy, sensitivity, and specificity. a) Accuracy: Accuracy is defined as the measure of accurateness based on the BSO-Stacked Autoencoders, and is expressed as,

$$(4.1) \quad Accuracy = \frac{X + Y}{X + P + Y + Q}$$

where, denotes the true positives, and is the true negatives. denotes the false positives and is the false negatives. b) Sensitivity: Sensitivity is also called as a True positive Rate (TPR), which defines the measure of positiveness and is computed using the below expression.

$$(4.2) \quad Sensitivity = \frac{Y}{P + Y}$$

c) Specificity: Specificity is otherwise called as a True Negative Rate (TNR), which is the measure of false negatives. Specificity is represented as,

$$(4.3) \quad Specificity = \frac{X}{X + Q}$$

4.3. Experimental results. The experimental results obtained by the proposed BSO-Stacked Autoencoders are given in this section. Figure 4.1 depicts the sample results of the umpire. Figure 4.1.a) shows the input image 23, figure 4.1.b) depicts the input image 42, figure 4.1.c) shows the input image 69, and figure 4.1.d) depicts the input image 60. Figure 4.1.e) shows the predicted output of input image 23, figure 4.1.f) depicts the predicted output of input image 42, figure 4.1.g), and figure 4.1.h) demonstrate the predicted output of input image 69, and 60.

The sample results of non-umpire are depicted in figure 4.2 depicts the sample results. Figure 4.2.a) shows the input image 1, figure 4.2.b) depicts the input image 13, figure 4.2.c) shows the input image 15, and figure 4.2.d) depicts the input image 61. Figure 4.2.e) shows the predicted output of input image 1, figure 4.2.f) depicts

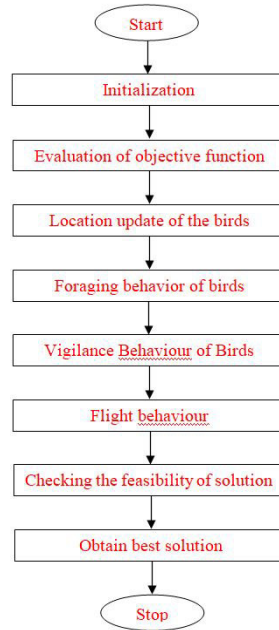


FIG. 3.3. Flowchart of the proposed BSO-Stacked Autoencoders



FIG. 4.1. Sample results of umpire a) Input image 23, b) Input image 42, c) Input image 69, d) Input image 60, e) predicted output of input image 23, f) predicted output of input image 42, g) predicted output of input image 69, h) predicted output of input image 60

the predicted output of input image 13, figure 4.2.g), and figure 4.2.h) demonstrate the predicted output of input image 15, and 61.

4.4. Performance analysis. The performance analysis of the developed BSO-Stacked Autoencoder method is carried out with respect to the sensitivity, specificity, and accuracy. These parameters are necessary to be improved to enhance the performance of the technique.

a) Analysis based on training percentage Figure 4.3 shows the performance analysis of the proposed BSO-Stacked Autoencoder method. Figure 4.3.a) shows the performance analysis in terms of accuracy by varying the number of iteration. In this figure, the accuracy value has its limit ranging from 0 to 90 and the training percentage value varies between 40 and 90. For 40% of training data, the accuracy values measured by the BSO-Stacked Autoencoders for the number of iterations 50, 100, 150, 200, and 250 are 61.182%, 62.756%, 63.014%, 68.125%, and 78.191%, respectively. Similarly, when the training data percentage is 90, the accuracy value measured by BSO-Stacked Autoencoders with iteration 50 is 71.074%, BSO-Stacked Autoencoders with

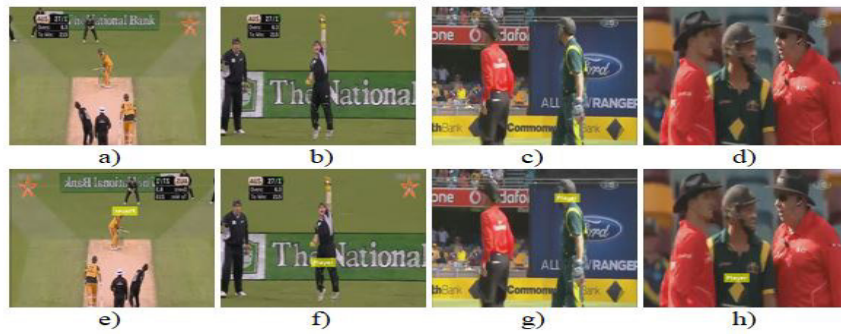


FIG. 4.2. Sample results of non-umpire a) Input image 1, b) Input image 13, c) Input image 15, d) Input image 61, e) predicted output of input image 1, f) predicted output of input image 13, g) predicted output of input image 15, h) predicted output of input image 61

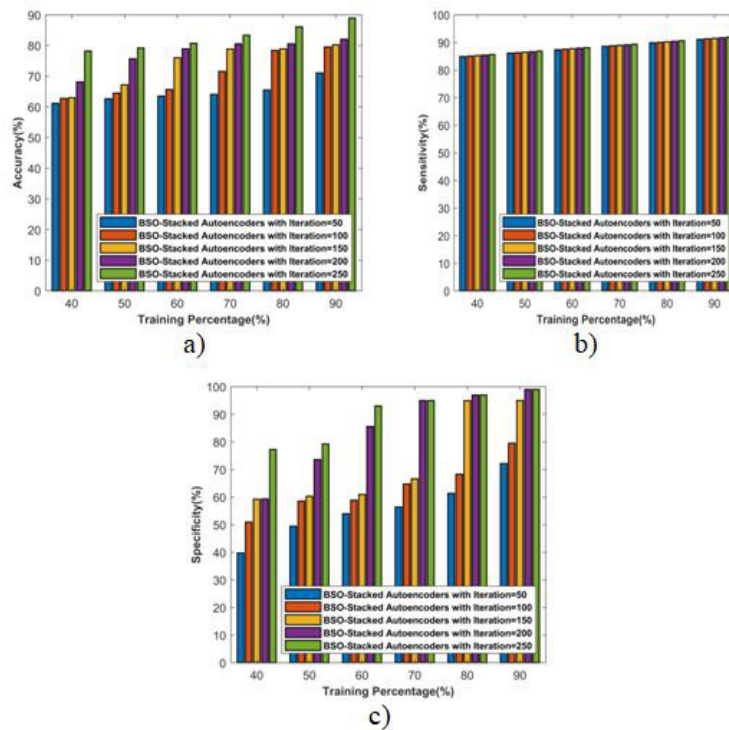


FIG. 4.3. Performance analysis of the proposed BSO-Stacked Autoencoders (a) Accuracy, (b) Sensitivity, and (c) Specificity

iteration 100 is 79.502%, BSO-Stacked Autoencoders with iteration 150 is 80.246%, BSO-Stacked Autoencoders with iteration 200 is 82.055%, and BSO-Stacked Autoencoders with iteration 250 is 88.941%. From the above figure, it is clearly shown the maximum accuracy measure of 88.941% is attained with the iteration 250 for the training data percentage 90.

Figure 4.3.b) depicts the performance analysis in terms of sensitivity. When the training data percentage is 50, the sensitivity values measured by the proposed BSO-Stacked Autoencoders with iteration 50, 100, 150, 200, and 250 are 86.152%, 86.33%, 86.508%, 86.686%, and 86.864%, respectively. Likewise, when the training data percentage is 80, the sensitivity value measured by BSO-Stacked Autoencoders with iteration 50 is 89.884%, BSO-Stacked Autoencoders with iteration 100 is 90.068%, BSO-Stacked Autoencoders with iteration

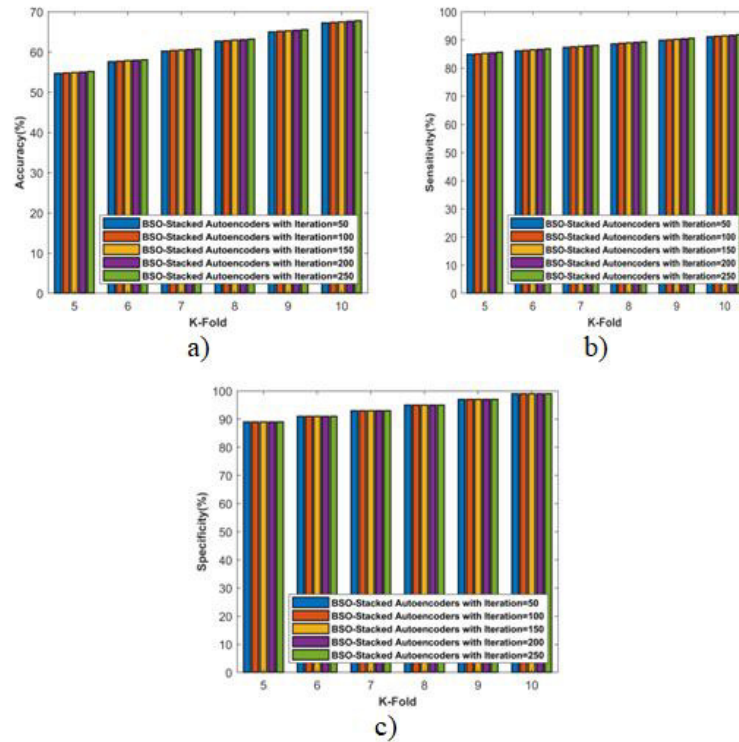


FIG. 4.4. Performance analysis of the proposed BSO-Stacked Autoencoders (a) Accuracy, (b) Sensitivity, and (c) Specificity

150 is 90.252%, BSO-Stacked Autoencoders with iteration 200 is 90.436%, and BSO-Stacked Autoencoders with iteration 250 is 90.62%. From the figure, it is concluded that when the training data percentage is 90, with the iteration 250, the proposed method acquired better performance.

Figure 4.3.c) depicts the performance analysis in terms of specificity. When the training data percentage is 60, the specificity values measured by the proposed BSO-Stacked Autoencoders with iteration 50, 100, 150, 200, and 250 are 53.975%, 58.910%, 60.983%, 85.592%, and 93%, respectively. Likewise, for 90% training data, the specificity value measured by BSO-Stacked Autoencoders with iteration 50 is 72.174%, BSO-Stacked Autoencoders with iteration 60 is 79.529%, BSO-Stacked Autoencoders with iteration 70 is 95%, BSO-Stacked Autoencoders with iteration 80 is 99%, and BSO-Stacked Autoencoders with iteration 90 is 99%. From the figure, it is clearly shown that when the training data percentage is 90 with iteration 200, and 250, the proposed method acquired better performance.

b) Analysis based on K-Fold Figure 4.4 depicts the performance analysis of the proposed BSO-Stacked Autoencoder method. Figure 4.4.a) shows the analysis based on accuracy by varying the K-Fold. For K-Fold=7, the accuracy values measured by the BSO-Stacked Autoencoder method for the number of iterations 50, 100, 150, 200, and 250 are 60.250%, 60.374%, 60.499%, 60.623%, and 60.748%. Similarly, when K-Fold is 10, the accuracy value measured by BSO-Stacked Autoencoders with iteration 50 is 67.235%, BSO-Stacked Autoencoders with iteration 100 is 63.374%, BSO-Stacked Autoencoders with iteration 150 is 67.512%, BSO-Stacked Autoencoders with iteration 200 is 67.650%, and BSO-Stacked Autoencoders with iteration 250 is 67.788%.

Figure 4.4.b) shows the performance analysis based on sensitivity by varying the K-Fold. In this figure, the sensitivity value is plotted against the K-Fold varies between 5 and 10. For K-Fold=8, the sensitivity values measured by the BSO-Stacked Autoencoder method for the number of iterations 50, 100, 150, 200, and 250 are 88.634%, 88.816%, 88.998%, 89.18%, and 89.362%. Similarly, when the K-Fold value is 10, the sensitivity value measured by the BSO-Stacked Autoencoders with iteration 50 is 91.14%, BSO-Stacked Autoencoders with

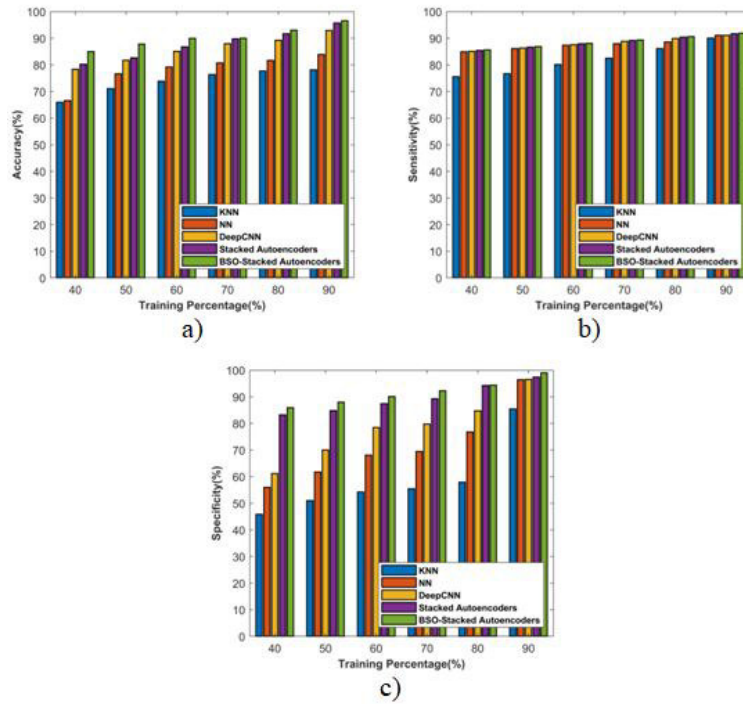


FIG. 4.5. Comparative analysis by varying the training percentage a) Accuracy, b) Sensitivity, and c) Specificity

iteration is 91.326%, BSO-Stacked Autoencoders with iteration 150 is 91.512%, BSO-Stacked Autoencoders with iteration 200 is 91.698%, and BSO-Stacked Autoencoders with iteration 250 is 91.884%. The maximum sensitivity measure of 91.884% is attained when the iteration is 250, with 10% of K-Fold.

Figure 4.4.c) depicts the performance analysis in terms of specificity. When the K-Fold is 5, the specificity values measured by BSO-Stacked Autoencoders with iterations 50, 100, 150, 200, and 250 are found to be 89%, respectively. Likewise, when the K-Fold is 9, the specificity value measured by BSO-Stacked Autoencoders with iterations 50, 100, 150, 200, and 250 are found to be 89%.

4.5. Comparative analysis. The comparative analysis of the proposed BSO-Stacked Autoencoder method by evaluating the performance of other comparative techniques is elaborated in this section. The comparative analysis is performed by varying the training data percentage, and K-fold, and the results are evaluated based on the metrics, like accuracy, specificity, and sensitivity.

4.6. Competing methods. The methods, such as KNN [26], Neural Network (NN) [27], DCNN [25], and Stacked Autoencoders, are used for the comparison with the proposed BSO-Stacked Autoencoder method for the analysis.

a) Comparative analysis based on training data percentage: The analysis of the comparative methods based on accuracy, sensitivity, and specificity is depicted in figure 4.5. Figure 4.5.a) shows the analysis based on accuracy by varying the percentage of training data. For the training data=40%, the existing techniques, such as KNN, NN, DCNN, and Stacked Autoencoders, possesses the accuracy of 65.937%, 66.593%, 78.345%, and 80.262%, respectively, which is comparatively lower than the proposed BSO-Stacked Autoencoders. For the same training data, the proposed BSO-Stacked Autoencoders acquired the accuracy of 84.99%. Similarly, when the training percentage increased to 90%, the methods, KNN, NN, DCNN, and Stacked Autoencoders, attained the accuracy of 78.114%, 83.903%, 92.936%, and 95.717%, respectively, whereas the accuracy of the proposed method is 96.56%. From the above interpretation, it is seen that the proposed BSO-Stacked Autoencoders achieved improved accuracy of 96.56% at 90% training data.

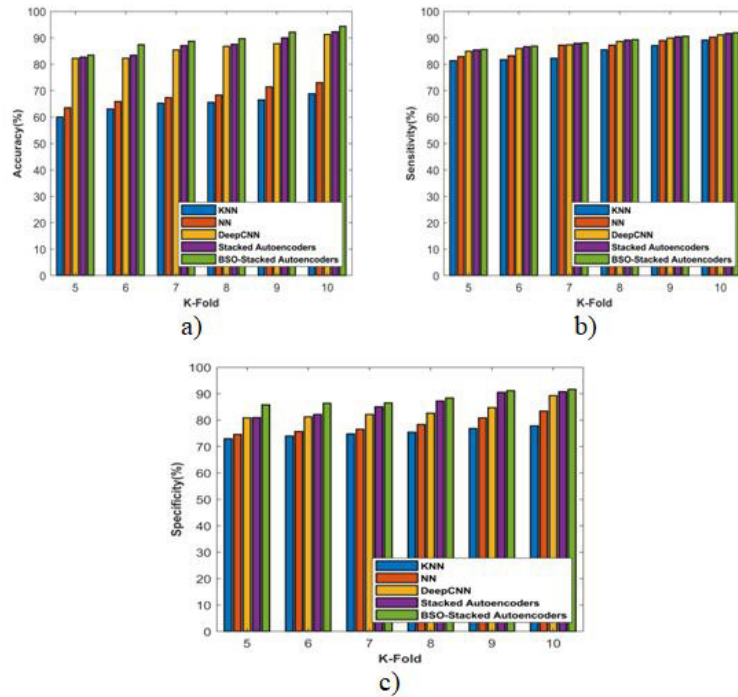


FIG. 4.6. Comparative analysis by varying the K-Fold a) accuracy b) sensitivity and c) specificity

The comparative analysis based on the sensitivity metric is depicted in figure 4.5.b). When 60% of training data is considered, the existing methods, like KNN, NN, DCNN, and Stacked Autoencoders, acquire the sensitivity value of 80.134%, 87.39%, 87.57%, and 87.93%, respectively. Meanwhile, the proposed BSO-Stacked Autoencoders obtained a sensitivity value of 88.11%. When 80% of training data is considered, the sensitivity of the existing methods, like KNN, NN, DCNN, and Stacked Autoencoders, is 86.159%, 88.634%, 89.884%, and 90.436%, respectively, whereas the proposed BSO-Stacked Autoencoders attained the sensitivity of 90.62%.

The analysis based on specificity by varying the training data percentage is depicted in figure 4.5.c). Here, for the 70% training data, the existing techniques, like KNN, NN, DCNN, and Stacked Autoencoders achieved the specificity of 55.454%, 69.450%, 79.771%, and 89.270%, but the proposed BSO-Stacked Autoencoders acquired the specificity of 92.245%. While considering 80% of training data, the existing techniques, like KNN, NN, DCNN, and Stacked Autoencoders, achieved the specificity of 57.953%, 76.818%, 84.736%, and 94.284%, respectively. Meanwhile, the proposed BSO-Stacked Autoencoders attained the specificity of 94.381%. From figure 8 c), the proposed BSO-Stacked Autoencoders is found to possess the maximum specificity of 99% at 90% training data.

b) Comparative analysis based on K-Fold: The analysis of the comparative methods based on accuracy, sensitivity, and specificity is depicted in figure 4.6. Figure 4.6.a) shows the analysis based on accuracy by varying the percentage of training data. For the K-Fold =5, the existing techniques, such as KNN, NN, DCNN, and Stacked Autoencoders, possesses the accuracy of 60.063%, 63.530%, 82.231%, and 82.717%, respectively, which is comparatively lower than the proposed BSO-Stacked Autoencoders. For the same training data, the proposed BSO-Stacked Autoencoders achieved an accuracy value of 84.49%. Similarly, when the K-Fold is increased to 9, the methods, KNN, NN, DCNN, and Stacked Autoencoders, attained the accuracy of 66.500%, 71.447%, 87.798%, and 90.001%, respectively, whereas the accuracy of the proposed BSO-Stacked Autoencoders is 92.135%. From the above interpretation, it is seen that the proposed method achieved an improved accuracy value of 94.356% at 10% of K-Fold.

The comparative analysis in terms of the sensitivity metric is depicted in figure 4.6.b). When K-Fold=6 is

TABLE 4.1
Comparative analysis based on training data percentage

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
KNN	78.114	90.065	85.444
NN	83.903	91.14	96.426
DCNN	92.936	91.151	96.525
Stacked Autoencoders	95.717	91.698	97.381
Proposed BSO-Stacked Autoencoders	96.562	91.884	99

TABLE 4.2
Comparative analysis based on K-Fold

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
KNN	68.811	89.177	77.810
NN	73.068	90.255	83.448
DCNN	91.305	91.14	89.287
Stacked Autoencoders	92.294	91.698	90.792
Proposed BSO-Stacked Autoencoders	94.356	91.884	91.663

considered, the existing methods, like KNN, NN, DCNN, and Stacked Autoencoders acquire the sensitivity of 81.707%, 83.246%, 85.993%, and 86.686%, respectively. Meanwhile, the proposed BSO-Stacked Autoencoders acquired the sensitivity value of 86.864%. For K-Fold=10, the sensitivity of the comparative methods, like KNN, NN, DCNN, and Stacked Autoencoders is 89.177%, 90.255%, 91.14%, and 91.698%, respectively, whereas the proposed BSO-Stacked Autoencoders attained the sensitivity of 91.884%.

The analysis in terms of specificity metric is depicted in figure 4.6.c). Here, for K-Fold=7, the existing techniques, like KNN, NN, DCNN, and Stacked Autoencoders, achieved the specificity of 74.806%, 76.551%, 82.167%, and 85.055%, respectively, but the BSO-Stacked Autoencoders acquired the specificity of 86.530%. While considering K-Fold=10, the existing techniques, such as KNN, NN, DCNN, and Stacked Autoencoders, achieved the specificity of 77.810%, 83.448%, 89.287%, and 90.792%, respectively. Meanwhile, the proposed BSO-Stacked Autoencoders attained the specificity value of 91.663%. From figure 9 c), the BSO-Stacked Autoencoders is found to possess the maximum specificity value of 91.663% at 10% K-Fold.

4.7. Comparative discussion. Table 4.1 depicts the comparative discussion of the existing KNN, NN, DCNN, and Stacked Autoencoders and the proposed BSO-Stacked Autoencoders in terms of sensitivity, specificity, and accuracy parameters by varying the training data percentage. The maximum performance measured by proposed BSO-Stacked Autoencoders in terms of accuracy parameter is 96.562%, whereas the accuracy values of existing KNN, NN, DCNN, and Stacked Autoencoders are 78.114%, 83.903%, 92.936%, and 95.717%, respectively. The maximal sensitivity is computed by proposed BSO-Stacked Autoencoders with a value of 91.884%, whereas the existing KNN, NN, DCNN, and Stacked Autoencoders acquired the sensitivity of 90.065%, 91.14%, 91.151%, and 91.698%, respectively. The specificity value computed by proposed BSO-Stacked Autoencoders is 99%, whereas the existing KNN, NN, DCNN, and Stacked Autoencoders methods acquired the specificity of 85.444%, 96.426%, 96.525%, and 97.381%, respectively.

Table 4.2 depicts the comparative discussion of the existing KNN, NN, DCNN, and Stacked Autoencoders and proposed BSO-Stacked Autoencoders in terms of accuracy, sensitivity, and specificity parameters based on K-Fold. The maximum performance measured by proposed BSO-Stacked Autoencoders in terms of accuracy parameter is 94.356%, whereas the accuracy values of existing KNN, NN, DCNN, and Stacked Autoencoders are 68.811%, 73.068%, 91.305%, and 92.294%, respectively. The maximal sensitivity is computed by proposed BSO-Stacked Autoencoders with a value of 91.884%, whereas the existing KNN, NN, DCNN, and Stacked Autoencoders acquired the sensitivity of 89.177%, 90.255%, 91.14%, and 91.698%, respectively. The specificity value computed by proposed BSO-Stacked Autoencoders is 91.663%, whereas the existing KNN, NN, DCNN, and Stacked Autoencoders methods acquired the specificity of 77.810%, 83.448%, 89.287%, and 90.792%, respectively.

5. Conclusion. This paper presents an approach for umpire detection and classification by proposing an optimization algorithm. The proposed model undergoes three steps for the umpire classification and detection, namely segmentation, feature extraction, and classification. At first, the cricket video is converted into frames, and the segmentation is done using the Viola-Jones algorithm. After the segmentation, the feature extraction is performed by extracting features, like HOG, and Fuzzy LGP. Finally, the umpire classification is carried out using the proposed BSO-Stacked Autoencoders deep learning classifier. Experimentation is carried out using a manually collected dataset. The performance of the BSO-Stacked Autoencoders is evaluated using accuracy, sensitivity, and specificity by varying the training percentage and K-Fold. The proposed method produces the maximal accuracy of 96.562%, maximal sensitivity of 91.884%, and the maximal specificity of 99%, which indicates the superiority of the proposed method. In future, the proposed BSO-Stacked Autoencoders will be further improved with a hybrid optimization approach for better results.

REFERENCES

- [1] A. RAVI, H. VENUGOPAL, S. PAUL, H. R. TIZHOOSH, *A Dataset and Preliminary Results for Umpire Pose Detection Using SVM Classification of Deep Features*, In proceedings of IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1396-1402, 2018.
- [2] A. SASITHRADEVI, S. M. M. ROOMI, AND G. MARAGATHAM, *Content Based Video Retrieval via Object Based Approach*, In proceedings of 10th IEEE conference, pp.5-8, 2017.
- [3] M. RAVINDER AND T. VENUGOPAL, *Pitch Frames Classification in a Cricket Video Using Bag-of-Visual-Words*, Artificial Intelligence and Evolutionary Computations in Engineering Systems, pp. 793-801, 2016.
- [4] M. Z. KHAN, M. A. HASSAN, AND A. FAROOQ, *Deep CNN based Data-driven Recognition of Cricket Batting Shots*, In proceedings of international conference on Applied and engineering mathematics, 2018.
- [5] S. PAUL, S. ROY AND A.K. ROY-CHOWDHURY, *W-TALC: Weakly-supervised Temporal Activity Localization and Classification*, In Proceedings of the European Conference on Computer Vision (ECCV), pp. 563-579, 2018.
- [6] M. A. RUSSO, L. KURNIANGGORO, AND K.-H. JO, *Classification of sports videos with combination of deep learning models and transfer learning*, In proceedings of International Conference on Electrical, Computer and Communication Engineering, pp.7-9, February 2019.
- [7] M.M. GOYANI, S. K. DUTTA, AND P. RAJ, *Key Frame Detection Based Semantic Event Detection and Classification Using Hierarchical Approach for Cricket Sport Video Indexing*, In proceedings of International Conference on Computer Science and Information Technology, pp. 388-397, 2011.
- [8] S. B. JAYANTH AND G. SRINIVASA, *Automated Classification of Cricket Pitch Frames in Cricket Video*, Electronic Letters on Computer Vision and Image Analysis, vol.13, no.1, pp.33-49, 2014.
- [9] M. H. KOLEKAR, S. SENGUPTA, *Semantic concept mining in cricket videos for automated highlight generation*, vol.47, pp.545-579, 2010.
- [10] M. GOYANI, S. DUTTA, G. GOHIL, AND S. NAIK, *wicket fall concept mining from cricket video using a-priori algorithm*, The International Journal of Multimedia & Its Applications (IJMA), vol.3, no.1, February 2011.
- [11] M. H. KOLEKAR AND S. SENGUPTA, *Event-importance based customized and automatic cricket highlight generation*, In proceedings of International Conference on Multimedia and expo, 2006.
- [12] P. XU, L. XIE, S. CHANG, A. DIVAKARAN, A. VETRO, AND H. SUN, *Algorithms and system for segmentation and structure analysis in soccer video*, In IEEE ICME, 2001.
- [13] N. HARIKRISHNA, S. SATHEESH, S. D. SRIRAM, AND K.S. EASWARAKUMAR, *Temporal Classification of Events in Cricket Videos*, In proceedings of National Conference on Communications (NCC), pp. 1-5, 2011.
- [14] VIJAYAKUMAR.V, AND NEDUNCHEZHIAN.R, *Event detection in cricket video based on visual and acoustic features*, vol.3, no.8, August 2012.
- [15] Y S. KUMAR, S. K. GUPTA, B R. KIRAN, K R RAMAKRISHNAN, AND C. BHATTACHARYYA, *Automatic summarization of broadcast cricket videos*, In proceedings of 15 international symposium on consumer electronics, 2011.
- [16] N. D. LAKSHMI, Y. M. LATHA, AND A. DAMODARAM, *SILHOUETTE Extraction of a human body based on fusion of hog and graph-cut segmentation in dynamic backgrounds*, 2013.
- [17] S. ABBURU, *Context Ontology Construction For Cricket Video*, International Journal on Computer Science and Engineering, vol.2, no. 8, pp.2593-2597, 2010.
- [18] C. K.MOHAN, B. YEGNANARAYANA, *Classification of sport videos using edge-based features and autoassociative neural network models*, Signal, Image and Video Processing, vol.4, no.1, pp.61-73, 2010.
- [19] A. KOKARAM, N. REA, R. DAHYOT, A. M. TEKALP, P. BOUTHEMY, P. GROS, AND I. SEZAN, *Browsing Sports Video*, IEEE signal processing magazine, March 2006.
- [20] P. CHAUDHARI, AND K. S. BHAGAT, *Video Frame Segmentation and Object Tracking using ANN for Surveillance System*, International Journal of Research in Advent Technology (IJRAT), April 2017.
- [21] K. JAYAPRIYA, N. A. B. MARY, *Employing a novel 2-gram subgroup intra pattern (2GSIP) with stacked auto encoder for membrane protein classification*, Molecular Biology Reports, February 2019.
- [22] X.-B. MENGAB, X.Z. GAOC, L. LUDE, Y. LIUB, AND H. ZHANGA, *A new bio-inspired optimisation algorithm: Bird Swarm Algorithm*, Journal of Experimental & Theoretical Artificial Intelligence, vol.28, no.4, pp.673-687, 2016.

- [23] P. VIOLA, AND M. J. JONES, *Robust Real-Time Face Detection*, International Journal of Computer Vision, vol.57, no.2, pp.137-154, 2004.
- [24] A. G. BINSAAADOON AND E.-SAYED M. EL-ALFY, *Gait-based Recognition for Human Identification using Fuzzy Local Binary Patterns*, In ICAART, vol.2, pp. 314-321, 2016.
- [25] A. RAKHLIN, A. SHVETS, V. IGLOVIKOV, AND A. A. KALININ, *Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis*, In proceedings of International Conference on Image Analysis and Recognition ICIAR, Image Analysis and Recognition, pp. 737-744, 2018.
- [26] S. ZHANG, Z. DENG, D. CHENG, M. ZONG, AND X. ZHU, *Efficient kNN Classification Algorithm for Big Data*, Neurocomputing, vol.195, pp.143-148, 2016.
- [27] CH. MAMATHA, P. B. REDDY, M.A. R. KUMAR, S. KUMAR, *Analysis of Big Data With Neural Network*, International Journal of Civil Engineering and Technology (IJCIET), vol. 8, no.12, December 2017.
- [28] C. DIAMANTINI, D. POTENA, AND E. STORTI, *Multidimensional query reformulation with measure decomposition*, Information Systems, vol. 78, pp. 23-39, November 2018.
- [29] T. CONG N. D. H. VO, P. V. NGUYEN, H. M. NGUYEN, AND A. T. VO, *Derivatives market and economic growth nexus: Policy implications for emerging markets*, North American Journal of Economics and Finance, 2019.
- [30] B. T. KHOA, H. M. NGUYEN, *The Relationship between the Perceived Mental Benefits, Online Trust, and Personal Information Disclosure in Online Shopping*, The Journal of Asian Finance, Economics and Business, vol. 6, no. 4, pp. 261-270, November 2019.
- [31] V.H. ARUL, V.G. SIVAKUMAR, R. MARIMUTHU, AND B. CHAKRABORTY, *An Approach for Speech Enhancement Using Deep Convolutional Neural Network*, Multimedia Research (MR), vol. 2, no. 1, pp. 37-44, 2019.
- [32] R. HARI, AND M. WILSCY, *Event Detection in Cricket Videos Using Intensity Projection Profile of Umpire Gestures*, In proceeding of 2014 Annual IEEE India Conference (INDICON), Pune, India, 2014.
- [33] M. M. GOYANI, S. K. DUTTA, AND P. RAJ, *Key Frame Detection Based Semantic Event Detection and Classification Using Heirarchical Approach for Cricket Sport Video Indexing*, Communications in Computer and Information Science, vol.131, pp. 388-397, 2011.
- [34] S. K. NERELLA, K. V. GADI, AND R. S. CHAGANTI, *Securing Images Using Colour Visual Cryptography and Wavelets*, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, no. 3, pp. 163-168, 2012.
- [35] R. V. BOPANA, R. CHAGANTI, AND V. VEDULA, *Analyzing the Vulnerabilities Introduced by DDoS Mitigation Techniques for Software-Defined Networks*, National Cyber Summit (NCS) Research Track, pp 169-184, 2019.

Edited by: P. Vijaya

Received: Dec 7, 2019

Accepted: Jun 23, 2020



ENHANCED DBSCAN WITH HIERARCHICAL TREE FOR WEB RULE MINING

NEELIMA GULLIPALLI* AND SIREESHA RODDA†

Abstract. Like other mining, web mining is also necessary to increase the power of web search engine to identify the intended web page and web document. While processing with large datasets, there arises several issues associated with space availability, similarity relationships between different webpage's and running time. Hence, this paper intends to develop an enhanced web mining model based on two contributions. At first, the hierarchical tree is framed, which produces different categories of the searching queries (different web pages). Next, to hierarchical tree model, enhanced Density-Based Spatial Clustering of Applications with Noise (DBSCAN) technique model is developed by modifying the traditional DBSCAN. This technique results in proper session identification from raw data. Moreover, this technique offers the optimal level of clusters necessitated for hierarchical clustering. After hierarchical clustering, the rule mining is adopted. The traditional rule mining technique is generally based on the frequency; however, this paper intends to enhance the traditional rule mining based on utility factor as the second contribution. Hence the proposed model for web rule mining is termed as Enhanced DBSCAN-based Hierarchical Tree (EDBHT). It benefits in providing the search results depending on high level information (e.g., location), so that the ability of search engine in providing the interesting association rules can be improved. Next, to the implementation, the performance of proposed EDBHT is found to be enhanced when compared over several traditional models.

Key words: Rule Mining; Hierarchical Clustering; Searching Behaviour; DBSCAN; A priori Algorithm.

AMS subject classifications. 68M11

1. Introduction. With the increase in usage of devices [35], weblog analysis software can be used to analyze the server logfile obtained from web server and depending on the standards present in the log file, insights into the manner in which pages are accessed, the user accessing the relevant webpages and the duration for which particular webpage is accessed, are gained [1]. The web server normally generates log files earlier; hence the original data is available in advance. Also, the web server consistently files each deal it makes [2]. Log files include details on visits from search engine spiders. Companies and organizations depend on the corresponding websites to communicate with their clients [3,38]. Maintaining present customers and drawing effective websites thrust such companies, associations, and foundations to come across the striking way to create their websites helpful and capable [4,34,37]. To attain this objective, several reviewing efforts have to be made [5]. These tasks can be done in two random modes. Clients of a particular website could be sought to assess their practice of browsing [5,6]. Subsequently, the performance will be engaged in progressing the construction and/or content depending on the response that is arrived to offer a feedback [2,7,36]. The involuntary navigational account recorded by clients' is also checked up consequently [1,8].

Web Mining [9,10] can be categorized into three diverse categories, based on the types of data to be mined. They include web content mining, web usage mining and web structure mining [11,12]. A lot of web testing equipment subsist however they were restricted, and the effectiveness of such equipments is in a state of excellence [13,14]. Several data mining algorithms have been successfully employed to weblog analysis to better understanding of the user behaviour. Clustering and classification are proving helpful with such issues [3,15,16]. There exist a lot of schemes for producing association rules. A priori algorithm [17] is a well-known and significant approach to discover association rules [5,18]. Improvements obtained from the research that are held over years are integrated to existing web systems for acquiring more successful suggestions.

In addition, data mining approaches have been helpful to deal with sparsity and presentation issues as they were not only dependent on invention assessments but also on various other attributes [19,20]. Hence,

*Associate Professor Vignana's Institute of Information Technology (VIIT), Visakhapatnam (gullipalli.neelima@gmail.com).

†Professor GITAM University, Visakhapatnam

it is essential to discover such algorithms that are considerably perceptive to sparsity for obtaining accurate recommendations [6,21,22]. The algorithm scrutinizes the training dataset once more to construct a frequent pattern tree (FP-tree) [6,23,24]. These trees are well-organized data structures to offer linear time solutions to risky problems in string [2]. A tree for all the data record can be obtained by adding a non-natural parent node in addition to nodes of data record [1,25,26,27]. Decision tree can also provide solution to complex issues in a string [28]. A decision tree is a decision support tool that adopts a tree-like design or graph of decisions and their possible effects, including utility, resource costs, and chance event outcomes [29,30]. Though several methods exist for supporting web mining, still need to be properly addressed.

This paper contributes an improved web mining scheme depending on two contributions. Initially, the hierarchical tree is framed that generates various categories of searching queries. Then, a relevant model is implemented by modifying the conventional clustering method known as DBSCAN that can also be referred as enhanced DBSCAN technique. This model results in appropriate session identification from raw data. From the DBSCAN, the optimal level of clusters can be obtained, which is then provided to hierarchical model. Subsequent to the formation of clusters from hierarchical tree model, the A priori algorithm is adopted from which the interesting association rules can be obtained. The paper is organized as follows. Section II analyses the related works and reviews done on this topic. In addition, section III describes the newly adopted web rule mining model and section IV explains Enhanced DBSCAN-based Hierarchical Tree. Section V confirms the results. At last, section VI concludes the paper.

2. Literature Survey.

2.1. Related works. Sheng sheng Shi et al. [1] have presented a paper based on a novel scheme said to be AutoRM that mines data accounts from a Web page without human intervention. This includes three steps, namely, building the DOM tree of the specified web page, mining the entire groups of neighboring Candidate data Records (C-Records) from the DOM tree, drawing out real data accounts from C-Records. In several pages of web, analogous data accounts are dispersed in larger and neighboring objects that are similar. The results obtained from experiments show that AutoRM is very efficient, and outperforms conventional approaches.

Abdelghani Guerbas et al. [2] has presented a novel algorithm called DBSCAN algorithm to enhance the weblog mining procedure and online navigational pattern assumption. The procedure involves three diverse mechanisms, (1) implementing an advanced time-out dependent heuristic for session detection. (2) Suggestion of using a particular density-dependent technique for navigational pattern discovery. (3) At last, an innovative mechanism for well-organized online assumption was provided. The performed analysis reveals the significance and usefulness of the presented method.

V. Sujatha et al. [3] presented a novel saliency detection algorithm called Patterns Using Clustering and Classification (PUCC). which offers the assumption from web log data. In the initial step PUCC concerns on partitioning the probable clients in weblog data, and in the subsequent step, clustering is adopted to assemble the effective users with equal attention, and in the third step, the outcomes of categorization and clustering are employed to guess the prospect of requests from users. The analyzed outcomes signified that this mechanism could produce a better quality of clustering for customer navigation pattern. Rui Liu [4] has presented web-video-mining-supported surgical workflow modeling (webSWM) that offers workflow model with a little cost and labor efficient methods. Also, a testing method for determining the quality of video depending upon study and sentiment investigation methods is produced to choose videos with high-quality from large and noisy videos in the web. Moreover, a numerical learning technique is adopted to construct the workflow representation depending on certain videos. Better outcomes confirmed the exactness of the produced SWM and SWM-associated skills.

Mingxing Wu et al. [5] suggested an approach called A priori algorithm that depends on web mining to study the product usability process. This method adopts the enormous online consumer reviews on similar products and characteristics as data basis that is simple to obtain from web and can reproduce the most efficient client opinions on the usage of products. Association rule mining methods are utilized to take out the opinions of consumers about the use of product and its characteristics. Moreover, it is employed for mining organization rules, depending on which a usability valuation technique is offered. Maria N. Moreno et al. [6] presented a paper on MovieLens database that proposes a total structure to agree with certain significant sparsity, scalability, first rater and cold start issues. Even though the structure is represented to movies' suggestion and considered

in the corresponding framework, it can be simply comprehended in various domains. It maintains diverse assumption designs for constructing recommendations based on specific situations. Such designs are induced by data mining approaches that are used as data for input in both product and consumer attributes ordered based on specific domain ontology. Experimental results on a various multispectral data sets and an evaluation with other methods exhibit the strength of the proposed method in identifying objects.

Ziang Li et al. [7] has presented effective Support Vector Regressions (SVR) architecture which is underpinned by a domain ontology that obtains joblessness associated concepts and their dealings to assist the mining of helpful assumption characteristics from related search engine queries. Further, conventional feature selection techniques and data mining designs like neural networks (NN) and SVR are utilized to improve the efficiency of unemployment rate assumption. The results obtained from experiments demonstrate that the presented method performs better than other techniques that were deployed for unemployment rate prediction.

Dirk Thorleuchter et al. [8] presented a paper based on Latent Semantic Indexing (LSI) algorithm that focuses on automated recognition of weak signals. This enhances prevailing knowledge dependent methods, as LSI considers the aspects of meaning and thus, related textual patterns in diverse contexts can be identified. A new weak signal maximization approach is established that replaces the usually adopted prediction modeling in LSI. It measures the numerous related weak signals that are addressed in singular value decomposition (SVD) dimensions. Thus, it is revealed that the proposed method enables organizations to recognize weak signals from the internet for a known suggestion.

2.2. Review. Table 2.1 shows the methods, features, and challenges of conventional techniques based on web block data mining. At first, Auto RM algorithm exhibits certain unique properties that require only one Web page. Further, there is no any necessity for vertically distributed data records, but it requires a area containing two data records and also it could not obtain and align data items between unique data records [1]. Moreover, DBSCAN algorithm is presented where the log file can be accessed without knowing the user's identity. The lacking factor in this algorithm is that it is highly complex and there is no any consideration of combined density-based clustering [2]. PUC algorithm separates the potential users in weblog data, and it also improves the quality of clustering for user navigation pattern. However, there is no analyzation of performance efficiency, and there are no suggestions in the direction of exploiting association rules for prediction engine [3]. Furthermore, webSWM algorithm solves the knowledge scalability issue in surgical workflow design and selects videos from enormous, noisy web videos. Here, there is no consideration of computer-vision-based scheme to fragment the surgical video and moreover, there is no implementation of produced SWM knowledge on the prediction of phase which is said to be a challenge in this paper [4]. On the contrary, A priori algorithm helps modelers to analyze the usability of product features and also it provides decision supports to alleviate FF in product development. However, there are limitations for some unpopular products in the related online reviews, and this method is computationally expensive [5]. In addition, MovieLens database possesses the ability to handle various predictive models for generating suggestions depending on particular circumstances, and it can be extended to various other domains, anyhow, it is more complex due to the application of several procedures [6]. SVR is much suited for obtaining the lowest average RMSE and MAE, and further, it achieves the best prediction performance. However, it should be combined with wrapper-based forward selection for better achievement [7]. LSI algorithm enables an organization to identify weak signals and helps strategic planners to react ahead of time, yet, the occurrence of new weak signals is not visible, and there are no any applications indicating parameter selection procedure [8]. Thus the challenges in various techniques enforce to improve the web mining more effectively in the current work.

3. Newly Contributed Web Rule Mining Model.

3.1. Proposed architecture. The overall architecture of the proposed model is given by Fig. 3.1. for enhancing the ability of search engine in offering the interesting association rules. Initially, the raw data is processed using four processing steps which include data cleaning, user identification, session identification and path completion. Data cleaning is the process of eliminating the irrelevant items from the log file. After cleaning, the cleaned data is forwarded for user identification; there how many users visited the website is identified with the help of IP address. Next to user identification, session identification process takes place. Session is the time between logged in and logged out. This activity is to find the sequence of pages and trace

TABLE 2.1
Review on state-of-the-art of web rule mining techniques

Author [citation]	Adopted methodology	Features	Challenges
Shengsheng Shi et al. [1]	AutoRM algorithm	Requires single Web page as input No need for vertically distributed data records	Requires at least two data records No extraction of data items among unique data records
Abdelghani Guerbas et al. [2]	DBSCAN	Log file can be accessed No necessity to know about user's identity	Highly complex No consideration of combined density-based clustering
V. Sujatha et al. [3]	PUCC	Separates the potential users in weblog data Improves the quality of clustering	No analyzation of effective computation. No investigation in the direction of prediction engine
Rui Liu et al. [4]	webSWM	Resolves the scalability crisis in surgical workflowdesign Chooses high-quality videos	No consideration infragmenting the surgical video No implementation of phase predicted SWM knowledge on
Mingxing Wu et al. [5]	A priori algorithm	Helps researchers to investigate the usability of product characteristics Provides decision supports to improve FF	Limitation for certain unpopular products in the related online reviews Computationally expensive
Maria N.Morenoet al.[6]	MovieLens data base	Extends over various domains easily. Handles various predictive models for on specific situations	Highly complex
Ziang Li et al. [7]	SVR	Achieves the lowest average RMSE and MAE Achieves the best prediction performance	It should be combined with wrapper-based forward selection for better achievement
Dirk Thorleuchteret al. [8]	LSI	Enables an organization to identify weak signals Assists deliberated planners to respond ahead of time	No application of parameter selection procedure. The occurrence of new weak signals is not visible

the user activity. This is because the user may visit many pages at a time. Moreover, the final phase of pre-processing is path completion. The path completion is the process of discovering the users travel pattern. After pre-processing, the processed data is clustered using hierarchical clustering model. In hierarchical clustering, the item sets are first identified and then in order In order to decide which clusters should be combined or where a cluster should be split, a measure of dissimilarity between sets of observations is required. This is achieved by measuring the distance between pairs of observations (Refer: https://en.wikipedia.org/wiki/Hierarchical_clustering). The distance matrix is computed followed by the computation of mean and standard deviation. Through this model, the number of clusters has to be determined optimally, for which DBSCAN algorithm is adopted. This is because DBSCAN has the property of grouping together the points that are close to each other based on a distance measurement. In DBSCAN, the cluster computation is performed based on the density reachable points with respect to equilibrium point and minimum points (min points) that results in proving maximum possible clusters. After obtaining the accurate number of clusters from DBSCAN, it is given to hierarchical clustering, and the clustered web pages are obtained with associated level of clustering. Hence, the searching of web pages in each area can became to know. Further, Apriori algorithms help in determining the searching behavior on each web page in each area.

3.2. Notations. The dataset comprises of fields and records, indicated by D_{ij} , where d_{ij} , indicates the data of i^{th} , and j^{th} , elements. The record i can be referred as $i = 1, 2, \dots, N_r$, in which N_r is the number of records and the fields are denoted by j , that can be referred as $j = 1, 2, \dots, N_f$, where N_f is the total number of fields. The data of each field belongs to each record i , i.e. $d_j \forall i, I \in A$, includes a set of attributes denoted by A , that can be represented as a_1, a_2, \dots, a_n , where n is the total set of attributes. The spatial information and searching behavior of people in various locations can be obtained as given by the following sections.

4. Enhanced DBSCAN-Based Hierarchical Tree.

4.1. Location Centric Equilibrium Point Estimation. Determining the equilibrium points ϵ is a non-trivial issue even in the nonexistence of uncertainty as it amounts to resolving a system of nonlinear equations.

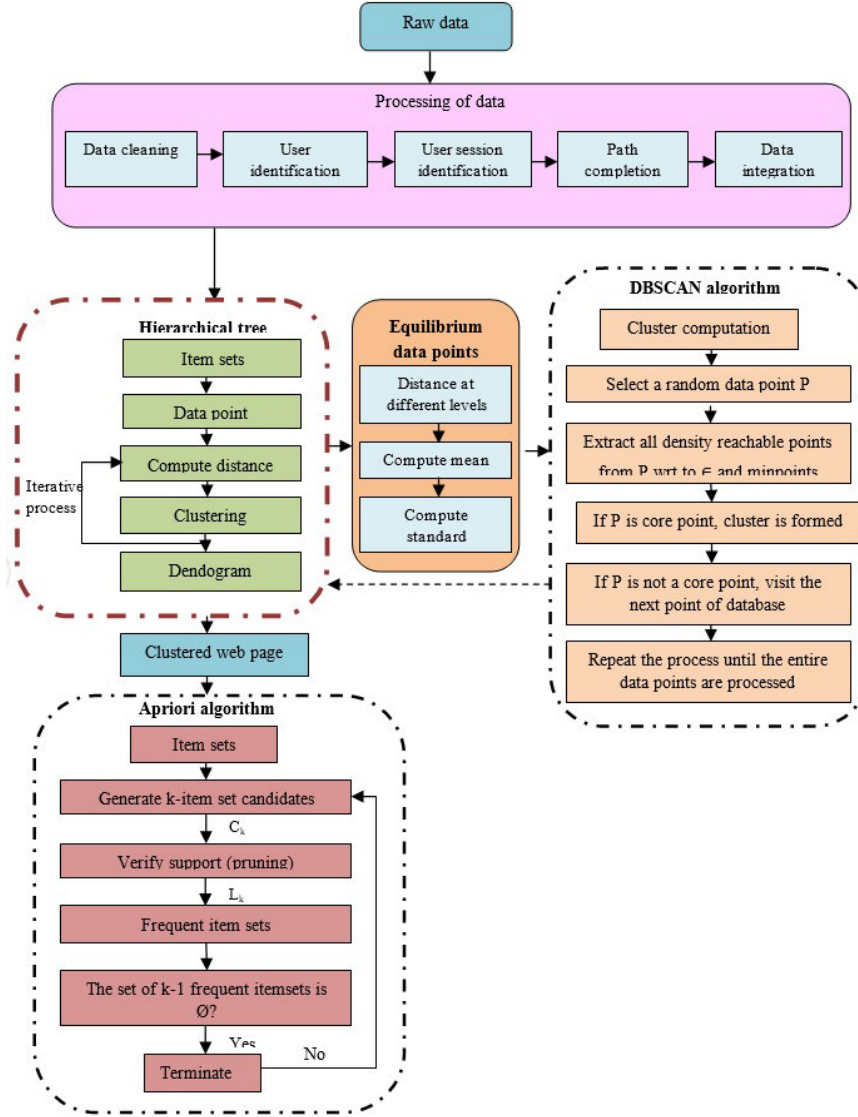


FIG. 3.1. Overall architecture of the proposed Web rule mining model

Moreover, the (ϵ) estimation offers the location-based information. In this research work, web transaction, and the IP address is necessary for performing the clustering process. Web transaction is nothing but the web pages from which the corresponding IP address visited.

For example, consider 5 levels with IP address and corresponding web pages as given by Table 4.1. A user from this IP address visited web page 1, web page 2 and so on.

The main steps of hierarchical tree are given below.

Step 1: Distance matrix computation. Let i indicates the records and j denotes the fields in the dataset. Here, N_w indicates the total number of web pages. The levels of clustering are denoted as K , where $K = 1, 2, \dots, N_L$, where N_L indicates the total number of levels.

The distance matrix for K^{th} level is expressed in Eq. (4.1), where d_{ij}^k represents the distance between i^{th} and j^{th} web pages at k^{th} level.

$$DI_k = d_{ij}^{(k)} \quad (4.1)$$

TABLE 4.1
Various Levels Of Web Transactions

Sl.no	IP address	Web transactions
1	199.55.6.7	1 2
2	199.45.4.5	1 2 3
3	199.48.5.8	1 4 5
4	199.36.9.8	1 2
5	199.57.5.6	1 3

TABLE 4.2
Model of Web Transaction for Distance Computation

Sl.no	IP address	Web transactions
1	1	1 2
2	2	1 3
3	3	1 3
4	6	4 2
5	7	3 4

$$d_{ij} = 0 \quad \text{if } i = j \tag{4.2}$$

The number of levels may increase or decrease based on number of web pages, which is given in Eq. (4.3).

$$N_L = N_w - 1 \tag{4.3}$$

Moreover, the number of clusters C_i formed at K^{th} level is also based on number of web pages, as given by Eq. (4.4).

$$N_c^{(k)} = N_w - k \tag{4.4}$$

$$DI_k = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1N_w} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2N_w} \\ d_{31} & d_{32} & d_{33} & \dots & d_{3N_w} \\ d_{41} & d_{42} & d_{43} & \dots & d_{4N_w} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{N_w 1} & d_{N_w 2} & d_{N_w 3} & \dots & d_{N_w N_w} \end{bmatrix} \tag{4.5}$$

There is a procedure for determining the distance between the IP addresses of various web pages. Let A be the set of web pages, which is searched by different IP addresses. Similarly, B be another set, which is searched by different IP addresses. If the similar web pages are detected in both A and B , i.e. $A \cap B$, then the similar data in both the web transactions have to be discarded, as given by Eq. (4.6) and Eq. (4.7).

$$\bar{A} = A - (A \cap B) \tag{4.6}$$

$$\bar{B} = B - (A \cap B) \tag{4.7}$$

An example for finding the distance between the web pages for various IP addresses is given by Table. 4.1.

Let us assume the web transactions 1 and 2, which can be determined as given in the subsequent steps. For example, let us consider the web address 199.55.10.8 as 1, 199.16.1.0 as 2, 199.55.6.7 as 3, 201.50.60.1 as 6 and 199.55.6.7 as 7. From Table 4.1, the web page transactions from page 1 is (1, 2, 3). Similarly, the web page transaction from web page 2 is (1, 6). The common IP address, found in both the transactions can be eliminated. Thus, the unique address selection for web page 1 can be represented as (2, 3) and the unique

Levels	1	2	3	4	5
1	0	1	0	2.23	2.64
2	1	0	1	3.46	0
3	2	1	0	2.64	3
4	2.23	3.46	2.64	0	2
5	2.64	0	3	2	0

FIG. 4.1. Representation of distance matrix

address selection for web page 2 can be indicated as 6. Consider 2 as 199.16.1.0 and 3 as 199.55.6.7 and 6 as 201.50.60.1. On considering the Euclidean distance between (2, 6), and (3,6) the Eq. (4.8) can be obtained.

$$T = \sqrt{(199-201)^2+(16-50)^2+(6-60)^2+(0-1)^2}, \quad T = \sqrt{(199-201)^2+(55-50)^2+(6-60)^2+(0-1)^2} \quad (4.8)$$

Thus the final distance T_d can be evaluated by dividing the distance between (2, 6) and (3, 6) by the average distance as given by Eq. (4.9):

$$T_d = \frac{T(2,6)}{T_A} = \frac{68.13}{61.364}, \quad T_d = \frac{T(3,6)}{T_A} = \frac{54.598}{61.364} \quad (4.9)$$

On assuming 5*5 levels of data points, for the given web pages, the distance matrix can be computed as given by Fig. 4.1.

Step 2: Computation of mean and standard deviation. The mean $\bar{\mu}$ is computed for every field j and is repeated for different levels. Hence at the end, $\mu_1\mu_2\mu_3\mu_{NL}$ will be calculated, where μ_1 is the mean of the first level, μ_2 is the mean of the second level and μ_{NL} is the mean of the last level. The overall mean of the entire levels can be computed as given by Eq. (4.10).

$$\bar{\mu} = \frac{1}{N_L} \sum_{k=1}^{N_L} \mu_k \quad (4.10)$$

Similarly, the standard deviation is evaluated for every field j , and then the standard deviation is computed. This overall standard deviation can be determined as given by Eq. (4.11).

$$\bar{\delta} = \sqrt{\frac{\sum_{k=1}^{N_L} (\mu_k - \bar{\mu})^2}{N_L - 1}} \quad (4.11)$$

The computation of mean and standard deviation for the various levels is given by Fig. 4.2. From the Fig. 4.2, the computation of mean is obtained for all the five levels. Fig. 4.2.(a) gives the value of μ_1 as 1.52, similarly, Fig. 4.2.(b) gives the value of μ_2 as 1.41, also, from Fig. 4.2.(c) the mean value of μ_3 can be obtained as 0.33 and from Fig. 4.2.(d), the mean value of μ_4 can be attained as 0.5. On taking the average of all the mean values, $\bar{\mu} = 0.75$ can be achieved. Moreover, the standard deviation can be formulated as given by Eq. (4.11). Accordingly, the standard deviation $\bar{\delta}$ can be obtained as 0.38.

Step 3: Computation of Equilibrium points. The equilibrium point is evaluated further using the overall mean and overall standard deviation. In the proposed architecture, the value of ϵ is based on the hierarchical clustering model, which can be evaluated by summation of $\bar{\mu}$ and $\bar{\delta}$ as shown by Eq. (4.12).

$$\epsilon = \bar{\mu} + \bar{\delta} \quad (4.12)$$

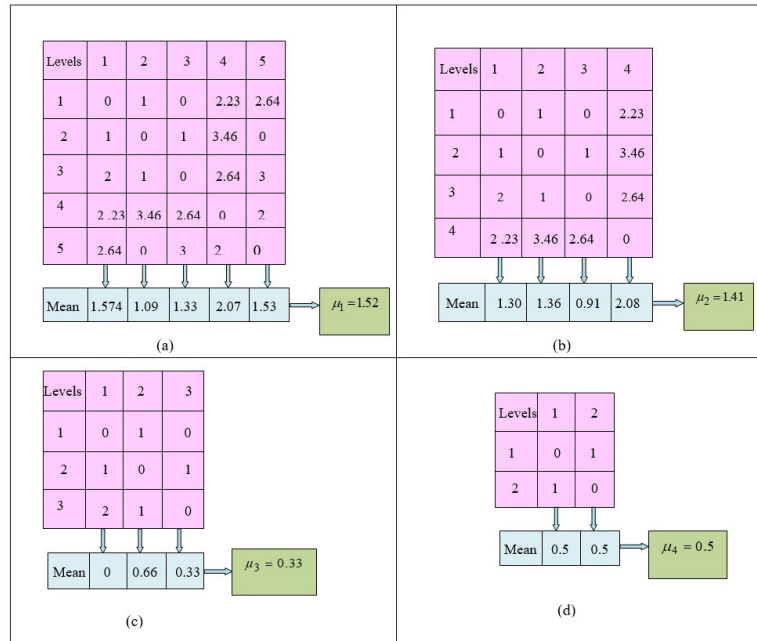


FIG. 4.2. Representation of mean computation in hierarchical clustering for (a) First low level (b) second low level (c) Third level (d) Fourth level

Let us consider the field and records as 5×5 where the distance for the entire fields are represented by d_{ij} . Moreover, the mean of each level is given by μ_{C1} to μ_{C5} . Accordingly, the overall mean for all the levels are evaluated and considered as $\bar{\mu}=0.75$. Similarly, the standard deviation of the entire levels are calculated and regarded as $\bar{\delta}=0.38$. Now, the ϵ evaluates to 1.13 based on Eq. (4.12).

4.2. DBSCAN Model. Using DBSCAN, the optimal number of clusters for the entire levels can be obtained, by which hierarchical clustering takes place. DBSCAN algorithm describes the cluster as an area of densely linked points partitioned by areas of non-dense points. If the equality evaluation is considered as Euclidean distance, the region is a hypersphere of radius at the specified point as center denoted by p .

ϵ -neighbourhood: For a point, the ϵ - neighborhood indicates the group of points, where the distance from $x \leq \epsilon$. The cardinality of ϵ -neighbourhood explains the threshold density of x .

ϵ -connected: On considering a pair of points, if $x, y \in D$, if $\|x-y\| \leq \epsilon$, then x, y are ϵ -connected points.

In DBSCAN technique, all the points in the dataset would rely on either border point or core point. Moreover, a border point may be density connected point or noise point.

Core point: At this point, the condition threshold density $\geq \text{min pts}$ is followed.

Border point: At this point, threshold density min pts is followed.

Noise point: At this point, p is a noise point if the threshold density (p) $< \text{min pts}$ and the entire points in the ϵ -neighbourhood of p are border points.

Density-connected point: It is also a border point with a minimum of one core point in its ϵ -neighbourhood.

The algorithm for DBSCAN is in Algorithm 1.

4.3. Extracting dendrogram. Based on the number of clusters obtained from the DBSCAN technique, the dendrogram is designed from which the desired level can be attained. E.g., consider the number of clusters as 3 that is obtained from DBSCAN. Therefore, the dendrogram can be modeled as given by Fig. 4.1, where the level 4 includes one cluster, level 3 includes two clusters, level 2 includes three clusters, and level 1 includes four clusters. From the Fig. 4.1, for $k = 3$, the cluster 1 consists of web pages 1 and 2. Similarly, cluster 2 includes web pages 3 and 4. Also, cluster 3 comprises of web page 5. Therefore, it can be concluded that web page combinations of (1, 2) are visited by users of similar or nearby locations. Also, web page combination of

Algorithm 2: DBSCAN Technique

```

1 Mark all patterns in  $D$  as unvisited
2 Cluid  $\leftarrow 1$ 
3 for each unvisited pattern  $x$  in  $D$  do
4    $z \leftarrow$  Discover neighbors ( $x, \epsilon, \text{min pts}$ )
5   if  $|z| < \text{min pts}$  then
6     Point out  $x$  as noise
7   else
8     Point out  $x$  and every pattern of  $z$  with cluid
9     Queue-list  $\leftarrow$  every unvisited patterns of  $z$ 
10    repeat
11       $y \leftarrow$  Remove a pattern from Queue-list
12       $z \leftarrow$  Discover neighbors ( $x, \epsilon, \text{min pts}$ )
13      if  $|z| \geq \text{min pts}$  then
14        for all pattern do
15          if  $w$  in  $z$  then
16            Point out  $w$  with cluid
17          if  $w$  is unvisited then
18            Queue-list  $\leftarrow \cup$  Queue-list
19      Point out  $y$  as visited end until
20    until Until Queue-list is empty;
21    Point out  $x$  as visited cluid  $\leftarrow +1$ 
22 Output: every patterns in  $D$  pointed with cluid or noise

```

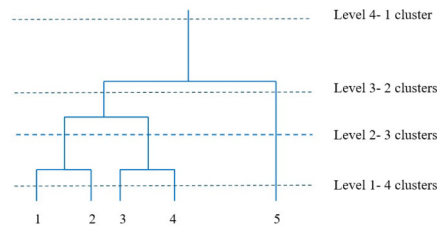


FIG. 4.3. Representation of dendrograms in Hierarchical clustering

(3, 4) is visited by users of similar or nearby locations and web page 5 is visited by users of similar or nearby locations. Fig. 4.3 shows the representation of dendrogram in hierarchical clustering model.

4.4. A priori algorithm. The spatial distribution of various web pages can be obtained using A priori algorithm. Association rule generation is generally partitioned into two phases [32]. The initial one is the minimum support, which is deployed to discover the entire frequent item sets contained in a database. The next one is the frequent itemsets and the minimum confidence constraint that are exploited to form rules.

The initial phase requires more consideration, but the second phase is straightforward. Discovering the entire frequent itemsets in a database is complex as it entails exploring the entire feasible combinations of items. The set of probable itemsets is the power set over and it includes a size of $2^n - 1$. Even though the size of the power set enlarges gradually in the count of items n in i , proficient search is made feasible by means of the downward-closure characteristics of support that assures that, for a frequent itemset, all the corresponding subsets are frequent and similarly, for an infrequent item set, its relative super sets should be infrequent. By deploying this feature, proficient techniques could be able to detect the entire frequent item sets. The pseudo

code of A priori algorithm is shown in Algorithm 2, which can easily determine the searching behavior of clustered web pages.

Algorithm 3: A priori algorithm

Data: D is the database and \min is the minimum support

- 1 Procedure A priori (D , \min support)
- 2 $L_1 =$ frequent items
- 3 **for** $k=2; L_k - 1! = \Theta; k++$ **do**
- 4 $C_k =$ candidates generated from L_{k-1}
- 5 // that is Cartesian product $L_{k-1} * L_{k-1}$ and $k - 1$ size item set that is not frequent
- 6 **for every transaction** t **in the database do**
- 7 #increment the count of the entire candidates in C_k that are available in t
- 8 $L_k =$ candidates in C_k with \min support
- 9 Return $U_k L_k$;

4.5. Proposed EDBHT Algorithm. The proposed web rule mining algorithm is shown in Algorithm 3.

Algorithm 4: Proposed EDBHT model

Data: Web data D
Result: Web rules $R_i, i = 1, 2, \dots, N_c$

- 1 **for every level in Hierarchical tree** T_H **do**
- 2 Determine d_{ij}
- 3 Calculate μ_k
- 4 Calculate $\bar{\mu}$ and $\bar{\delta}$ from μ using Eq. (4.10) and Eq. (4.11), respectively
- 5 Estimate ε using Eq. (4.12)
- 6 Construct T_H
- 7 $N_L \leftarrow$ DBSCAN(ε , Min point)
- 8 Extract cluster from $T(K) : k = N_L$
- 9 **for every cluster do**
- 10 $R_i \leftarrow$ apriori(C_i)
- 11 Return R_i

5. Results And Discussion.

5.1. Simulation Procedure. The proposed scheme was implemented in JAVA, and the results were obtained. The proposed scheme was executed based on the NASA weblog dataset [33]. The dataset includes two traces that comprises of two month's worth of all HTTP requests to the NASA Kennedy Space Center WWW server in Florida. The proposed EDBHT model was compared with conventional algorithms such as DBHT 1 ($\epsilon=19$), DBHT 2 ($\epsilon=21$), DBHT 1 ($\epsilon=25$) and traditional algorithm and the outcomes were analyzed.

5.2. Spatial Index Analysis. The implemented scheme for extracting the interesting association rules was analyzed based on spatial index analysis. Accordingly, from Table 5.2, for cluster 1, when the support value is 0.05, the proposed model is 2.65% better than DBHT 1, 0.61% better than DBHT 2, and 10.41% better than DBHT 3 techniques. Similarly, from Fig. 5.1.(a), for cluster 1, it can be observed that, when the support value is 0.2, the implemented design is 2.65% superior to DBHT 1, 0.61% superior to DBHT 2, and 10.41% superior to DBHT 3 methods. Also from Fig. 5.1.(b), when the support value is 0.4, the suggested method is 2.65% better than DBHT 1, 0.61% better than DBHT 2 and, 10.41% better than DBHT 3 algorithms.

TABLE 5.1
Spatial Index: Computation Of Proposed Over Conventional Web Rule Mining With Respect To Support Value At 0.05

Support value 1=0.05			
Method	Cluster 1	cluster 2	cluster 3
Method	Cluster 1	cluster 2	cluster 3
Proposed	364.658	364.259	365.5082
DBHT 1	374.3258	373.8744	375.1143
DBHT 2	366.9055	370.9248	372.1198
DBHT 3	402.6452	385.5523	381.4983
Traditional	288.9584	-	-

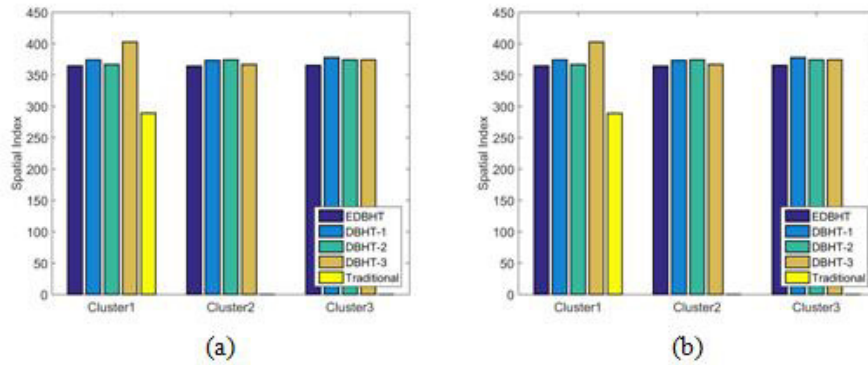


FIG. 5.1. Spatial index analysis for proposed over conventional rule mining with respect to support value at (a) 0.2 (b) 0.4

TABLE 5.2
Spatial index computation of proposed over conventional web rule mining with respect to confidence value at 0.5

Confidence value 1=0.5			
Method	Cluster 1	cluster 2	cluster 3
Method	Cluster 1	cluster 2	cluster 3
Proposed	364.658	364.259	365.5082
DBHT 1	374.3258	373.4833	378.0071
DBHT 2	366.9055	374.3258	374.3258
DBHT 3	402.6452	366.9055	374.3258
Traditional	288.7682	-	-

In addition, from Table 5.2, for cluster 1, when the confidence value is at 0.5, the presented model is 2.65% superior to DBHT 1, 0.61% superior to DBHT 2 and, 10.41% superior to DBHT 3 methods. Moreover, from Fig. 5.2.(a), when the confidence value is set at 0.3, the proposed design is 2.65% better than DBHT 1, 0.61% better than DBHT 2, and 10.41% better than DBHT 3 methods. Also, from Fig. 5.2.(b), when the confidence value is fixed at 0.1, the implemented mechanism is 2.65% superior to DBHT 1, 0.61% superior to DBHT 2, and 10.41% superior to DBHT 3 algorithms. The traditional technique was seemed to attain better results in some cases, but the process of clustering was not adopted in the conventional method, and hence the obtained results are not reliable. Thus the superiority of the implemented scheme was verified successfully.

5.3. Frequency rule analysis. The suggested EDBHT method for rule mining was analyzed in terms of the frequency rule analysis for database. From Table 5.3, the when the value of support value is 0.05, the implemented model is 46% superior to DBHT 1, 12% superior to DBHT 2, 56.7% superior to DBHT 3 techniques. Similarly, from Fig. 5.3.(a), it can be observed that, when the support value is 0.2, the implemented design is 46% better than DBHT 1, 12% better than DBHT 2 and, 56.7% better than DBHT 3 methods. Also from Fig. 5.3.(b), when the support value is 0.4, the suggested method is 46% superior to DBHT 1, 12% superior to DBHT 2, 56.7% better than DBHT 3 algorithms.

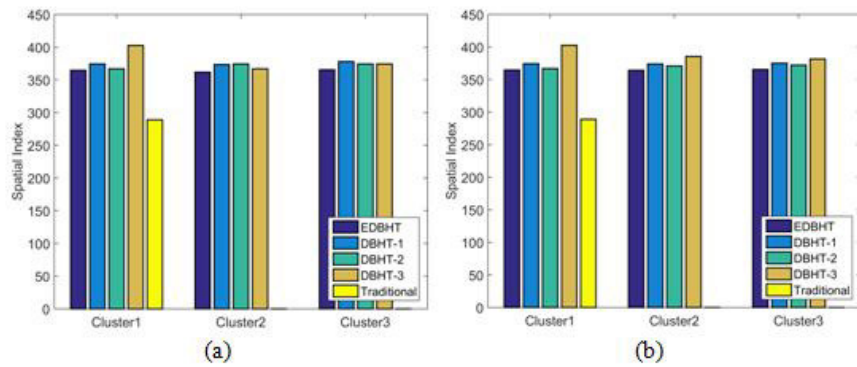


FIG. 5.2. Spatial index analysis for proposed over conventional rule mining with respect to confidence value at (a) 0.3 (b) 0.1

TABLE 5.3

Frequency rule computation of proposed over conventional web rule mining with respect to support value at 0.05

Support value 1=0.05			
Method	Cluster 1	cluster 2	cluster 3
Proposed	0.003	0.001	0.0014
DBHT 1	0.004385	0.004143	0.004525
DBHT 2	0.003364	0.003917	0.004081
DBHT 3	0.004667	0.004043	0.004168
Traditional	0.126708	-	-

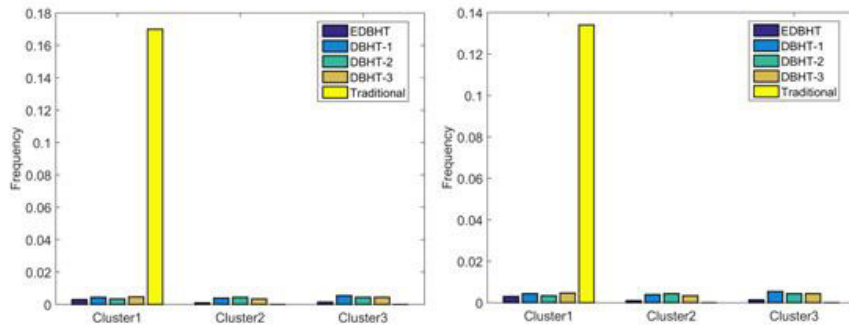


FIG. 5.3. Frequency rule analysis for proposed over conventional rule mining with respect to confidence value at (a) 0.2 (b) 0.4

Moreover, from Table 5.4, when the confidence value is at 0.5, the presented model is 46% better than DBHT 1, 12% better than DBHT 2, 56.7% better than DBHT 3 approaches. Moreover, from Fig. 5.4.(a), when the confidence value is set at 0.6, the proposed design is 46% superior to DBHT 1, 012% superior to DBHT 2, 56.7% superior to DBHT 3 techniques. Also, from Fig. 5.4.(b), when the confidence value is fixed at 0.7, the implemented mechanism is 46% better than DBHT 1, and 12% better than DBHT 2, and 56.7% better than DBHT 3 schemes. The compared traditional scheme is seemed to be better, but as clustering technique was not adopted in traditional scheme, the result could not be considered reliable.

6. Conclusions. This paper has presented an enhanced web mining method on the basis of two contributions. Primarily, the hierarchical tree was developed that produced several categories of searching queries. Subsequently, the enhanced DBSCAN technique was adopted from which the required levels of clusters can be attained. Moreover, this scheme results in appropriate session identification from raw data. The level of clusters obtained from DBSCAN was again given to the hierarchical tree model. Following the formation of

TABLE 5.4
 Frequency Rule Computation Of Proposed Over Conventional Web Rule Mining With Respect To Confidence Value At 0.5

Confident value 1=0.5			
Method	Cluster 1	cluster 2	cluster 3
Method	Cluster 1	cluster 2	cluster 3
Proposed	0.003	0.001	0.0014
DBHT 1	0.004385	0.003933	0.005417
DBHT 2	0.003364	0.004385	0.004385
DBHT 3	0.004667	0.003364	0.004385
Traditional	0.141667	-	-

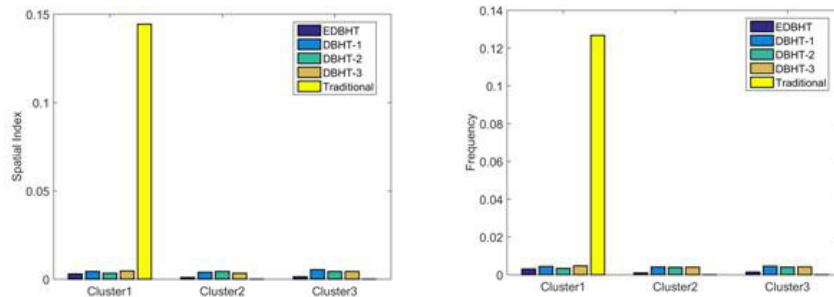


FIG. 5.4. Frequency rule analysis for proposed over conventional rule mining with respect to confidence value at (a) 0.6 (b) 0.7

hierarchical tree structure, the enhanced rule mining model was exploited from which the interesting association rules can be attained. Moreover, the proposed technique was compared with the conventional algorithms like DBHT 1, DBHT 2, DBHT 3 and traditional schemes and the results were obtained. From the spatial index analysis, when the support value is 0.05, the suggested model was 2.65% better than DBHT 1, 0.61% better than DBHT 2 and 10.41% better than DBHT 3 techniques. Thus the superiority of the implemented scheme was proved successfully over traditional web rule mining models.

REFERENCES

- [1] S. SHI, C.LIU, Y. SHEN, C.YUAN, Y. HUANG, *AutoRM: An effective approach for automatic Web data record mining*, Knowledge-Based Systems, vol. 89, pp. 314-331, November 2015
- [2] A. GUERBAS, O. ADDAM, O. ZAAROUR, M. NAGI, R.ALHAJJ, *Effective web log mining and online navigational pattern prediction*, Knowledge-Based Systems, vol.49, pp. 50-62, September 2013.
- [3] V. SUJATHA, PUNITHAVALLI, *Improved user Navigation Pattern Prediction Technique from Web Log Data*, Procedia Engineering vol.30, pp. 92-99, 2012
- [4] R. LIU, X. ZHANG, H. ZHANG, *Web-video-mining-supported workflow modeling for laparoscopic surgeries*, Artificial Intelligence in Medicine vol. 74, pp. 9-20, November 2016.
- [5] M. WU, L. WANG, M. LI, H. LONG, *An approach of product usability evaluation based on Web mining in feature fatigue analysis*, Computers & Industrial Engineering, vol. 75, pp. 230-238, September 2014
- [6] M.N. MORENO, S. SEGRERA, V. F. LÓPEZ, M. D. MUÑOZ, Á.L.SÁNCHEZ., *Web mining based framework for solving usual problems in recommender systems. A case study for movies recommendation*, Neurocomputing, vol. 176, pp. 72-80, 2 February 2016
- [7] Z. LI, W. XU, L. ZHANG, R.Y.K. LAU, *An ontology-based Web mining method for unemployment rate prediction*, Decision Support Systems, vol. 66, pp. 114-122, October 2014.
- [8] D.THORLEUCHTER, D. V. DEN POEL, *Weak signal identification with semantic web mining*, Expert Systems with Applications, vol. 40, no. 12, pp. 4978-4985, 15 September 2013.
- [9] C. ZAÍN, M. PRATAMA, E. LUGHOFFER, S. G. ANAVATTI, *Evolving type-2 web news mining*, Applied Soft Computing, Vol. 54, pp. 200-220, May 2017.
- [10] A.B. GAIA DO COUTO, L. F. A. MONTEIRO GOMES, *Multi-criteria Web Mining with DRSA*, Procedia Computer Science, vol. 91, pp. 131-140, 2016.
- [11] V. MEDVEDEV, O. KURASOVA, J. BERNATAVIČIENĖ, P.TREIGYS, G.DZEMYDA, *A new web-based solution for modelling data mining processes*, Simulation Modelling Practice and Theory, vol.76, pp. 34-46, August 2017.

- [12] J. A. IGLESIAS, A. TIEMBLO, A. LEDEZMA, A. SANCHIS, *Web news mining in an evolving framework*, Information Fusion, vol. 28, pp. 90-98, March 2016.
- [13] E. HABIBI, S.H. M. HOSSEINABADI, *Event-driven web application testing based on model-based mutation testing*, Information and Software Technology, vol. 67, pp. 159-179, November 2015.
- [14] F. SIMEONOV, N. PALOV, D. IVANOVA, D. KOSTOVA-LEFTEROVA, J. VASSILEVA, *Web-based platform for patient dose surveys in diagnostic and interventional radiology in Bulgaria: Functionality testing and optimisation*, Physica Medica, 4 May 2017.
- [15] H. TONGAL, B. SIVAKUMA, *Cross-entropy clustering framework for catchment classification*, Journal of Hydrology, Vol. 552, pp. 433-446, September 2017.
- [16] J. LUO, L. JIAO, R. SHANG, F. LIU, *Learning simultaneous adaptive clustering and classification via MOEA*, Pattern Recognition, vol. 60, pp. 37-50, December 2016.
- [17] L. HANGUANG, N. YU, *Intrusion Detection Technology Research Based on Apriori Algorithm*, Physics Procedia, vol. 24, Part C, pp. 1615-1620, 2012.
- [18] A. BHANDARI, A. GUPTA, D. DAS, *Improvised Apriori Algorithm Using Frequent Pattern Tree for Real Time Applications in Data Mining*, Procedia Computer Science, vol. 46, pp. 644-651, 2015.
- [19] M.R.F. COELHO, J. M. SENA-CRUZ, L.A.C. NEVES, M. PEREIRA, T. MIRANDA, *Using data mining algorithms to predict the bond strength of NSM FRP systems in concrete*, Construction and Building Materials, vol. 126, pp. 484-495, 15 November 2016.
- [20] S. GARCÍA, J. LUENGO, F. HERRERA, *Tutorial on practical tips of the most influential data preprocessing algorithms in data mining*, Knowledge-Based Systems, vol. 98, pp. 1-29, 15 April 2016.
- [21] D. APILETTI, E. BARALIS, T. CERQUITELLI, P. GARZA, L. VENTURINI, *Frequent Itemsets Mining for Big Data: A Comparative Analysis*, Big Data Research, 24 August 2017.
- [22] F. ALAM, R. MEHMOOD, I. KATIB, A. ALBESHRI, *Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)*, Procedia Computer Science, vol. 98, pp. 437-442, 2016.
- [23] F. ALAVI, S. HASHEMI, *DFP-SEPSF: A dynamic frequent pattern tree to mine strong emerging patterns in streamwise features*, Engineering Applications of Artificial Intelligence, vol. 37, pp. 54-70, January 2015.
- [24] J. ZHANG, X. ZHAO, S. ZHANG, S. YIN, *Interrelation analysis of celestial spectra data using constrained frequent pattern trees*, Knowledge-Based Systems, vol. 41, pp. 77-88, March 2013.
- [25] F. M. BIANCHI, A. RIZZI, A. SADEGHIAN, C. MOISO, *Identifying user habits through data mining on call data records*, Engineering Applications of Artificial Intelligence, vol. 54, pp. 49-61, September 2016.
- [26] M. KOOL, E. BASTIAANNET, C.J.H. VAN DE VELDE, P.J. MARANG-VAN DE MHEEN, *Reliability of self-reported treatment data by breast cancer patients compared with medical record data*, Clinical Breast Cancer, 18 August 2017.
- [27] C. BINGEN, C. E. ROBERT, K. STEBEL, C. BRÜHL, S. PINNOCK, *Stratospheric aerosol data records for the climate change initiative: Development, validation and application to chemistry-climate modelling*, Remote Sensing of Environment, 8 July 2017.
- [28] K. KIM, J. HONG, *A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis*, Pattern Recognition Letters, vol. 98, pp. 39-45, 15 October 2017.
- [29] A. SAETTTLER, E. LABER, F. D. A. MELLO PEREIRA, *Decision tree classification with bounded number of errors*, Information Processing Letters, vol. 127, pp. 27-31, November 2017.
- [30] R. YAN, Z. MA, Y. ZHAO, G. KOKOGIANNAKIS, *A decision tree based data-driven diagnostic strategy for air handling units*, Energy and Buildings, vol. 133, pp. 37-45, 1 December 2016.
- [31] K. MAHESH KUMAR, A. RAMA MOHAN REDDY, *A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method*, Pattern Recognition, vol. 58, pp. 39-48, October 2016.
- [32] Y. DJENOURI, M. COMUZZI, *Combining Apriori heuristic and bio-inspired algorithms for solving the frequent itemsets mining problem*, Information Sciences, vol. 420, pp. 1-15, December 2017.
- [33] <http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>
- [34] G. SINGH, V.K. JAIN, A. SINGH, *Adaptive network architecture and firefly algorithm for biogas heating model aided by photovoltaic thermal greenhouse system*, Journal of Energy Environment, pp.1-25, 2018.
- [35] SHERIFI AND SENJA, *Internet Usage On Mobile Devices And Their Impact On Evolution Of Informative Websites In Albania* Vol. 3, no. 6, pp. 37-43, 2018
- [36] A. H. SABLE AND K. C. JONDALE, *Modified Double Bilateral Filter for Sharpness Enhancement and Noise Removal*, 2010 International Conference on Advances in Computer Engineering, Bangalore, 2010, pp. 295-297, doi: 10.1109/ACE.2010.76
- [37] R. REMMIYA AND C. ABISHA, *Artifacts Removal in EEG Signal Using a NARX Model Based CS Learning Algorithm*, Multimedia Research, Vol. 1, no. 1, pp. 1-8.
- [38] Q. N. L. HOANG THUY TO, T.N. PHONG, DBH VY, *A hybrid multi criteria decision analysis for engineering project manager evaluation*, International Journal of Advanced and Applied Sciences, 4 (4), 49-52, 2017.

Edited by: P. Vijaya

Received: Nov 10, 2019

Accepted: Apr 1, 2020



A COMPREHENSIVE SURVEY OF THE ROUTING SCHEMES FOR IOT APPLICATIONS

DIPALI K. SHENDE*, YOGESH ANGAL, AND S. S. SONAVANE†

Abstract. Internet of Things (IoT) is with a perception of ‘anything’, ‘anywhere’ and provides the interconnection among devices with a remarkable scale and speed. The prevalent intention of IoT is the data transmission through the internet without the mediation of humans. An efficient routing protocol must be included in the IoT network for the accomplishment of its objectives and securing data transmission. Accordingly, the survey presents various routing protocols for secure data communication in IoT for providing a clear vision as the major issue in the IoT networks is energy consumption. Therefore, there is a need for devising an effective routing scheme to provide superior performance over the other existing schemes in terms of energy consumption. Thus, this review article provides a detailed review of 52 research papers presenting the suggested routing protocols based on the content-based, clustering-based, fuzzy-based, Routing Protocol for Low power and Lossy Networks, tree-based and so on. Also, a detailed analysis and discussion are made by concerning the parameters, simulation tool, and year of publication, network size, evaluation metrics, and utilized protocols. Finally, the research gaps and issues of various conventional routing protocols are presented for extending the researchers towards a better contribution of routing protocol for the secure IoT routing.

Key words: Internet of Things, routing protocol, energy consumption, routing protocol for low power, security, lossy networks.

AMS subject classifications. 68M12

1. Introduction. One of the emerging technologies, which enabled communication among things and people and between things, is the IoT [2]. IoT [27] is a new paradigm used for the pervasive communication establishment among the objects connected to the internet by employing distinct approaches [53]. The interconnection of networks connecting numerous devices for exchanging of information and services is called as ‘Internet’. Any type of object, device or gadget playing the designated role for the provision of the effective communication between the objects and people is mentioned as the ‘Thing’ [54]. The evolution of IoT has brought out the device’s proactive behavior rather than the reactive behavior of devices. The IoT applications are combined with the machine learning algorithms to makes the feature smarter [66]. The IoT comprises of a wide range of heterogeneous networks of varying processing powers, platforms, and capacities for expanding into the unreachable places [11]. The set of objects/ things enabled by IoT are Pervasive, Identification using unique address and Co-operation between things [55]. IoT has its application in a wide range of areas like smart environments, healthcare [63], environment monitoring [65], city surveillance, transportation, various firms [64], and energy monitoring, etc [3].

Over the past few years, the IoT has gained appealing research interest. The simplified IoT architecture has two layers, namely the perception layer and network layer, each provided with two sublayers. The perception layer has all the methods, which permit the gathering and perceiving of data. The data transmission in a transparent nature by utilizing the appropriate communication standards is handled by the network layer. The critical layer of this IoT architecture is the data management sub-layer, which is also known as a middleware layer. The application service sub-layer is preferred over the data management sub-layer for the management of the data transmission and provision of user application interface [56]. The vital network systems [2] available for the communication between the distinct objects in IoT are Wireless Sensor Network (WSN), Radio-Frequency Identification (RFID) systems and RFID Sensor Network (RSN). In these networks, the nodes are located in a specified range concerning the application for collecting the necessary information, such as physical change, temperature, and motion [57]. As the node’s transmission range is limited the collected information is forwarded

*Composition Department, Society for Industrial and Applied Mathematics, 3600 University City Science Center, Philadelphia, Pennsylvania, 19104-2688 (duggan@siam.org).

†This work was supported by the Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

to the intermediate nodes and this leads to higher energy consumption by the nodes [58]. Thus, the node's energy efficiency is concerned as a vital factor, which influences the performance of distributed IoT networks [2]. The IoT platform needs a potential networking framework and an appropriate routing scheme for supporting secure data communication and interconnection among devices.

One of the important aspects to be considered for improving the communication efficiency in IoT is the routing scheme adopted for the transmission path selection. The main aim of the routing protocol is to design and maintain interactions between the devices in the dynamic IoT network. The routing protocols are broadly classified into three, namely flat routing, hierarchical routing and geographic position assisted routing, based on the routing principle. In the flat routing, all the nodes are at the same level and they similarly receive their routing information; it is further classified into reactive protocols and proactive protocols. In the hierarchical routing, the network layer is partitioned into different levels based on the specific rules and this routing makes the network scale expansion easier. The geographic position assisted routing utilizes the node location information gathered from the users and it reduces the routing cost. The three primitive concerns [3] influencing the IoT routing are energy consumption, type of IoT middleware and mobility of devices. An effective protocol enhances the transmission efficiency and assures the judicious utilization of the available network resources [26].

The primary intention of this paper is to provide a detailed survey of the various routing protocols for the distinct IoT applications. This review deliberates the existing routing schemes for secure routing in the IoT network. The survey is made by considering the various methodologies utilized, implementation tools and the evaluation metrics, in the existing protocols. Additionally, the number of nodes considered for the simulation is concerned with the performance evaluation of the suggested routing protocols. The existing approaches have been categorized into distinct schemes, and then, the further survey is performed for the exploitation of research gaps and issues. Thus, it acts as the motivation for the future extension of improved secure IoT routing protocols.

The rest of the paper is organized as follows: Section 2 discusses the existing routing schemes under nine categories. Section 3 discusses the analysis and discussion of the survey. In section 4, the research gaps and the future works are elaborated and the conclusion of this paper is given in Section 5.

2. Literature Review. This section extensively discusses the review of the different IoT routing protocols for the secure routing in the IoT network. The categorization of distinct IoT protocols for the varying applications is pictorialized in figure 2.1. The routing protocols are categorized into nine routing schemes namely, Ad-hoc On-demand Distance Vector (AODV) Routing, Content-based Routing, Dynamic Source Routing (DSR), Tree-based Routing, Software Defined Network (SDN) based Routing, fuzzy-based Routing, RPL, Intelligent method based routing and Clustering-based Routing. These routing schemes give the various algorithms and techniques opted for the secure routing in IoT. The investigation on the several routing schemes provides a clear view of the recommended methods along with its advantages and disadvantages.

The distinct methodologies employed in the research papers for the IoT routing is depicted in figure 2.2. From the considered research papers, 6% of the research papers used the Clustering-based routing protocol, 9% of the paper used Tree-based protocol, 6% of the research papers employed the content-based routing, and 8% of the research paper utilized the AODV protocol. The intelligent method based routing is adopted in 9% of the research papers, SDN based routing is practiced in 6% of the research works and the Fuzzy based routing is utilized in 9% of the research works. The DSR routing is adopted in 3% of the research papers and the RPL routing is practiced in the remaining 44% research works.

2.1. Using AODV Routing Protocol. AODV is a reactive protocol, which provides the secure route for the data by concerning the next hop and the routing table values of every node. The research papers employing the AODV routing protocol are discussed below,

Sang-Hyun Park et al. [2] designed an Energy-Efficient Probabilistic Routing (EEPR) algorithm for controlling the routing request packets transmission. EEPR enhanced the network lifetime, while minimizing the Packet Loss Ratio (PLR). The probabilistic control was made by utilizing the Expected Transmission Count (ETX) metric in the AODV protocol context and node's residual energy. The results revealed that the EEPR algorithm provided better network lifetime along with even consumption of node's residual energy.

Hamoud M. Aldosari et al. [28] modeled an independent single security layer for the management of security mechanisms within the network layers. AODV is the routing protocol used, and the security layer

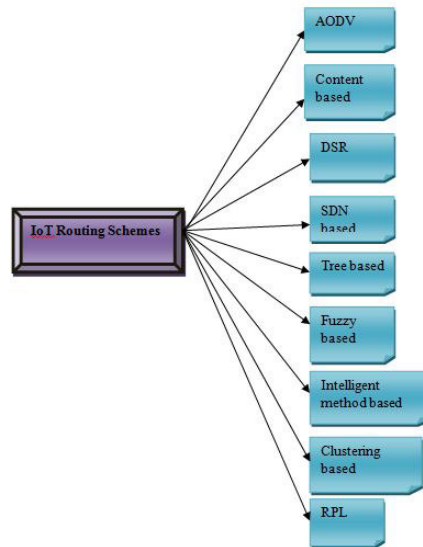


FIG. 2.1. Categorization of distinct IoT protocols

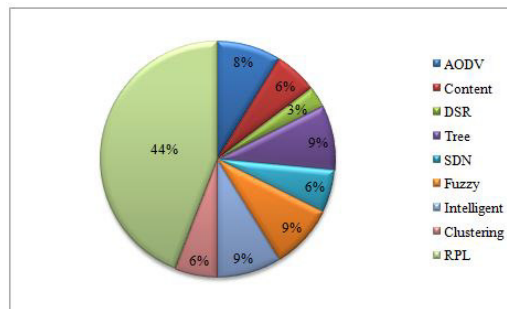


FIG. 2.2. Distinct Methodologies for Routing in IoT

was incorporated for cross-validating the sender and the receiver to eradicate the attacks in the network. The simulation results revealed that this security layer assured better performance than the other models in the aspect of throughput, End to End Delay (EED), Packet Delivery Ratio (PDR), Normalized Routing Load (NRL) and the number of dropped packets.

Greg Kuperman et al. [33] made research on the importance of the routing protocols in the efficient and reliable data communication in the multi-hop wireless network of the IoT environment. The link-based routing leads to the PLR, high maintenance cost, and unreliable data transmission within the network. The unfit nature of the link-based routing protocols for the wireless networks was explored in this research by comparing the performance of AODV and Optimized Link State Routing (OLSR) in terms of mobility, Routing Success Probability (RSP), the distance between users, PDR and overhead.

2.2. Using Content-based Routing protocol. In the content-based routing protocols, the routing information is differentiated by concerning the content. The distinct research works employing the content-based routing are presented below.

Samia Allaoua Chelloug [3] modeled an Energy-Efficient Content-Based Routing (EECBR) protocol for reducing the energy consumption in the IoT applications. This was a context and content-based routing protocol utilizing a centralized virtual topology constructed for the event routing in a distributed manner from the publishers to the destined subscribers. The results of simulation deliberated that EECBR outperformed the other protocols with respect to the variance of energy.

Yichao Jin et al. [8] designed the Content-Centric Routing (CCR) protocol for resolving the traffic congestion issue in IoT applications. In this protocol, the routing paths were predicted based on the content; this improved the data aggregation ratio and thereby, minimized the network traffic. The simulation results justified that the CCR has superior reliability with improved energy conservation and minimum network latency.

2.3. Using DSR protocol. The DSR protocol is an on-demand mode protocol following the principle of the shortest path for the secure path selection. The research paper utilizing the DSR protocol is presented as follows,

Ying Wei [30] modeled an improved DSR protocol (DSR-s), for the efficient channel utilization in IoT networks by concerning the packet rate and shortest-path routing. The secure routing path was selected by considering the data transmission rate and the hop count. The DSR-s enhanced the performance by a complete utilization of channel and avoiding the congestion problem within the network was deliberated through the simulation outcome.

2.4. Using Tree-based Routing protocol. The tree-based routing protocol is devised based on the tree network formed by the integration of the bus network and the star network. The research works adopting the tree-based routing protocols are discussed below,

Tie Qiu et al. [4] designed an Efficient Tree-Based Self-Organizing Protocol (ETSP) concerning energy for the sensor networks of IoTs. In this protocol, all the nodes were classified into two types, namely the network nodes for broadcasting the packet to neighboring nodes and the non-network nodes for gathering the broadcasted packets. In order to improve the network lifetime and consumption of energy balance, the topology was subjected to dynamic modifications. The simulation results revealed that this protocol assures the superior RSP, but the average hop, PLR and self-organization time of this protocol didn't increase with the increase in network scaling.

Zhangbing Zhou et al. [9] designed an energy-efficient index tree (EGF-Tree) for minimizing the energy consumption in the IoT devices. This protocol was based on the minimal merging principle in sensor node skewness distribution and grid division. This protocol has a prevalent concentration on the recombination of the region. The simulation results deliberated that the EGF-Tree outperformed the other original index tree regarding energy consumption.

S. Lana Fernando and A. Sebastian [15] introduced an effective clustering scheme, called Minimum Spanning Tree Particle Swarm Optimization (MST-PSO). The primitive intention of this scheme was the extension of network lifetime and minimization of the energy consumption and router dependency. The better performance of this algorithm over the other protocols with respect to energy consumption and network lifetime was illustrated by the simulation results.

G. Li et al. [62] introduced a multicast protocol, called heuristic algorithms for the solution of QoS constrained multicast routing problem, with incomplete information in WSN. Here, link measures were considered as the random variables for information aggregation. The major aim of this method was that it transformed the original probabilistic link descriptors, which reduced the tree selection to a deterministic problem. Then, the Hop Neural Networks (HNN) was applied which ensured the fast convergence to a suboptimal solution.

2.5. Using SDN based Routing protocol. The SDN controller utilizes a centralized routing protocol for collecting the network information. The research works practicing the SDN based routing protocols are discussed below,

Carynthia Kharkongor et al. [11] devised a routing mechanism using the SDN controller for providing secure data transmission among the heterogeneous devices. The primary intention of this mechanism was to minimize the energy consumption of heterogeneous devices by eliminating the approach of selfish nodes. The simulation results revealed that the routing by SDN controller provided better performance than the other protocols, like AODV, Destination Sequence Distance Vector (DSDV) and DSR with respect to average EED, throughput, and PDR.

Chiara Buratti et al. [25] performed the comparative performance analysis of SDN entitled Software-Defined Wireless Networking (SDWN), ZigBee and IPv6 over Low power Wireless Personal Area Networks (6LoWPAN). The SDWN utilized the centralized network layer protocol, whose routing schemes were described by an external controller, whereas ZigBee and 6LoWPAN utilized the distributed routing protocol. The results demonstrated

that the SDWN provided superior performance than the other two protocols in terms of traffic, payload size and network size. Jun Huang et al. [59] introduced two algorithms to construct the multicast routing tree for multimedia data transmissions. These algorithms leveraged an entropy-based process to combine all weights into a comprehensive metric, and utilized it to search a multicast tree through the shortest path tree and spanning tree algorithms. These algorithms assisted multimedia communications in an IoT environment. Here, collecting and updating the network topology and QoS constraint information were facilitated by the application of SDN technologies in the IoT environment.

2.6. Using Fuzzy based Routing protocol. The fuzzy logic can predict the tradeoffs between the distinct network parameters, the research papers employed with the fuzzy-based routing are presented as follows,

Ning Li et al. [5] devised a Cross-Layer and Reliable Opportunistic Routing Algorithm (CBRT) by the inclusion of fuzzy logic and humoral regulation stimulated topology control with the opportunistic routing algorithm. The relative variance was used as the fuzzy logic system input and the increase in the number of inputs didn't affect the number of fuzzy rules. Here, the relaying priority concerning the utility of reliable nodes was determined by the source node. The CBRT protocol has improved network performance over the Extremely Opportunistic Routing (ExOR) protocol and there was no up-gradation in the computational complexity.

Aljawharah Alnasser And Hongjian Sun [23] designed a fuzzy logic-based trust model for the detection of intruder nodes in the smart grid networks for enhancing the security. By the utilization of this model, the detection rate and routing efficiency can be enhanced for all the regarded destructive behavior. The simulation results demonstrated that the model outperformed the lightweight and dependable trust system model in terms of PDR by up to 90%.

Dong Chen et al. [24] modeled a trust and reputation model for IoT (TRM-IoT) based on the trust establishment scheme for maintaining the cooperation between the network things concerning their behavior. The fuzzy set was incorporated into this scheme for the effective management of the relationship between trust and reputation. This model had eradicated the data packet forwarding failure, achieved a good PDR and minimized energy consumption.

2.7. Using RPL Routing protocol. RPL is a distance-vector tree-based standardized routing protocol designed for the secure routing in most of the IoT applications. The different research works practicing this routing scheme are elaborately discussed as follows,

Maha Bouaziz et al. [6] modeled an Energy-efficient and Mobility aware Routing Protocol (EC-MRPL) based on RPL. This protocol assured better energy conservation and preserved the connectivity between mobile nodes. EC-MRPL protocol concatenated an enhanced mobility detection scheme with the prediction based on the point of attachment and replacement scheme with insight about resource restrictions. The protocol mitigated the mobility issues and provided better performance than the RPL and the Mobility Aware Routing Protocol (MRPL) with respect to signaling cost, energy consumption, handover delay, and data loss rate.

Mai Banh et al. [13] designed an energy balancing RPL along with the other routing metrics to deal with the link quality diversity. The diverse combination of energy consumption and ETX was concerned as the routing metrics. The method based on Radio Duty Cycle (RDC) was utilized for estimating the consumption of energy. This protocol assured better energy balance, maintaining good PDR and energy efficiency.

Harith Kharrufa et al. [16] modeled an enhanced Dynamic RPL (D-RPL) for distinct applications with dynamic network and dynamic mobility. A dynamic Objective-Function (D-OF) was included for enhancing the end-to-end delay, energy consumption, and PDR. Moreover, D-RPL had retained the avoidance of loop and low packet overhead. The simulation results revealed the fact that this protocol has higher PDR, minimal EED along with acceptable energy consumption.

Sheeraz A. Alvi et al. [20] devised an enhanced RPL protocol for the Internet of Multimedia Things (IoMT), where the multimedia equipment provided the sensed data. This protocol reduced the energy consumption and carbon footprint emissions with the added inclusion of QoS requirements. The simulation outcome deliberated the beneficial gain with respect to the latency and energy efficiency.

Zeeshan Ali Khan et al. [21] designed a trust-based RPL routing (t-RPL) for the IoT devices in distinct applications. The primitive intention concerning the trust of every node was for the prominent IoT network management. The scheme enhanced the network flexibility, average delivery ratio and reduced the number

of paths with malicious nodes compared to the other variants, like resilient RPL (r-RPL) and classical RPL (c-RPL).

Nabil Djedjig et al. [22] devised a Metric-based RPL Trustworthiness Scheme (MRTS) for improving the security in RPL by resolving the trust inference issues. The Destination Oriented Directed Acyclic Graph (DODAG) Information Object (DIO) was extended by the inclusion of a trust-based Extended RPL Node Trustworthiness (ERNT) metric and a Trust Objective Function (TOF). The ERNT was computed through the collaboration of nodes by concerning the behavior of nodes, such as energy, honesty, and selfishness. The simulation revealed enhanced performance in terms of security.

Natanael Sousa et al. [27] devised an Energy-Efficient and Path Reliability Aware Objective Function (ERAOF) and an objective function for RPL to be employed in the distinct IoT applications. This ERAOF was designed based on the routing metrics, like energy and link quality. The simulation outcome demonstrated that this method improved the energy efficiency, reliability of communication along with maintaining a high PDR.

Emilio Ancillotti et al. [37] presented research work for the exploration of RPL application in the Advanced Metering Infrastructure (AMI) arrangement. All the characteristics were evaluated for discovering the limitations and significance of the RPL protocol. The investigation deliberated that the RPL has good scalability and it suffers from unreliability problems because of the lack of link quality insight.

Mamoun Qasem et al. [38] presented a load-balanced objective function for the RPL to ensure the maximization of node lifetime. This objective function managed the total number of children nodes in a network and eliminated the potential extra overhead. The simulation results demonstrated that this objective function outperformed the other schemes with respect to the PDR, network lifetime and power consumption.

Ming Zhao et al. [39] made an exhaustive study on the RPL protocol and its standardized version Point-To-Point RPL (P2P-RPL) for supporting the Low Power Lossy Network (LLN) applications. The study explored the routing specifications, challenges and performance evaluation of both the protocols. The P2P-RPL protocol outperformed the standard RPL protocol in terms of flexibility, PDR, EED and control overhead.

José V. V. Sobral et al. [40] made the performance analysis of the reactive protocol Lightweight On-Demand Ad hoc Distance vector Routing (LOADng) protocol in two distinct types of traffics, namely P2P and MP2P. The protocol was a lighter version of the AODV protocol depending on the limited IoT resources. This protocol assured superior performance when employed in the MP2P applications rather than the P2P applications for the chosen scenarios. One of the limitations of this protocol is that with the increase in the network size, the performance decreases.

Olfa Gaddour et al. [41] modeled a fuzzy-logic objective function (OF-FL) for the RPL based LLNs. This objective function was a suitable approach as it integrated the four nodes with its link parameters, like EED, hop count, battery level and ETX, by utilizing the fuzzy logic. The simulation results demonstrated the performance enhancement in RPL with respect to the network lifetime, EED and PLR.

Baraq Ghaleb et al. [46] designed an Enhanced-RPL for alleviating the storage limitation issue in the preferred parent of the node. A Downward Advertisement Object (DAO) was modeled for eliminating under-specification mechanism problem. This Enhanced-RPL provided better performance than RPL by 64 and 30 in the aspect of both control plane overhead and PDR, respectively.

Mai Banh et al. [48] explored the knowledge of RPL under the multiple RPL routing tree instance conditions. This RPL constructed a Directed Acyclic Graph (DAG) representation for the IoT network, there can be many DAG's for the same network but with varying routing metrics, like ETX, hop count and node power. The simulation outcome revealed that the scheme enhanced the performance in terms of PDR and latency by differentiating the network traffic and allowing them into distinct DAG's of the multiple RPL instances.

Ayman El Hajjar et al. [50] devised a Shared the Identifier Secure Link Objective Function (SISLOF) for the RPL assuring that only the nodes, which share an appropriate key, can take part in the RPL routing table. The simulation outcome demonstrated that the scheme enhanced the security within the IoT network, reduced the storage size along with maintaining a small ring size.

Quan Le et al. [60] introduced three multipath methods, Fast Local Repair (FLR), Energy Load Balancing (ELB), and combination of ELB-FLR depending on RPL and integrated them in a modified IPv6 communication stack for IoT. These methods were implemented in the OMNET++ simulator and the experiment results showed that these methods attained better network load balance, packet delivery rate, end-to-end delay, and energy

efficiency.

2.8. Using Intelligent method based Routing protocol. The intelligent method based routing protocols are usually incorporated with distinct optimization algorithms for secure routing. The research works utilizing the intelligent method based routing protocols are presented below,

Praveen Kumar Reddy and Rajasekhara Babu [7] devised a protocol by integrating the Optimal Secured Energy-Aware Protocol (OSEAP) and Improved Bacterial Foraging Optimization (IBFO) algorithm. The protocol was mainly employed for improving the energy conservation and security between nodes in the IoT devices. The Fuzzy C-Means (FCM) clustering algorithm was included for the selection of an effective cluster head. The protocol provided better simulation results than the Secure Energy-aware routing protocol (SEAP) with respect to energy, throughput, and delay.

Sofiane Hamrioui and Pascal Lorenz [14] devised a routing algorithm, entitled Efficient IoT Communications based on Ant System (EICAntS), for enhancing the route selection process by the exploitation of ant colony system significances. This routing algorithm reduced the latency and enhanced the network lifetime, throughput and energy conservation.

Chengjie Wu et al. [32] designed an optimal algorithm based on integer programming and a greedy heuristic algorithm for extending the lifetime of Wireless HART (Highway Addressable Remote Transducer) networks. The integer programming was a linear programming relaxation algorithm. The graph routing was incorporated for prolonging the network lifetime. The experimentation was carried out in the physical testbed and the results elucidated that this algorithm enhanced the network lifetime by 60 maintaining the graph routing reliability.

2.9. Using Clustering-based Routing protocol. The clustering based routing protocol is developed by concerning the node with the highest degree as the cluster head. The research papers employed with the clustering based routing protocol are elaborated below,

Li Qing and Li Cong [26] modeled a node position based optimized clustering routing protocol utilizing the minimum distance routing competition scheme. The simulation results of this method deliberated the superior performance over the other routing protocols, such as AODV and DSR, in terms of network load and transmission delay. This method assures a stable cluster size in contrast to the other traditional clustering routing protocols.

Jau-Yang Chang [52] devised a distributed cluster computing energy-efficient routing scheme for minimizing the sensing node's data transmission distances by utilizing the concept of cluster structure. For the selection of an appropriate cluster head node, the center of gravity of the sensing nodes and the residual energy were computed. The simulation outcome deliberated that this scheme provided better performance than other methodologies in the aspects of network lifetime and energy conservation.

2.10. Using other Routing protocols. The other routing protocols utilized for the secure routing in the IoT environment are elaborated below,

Kássio Machado et al. [1] devised a routing protocol concerning the energy and link quality entitled Routing by Energy and Link quality (REL) for the applications of IoT. The route selection in REL was carried out based on the residual energy, end-to-end link quality estimator mechanism and hop count. The simulation results deliberated that REL increases the service availability, quality of service and network lifetime. This protocol assured uniform distribution of scarce network resources and minimized the PLR in comparison with the other eminent protocols.

Shahid Raza et al. [10] modeled a protocol entitled Lithe formed by the concatenation of the Constrained Application Protocol (CoAP) and Datagram Transport Layer Security (DTLS) for the resource-constrained IoT devices. The protocol provided authentication and confidentiality during communication. A DTLS header compression method was included with Lithe for minimizing energy consumption. The simulation results revealed that this protocol assures significant improvement in packet size, processing time, network-wide response times and energy consumption.

Shu-Chiung Hu et al. [12] modeled a ZigBee compatible energy-efficient multicast protocol for permitting the node to implement the devised procedure in a distributed manner. The management of neighbors by avoiding the unwanted data transmission was carried out with the support of the designed procedures and

backoff mechanism. This protocol enhanced the lifetime of the network, minimized the redundant packet and maintained the reliability of the network.

Omar Said et al. [17] devised the adaptive versions of the Real-time Transport Protocol (RTP) and Real-Time Control Protocol (RTCP), i.e., IoT-RTP and IoT-RTCP, for the IoT applications. The primitive principle employed in these devised versions was the division of large multimedia sessions into simple sessions with the network status insight. The superior performance of these protocols was deliberated by the obtained simulation results by concerning the number of receiver reports, EED, PLR, delay jitter, energy consumption, and throughput.

Rehmat Ullah et al. [18] designed the composite multi-routing metric for the RPL network. These metrics were developed by concerning the queue utilization, minimal hop count, minimum ETX value and residual energy of node. This scheme outperformed the Queue Utilization based RPL (QU-RPL) regarding the average power consumption. The simulation results conveyed that this protocol minimized power consumption and improved the network's lifetime.

Zhikui Chen et al. [19] modeled a Context-Awareness in Sea Computing Routing Protocol (CASCR) for IoT applications concerning the insight on context. The data structure, quantitative algorithm, and workflow of the devised CASCR protocol were also discussed in this research. The improved network lifetime and energy efficiency were illustrated through the obtained simulation outcome.

Pedro Henrique Gomes et al. [29] designed a Time Slotted Channel Hopping (TSCH) mode with a controlled flooding-based routing protocol for participating in the competition of EWSN Dependability. In the network based on TSCH, the time was split into time slots and the channel hopping was utilized for eradicating the fading and interference issues. The protocol employed here increased the PBR and flexibility along with maintaining minimum latency and overhead.

M. Vellanki et al. [31] designed a Node Level Energy Efficient (NLEE) routing protocol for enhancing the energy efficiency in the IoT networks. The protocol acted as the deciding authority for determining the shortest hop count by considering the hop count of node path and residual energy. This scheme determined the secure shortest path between the source and the destination and maintained better energy conservation by the effective utilization of the energy of nodes.

Hicham Lakhlef et al. [34] designed an agent-based efficient broadcast protocol for mobile with few communication channels. The ideology behind this protocol was to split things managed by the agents into groups based on the channel number. This protocol assured effective performance without collision or conflict in the communication channels in terms of the number of broadcast rounds and runs.

Samad Riaz et al. [35] designed an energy harvesting scheme for the routing protocols and made a comparative performance analysis between the two routing protocols AODV and OLSR in an Energy Harvest enabled Device to Device communication network. The simulation results deliberated that the energy harvesting was a feasible process for improving the residual energy of the network; thereby, assuring prolonged network lifetime. The available residual energy improved the goodput of both the considered routing protocols.

Joanna Glowacka et al. [36] presented a cognitive mechanism based on trust for the OLSR to enable the awareness of the situation in the IoT networks for intrusion detection. The main intention of this protocol was to explore the efficiency and robustness of the devised method through simulation. The results demonstrated that the protocol was efficient, had improved the security and total trusted traffic in the network.

Badis Djamaa et al. [42] modeled a centralized/distributed resource discovery architecture employing a CoAP for the IoT application. By employing this architecture, the nodes in the network can identify the availability of Resource Directories (RDs) using a proactive RD discovery scheme. The performance analysis deliberated the enhanced performance of this architecture, in terms of resource economy, reliability and time efficiency, over the other traditional architectures.

Nawel Alioua et al. [43] modeled a Uniform Stress Routing Protocol (USR) for exploiting the routing issues in Low power and Lossy Networks (LLNs). This protocol utilized the axiom of uniform stress in the routing for the encouraging adaptation. This USR restricted the upper bound of required memory and it didn't acquire exaggerated control traffic.

Youhua Xia et al. [44] developed a Privacy-Aware Routing Protocol (PALXA) for secure communication in the IoT network. This protocol concatenated the Arrow-d'Aspremont-Gerard-Varet (AGV) mechanism depend-

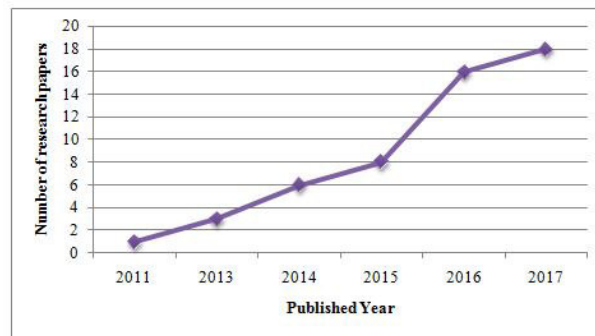


FIG. 3.1. Analysis based on the year of publication

ing on the theory of machine design with the reputation mechanism based on the subject logic for eradicating the internal attacks. This protocol outperformed the other protocols, like PALX and Pruned Adaptive IoT Routing (PAIR), in the aspect of node survival percentage, throughput, and successful transmission rate.

Lutando Ngqakaza and Antoine Bagula [45] presented a frugal protocol entitled Least Path Interference Beaconing (LIBP) for the dissemination of sensor readings in IoT. LIBP was a lightweight path selection model, which constructed a routing spanning tree with its root at the sink node through a seasonal beaconing mechanism. This protocol assured enhanced performance with respect to the throughput, scalability, and failure recovery and power consumption over the other protocols, like RPL and Collection Tree Protocol (CTP).

Piergiuseppe Di Marco et al. [47] analyzed the interdependence between the MAC and the Internet Engineering Task Force (IETF) RPL routing protocol in IoT. Then, a mathematical framework was modeled for improving the standard by the cooperative optimization of the MAC and the parameters of the routing protocol. The routing layer parameters considered for the simulation were R-metric and Q-metric, and the node energy consumption was reduced by 20%.

Mohsen Hallaj Asghar and Nasibeh Mohammadzadeh [49] devised a critical routing protocol, entitled Message Queuing Telemetry Transport (MQTT), for building the connection between the physical world and the real world. The primary intention of this research was to improve the Message Queuing (MQ) Service components utilized for connecting the distinct software applications. The main significance of this scheme was easy implementation, light, reduced delay, and low bandwidth system.

Yuxin Liu et al. [51] designed an energy-efficient Fast data collection for nodes Far away from the sink and Slow data collection for nodes Close to the sink (FFSC) approach, for reducing the EED and configuration complexity of IoT network. This FFSC assured the better network lifetime and energy conservation when compared to the other direct forwarding and single fixed threshold methodologies.

Meng-Shiuan Pan and Shu-Wei Yang [61] introduced a lightweight and distributed geographic multicast routing protocol that had three phases. The first phase selected the intermediate nodes to reach the multicast destinations. The second phase eliminated the loops and trims routes constructed in the previous phase. The third phase ensured, if the selected multicast links could be merged. This technique decreased the transmission links and shortens path lengths in the constructed multicast paths. Also, it reduced the multicast latency.

3. Analysis and discussion. This section presents the analysis and discussion of the parameters considered for the secure routing, tools used and metrics used in the research papers for exploring the distinct routing protocols in the IoT platform.

Analysis based on the year. The analysis of 52 research papers utilized for the secure IoT Routing carried out in the aspect of the published year is elaborated in this subsection. The analysis made by concerning the year of publication is depicted in figure 3.1. Among the 52 papers chosen for the analysis, most of the research papers were published in 2017.

Analysis based on the objective parameters. This subsection deliberates the analysis carried out by considering the objective parameters, like energy, trust, and link quality, based on their different combinations. Figure 3.2 displays the analysis chart based on the considered parameters. Through this analysis, it is clearly shown

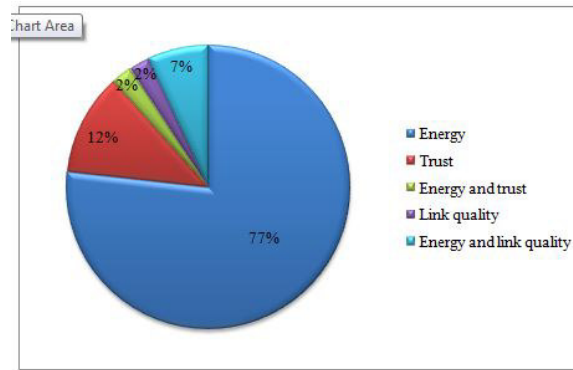


FIG. 3.2. Analysis based on the objective parameters considered

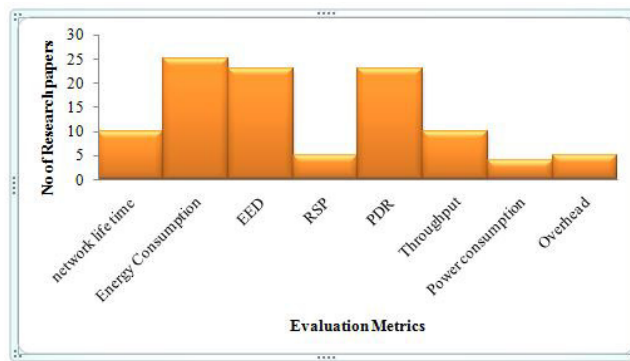


FIG. 3.3. Analysis based on the evaluation metrics

that nearly 77% of the research papers have used energy as its objective. 12% of the research papers have used trust and 2% of the research papers have utilized the link quality as the objective parameter. Nearly 7% of the research papers are based on the combination of two objectives, i.e., energy and link quality and the remaining 2% of the research papers are based on the parameters, energy, and trust.

Analysis based on the evaluation metrics. This subsection demonstrates the analysis based on the different evaluation metrics used in the research papers. The analysis chart built based on the evaluation metrics, such as network lifetime, energy consumption, EED, RSP, PDR, Throughput, Power Consumption, and overhead are depicted in figure 3.3.

From figure 3.3, it is conveyed that nearly in 50% of the research papers, the simulation is done in terms of energy consumption, EED, and PDR. In the remaining papers, the simulation is carried out concerning the network lifetime, RSP, throughput, overhead, and power.

Analysis based on Simulation Tool. The analysis carried out regarding the simulation tool used in the research works is presented in this subsection. As displayed in table 3.1, the various simulation tools utilized are C, NS, NS2, NS3, Cooja, Omet ++, Matlab, C++, OPNET modeler, and Castalia. This table clearly elucidates that the Cooja simulator is employed in most of the research papers when compared to the researches.

Analysis based on the network size. The analysis carried out by concerning the network size in different research works is demonstrated in this subsection. Table 3.2 shows the analysis based on the network size. From the table, it is deliberated that most of the research works have used nearly 50 to 100 nodes as their network size for the simulation.

4. Research gaps and issues. This section deals with the various research gaps and issues in the different IoT Routing protocols.

The simulation outcome illustrates that the routing setup delay of the EEPR algorithm [2] is marginally

TABLE 3.1
Analysis based on the Simulation Tool

Simulation Tool	ResearchPapers
C	[52]
NS	[30]
NS2	[2], [4],[5], [11],[17],[28],[48]
NS3	[24],[35],[39]
Cooja	[6],[13],[16],[18],[20],[22],[27],[37],[38],[41],[45],[46],[47],[48],[50]
Omnet ++	[1],[3],[14],[51]
[Matlab]	[7], [15], [19], [44]
[C++]	[12], [31], [32], [34]
OPNET modeler	[26],[33], [36]
Castalia	[40]

TABLE 3.2
Analysis based on Network Size

Number of nodes	ResearchPapers
0-10 nodes	[29], [42]
10-50 nodes	[[2],[11], [13], [16], [22], [23], [36], [45], [47]
50-100 nodes	[4], [7], [15], [18], [19], [20], [25], [27],[28], [30], [33], [35], [38], [40], [41]]
100-500 nodes	[1], [5], [8], [12], [14], [24], [26], [31],[37], [39]
500-1500 nodes	[21], [51]
1500-2500 nodes	[9], [50]

increased by 0.4ms and thereby, reducing the RSP of routing by 1.8% compared to the AODV protocol. In the routing tree based optimization algorithm, MST-PSO [15], the implementation time is higher when compared to the traditional networks. This could be resolved by the employment of the end devices instead of the base station for the cluster head selection. In the fuzzy logic-based CBRT [5], the main limitation is that the network lifetime, transmission range and throughput reduces with the increase in the number of nodes. In the content-based CCR protocol [8], the distinct types of content can be regarded by the improvement in the content definition by the incorporation of the emerging processing schemes. The SDWN routing protocol [25] are not preferred for the dynamic scenarios as it is mainly suitable for the static network environment with fixed node location with minimum mobility. The RPL routing protocol [37] is affected by unreliability issues and high PLR because of the lack of link quality insight, which thereby, results in the selection of unreliable links as suboptimal paths. In the link based routing [33], the routing delay is escalated along with the reduction in the RSP. The other drawbacks are the packet loss is severe because of the unreliability of control data, the message delivery is not ensured and the route maintenance and restoration is expensive.

The LOADng Routing Protocol [40] is not chosen as a scalable alternate as the protocol performance decreases with the increase in the network size. Hence, for the conclusive definition, a detailed study is to be made. By the employment of the RPL protocol provided with SISLOF [50] as an objective function, the IoT security is enhanced in the averaged sized universities. However, the storage size increases with the improvement in the network size. The research can be further extended by concerning the generated overhead for the network suitability discovery and the utilization of multiple DODAGs, for the secure routing between roots. In the NLEE algorithm [31], the hop count and the residual energy of nodes are concerned for the shortest path discovery. This algorithm leads to enhanced consumption of energy, which leads to an increase in setup delay for routing and reduces the RSP. The QoS metric can also be considered for performance improvement. The CASCR routing protocol [19] gathers the context knowledge from the neighboring nodes, leading to enhanced energy consumption by the nodes. This algorithm can be made self-optimizing by the provision of satisfactory context insight into the nodes.

Due to the conservativeness of the Chern off approximation in [62], the delay bound always met at the expense of consuming more transmit power. In [60], the FLR had the problem of large end-to-end delay which was caused by the increasing number of packets and the hop to transfer the packet to root. The results are good in [61], but it has limited effect. The data packet sizes did not bring much effect on the result since the

transmitting and receiving times on data packets are short, and most energy was consumed while nodes are idle in their active mode. Issues, like an injection of false information into the network, wireless broadcast of messages, and eavesdropping greatly negotiate the integrity of IoT communication. Moreover, due to the constrained nature and self-organizing attribute of IoT sensor nodes, the utilization of a solution centred on Certification Authorities (CA) via connected servers causes excessive obscurity for secure routing among IoT nodes.

In IoT, excessive energy consumption is a crucial problem, which is ignored by several existing methods. Also, the existing works based on energy and trust aware multicast routing in IoT have some drawbacks, like increased packet delay, link constrained problem, link optimization problem, tree optimization problem, and so on. Moreover, a few of literature works that are based on optimization algorithms are available in multicast routing and those optimization methods are not a very recent one and effective.

5. Conclusion. This paper provides a review of the classification of various routing protocols employed in the distinct IoT applications in the research papers. The primary aim of this article is to review and learn the several IoT routing protocols by analyzing the 52 research papers from IEEE (Institute of Electrical and Electronics Engineers), Google Scholar and Science Direct. The analysis is carried out in terms of year of publication, parameters, evaluation metrics, network size, simulation tool, and the adopted routing protocol. The research gaps and issues are also included in this article for directing the research towards the utilization of an effective routing protocol. Based on the discussion and analysis, it can be summarised that the energy consumption is the vastly concerned objective parameter in most of the research works for the discovery of better protocol. The evaluation metrics regarded in more than 50% of the research papers are energy consumption, EED and PDR.

REFERENCES

- [1] K., MACHADO, D., ROSÁRIO, E., CERQUEIRA, A.A.F., LOUREIRO, A., NETO AND J.N., SOUZA, *A routing protocol based on energy and link quality for internet of things applications*, sensors, 13 (2013), 1942-1964.
- [2] S-H. PARK, S., CHO AND J-R., LEE, *Energy-efficient probabilistic routing algorithm for internet of things*, Journal of Applied Mathematics (2014).
- [3] S.A., CHELLOUG, *Energy-Efficient Content-Based Routing in Internet of Things*, Journal of Computer and Communications, 3 (2015).
- [4] T. QIU, X. LIU, L. FENG, Y. ZHOU AND K. ZHENG, *An efficient tree-based self-organizing protocol for internet of things*, IEEE Access, 4, (2016), pp.3535-3546,
- [5] N. LI, J-F. MARTINEZ-ORTEGA AND V.H. DIAZ, *LU-Intelligent Cross-layer and Reliable Opportunistic Routing Algorithm for Internet of Things*, Networking and Internet Architecture, (2017).
- [6] M. BOUAZIZ, A. RACHEDI, B. ABDELFTTAH, *EC-MRPL: An energy-efficient and mobility support routing protocol for Internet of Mobile Things*, In the proceedings of the 14th Annual IEEE Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, January (2017).
- [7] P.K. REDDY AND R. BABU, *An Evolutionary Secure Energy Efficient Routing Protocol in Internet of Things*, International Journal of Intelligent Engineering and Systems, 10 (2017), 337-346.
- [8] Y. JIN, S. GORMUS, P. KULKARNI AND M. SOORIYABANDARA, *Content centric routing in IoT networks and its integration in RPL*, Computer Communications, vol.89 (2016), 87-104.
- [9] Z. ZHOU, J. TANG, L-J ZHANG, K. NING AND Q. WANG, *EGF-tree: an energy-efficient index tree for facilitating multi-region query aggregation in the internet of things*, Personal and ubiquitous computing, 18 (2014), pp.951-966.
- [10] S. RAZA, H. SHAFAGH, K. HEWAGE, R. HUMMEN AND T. VOIGT, *Lite: Lightweight secure CoAP for the internet of things*, IEEE Sensors Journal, 13 (2013), 3711-3720.
- [11] C. KHARKONGOR, T. CHITHRALEKHA AND R. VARGHESE, *A SDN Controller with Energy Efficient Routing in the Internet of Things (IoT)*, Procedia Computer Science, 89 (2016), 218-227.
- [12] S-C. HU, C-H. TSAI, Y-C. LU, M-S. PAN AND Y-C. TSENG, *An energy-efficient multicast protocol for ZigBee-based networks*, In the proceedings of the IEEE international conference on Wireless Communications and Networking Conference (WCNC), Doha, Qatar (2016), pp. 1-6.
- [13] M. BANH, N. NGUYEN, K-H. PHUNG, L. NGUYEN, N.H. THANH AND K. STEENHAUT, *Energy balancing RPL-based routing for Internet of Things*, In the proceedings of the Sixth IEEE International Conference on Communications and Electronics (ICCE), Ha Long, Vietnam (2016), pp.125-130.
- [14] S. HAMRIOUI AND P. LORENZ, *Bio-Inspired Routing Algorithm and Efficient Communications within IoT*, IEEE Network, 31 (2017), pp.74-79.
- [15] S. L. FERNANDO AND A. SEBASTIAN, *IoT: Smart Homeusing Zigbee Clustering Minimum Spanning Tree and Particle Swarm Optimization (MST-PSO)*, International Journal of Information Technology (IJIT), 3 (2017).

- [16] H. KHARRUFA, H. AL-KASHOASH, Y. AL-NIDAWI, M. Q. MOSQUERA AND A.H. KEMP, *Dynamic RPL for multi-hop routing in IoT applications*, In the proceedings of the 13th IEEE Annual Conference on Wireless On-demand Network Systems and Services (WONS), Jackson, WY, USA, pp. (2017), 100-103.
- [17] O. SAID, Y. ALBAGORY, M. NOFAL AND F.A. RADDADY, *IoT-RTP and IoT-RTCP: Adaptive Protocols for Multimedia Transmission over Internet of Things Environments*, IEEE Access, 5 (2017), pp.16757-16773.
- [18] R. ULLAH, T.D. HIEU, AND B-S. KIM, *A Multi-Metric Routing Protocol for Low-Power and Lossy Networks*, KICS Conference (2017), 305-306.
- [19] Z. CHEN, H. WANG, Y. LIU, F. BU AND Z. WEI, *A context-aware routing protocol on internet of things based on sea computing model*, Journal of computers, 7 (2012), 96-105.
- [20] S. A. ALVI, G.A. SHAH, AND W. MAHMOOD , *Energy efficient green routing protocol for internet of multimedia things*, In the proceedings of Tenth IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore (2015), 1-6.
- [21] Z. A. KHAN, J. ULLRICH, A. G. VOYIATZIS AND P. HERRMANN, *A Trust-based Resilient Routing Mechanism for the Internet of Things*, In the Proceedings of the 12th ACM International Conference on Availability, Reliability and Security ARES'17 (2017), pp. 1-6.
- [22] N. DJEDJIG, D. TANDJAQUI, F. MEDJEK AND I. ROMDHANI , *New trust metric for the RPL routing protocol,* In the Proceedings of the 8th IEEE International Conference on Information and Communication Systems (ICICS), Irbid, Jordan (2017), pp.328-335.
- [23] A. ALNASSER AND H. SUN, *A Fuzzy Logic Trust Model for Secure Routing in Smart Grid Networks*, IEEE Access, 5 (2017), pp.17896-17903.
- [24] D. CHEN, G. CHANG, D. SUN, J. LI, J. JIA AND X. WANG, *TRM-IoT: A trust management model based on fuzzy reputation for internet of things*, Computer Science and Information Systems, 8 (2011), pp.1207-1228.
- [25] C. BURATTI, A. STAJKIC, G. GARDASEVIC, S. MILARDO, M. D. ABRIGNANI, S. MIJOVIC, G. MORABITO AND R. VERDONE, *Testing protocols for the internet of things on the EuWIn platform*, IEEE Internet of Things Journal 3 (2016), 124-133.
- [26] L. QING AND L. CONG, *Efficient Cluster Routing Design under the Environment of Internet of Things Based on Location*, In the Proceedings of the IEEE International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China (2016), pp. 318-323.
- [27] N. SOUSA, J.V.V. SOBRAL, J.J. RODRIGUES, R.A.L. RABELO AND P. SOLIC, *ERAOF: A new RPL protocol objective function for Internet of Things applications*, In the Proceedings of the 2nd IEEE International Multidisciplinary Conference on Computer and Energy Science (SpliTech), Split, Croatia (2017), pp. 1-5.
- [28] H.M. ALDOSARI, V. SNASEL AND A. ABRAHAM, *A New Security Layer for Improving the security of internet of things (IoT),*, International Journal of Computer Information Systems and Industrial Management Applications 8 (2016), 275-283.
- [29] P. H. GOMES, T. WATTEYNE, P. GHOSH AND B. KRISHNAMACHARI. , *Competition: Reliability through Timeslotted Channel Hopping and Flooding-based Routing*, In the Proceedings of the ACM International Conference on Embedded Wireless Systems and Networks (EWSN '16), Graz, Austria (2016), pp.297-298,
- [30] Y. WEI, *The Congestion Control Based on Routing Protocol in the Internet of Things*, In the Proceedings of the International Conference on Electronic Information Technology and Intellectualization (ICEITI) (2016).
- [31] M. VELLANKI, S. P. R. KANDUKURI AND A. RAZAQUE, *Node Level Energy Efficiency Protocol for Internet of Things*, Journal of Theoretical and Computational Science, 3 (2016).
- [32] C. WU, D. GUNATILAKA, A. SAIFULLAH, M. SHA, P.B. TIWARI, C. LU AND Y. CHEN, *Maximizing network lifetime of wireless hart networks under graph routing*, In the Proceedings of First IEEE International Conference on Internet-of-Things Design and Implementation (IoTDI), Berlin, Germany (2016), pp. 176-186.
- [33] G. KUPERMAN, S. MOORE, B-N. CHENG AND A. NARULA-TAM, *Characterizing deficiencies of path-based routing for wireless multi-hop networks*, In Proceedings of the IEEE International Conference on Aerospace, USA (2017), 1-9.
- [34] H. LAKHLEF, MICHEL RAYNAL AND JULIEN BOURGEOIS , *Efficient Broadcast Protocol for the Internet of Things*, In the Proceedings of the 30th IEEE International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, Switzerland (2016), pp. 998-1005.
- [35] S. RIAZ, H. K. QURESHI AND M. SALEEM , *Performance evaluation of routing protocols in energy harvesting D2D network*, In the Proceedings of the IEEE International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan (2016), pp. 251-255.
- [36] J. GŁOWACKA, J. KRYGIER AND M. AMANOWICZ, *A trust-based situation awareness system for military applications of the internet of things*, In the Proceedings of the 2nd IEEE World Forum on Internet of Things (WF-IoT), Milan, Italy (2015), pp. 490-495.
- [37] E. ANCILLOTTI, R. BRUNO AND M. CONTI, *The role of the RPL routing protocol for smart grid communications*, IEEE Communications Magazine, 51 (2013), pp.75-83.
- [38] M. QASEM, A. AL-DUBAI, I. ROMDHANI, B. GHALEB AND W. GHARIBI , *A new efficient objective function for routing in Internet of Things paradigm*, In the Proceedings of the IEEE Conference on Standards for Communications and Networking (CSCN), Berlin, Germany (2016), pp. 1-6.
- [39] M. ZHAO, A. KUMAR, P. H. J. CHONG AND R. LU, *A comprehensive study of RPL and P2P-RPL routing protocols: Implementation, challenges and opportunities*, Peer-to-Peer Networking and Applications, 10 (2017), pp. 1232-1256.
- [40] J. V. V SOBRAL, J.J. RODRIGUES, K. SALEEM AND J. AL-MUHTADI, *Performance evaluation of LOADng routing protocol in IoT P2P and MP2P applications*, In the Proceedings of IEEE International Multidisciplinary Conference on Computer and Energy Science, Split, Croatia (2016), pp. 1-6.
- [41] O. GADDOUR, A. KOUBÁA, N. BACCOUR AND M. ABID, *OF-FL: QoS-aware fuzzy logic objective function for the RPL routing protocol*, In the Proceedings of 12th IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc,

- and Wireless Networks (WiOpt), Hammamet, Tunisia (2014), 365-372.
- [42] B. DJAMAA, A. YACHIR AND M. RICHARDSON, *Hybrid CoAP-based resource discovery for the Internet of Things*, Journal of Ambient Intelligence and Humanized Computing, 8 (2017), pp.357–372.
- [43] N.ALILOUA, K. NAKAMURA, M. KAMIO, N. KOSHIZUKA AND K. SAKAMURA, *USR: Uniform stress routing protocol for constrained networks*, In Proceedings of 5th IEEE Global Conference on Consumer Electronics, Kyoto, Japan (2016), 1-3.
- [44] Y. XIA, H. LIN AND L. XU., *An AGV Mechanism Based Secure Routing Protocol for Internet of Things*, In the Proceedings of IEEE Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), Liverpool, UK (2015), pp. 662-666.
- [45] L. NGQAKAZA AND A. BAGULA, *Least Path Interference Beaconing Protocol (LIBP): A Frugal Routing Protocol for the Internet-of-Things*, In the Proceedings of Springer International Conference on Wired/Wireless Internet Communications, 8458 (2014), 148-161.
- [46] B. GHALEB, A. AL-DUBAI, E. EKONOMOU AND I. WADHAJ, *TA new enhanced RPL based routing for Internet of Things*, In the Proceedings of IEEE International Conference on Communications Workshops (ICC Workshops), Paris, France (2017), pp. 595-600.
- [47] P. D. MARCO, G. ATHANASIOU, P-V. MEKIKIS AND C. FISCHIONE, *MAC-aware routing metrics for the internet of things*, Computer Communications, vol.74 (2016), pp.77-86.
- [48] M. BANH, H. MAC, N. NGUYEN, K-H. PHUNG, N. H. THANH AND K. STEENHAUT , *Performance evaluation of multiple RPL routing tree instances for Internet of Things applications*, In the Proceedings of IEEE International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, Vietnam (2015), pp. 206-211.
- [49] M. H. ASGHAR AND N. MOHAMMADZADEH, *Design and simulation of energy efficiency in node based on MQTT protocol in Internet of Things*, In the Proceedings of IEEE International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, India (2015), 1413-1417.
- [50] A. E. HAJJAR, G. ROUSSOS, AND M. PATERSON. , *Secure routing in IoT networks with SISLOF*, In the Proceedings of the IEEE International Conference on Global Internet of Things Summit (GIoTS), Geneva, Switzerland (2017), 1-6.
- [51] Y. LIU, A. LIU, Y. HU, Z. LI, Y-J. CHOI, H. SEKIYA AND J. LI, *FFSC: an energy efficiency communications approach for delay minimizing in internet of things*, IEEE Access, vol.4 (2016), pp.3775-3793.
- [52] J-Y. CHANG, *A Distributed Cluster Computing Energy-Efficient Routing Scheme for Internet of Things Systems*, Wireless Personal Communications, 82 (2015), pp.757-776.
- [53] A. ZANELLA, N. BUI, A. CASTELLANI, L. VANGELISTA AND M. ZORZI, *Internet of things for smart cities.*” IEEE Internet of Things journal, Internet of things for smart cities,” IEEE Internet of Things journal, 1 (2014), pp.22–32.
- [54] C. LU, *Overview of Security and Privacy Issues in the Internet of Things*, Internet of Things (IoT): A vision, Architectural Elements, and Future Directions, 2014.
- [55] O. VERMESAN AND P. FRIESS, *Internet of Things-From Research and Innovation to Market Deployment*, River Publishers, Aalborg (2014), vol.29.
- [56] X. JIA, Q. FENG, T. FAN AND Q. LEI, *RFID Technology and Its Applications in Internet of Things (IoT)*, In the proceedings of 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang (2012), pp.1282-1285.
- [57] G. LEE, *A cluster-based energy-efficient routing protocol without location information for sensor networks*, International Journal of Information Processing Systems, 1 (2005), pp. 49–54.
- [58] X. LI, *Achieving load awareness in position-based wireless adhoc routing*, KITCS/FTRA Journal of Convergence, 3 (2012).
- [59] J. HUANG, Q. DUAN, Y. ZHAO, Z. ZHENG, AND W. WANG, *Multicast Routing for Multimedia Communications in the Internet of Things*, IEEE Internet of Things Journal, 4 (2017), pp. 215 – 224.
- [60] Q. LE, T. NGO-QUYNH, AND T. MAGEDANZ, *RPL-based Multipath Routing Protocols for Internet of Things on Wireless Sensor Networks*, International Conference on Advanced Technologies for Communications (ATC), Hanoi, Vietnam, October 2014.
- [61] M-S. PAN, AND S-W. YANG, *A Lightweight and Distributed Geographic Multicast Routing Protocol for IoT Applications*, Computer Networks, 112 (2017), 95-107.
- [62] G. LI, D. G. ZHANG, K. ZHENG, X. C. MING, Z. H. PAN, AND K. W. JIANG, *A Kind of New Multicast Routing Algorithm for Application of Internet of Things*, Journal of Applied Research and Technology, 11 (2013), 578-585.
- [63] N. M. HA AND N. T. AN, *Impact of work-family conflict on job performance of nurses working for hospitals in Ho Chi Minh city*, Science Journal 4 (2015).
- [64] N. M. HA, *The effect of growth on firm survival in vietnam*, April 2016.
- [65] R. R. ANDRADE, I.F.F. TINÔCO, F.C. BAÊTA, M. BARBARI, L. CONTI, P.R. CECON, M.G.L. CÂNDIDO, I.T.A. MARTINS, C.G.S.T. JUNIOR, *Evaluation of the surface temperature of laying hens in different thermal environments during the initial stage of age based on thermographic images*, Agronomy Research 15 (2017), 629-638.
- [66] V.H. ARUL. V.G. SIVAKUMAR, R. MARIMUTHU, AND B. CHAKRABORTY, *An Approach for Speech Enhancement Using Deep Convolutional Neural Network*, Multimedia Research (MR), 2 (2019), 37-44.

Edited by: P. Vijaya

Received: Dec 9, 2019

Accepted: Apr 1, 2020



CHICKEN-MOTH SEARCH OPTIMIZATION-BASED DEEP CONVOLUTIONAL NEURAL NETWORK FOR IMAGE STEGANOGRAPHY

RESHMA V K*, VINOD KUMAR. R. S[†] SHAHI D[‡] AND SHYJITH M.B[§]

Abstract. Image steganography is considered as one of the promising and popular techniques utilized to maintain the confidentiality of the secret message that is embedded in an image. Even though there are various techniques available in the previous works, an approach providing better results is still the challenge. Therefore, an effective pixel prediction based on image steganography is developed, which employs error dependent Deep Convolutional Neural Network (DCNN) classifier for pixel identification. Here, the best pixels are identified from the medical image based on DCNN classifier using pixel features, like texture, wavelet energy, Gabor, scattering features, and so on. The DCNN is optimally trained using Chicken-Moth search optimization (CMSO). The CMSO is designed by integrating Chicken Swarm Optimization (CSO) and Moth Search Optimization (MSO) algorithm based on limited error. Subsequently, the Tetrolet transform is fed to the predicted pixel for the embedding process. At last, the inverse tetrolet transform is used for extracting the secret message from an embedded image. The experimentation is carried out using BRATS dataset, and the performance of image steganography based on CMSO-DCNN+tetrolet is evaluated based on correlation coefficient, Structural Similarity Index, and Peak Signal to Noise Ratio, which attained 0.85, 46.981dB, and 0.6388, for the image with noise.

Key words: Image steganography, Deep Convolutional Neural Network, Tetrolet transform, Chicken swarm optimization, Moth search optimization algorithm.

AMS subject classifications. 68T05

1. Introduction. Nowadays, steganography plays a very significant role in various areas, which is utilized for the protection of data from unauthorized access. Steganography becomes very popular in secret message communication. In the steganography process, the sender must choose a suitable message, and the effective information needs to be hidden [38, 39, 40]. The object employed for carrying the message is known as the carrier image or the message carrier. The secret message is nothing, but the image, which is to be embedded in the carrier image. Stego-image is employed for carrying the hidden message. Hence, given the secret and the carrier image, the main aim of steganography is to build the stego-image for carrying the hidden message [9]. Steganography is the security approach for hiding the secret data. It has worked on concealing the image, text, video, or audio (sensitive data) inside other images, video or audio, text (cover) [10]. Steganography approaches are divided into video, text, protocol, Deoxyribonucleic Acid (DNA), and video steganography.

Image steganography is the technique for hiding the unnatural hidden message in carrier image, so the quality of carrier image has a small change, and hence no one can find it [11]. In the current, most of the steganographic approaches are employed for reducing the distortion function correlated with the statistical detectability [13]. Image steganography is defined by hiding the secret message in the carrier image so that the receiver can recover the watermark message when the warder does not detect the secret information. Most of the image steganography approaches attained their goals by embedding process that leaves evidence for distortion [8]. The image-based stego is considered as the effective cover medium due to its easy human remembrance possibility, and popularity [10]. Image steganography is also utilized for various applications, like safe communication between two parties [12, 14], captioning and contents protection [12, 16, 17], securing

*1Research Scholar in Noorul Islam Centre For Higher Education, Kumaracoil, Tamil Nadu, and Assistant Professor, Department of CSE, Jawaharlal College of Engineering and Technology, Palakkad, Kerala (reshmavk14@gmail.com).

[†]Head of Department, Electronics and Communication Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu

[‡]Assistant Professor, Dept. of ECE, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu

[§]Assistant Professor, Dept of CSE, Jawaharlal College of Engineering and Technology, Palakkad, Kerala

online voting systems, secure mobile computing [12, 15], privacy-protection of medical records, personalized secure image retrieval, and secure surveillance systems [12, 18].

Image Steganography allows two parties for secure communication. Image steganography approaches are divided into two, namely, frequency, and spatial domain [19]. The process of the spatial domain is to allow whole pixels of carrier image, and the embedding of the secret message into the frequency domain is performed after numerous conversions for converting the image to the frequency domain. The name “stego” comes from a concealed secret message in the image. Several methods are related to the spatial domain, like Gray level modification (GLV), Least Significant Bit (LSB), and so on [20]. The frequency domain-based methods are Discrete Fourier Transform (DFT), and Discrete cosine transform (DCT), and so on [11].

In this research work, the image steganography is performed using the proposed pixel-prediction approach. The developed pixel prediction-based approach employs three phases for hiding the watermark information in the image as, identification, embedding, and extraction phase. Initially, the input image is subjected to the pixel prediction phase in which the appropriate pixels are extracted from the image. The features, like edge information, pixel coverage, texture features, wavelet energy, Gabor feature value, and scattering values are extracted from the pixels using DCNN classifier, which is trained by CSO and MSO based on error. In the third phase, the watermark message is embedded in the input image using tetrolet transform that assures the embedding strength. Finally, the extraction of a watermark image is done by applying Tetrolet Transform (TT).

The main contribution of the research paper is enlisted below:

- The pixel identification is carried out based on the DCNN classifier using error, which is trained by applying CSO and MSO for finding the effective pixel employed for the embedding process.
- The features are extracted effectively from the pixel for generating the prediction map, which is followed with the embedding procedure to embed the watermark message in the input image using TT. The inverse TT is employed to extract the secret message from the input image.

The paper is structured as follows: Section 2 discusses existing methods of image steganography with challenges of the methods. The proposed method of CMSO-DCNN+tetrolet is demonstrated in section 3, and section 4 provides the results and discussion. At last, section 5 concludes the research work.

2. Literature Survey. This section presents the literature survey of several methods utilized for the image steganography and the challenges of the existing works.

Several methods related to image steganography are described, and analyzed as follows: Yuileong Yeung et al. [20] developed binary image steganography to reduce the flipping distortion on Local Texture Pattern (LTP) and designed flexible carriers of syndrome-trellis code (STC). Here, the security for both vision and statistics was found better, but it failed to access the scalable and nearly continuous capacity upper bound. Adnan Gutub and Maimoona Al-Ghamdi [10] presented counting-based secret sharing method to improve the shares reconstruction and distribution. The method failed to define the maximum and minimum shares for reconstructing the secret. Weixuan Tang [21] employed a Convolutional Neural Network based on adversarial embedding (ADV-EMB) for image steganography. Here, the security was found better, but failed to detect the stego images in the current iteration. Soumendu Chakraborty et al. [22] developed Predictive Edge Adaptive image steganography in which the selected area of the cover image was determined based on Modified Median Edge Detector (MMED) to embed the binary payload (data). The developed method achieved limited level of distortion and best embedding rate. More bits were needed in sharper edges for adaptive selection.

Dipti Kapoor Sarmah and Anand J. Kulkarni [5] presented Multi Random Start Local Search (MRSLS) for achieving a better balance between secret text capacity, security, and image quality. This framework employs Cohort Intelligence (CI) for attaining the enhanced quality in the image. The method failed to consider the cohort intelligence approach was integrated with a cuckoo search algorithm for enhancing efficiency. Aref Miri and Karim Faez [23] developed an integer wavelet transform for mapping cover images with a specific frequency domain. The Most Significant Bit (MSB) bit was utilized for categorizing the edge coefficients in the frequency domain. Embedding the secret bits in the frequency coefficients are needed for obtaining the stego image. S.I. Nipanikar and V. Hima Deepthi [6] presented edge and wavelet transformation approach to finding the accurate location to embed the message. Here, the edges are detected accurately based on the wavelet coefficient, and the intensity of pixel, but does not consider another optimization algorithm for improving the cost estimation of the pixel. Tomas Denmark and Jessica Fridrich [7] developed an approach for inferring the accurate direction

of changes made in steganographic embedding. This change was incorporated with cost-based steganography for reducing the embedding costs based on the multiplicative modulation factor, but more than two acquisitions were needed for the embedding process.

2.1. Challenges. This section deals with the challenges faced by the existing techniques of image steganography.

- Once the data embedding is done, the changes in the image statistics are predicted using the steganalyzers, and the bit planes employed in the typical image are less correlated, which is one of the challenges faced by the image steganography [24].
- Other challenges face by image steganography are the visibility of extracted images at the receiver side by changing the value of the blending coefficient is very sensitive to modification [8].
- The stego image quality and the capacity of embedding is the challenging task for designing the secure binary image by enhancing the undetectability [14].
- In real-time, if there is additive noise in the cover image, there is the possibility for the poor PSNR or in other words, the image quality is affected.
- The embedding capacity associated with the steganography is not significant using the color parity, which depends on the order of the colors, and the capacity of the image is very less, which is the major challenge in embedding the stego images from the cover image [16].

3. Image Steganography using the proposed Chicken-Moth search optimization-based deep convolutional neural network. This section presents the Image Steganography using Chicken-Moth search optimization algorithm. Figure 3.1 deliberates the block diagram of the proposed Chicken-Moth search optimization algorithm for Image Steganography. At first, the input medical image is given to the pixel prediction phase in which the features, like wavelet, pixel coverage, edge details, and so on will be extracted from the image, which is subjected to the pixel identification using the proposed CMSO-based DCNN. Thus, the proposed CMSO algorithm is developed by integrating CSO [37] with the MSO [36]. Thus, the pixels to be embedded will be decided using the Deep CNN that will be trained using the CMSO algorithm. Here, the embedding of the secret message is done by applying TT such that it offers fast and efficient image representation. The extraction of the secret image is done at last by applying the inverse transform.

Let us consider the input medical image M with dimension $X \times Y$. Here, the term U represents the watermark message of dimension $P \times Q$. The watermark image is in the form of binary. The representation of the input, and embedded watermark image is given by

$$M = \{m_{uv}\}; 1 \leq u \leq X; 1 \leq v \leq Y \quad (3.1)$$

$$U = \{S_{xy}\}; 1 \leq x \leq P; 1 \leq y \leq Q \quad (3.2)$$

where, the term m_{uv} refer to the input image pixels, ranging between 0 to 255, and S_{xy} denotes the watermarked image pixels, having the binary value as 0 or 1.

3.1. Pixel prediction. Extraction of features. In the pixel prediction phase, the feature extraction is performed based on seven features, such as wavelet energy, edge information, scattering value, Gabor filter, pixel coverage, Local binary pattern (LBP), and Tetrolet transform. The feature extraction step carried out in this paper is explained as follows.

3.1.1. Wavelet energy. The TT [25] is subjected to the input medical image to achieve the wavelet energy. The wavelet transform is employed for decomposing the original image to four sub-bands, which includes HL, HH, LL, and, LH. The four bands are represented as, $\{L_1, L_2, L_3, L_4\}$. The tetrolet energy for the pixel is expressed as

$$TT(m_{uv}) = \{L_1, L_2, L_3, L_4\} \quad (3.3)$$

where $TT(m_{uv})$ represents the tetrolet function. In this paper, the embedding phase focussed on HL band, hence the energy consumed in HL band is taken as the tetrolet energy feature and is represented as $F_1 = L_3(m_{uv})$.

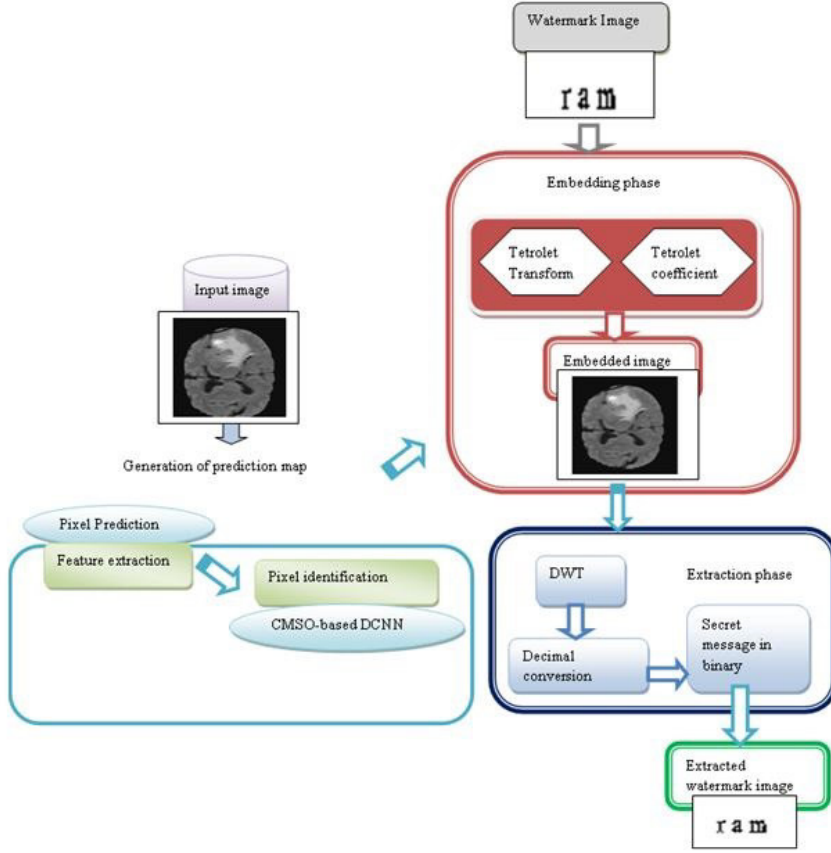


FIG. 3.1. Schematic diagram of image steganography using the proposed CMSO-based DCNN

3.1.2. Pixel coverage. The pixel coverage is computed based on the mean value of neighborhood pixel to provide the information about the coverage value of each pixel. Assume the pixel m_{uv} in image M with N number of neighboring pixels, and the pixel coverage is expressed as,

$$F_2 = \frac{1}{N} \sum_{a=0}^{N-1} m_{uv}^a \quad (3.4)$$

where m_{uv}^a is the symbol indicates the a^{th} neighbor pixel of m_{uv} and N denotes the neighboring pixel.

3.1.3. Edge detection. It is the process of determining whether the pixel is available in the corner edge or not. In edge information feature, if the pixel is in the edge then the value is fixed as one or otherwise the value is set to zero. The edge information for the pixel is given by

$$g(u, v) = A(m_{uv}) \quad (3.5)$$

where $A(m_{uv})$ refer to the edge information, and the output of the edge information is represented as $F_3 = g(u, v)$.

3.1.4. Scattering value. The scattering transform [26] is employed for obtaining scattering coefficients. This transform finds the texture information by applying the filter convolution in the pixel. The scattering coefficient is given by

$$K[M] = |||M \otimes \eta_{b1} \otimes \eta_{b2} |K| \otimes \eta_{bh} \otimes A(e) ||| \quad (3.6)$$

where the average filter is denoted as $A(e)$, and the term indicates the filter banks. The output obtained from the scattering transform is denoted as $F_4 = K(m_{uv})$.

3.1.5. Gabor feature. The Gabor filter [27] is employed for identifying the time-frequency location of the pixel and to achieve the robust against various brightness or contrast of the image. For the feature extraction, 2D Gabor filter is the broadly utilized filter and the filter function is given by,

$$P(u, v, \varphi, h, \chi) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + V^2}{2\sigma^2}\right) * \exp\{2\pi u_m(hu \cos \theta + hv \sin \theta)\} \quad (3.7)$$

where $P(u, v, \varphi, h, \chi)$ denote the Gabor filters with pixel. The term $u - m$ denotes the imaginary part and the value is $\sqrt{-1}$. The sinusoidal wave frequency in the Gaussian filter is represented as h . The output of the Gabor filter is denoted as F_5 , and is expressed as $D(\{u, v\} = P(m_{uv})$. Therefore, $F_5 = D\{u, v\}$.

3.1.6. Local Binary Pattern. The crisp form of LBP [28] utilizes the neighbourhood pixel properties to explain each pixel. It is more resistant, efficient, and simple to make changes in gray-level using lighting variations. The obtained feature is indicated as, $F_6 = V(u, v)$. The texture features obtained from the extraction phase is expressed by

$$V(u, v) = LBP)m_{uv} = \sum_{a=0}^{N-1} r(t_a - t_f)2^n \quad (3.8)$$

where the term t_f and t_a denotes the centre and neighbor pixels gray value. Then, the membership function is calculated as

$$r(b) = \begin{cases} 1; & b \geq 0 \\ 0; & \text{otherwise} \end{cases} \quad (3.9)$$

3.1.7. Tetrolet Transform. The tetrolet descriptor [25] is an adaptive Haar wavelet transform, to support tetrominoes, which is formed by joining four squares with similar size. Here, the input low-pass image is divided into blocks and local tetrolet basis are generated based on the geometry of the image. The steps involved in the tetrolet transform are illustrated below.

i) Initialization: The input image is split into blocks of size.

ii) Representation of image blocks as the sparsest tetrolet: Each of the image blocks is subjected to the sparsest tetrolet representation and for every individual blocks, a total of 117 tetromino coverings are admitted each of which is given to the Haar wavelet transform along with four low pass coefficients for generating 12 Tetrolet coefficients. For the individual block, the tetrolet decomposition is done at the optimum based on 12 tetrolet coefficients to obtain the final sparse image.

iii) Representation of the high pass and Low pass coefficients: The steps involved in the Tetrolet decomposition algorithm is preceded with the arrangement of matrix based on reshape function.

iv) Tetrolet Coefficients: After representing the sparse matrix for the individual blocks, the high pass as well as the low pass matrices is kept safe for the usage of future.

v) Termination: The steps (ii) to (iii) are repeated for the low pass image and the output obtained in the binary image, which is denoted as F_7 . Therefore, the extracted features are represented as

$$J^{red} = \{F_1, F_2, F_3, F_4, F_5, F_6, F_7\} \quad (3.10)$$

The size of the extracted features is denoted as $[1 \times 7]$.

3.2. Pixel identification using DCNN. Once the features are extracted, DCNN is utilized for pixel identification. The DCNN is utilized for generating the prediction map for image pixels. The DCNN classifier [29] uses the extracted features J^{red} as input and generates the prediction map based on input image. The architecture of DCNN and the algorithmic steps of the CMSO-based DCNN are described bellow.

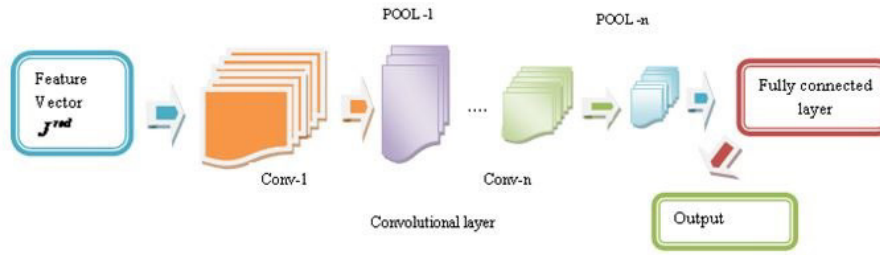


FIG. 3.2. Architecture of DCNN for the construction of prediction map

3.2.1. Architecture of the DCNN. The basic architecture of the DCNN [29] is discussed in this section with its architecture in figure 3.2. The DCNN comprises of the number of convolutional (conv) layers, pooling (POOL). The architecture of the Deep CNN [34,35] is deliberated in figure 3.2 and the architecture of DCNN consists of three layers, such as pooling (POOL), convolutional (conv), and Full Connected (FC) layers. Among the three layers of DCNN, each of the layers constitutes specific function. The main function of the conv layers is to generate the feature maps from the segments of the pre-processed image and these feature maps are further sub-sampled down in the pool layers. The third layer is the FC layer, where the classification is progressed. The convolutional layer engages in mapping the input such that the input maps undergo convolution with the convolutional kernels in order to develop the output map. The size of the output map is similar to the kernel number, and the size of the kernel matrix is $[3 \times 3]$. Thus, making it clear that the conv layers is the multilayer loop of input maps, kernel weights, and output maps. In the first conv layer, there are a number of inputs and outputs, whose size reduces in the successive conv layers such that the classification accuracy of the objects depends on the number of the layers in the DCNN.

Conv layers: The responsibility of the conv layers relied on obtaining the patterns buried in the input feature vector using the conv filters that are connected using the receptive fields, which act as an interconnection between the neurons in the previous layer with the successive conv layers through the trainable weights. The feature maps are developed through the convolution of the input feature vector with the trainable weights in such a way that the trainable weights are derived using the hybrid optimization algorithm. The neurons of the single layer engage themselves in extracting the variable features available in various location based on the variable weights of the single layer. Let us assume the input to the deep CNN is G and hence, the output from the conv layer is given as

$$(G_c^b)_{i,j} = (B_c^b)_{i,j} + \sum_{d=1}^{\varpi_1^{d-1}} \sum_{k=\varpi_1^1}^{\varpi_1^1} \sum_{n=\varpi_1^1}^{\varpi_2^1} (\vartheta_{c,d}^b)_{k,n} * (J^{red})_{i+k,j+n} \quad (3.11)$$

where the symbol $*$ refers to the convolutional operator that paves way for obtaining the local patterns from the alternative conv layers, J^{red} is the extracted features, $(B_c^b)_{i,j}$ indicates the fixed feature map or the output from the b^{th} conv layers centered as (i, j) . The output from the previous $(b-1)^{th}$ layer forms the input to the l^{th} conv layer. Let the weights of the conv layers be denoted as, $\vartheta_{c,d}^b$, which is the weights of b^{th} conv layer and the bias of b^{th} conv layer is denoted as (B_c^b) . Let us consider d, k and n as the notations of feature maps.

ReLU layer: ReLU is abbreviated as Rectified Linear Unit that applies non-saturating activation function. It eliminates the negative values effectively from the activation map by fixing them to zero, also improves the nonlinear properties of decision function without affecting the receptive fields of the convolution layer. The neurons in conv layers are arranged in 3-dimensions along the depth, height and width, so as for extracting the features from all the dimensions of ReLU layer, which uses an element-wise activation function to simplify the computation using the removal of negative values. The output from the l^{th} layer is the activation function of the preceding $(k-1)^{th}$ layer, and is expressed as

$$G_c^b = Afn(G_c^{b-1}) \quad (3.12)$$

The importance of ReLU layer is regarding the speed of DCNN, which is enhanced and offers the ability to deal with large number of networks.

POOL layers: It is a non-parametric layer with no weights, and bias, undergoing a fixed operation. The importance of POOL layer is to mitigate the spatial dimensions of the input and minimizes the computational complexity.

FC layers: The patterns generated using the pooling and the conv layers form the input to the fully connected layers that are subjected to high-level reasoning. The output from the fully-connected layers is given as

$$(H_c^b) = \delta(G_c^{b-1}) \text{with}(G_c^b) = \sum_{d=1}^{\varpi_1^{d-1}} \sum_{k=\varpi_1^1}^{\varpi_1^1} \sum_{n=\varpi_1^1}^{\varpi_2^1} (\vartheta_{c,d}^l)_{k,n} * (J^{red})_{i+k,j+n} \quad (3.13)$$

where $(\vartheta_{c,d}^l)_{k,n}$ denotes the weight.

3.2.2. Training of DCNN based on Chicken-Moth search optimization. The proposed CMSO is the integration of CSO algorithm and MSO. The CSO [37] is a bio-inspired optimization approach, which mimics the hierarchy of chickens swarm and the behavior of searching the food, in which each chicken denotes the potential solution for optimization issue. The hierarchical order is very important in social lives of chicken. Here, the chicken swarms are categorized into many hens, one rooster, and chicks. For each group, the chicken finds the rooster, hen, and chick based on fitness value of chicken. MSO [36] is an advanced meta-heuristic algorithm motivated with photo axis and levy flights of the moths. MS algorithm can search the best solution effectively with improved accuracy. Moreover, the algorithm negotiates complex operations, and thus, the execution of MS algorithm is easy and flexible. The steps involved in the proposed CMSO are described as follows.

a) Initialization: In the first step, the position of the moths is randomly initialized, represented as $\{X_{ef}, 1 \leq e \leq t; 1 \leq f \leq p\}$ where t is the population size, and p denotes the dimension. $X \in \{(\vartheta_{c,d}^l, B_c^b, \vartheta_{c,d}^b)\}$.

b) Evaluation of the objective function: The selection of the optimal location of the chicken is performed based on minimization problem. The minimal value of the objective function describes the better solution and therefore, the solution with the minimum value of the error is chosen as the best solution. The error is determined as

$$MSE = \frac{1}{X} \left[\sum_{h=1}^X H_{target} - H_c^b \right] \quad (3.14)$$

where H_{target} , H_c^b and are the estimated and target output of the classifier. The term X denotes the total number of samples.

c) Location update using levy flights: After evaluating the objective function, the solution undergoes position update based on the levy flight update, and it is mentioned as follows:

$$X_{e,f}^{\tau+1} = X_{e,f}^{\tau} + \varepsilon.F(z) \quad (3.15)$$

Rearranging the above equation,

$$X_{e,f}^{\tau} = X_{e,f}^{\tau+1} - \varepsilon.F(z) \quad (3.16)$$

where $X_{e,f}^{\tau}$ specifies the location of the moth at the iteration τ , and the term $F(Z)$ signifies to the step drawn due to the movement of levy flight. The parameter ε indicates the scaling factor and is expressed as,

$$\varepsilon = \frac{W_{max}}{\tau^2} \quad (3.17)$$

where W_{max} refer to the maximum step walk. Then, the levy distribution $H(t)$ is represented as

$$H(t) = \frac{(\alpha - 1) \lceil (\alpha - 1) \sin(\frac{\pi(\alpha - 1)}{2}) \rceil}{\pi a^2} \quad (3.18)$$

where q is greater than 0. $\Gamma(y)$ is the gamma function.

d) Fly straightly: The location of the moth is also influenced by light source, and the upgrade solution is represented as follows:

$$X_{e,f}^{\tau+1} = \lambda \times (X_{e,f}^{\tau} + \beta \cdot (X_{best}^{\tau} - X_{e,f}^{\tau})) \quad (3.19)$$

where X_{best}^{τ} denotes the best location of the moth, and the term β signifies the acceleration factor. The scaling factor is represented as λ . During the fly straightly movement, the location of the moth is influenced by the location of the light source. Here, the acceleration constant influences the convergence speed of algorithm. In some cases, the position of the moth goes beyond the position of the light source. Then, the equation (16) is modified with the CSO for enhancing the effectiveness of the approach and to find solutions to various optimization issues. The standard equation of the movement of chick is given by

$$X_{e,f}^{\tau+1} = X_{e,f}^{\tau} + AB * (X_{w,f}^{\tau} - X_{e,f}^{\tau}) \quad (3.20)$$

$$X_{e,f}^{\tau+1} = X_{e,f}^{\tau} [1 - AB] + AB * X_{w,f}^{\tau} \quad (3.21)$$

Substituting equation (16) in equation (21),

$$X_{e,f}^{\tau+1} = [X_{e,f}^{\tau+1} - \varepsilon \cdot F(z)] [1 - AB] + AB * X_{w,f}^{\tau} \quad (3.22)$$

$$X_{e,f}^{\tau+1} - [1 - AB] X_{e,f}^{\tau+1} = (AB - 1) \varepsilon \cdot F(z) + AB * X_{w,f}^{\tau} \quad (3.23)$$

$$X_{e,f}^{\tau+1} = \frac{1}{AB} [(AB - 1) \varepsilon \cdot F(z) + AB * X_{w,f}^{\tau}] \quad (3.24)$$

Thus by equation (24), the position update of the moths can be obtained, using the location of the moths in its preceding iteration, light absorption coefficient, attractiveness, and the distance between the moths.

e) Finding the best solution: The feasibility of the solution is computed based on the objective function. If the newly generated solution is best than the previous one, then it is changed by the new solution.

f) Termination: After a certain iteration limit, the algorithm terminates, and the optimal solution is retained at the end of the procedure. Thus, the best solution, chosen using the proposed CMSO algorithm is utilized for embedding the secret message. The predicted map is denoted as O .

3.2.3. Embedding using Tetrolet transform. This section presents the embedding phase using Tetrolet transform for image steganography. The embedding is utilized for hiding the secret message in the HL band, and the tetrolet transform is given for extraction and embedding process. During embedding the input image is divided into sub-bands and the watermark is embedded using tetrolet coefficient. The embedding process of pixel prediction approach is depicted in figure 3.3.

At first, the sub-bands of the input medical image are generated based on tetrolet transform in the embedding phase. The tetrolet transform is employed for acquiring tetrolet coefficient. When tetrolet transform is applied to input medical image, four bands are generated, and is expressed as,

$$TT(M) = \{L_1, L_2, L_3, L_4\} \quad (3.25)$$

In the tetrolet coefficient, each band refers to frequency and energy. Embedding the watermark image is performed in band L_3 is denoted as

$$L_3^* = L_3 + U^* \gamma^* O \quad (3.26)$$

where the term U denotes the watermark message, and L_3 is the symbol of HL band. The term represents the HL band embedded by the watermark message, and γ be the embedding strength. The secret message U is embedded in L_3 using the predicted map and the embedding strength from the classifier. The predicted map in the classifier determines the appropriate pixels and the intensity is defined using the embedding strength. The watermark image embedded to the HL band is expressed as, $\{L_1, L_2, L_3^*, L_4\}$. After the embedded of watermark message, the inverse tetrolet transform is applied to obtain the embedded image. The embedded input medical image is expressed as

$$M^* = ITT\{L_1, L_2, L_3^*, L_4\} \quad (3.27)$$

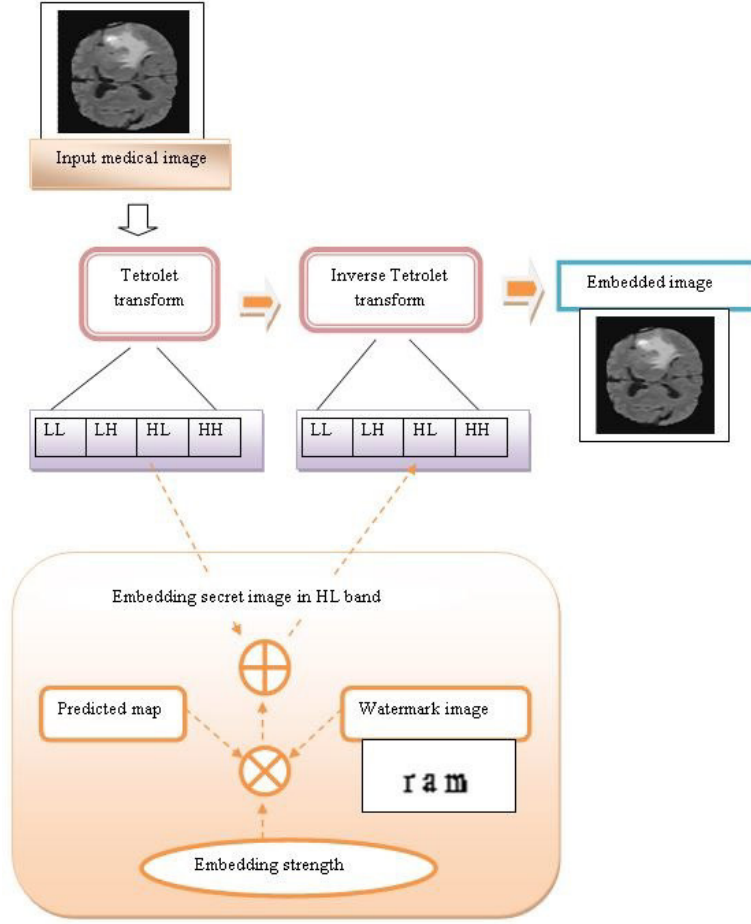


FIG. 3.3. Embedding process of the secret image

3.2.4. Watermark message retrieval. Once the embedded image is determined, then, the image is transmitted to the extraction phase. The TT is applied for obtaining tetrolet coefficients. At the extraction steps, the receiver extracts the watermark message. The extraction of secret message is expressed as

$$TT(M^*) = \{N_1^{ex}, N_2^{ex}, N_3^{ex*}, N_4^{ex}\} \quad (3.28)$$

where N_1^{ex} , N_2^{ex} , N_3^{ex*} and N_4^{ex} refers to the four bands obtained from the tetrolet process. The extraction process is formulated as,

$$T^{ex} = L_3^{E_{x*}} - L_3 \quad (3.29)$$

In the extraction phase, the watermarked message is represented as $U^{ex} = \{U_{xy}^{ex}\}$, which is utilized for steganography.

4. Results and Discussion. The results and discussion of the developed CMSO-DCNN+tetrolet for image steganography are demonstrated in this section with an effective comparative analysis to prove the effectiveness of proposed method.

4.1. Experimental Setup. The experimentation of image steganography method is performed in system with 2 GB RAM, Intel i-3 core processor, Windows 10 Operating System. The proposed method is executed in MATLAB.

4.2. Database description. The dataset is taken from the BRATS database [30] for image steganography. Here, the image of every patient is collected as four modalities, like T1, T1C, FLAIR, and T2. In this dataset, all the datasets are manually segmented, by one to four rates, which follow the similar annotation protocol, approved by experienced doctors.

4.3. Performance metrics. The evaluation of the developed model is performed based on three metrics namely, Correlation factor, SSIM, and PSNR.

a) PSNR: The quality of frame is determined using PSNR. The maximum value of PSNR assures that the system is better and it is represented in decibel (dB).

$$PSNR = 10 \log_{10} \left(\frac{m_{max}^2}{MSE} \right) \quad (4.1)$$

where the term m_{max} indicates the highest pixel value for M^{th} image.

b) SSIM index: For predicting the perceived quality of the video frame, SSIM is used. The SSIM value is maximal for the effective method. Here, the SSIM is measured by two windows, such as χ_1 , and χ_2 .

$$SSIM(\chi_1, \chi_2) = \frac{(2\eta_{\chi_1}\eta_{\chi_2} + \varphi_1)(2\kappa_{\chi_1, \chi_2} + \varphi_2)}{(\eta_{\chi_1}^2 + \eta_{\chi_2}^2 + \varphi_1)(\kappa_{\chi_1}^2 + \kappa_{\chi_2}^2 + \varphi_2)} \quad (4.2)$$

where η_{χ_1} and η_{χ_2} represents the mean value of pixels for two windows, and the variance of pixels are denoted as κ_{χ_1} and κ_{χ_2} denotes the variance of pixels. The terms φ_1 and φ_2 are utilized for stabilization.

c) Correlation factor: The correlation coefficient offers statistical relationship among the original image and embedded video image.

$$CF(\chi_1, \chi_2) = \frac{Cov(\chi_1, \chi_2)}{\kappa_{\chi_1}, \kappa_{\chi_2}} \quad (4.3)$$

where the term $Cov(\chi_1, \chi_2)$ represents the covariance factor.

4.4. Experimental Results. The experimental results obtained by the developed technique are discussed in this section. Figure 4.1 depicts the experimental results obtained from the proposed method without using noise in the image, and salt and pepper noise, impulse noise, and Gaussian noise added in the image. Figure 4.1 a) depicts the input image, and figure 4.1 b) depicts the watermark message. The embedded image is shown in figure 4.1 c), and figure 4.1 d) depicts the final extracted message.

4.5. Comparative techniques. The methods, such as random [31], sequential [32], optimal order [33], SVNN-wavelet [34], Cost Function for Image Steganography Using Wavelet (CWSM) [35], SVNN-Contourlet, Moth+tetrolet, DCNN+Contourlet, and are used for the comparison with the proposed CMSO-DCNN+tetrolet for the analysis.

4.6. Comparative analysis.

4.6.1. Analysis using image without noise. The comparative analysis of the developed method is analyzed based on correlation coefficient, PSNR, and SSIM without adding noise in the image is shown in figure 4.2. Figure 4.2 a) illustrates the analysis based on correlation coefficient by varying the number of images. When the number of image is 1, then the corresponding correlation coefficient values computed by existing random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, DCNN+Contourlet, and the proposed CMSO-DCNN+tetrolet are found to be 0.93, 0.98, 0.98, 0.98, 0.98, 0.98, 0.978, 0.98, and 0.98, respectively. The comparative analysis based on PSNR is depicted in figure 4.2 b). For image 2, the PSNR values achieved by random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, and the proposed model are 39.15dB, 39.533dB, 40.08dB, 40.02dB, 42.81dB, 42.81dB, 46.43dB, 46.71dB, and 47dB, respectively. The analysis in terms of SSIM is depicted in figure 4.2 c). For image 3, the existing techniques, like random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, possesses the SSIM of 0.484, 0.497, 0.959, 0.959, 0.959, 0.96, 0.956, and 0.952, respectively, which is comparatively lower than the CMSO-DCNN+tetrolet. For the same image, the developed CMSO-DCNN+tetrolet acquired the SSIM of 0.96.

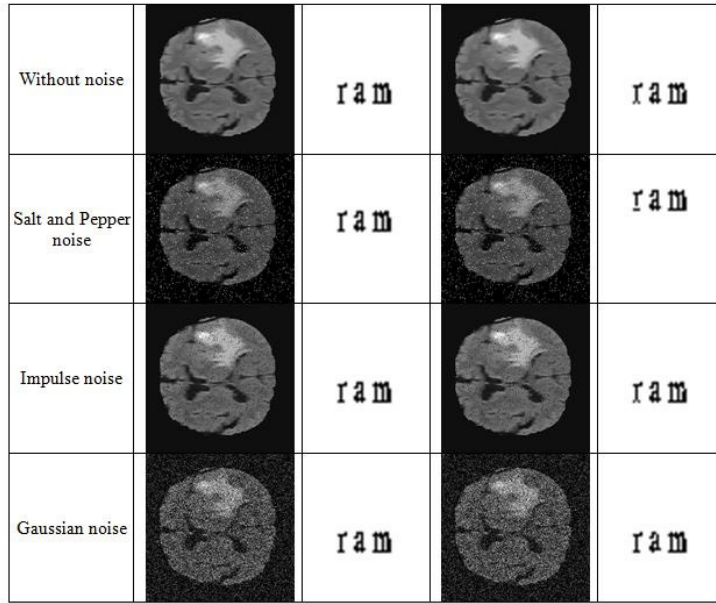


FIG. 4.1. Sample results a) Input image b) Watermark message c) Embedded image, and d) Final extracted message.

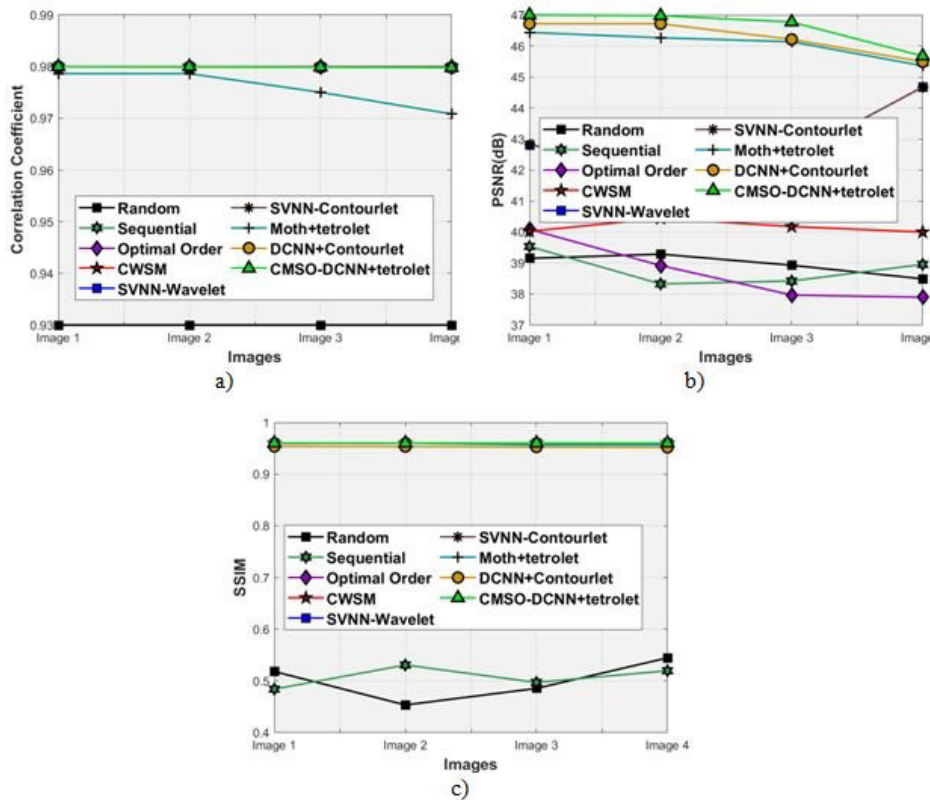


FIG. 4.2. Comparative analysis using image without noise (a) Correlation coefficient, b) PSNR, and (c) SSIM

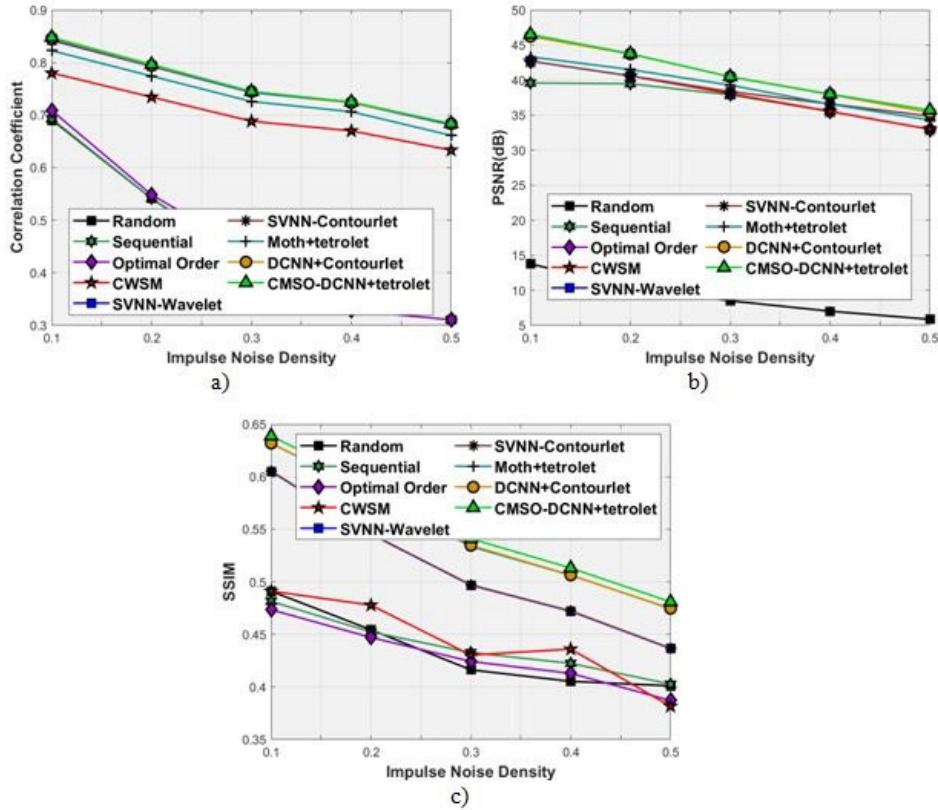


FIG. 4.3. Comparative analysis by adding impulse noise (a) Correlation coefficient, b) PSNR, and (c) SSIM

4.6.2. Analysis using image with Impulse noise. The comparative analysis of the developed method is analyzed based on correlation coefficient, PSNR, and SSIM with impulse noise is shown in figure 4.3. Figure 4.3 a) shows the analysis in terms of correlation coefficient by varying the impulse noise density. Similarly, when the impulse noise density is increased to 0.5, the methods, random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, attained the correlation coefficient of 0.3105, 0.3105, 0.3105, 0.6333, 0.6816, 0.6826, 0.6613, and 0.6837, whereas the correlation coefficient of the developed method is 0.683. The comparative analysis based on PSNR is depicted in figure 4.3 b). When the impulse noise density=0.2, the PSNR values achieved by random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, and the proposed model are 10.53dB, 39.48dB, 40.59dB, 40.59dB, 40.601dB, 40.608dB, 41.512dB, 43.71dB, and 43.77dB, respectively. The analysis in terms of SSIM is depicted in figure 4.3 c). When the impulse noise density is 0.3, the existing techniques, like random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, possesses the SSIM of 0.416, 0.432, 0.424, 0.430, 0.496, 0.497, 0.533, and 0.534, respectively, which is comparatively lower than the CMSO-DCNN+tetrolet. For the same impulse noise density, the developed CMSO-DCNN+tetrolet acquired the SSIM of 0.541.

4.6.3. Analysis using image with salt and pepper noise. The comparative analysis of the developed method is analyzed based on correlation coefficient, PSNR, and SSIM with salt and pepper noise is depicted in figure 4.4. Figure 4.4 a) illustrates the analysis based on correlation coefficient by varying the number of salt and pepper noise density. When the salt and pepper noise density=0.3, the existing techniques, like random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, DCNN+Contourlet, and the proposed CMSO-DCNN+tetrolet possesses the correlation coefficient of 0.4129, 0.4167, 0.4379, 0.688, 0.7434,

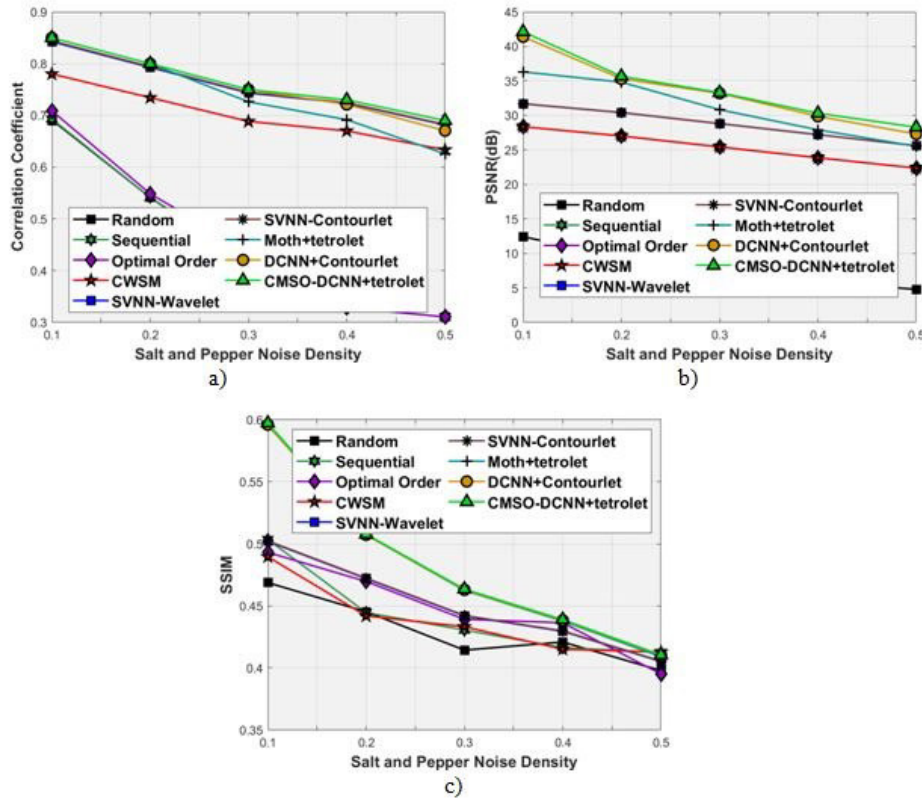


FIG. 4.4. Comparative analysis by adding salt and pepper noise (a) Correlation coefficient, b) PSNR, and (c) SSIM

0.7445, 0.726, 0.75, and 0.75, respectively. The comparative analysis based on PSNR is depicted in figure 4.4 b). When the density of salt and pepper noise is 0.3, the PSNR values achieved by random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, and the proposed model are 7.308 dB, 25.42 dB, 25.429 dB, 25.432 dB, 28.806 dB, 28.824 dB, 30.830 dB, 33.211 dB, and 33.30 dB respectively. The analysis in terms of SSIM is depicted in figure 4.4 c). When the salt and pepper noise density is 0.4, the existing techniques, like random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, possesses the SSIM of 0.421, 0.416, 0.436, 0.415, 0.429, 0.429, 0.437, and 0.438, respectively, which is comparatively lower than the CMSO-DCNN+tetrolet. For the same noise density, the developed CMSO-DCNN+tetrolet acquired the SSIM of 0.4389.

4.6.4. Analysis using image with Gaussian noise. The comparative analysis of the developed method is analyzed based on correlation coefficient, PSNR, and SSIM with Gaussian noise is depicted in figure 4.5. Figure 4.5 a) illustrates the analysis based on correlation coefficient by varying the Gaussian noise variance. When the Gaussian noise variance 0.4 is considered, the existing techniques, like random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, DCNN+Contourlet, and the proposed CMSO-DCNN+tetrolet possesses the correlation coefficient of 0.328, 0.328, 0.328, 0.670, 0.7130, 0.7140, 0.7389, 0.7189, and 0.7201, respectively. The comparative analysis based on PSNR is depicted in figure 4.5 b). When the Gaussian noise variance is 0.2, the PSNR values achieved by random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, and the proposed model are 8.121 dB, 34.687 dB, 34.701 dB, 34.709 dB, 39.479 dB, 39.497 dB, 43.231 dB, 43.480 dB, and 43.491 dB, respectively. The analysis in terms of SSIM is depicted in figure 4.5 c). When the Gaussian noise variance=0.3, the existing techniques, like random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, possesses the SSIM of 0.439, 0.422, 0.429, 0.417, 0.451, 0.4515, 0.456,

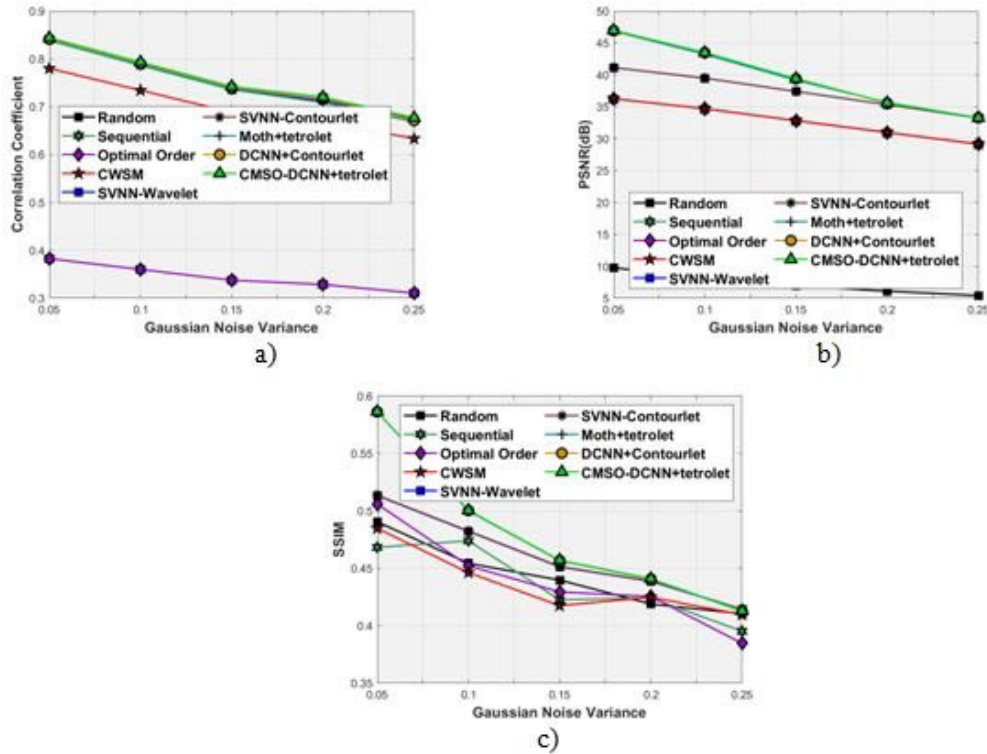


FIG. 4.5. Comparative analysis by adding Gaussian noise (a) Correlation coefficient, b) PSNR, and (c) SSIM

and 0.4567, respectively, which is comparatively lower than the CMSO-DCNN+tetrolet. For the same Gaussian noise variance, the developed CMSO-DCNN+tetrolet acquired the SSIM of 0.456.

4.7. Comparative discussion. Table 4.1 depicts the comparative discussion of the existing random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, and the proposed CMSO-DCNN+tetrolet in terms of correlation coefficient, PSNR, and SSIM parameters with impulse, salt and pepper, and Gaussian noise present in the image. The maximum performance measured by proposed CMSO-DCNN+tetrolet in terms of correlation coefficient parameter is 0.85, whereas the correlation coefficient values of existing random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, are 0.691, 0.693, 0.708, 0.780, 0.842, 0.844, 0.822, and 0.846, respectively. The maximal PSNR achieved by the proposed CMSO-DCNN+tetrolet is 46.981 dB, whereas the existing random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, acquired the PSNR of 13.748 dB, 39.6 dB, 42.699 dB, 42.704 dB, 42.705 dB, 42.713 dB, 46.86 dB, and 46.935 dB, respectively. The SSIM value computed by proposed CMSO-DCNN+tetrolet is 0.6388, whereas the existing random, sequential, optimal order, CWSM, SVNN-Wavelet, SVNN-Contourlet, Moth+tetrolet, and DCNN+Contourlet, methods acquired the SSIM of 0.490, 0.503, 0.505, 0.491, 0.605, 0.6316, and 0.6319, respectively. It is clearer that the proposed method acquired a maximal correlation coefficient, PSNR, and SSIM.

5. Conclusion. This paper presents the pixel prediction approach based on DCNN classifier and tetrolet transform. This framework employs the medical input image, and the tetrolet Transform is employed for hiding the sensitive information. At first, the best pixels are found out from the image using DCNN classifier, and are trained by CSO and MSO. The Tetrolet transform employs the embedding strength and the coefficient of Transform is employed for embedding the secret message to the input image. At last, the watermark message,

TABLE 4.1
Analysis based on the image with noise

Methods	Correlation coefficient	PSNR (dB)	SSIM
Random	0.691	13.748	0.490
Sequential	0.693	39.6	0.503
Optimal order	0.708	42.699	0.505
CWSM	0.780	42.704	0.491
SVNN-Wavelet	0.842	42.705	0.605
SVNN-Contourlet	0.844	42.713	0.605
Moth+tetrolet	0.822	46.86	0.6316
DCNN+Contourlet	0.846	46.935	0.6319
Proposed CMSO-DCNN+tetrolet	0.85	46.981	0.6388

and the input image gets extracted based on inverse Tetrolet Transform coefficient. The performance of the CMSO-DCNN+tetrolet is evaluated based on correlation coefficient, PSNR, and SSIM. The proposed method produces the maximal correlation coefficient of 0.85, maximal PSNR of 46.981dB, and the maximal SSIM of 0.6388, by applying the impulse, salt and pepper noise, and Gaussian noise in the image that indicates the superiority of proposed method. The future dimension of the research will be concentrated on extending the analysis using other standard databases with highly advanced features.

REFERENCES

- [1] P. EZE, U. PARAMPALLI, R. EVANS, AND D. LIU, *Integrity Verification in Medical Image Retrieval Systems using Spread Spectrum Steganography*, In Proceedings of International Conference on Multimedia Retrieval, (2019), pp. 53-57.
- [2] S. ARUNKUMAR, V. SUBRAMANIASWAMY, V. VIJAYAKUMAR, N. CHILAMKURTI, AND R. LOGESH, *SVD-based Robust Image Steganographic Scheme using RIWT and DCT for Secure Transmission of Medical Images*, Measurement, 139 (2019), pp. 426-437.
- [3] R. F. MANSOUR, ELSAID M. ABDELRAHIM, *An evolutionary computing enriched RS attack resilient medical image steganography model for telemedicine applications*, Multidimensional Systems and Signal Processing, 30(2) (2017). pp. 791-814.
- [4] X. LIAO, J. YIN, S. GUO, X. LI, A. K. SANGAIAH, *Medical JPEG image steganography based on preserving inter-block dependencies*, Computers and Electrical Engineering, (2017), pp. 1-10.
- [5] D. K. SARMAH, AND A. J. KULKARNI, *Image Steganography Capacity Improvement Using Cohort Intelligence and Modified Multi-Random Start Local Search Methods*, Arabian Journal for Science and Engineering, 43(8) (2018), pp. 3927-3950.
- [6] S. I. NIPANIKAR, AND V. H. DEEPTHI, *A Multiple Criteria-Based Cost Function Using Wavelet and Edge Transformation for Medical Image Steganography*, Journal of Intelligent Systems, 27(3) (2018), pp. 331-347.
- [7] T. DENEMARK, AND J. FRIDRICH, *Steganography with Multiple JPEG Images of the Same Scene*, IEEE Transactions on Information Forensics and Security, 12(10) (2017), pp. 2308-2319.
- [8] D. HU, L. WANG, W. JIANG, S. ZHENG, B. LI, *A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks*, IEEE Access, 6 (2018), pp. 38303-38314.
- [9] N. G. KINI, GAUTAM AND V. G. KINI, *A Parallel Algorithm to Hide an Image in an Image for Secured Steganography*, In Integrated Intelligent Computing, Communication and Security, (2019), pp. 585-594.
- [10] A. GUTUB, AND M. AL-GHAMDI, *Image Based Steganography to Facilitate Improving Counting-Based Secret Sharing*, 3D Research, 10(1) (2019), pp. 6.
- [11] M. M. HASHIM, M. S. M. RAHIM, F. A. JOHI, M. S. TAHA, AND H. S. HAMAD, *Performance evaluation measurement of image steganography techniques with analysis of LSB based on variation image formats*, International Journal of Engineering and Technology, 7(4) (2018), pp. 3505-3514.
- [12] K. MUHAMMAD, M. SAJJAD, I. MEHMOOD, S. RHO, AND S. W. BAIK, *Image steganography using uncorrelated color space and its application for security of visual contents in online social networks*, Future Generation Computer Systems, 86 (2018), pp. 951-960.
- [13] B. IN LI, M. WANG, X. LI, S. TAN, AND J. HUANG, *A Strategy of Clustering Modification Directions in Spatial Image Steganography*, IEEE Transactions on Information Forensics and Security, 10(9) (2015), pp. 1905-1917.
- [14] G. LINJIE, N. JIANGQUN, AND S. Y. QING, *Uniform Embedding for Efficient JPEG Steganography*, IEEE Transactions on Information Forensics and Security, 9 (2014), pp. 814-825.
- [15] W. MAZURCZYK AND L. CAVIGLIONE, *Steganography in Modern Smartphones and Mitigation Techniques*, Communications Surveys & Tutorials, IEEE, 17 (2014), pp. 334-357.
- [16] Z. LIU, F. ZHANG, J. WANG, H. WANG, AND J. HUANG, *Authentication and recovery algorithm for speech signal based on digital watermarking*, Signal Processing, 123 (2015), pp. 157-166.
- [17] J. LI, X. LI, B. YANG, AND X. SUN, *Segmentation-based image copy-move forgery detection scheme*, IEEE Transactions on Information Forensics and Security, 10 (2015), pp. 507-518.
- [18] A. KHAN, A. SIDDIQA, S. MUNIB, AND S. A. MALIK, *A recent survey of reversible watermarking techniques*, Information

- Sciences, 279 (2014), pp. 251-272.
- [19] S. M. M. KARIM, S. RAHMAN, I. HOSSAIN, *New Approach for LSB Based Image Steganography using Secret Key*, In Proceedings of 14th International Conference on Computer and Information Technology, (2011), pp. 22-24.
- [20] Y. YEUNG, W. LU, Y. XU, J. CHEN, AND R. LI, *Secure binary image steganography based on LTP distortion minimization*, Multimedia Tools and Applications, 2019, pp. 1-22.
- [21] W. TANG, B. LI, S. TAN, M. BARNI, AND J. HUANG, *CNN-based Adversarial Embedding for Image Steganography*, IEEE Transactions on Information Forensics and Security, 14(8) 2019, pp. 2074-2087.
- [22] S. CHAKRABORTY, A. S. JALAL, AND C. BHATNAGAR, *LSB based non blind predictive edge adaptive image steganography*, Multimedia Tools and Applications, 76(6) 2017, pp. 7973-7987.
- [23] A. MIRI, AND K. FAEZ, *An image steganography method based on integer wavelet transform*, Multimedia Tools and Applications, 77(11) 2018, pp. 13133-13144.
- [24] G. S. LIN, Y. T. CHANG, AND W. N LIE, *A framework of enhancing image steganography with picture quality optimization and anti-steganalysis based on simulated annealing algorithm*, IEEE Transactions on Multimedia, 12(5) (2010), pp. 345-357.
- [25] J. KROMMWEH, *Tetrolet Transform: A New Adaptive Haar Wavelet Algorithm for Sparse Image Representation*, Journal of Visual Communication and Image Representation, 21(4) 2010, pp. 364-374.
- [26] PANDEY, PRATEEKSHIT, R. SINGH, AND M. VATSA, *Face recognition using scattering wavelet under Illicit Drug Abuse variations*, In Proceedings of International Conference on Biometrics (ICB), Halmstad, Sweden, (2016), pp. 1-6.
- [27] W. K. KONG, D. ZHANG, AND W. LI, *Palmprint feature extraction using 2-D Gabor filters*, Pattern recognition, 36(10) (2003), pp. 2339-2347.
- [28] Z. GUO, L. ZHANG AND D. ZHANG, *A Completed Modeling of Local Binary Pattern Operator for Texture Classification*, IEEE Transactions on Image Processing, 19(6) (2010), pp. 1657-1663.
- [29] F. TU, S. YIN, P. OUYANG, S. TANG, L. LIU, AND S. WEI, *Deep Convolutional Neural Network Architecture With Reconfigurable Computation Patterns*, IEEE Transactions on very large scale integration (VLSI) systems, 25(8) 2017, pp. 2220 - 2233.
- [30] BRATS BRAIN TUMOR DATABASE, <https://www.smir.ch/BRATS/Start2015>, Accessed on (2019).
- [31] S. ISLAM, M. R. MODI AND P. GUPTA, *Edge-based image steganography*, EURASIP Journal on Information Security, 2014(1) (2014).
- [32] M. RAMALINGAM AND N. A. M. ISA, *A steganography approach for sequential data encoding and decoding in video images*, In proceedings of International Conference on Computer, Control, Informatics and Its Applications (IC3INA), Bandung, Indonesia, (2014), pp. 120-125.
- [33] B. LI, M. WANG, X. LI, S. TAN AND J. HUANG, *A Strategy of Clustering Modification Directions in Spatial Image Steganography*, IEEE Transactions on Information Forensics and Security, 10(9) (2015), pp. 1905-1917.
- [34] S. NIKOLOV, P. HILL, D. BULL, AND N. CANAGARAJAH, *Wavelets for image fusion*, In: Petrosian A.A., Meyer F.G. (eds) *Wavelets in Signal and Image Analysis*, Computational Imaging and Vision, 19 (2001), pp. 213-241.
- [35] S. NIPANIKAR, V. H. DEEPTHI, *A Multiple Criteria-Based Cost Function Using Wavelet and Edge Transformation for Medical Image Steganography*, Journal of Intelligent Systems, 27(3) (2016).
- [36] G. -G. WANG, *Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems*, Memetic Computing, 10(2) (2018), pp. 151-164.
- [37] X. MENG, Y. LIU, X. GAO, AND H. ZHANG, *A New Bio-inspired Algorithm: Chicken Swarm Optimization Xianbing*, In proceedings of International conference in swarm intelligence, (2018), pp. 86-94.
- [38] V. VINOLIN AND S. VINUSHA, *Edge-based Image Steganography using Edge Least Significant Bit (ELSB) Technique*, Multimedia Research, 1(1) (2018), pp. 9-16.
- [39] S. VINUSHA, *Secret Image Sharing and Steganography Using Haar Wavelet Transform*, Multimedia Research, (2019), 2(2), pp. 28-34.
- [40] M. MUKHEDKAR, P. POWAR, AND P. GAIKWAD, *Secure non real time image encryption algorithm development using cryptography & Steganography*, In Proceedings of the Annual IEEE India Conference (INDICON), (2015), pp. 1-6.

Edited by: P. Vijaya

Received: Dec 9, 2019

Accepted: Jun 22, 2020



AN EFFICIENT DYNAMIC SLOT SCHEDULING ALGORITHM FOR WSN MAC: A DISTRIBUTED APPROACH

MANAS RANJAN LENKA* AND AMULYA RATNA SWAIN†

Abstract. In the current scenario, the growth of IoT based solutions gives rise to the rapid utilisation of WSN. With energy constraint sensor nodes in WSN, the design of energy efficient MAC protocol along with timeliness requirement to handle collision is of paramount importance. Most of the MAC protocols designed for a sensor network follows either contention or scheduled based approach. Contention based approach adapts well to topology changes, whereas it is more costly in handling collision as compared to a schedule based approach. Hence, to reduce the collision along with timeliness, an effective TDMA based slot scheduling algorithm needs to be designed. In this paper, we propose a TDMA based algorithm named DYSS that meets both the timeliness and energy efficiency in handling the collision. This algorithm finds an effective way of preparing the initial schedule by using the average two-hop neighbors count. Finally, the remaining un-allotted nodes are dynamically assigned to slots using a novel approach. The efficiency of the algorithm is evaluated in terms of the number of slots allotted and time elapsed to construct the schedule using the Castalia simulator.

Key words: WSN, TDMA Slot Scheduling, Correlated Contention, Data Delivery Latency, Feasible Schedule.

Abbreviations. Internet of Things (IoT), Wireless Sensor Network (WSN), Media Access Control (MAC), Time Division Multiple Access (TDMA), Carrier Sense Multiple Access (CSMA), Randomized TDMA (RAND), Distributed RAND (DRAND), Randomized Distributed TDMA (RD-TDMA), Hybrid based Distributed Slot Scheduling (HDSS), and Dynamic Slot Scheduling (DYSS).

AMS subject classifications. 68M14

1. Introduction. In the real world, the widespread use of WSN mainly owes to monitor and control various devices in several industrial and home automation systems. A sensor network is unique in the sense that the nodes are battery powered and mostly can not be recharged from time to time. Hence, every application in a WSN environment must be designed in such a way that energy consumption must be minimal. Along with energy efficiency, other requirements like timeliness, collision handling, and reduced latency during data transmission also need to be considered. To handle collision along with the above requirements, the design of an efficient MAC protocol is of paramount importance. Studies carried out by Ergen et al. [1] revealed that TDMA base MAC protocols perform better than CSMA based MAC protocols in a WSN to satisfy these stringent requirements [14, 16, 19, 20, 21, 22].

In TDMA based MAC protocols [15, 17, 18, 24, 25], a schedule with several time slots is prepared where each sensor node is assigned to a particular time slot. The assignment of slots to each sensor node is carried out in such a manner that collision is handled with reduced data delivery latency. Further, to minimise energy consumption, the nodes are put to sleep for it's allocated time slot as proposed in [2, 3, 26]. The delivery latency is reduced through proper scheduling of the TDMA slots as proposed by Moriyama et al. [4], Ahmad et al. [5], and Ahmad et al. [6]. The collision during data transmission is handled in such a way that the nodes interfering with each other are not assigned to the same slot. Most of the TDMA based slot scheduling algorithms prepare an optimal schedule that normally takes more time to prepare a schedule. However, in the case of correlated contention, preparing an optimal schedule may not be of much use as the contention exists for a short duration. Hence, a feasible schedule needs to be prepared in a quick time to handle the correlated contention, as proposed by Lenka et al. [7], Lenka et al. [8], and Bhatia et al. [9].

*KIIT Deemed to be University, Bhubaneswar, India (manasy2k3@gmail.com).

†KIIT Deemed to be University, Bhubaneswar, India (swainamulya@gmail.com).

Preparing a feasible schedule in a quick time may lead to latency during data transmission. Therefore, design of a scheduling algorithm has to take care of both correlated contention and the latency during data transmission to enhance efficiency. Based on the above facts, this paper proposed an efficient TDMA based dynamic slot scheduling algorithm that handles both correlated contention and delivery latency. The proposed algorithm uses the average two-hop neighbors count as opposed to the maximum two-hop neighbors count used in earlier approaches. The maximum two-hop neighbors count varies significantly from the average two-hop neighbors count in most of the sensor networks due to the random deployment of the sensor nodes. As a result, the use of average two-hop neighbors count helps to reduce the number of slots to be attached to a schedule. However, the use of average two-hop neighbors count to prepare a schedule may leave a few nodes to remain un-allotted. The remaining un-allotted nodes are then dynamically assigned to the best possible slots using a novel message passing technique in a quick time to prepare the final schedule. Finally, the proposed algorithm is implemented using Castalia simulator to evaluate its performance. The performance analysis shows that our proposed scheme outperforms DRAND, RD-TDMA, and HDSS in terms of number of slots and time to allocate the slots in a schedule.

The rest of the paper is structured as follows. A review of the related works has been summarised in Section 2. Section 3 describes our proposed algorithm. The correctness of the algorithm has been analysed through various scenarios in Section 4. The simulation results are analysed in Section 5. The conclusion is drawn in Section 6.

2. Related Work. In today's world, the recent trend is to automate everything for living a relaxed and comfortable lifestyle. WSN plays a vital role in achieving the same. In a resource constraint sensor network, the design and development of an efficient application has to deal with several challenges such as energy efficiency, collision handling, reduction in latency, reliability, etc. In this paper, we mainly focus on TDMA based slot scheduling approach that helps to handle the collision and reduce the latency during data transmission.

Ahmad et al. [6] proposed a centralised TDMA scheduling for clustered based tree topology in WSN. This algorithm prepares a collision-free clustered based schedule to satisfy the timeliness for several data flows. The timeline for each data flow is expressed based on the length of the schedule period. The algorithm takes care of minimum utilisation of energy by putting the nodes into low power mode to the maximum possible extent. As opposed to the centralised approach, Ahmad et al. [10] proposed a distributed version of the clustered based TDMA algorithm, where each cluster is capable enough to prepare its time slot for the TDMA schedule. This distributed nature of the algorithm aid to the WSN as the resources in the WSN environment is scarce.

Severino et al. [11] proposed a self-adaptive clustered based dynamic scheduling algorithm for WSN. Based on the nature of the traffic flow, this algorithm adapts different required bandwidth and latency by changing the scheduling algorithm attached to the clusters. This adaption happens in quick time by exerting a small downtime for the WSN.

Wang et al. [18] proposed a deterministic TDMA based slot scheduling approach to avoid collision during data transmission. In this algorithm, each sensor node calculate it's own time slot in a distributed manner as per the available neighborhood information. However, each sensor node need synchronisation among each other to calculate their own time slot and as a result the collision handling model was not too realistic.

Long et al. [12] proposed a multi-hop TDMA scheduling algorithm for WSN that extends the one-hop TDMA scheduling to multi-hop scheduling. This extension helps in balancing the energy consumption among the nodes in a WSN which ultimately prolongs the network lifetime.

Rhee et al. [13] proposed a distributed version of RAND, a centralised random slot scheduling algorithm. In this algorithm, each node presents in one of the four states, i.e. REQUEST, RELEASE, IDLE, and GRANT. Each node starts with IDLE state and goes for a lottery. The node that wins the lottery sends a request message for allotment of a slot and enters into the REQUEST state. The node that receives this request message enters into GRANT state provided the node is in IDLE or RELEASE state. In case, the receiver node is either in REQUEST or GRANT state while receiving this request message then it sends back a reject message to the sender. When the sender node receives a reject message, it goes back to IDLE state. Once a node, who has started the slot allotment request, receives a grant message from all of its neighbors then it enters into the RELEASE state and the requested slot is allocated to that node.

Li et al. [23] proposed a distributed TDMA slot scheduling that improves upon the DRAND algorithm.

This algorithm prepares the TDMA schedule based on the residual energy and topology associated with a sensor network. In order to reduce the energy consumption and execution time of a schedule, initially it defines the energy-topology factor and then applies the same to arrange the priority of the time of a slot in a schedule to deal with the energy consumption and execution time.

Most of the slot scheduling algorithms focus on preparing an optimal schedule to handle the collision during data transmission. In case of co-related contention (i.e. the collision that occurs at the receiver end for a very short period), preparing an optimal schedule may not help a lot. Keeping the above requirement in mind Bhatia et al. [9] proposed a feasible slot scheduling algorithm named RD-TDMA that handles the co-related contention and at the same time reduces the latency during data transmission.

Lenka et al. [8] proposed a HDSS algorithm that initially prepares a feasible schedule based on maximum two-hop neighbors count using the DRAND algorithm. The number of slots attached to the prepared feasible schedule is further reduced by reallocating certain nodes based on the ratio of average two-hop neighbors count to the maximum two-hop neighbors count. Finally, a sub-optimal schedule, with less number of slots as compared to DRAND and RD-TDMA, is prepared in quick time which ultimately handles the collision due to co-related contention and at the same time reduces the latency during data transmission.

Although the existing HDSS algorithm performs better as compared to DRAND and RD-TDMA with respect to the number of slots attached to a feasible schedule, still this paper finds a way to further fine-tuned the slot scheduling approach in a novel way to reduce the number of slots. The proposed approach focuses on preparing a feasible schedule based on the average two-hop neighbors count instead of maximum two-hop neighbors count as used in HDSS. Moreover, the proposed approach uses a novel dynamic slot scheduling technique to allot slots for the remaining nodes in the best possible way to achieve the desired goal.

3. Proposed Dynamic Slot Scheduling Algorithm. In the earlier proposed HDSS algorithm for WSN MAC, a feasible schedule is prepared based on the maximum two-hop neighbors count. Each sensor node in WSN is assigned to a particular slot in the feasible schedule and all its two-hop neighbors are assigned to different slots so that the collision during data transmission will be avoided. Nevertheless, in the real scenario, during the deployment of the sensor nodes, there is every chance that some of the regions of WSN may have very high node density as compared to other regions. In such scenarios, there will be a significant difference between the maximum two-hop neighbors and the average two-hop neighbors count due to the presence of the outliers (i.e. specific regions with high node density). Therefore, preparing a feasible schedule based on the maximum two-hop neighbors count leads to a significant rise in the number of slots required for the same. In order to get rid of the above issue, in the proposed approach, the schedule is initially prepared based on the average two-hop neighbors instead of maximum two-hop neighbors count.

To start with, in the proposed approach a feasible schedule is initially prepared based on the average two-hop neighbors count. As per the earlier discussion, since there is more possibility of higher density in some regions of WSN, the actual two-hop neighbors count in those regions will be more than the average two-hop neighbors count in the whole network. Hence, there is every possibility that some sensor nodes may remain un-allotted due to the lack of availability of slots as the feasible schedule is initially prepared based on the average two-hop neighbors count. The nodes which remain un-allotted during the preparation of the feasible schedule, those nodes will be allotted to a feasible slot through these following three phases:

1. Slot Allotment Request
2. Slot Allotment Message Handler
3. Slot Allotment Status

The notations used for the set of information maintained at each node for slot allocation in the proposed algorithm is summarises in Table 3.1.

3.1. Phase 1: Slot Allotment Request. Let N_{Avg} is the average two-hop neighbors count calculated during the preparation of feasible schedule at the very beginning and a node i has not been allotted to a feasible slot during the preparation of feasible schedule. In order to get a feasible slot, the node i selects a slot k , where $k = N_{Avg} + 1$ and verifies at its end whether this slot k has already been allotted to any one of its one-hop or two-hop neighbor nodes. In case, the slot k has already been allotted then it tries with the next slot, i.e. $k + 1$ and so on till the chosen slot is fit for him. Once the node i finds a feasible slot say k then it stores the information, i.e. $Orig_{(i,k)}$, $Sac_{(j,k)}$, $HopCnt_{(i,k)}$, $SlotReqTime_{(i,k)}$ into a vector V at its own end. Finally, node

TABLE 3.1
Set of information maintained at each node for slot allocation of the remaining nodes.

Notation	Description
$Orig_{\{(i,k)\}}$	Node i , the originator of dynamic allotment procedure for slot k
$Sac_{\{(j,k)\}}$	Intermediate node j , who sacrificed the slot k for another node
$HopCnt_{\{(i,k)\}}$	Hop distance at node i for slot k
$SlotReq$	Slot requested for Allotment
$SlotReqTime_{\{(i,k)\}}$	Time at which the request for allotment of slot k has been initiated by node i
V	A Vector of size S containing $Orig_{(i,k)}, Sac_{(j,k)}, HopCnt_{(i,k)}, SlotReqTime_{(i,k)}$

i broadcasts a '**DynamicSlotAllocation**' message that contains the stored information to all its neighbors. The slot allotment request of i^{th} node is given in Algorithm 5.

Algorithm 5: Slot allotment request for remaining nodes at node i

```

1 Procedure SLOT_ALLOTMENT_REQUEST
2 begin
3    $k = N_{Avg} + 1$  ;
4   while ( $k$  already allotted to one of it's one-hop or two-hop neighbors) do
5      $k = k + 1$  ;
6    $HopCnt_{(i,k)} = 0$  ;
7   Store ( $Orig_{(i,k)}, Sac_{(j,k)}, HopCnt_{(i,k)}, SlotReqTime_{(i,k)}$ ) into vector  $v$  ;
8   Send(DynamicSlotAllocation( $Orig_{(i,k)}, Sac_{(j,k)}, HopCnt_{(i,k)}, SlotReqTime_{(i,k)}$ )) to all
   neighbors ;

```

3.2. Phase 2: Slot Allotment Message Handler. After receiving the 'DynamicSlotAllocation' message from node i , a node j checks whether there exist an entry in the stored vector V at it's end for the requested slot k and goes through these following steps.

- [**Step 1:**] In case the entry related to the requested slot k does not exist then the received information is pushed into the vector V at it's end but with a updated value for the hop count, i.e. $HopCnt_{(j,k)} = HopCnt_{(i,k)} + 1$. Then the j^{th} node broadcasts the updated received information to all of it's neighbors provided the $HopCnt_{(j,k)}$ is less than 2.
- [**Step 2:**] If the entry related to the requested slot k exist then increment the received $HopCnt_{(i,k)}$ by one and check whether the updated received $HopCnt_{(i,k)}$ is less than the stored $HopCnt_{(j,k)}$ for the requested slot k .
- [**Step 3:**] If the above condition holds good then update the existing entry for the slot k with the received information and broadcast the same to its neighbors.
- [**Step 4:**] If the above condition does not hold then it again checks whether the received $SlotReqTime_{(i,k)}$ is less than the stored $SlotReqTime_{(j,k)}$ for the slot k . If it is found true then it checks for whether the receiver node j is the originator of slot allotment procedure for the same slot k or not. In case, the receiver node j is the originator of slot allotment procedure for the slot k then update the entry in the vector V with the received information but with modified $Sac_{(j,k)}$ value.
- [**Step 5:**] If the receiver node j is not the originator of slot allotment process for the slot k then it checks whether $Sac_{(i,k)}$ information in the received message is either same as the information stored at the originator or the sacrificed intermediate node. If it matches then the existing entry for the slot k is updated with the received information but with increment the value of the $HopCnt_{(i,k)}$ by one and broadcast the same to it's neighbors provided the incremented $HopCnt_{(j,k)}$ value is less than 3.

Algorithm 6: Slot allotment process after receive of slot allotment request message at node i

```

1 Procedure SLOT_ALLOTMENT_MESSAGE_HANDLER
2 begin
3    $msg = \text{Receive\_Message}()$ ;
4   if ( $msg == \text{DynamicSlotAllocation}(\text{Orig}_{(i,k)}, \text{Sac}_{(j,k)}, \text{HopCnt}_{(i,k)}, \text{SlotReqTime}_{(i,k)})$ ) then
5     for  $m \leftarrow 0$  to  $S - 1$  do
6       if ( $V_{[m]}.SlotReq == k$ ) then
7          $\lfloor$  break;
8       if ( $(S == 0) \parallel (\text{requested slot } k \notin V)$ ) then
9          $\text{HopCnt}_{(i,k)} = \text{HopCnt}_{(i,k)} + 1$  ;
10        push the updated received information into the vector  $v$  ;
11        if ( $\text{HopCnt}_{(i,k)} < 2$ ) then
12           $\text{Send}(\text{DynamicSlotAllocation}(\text{Orig}_{(i,k)}, \text{Sac}_{(j,k)}, \text{HopCnt}_{(i,k)}, \text{SlotReqTime}_{(i,k)}))$  to
13           $\lfloor$  all neighbors ;
14        else
15           $\text{HopCnt}_{(i,k)} = \text{HopCnt}_{(i,k)} + 1$  ;
16          if ( $\text{HopCnt}_{(i,k)} < V_{[m]}.HopCnt_{(n,k)}$ ) then
17             $\lfloor$   $\text{func1}()$ ;
18          else
19            if ( $\text{SlotReqTime}_{(i,k)} < V_{[m]}.SlotReqTime_{(n,k)}$ ) then
20              if ( $V_{[m]}.HopCnt_{(n,k)} == 0$ ) then
21                 $\lfloor$   $\text{func2}()$ ;
22              else
23                if ( $\text{Sac}_{(j,k)} == V_{[m]}.Orig_{(i,k)} \parallel (\text{Sac}_{(j,k)} == V_{[m]}.Sac_{(j,k)})$ ) then
24                  if ( $\text{HopCnt}_{(i,k)} < 2$ ) then
25                     $\lfloor$   $\text{func3}()$ ;
26                  else
27                     $\lfloor$  remove the entry at index  $m$  from the vector  $v$  ;

```

```

1 Procedure func1()
2 begin
3    $\text{HopCnt}_{(i,k)} = \text{HopCnt}_{(i,k)} + 1$  ;
4   Update the existing entry in vector  $V$  for slot  $k$  with the received information but modified
5    $\text{HopCnt}_{(i,k)}$  value ;
6   if ( $\text{HopCnt}_{(i,k)} < 2$ ) then
7      $\text{Send}(\text{DynamicSlotAllocation}(\text{Orig}_{(i,k)}, \text{Sac}_{(j,k)}, \text{HopCnt}_{(i,k)}, \text{SlotReqTime}_{(i,k)}))$  to all
8      $\lfloor$  neighbors ;

```

3.3. Phase 3: Slot Allotment Status. After certain period of time, node i checks the content of vector V at its own end. In case the vector V contains an entry with a $\text{HopCnt}_{(i,k)} = 0$, then it indicates that the node i is the originator for the slot k and hence node i is allotted with the slot k . In case no entry with a $\text{HopCnt}_{(i,k)} = 0$ is found then try with the next slot.

```

1 Procedure func2()
2 begin
3    $V_{[m]}.Sac_{(j,k)} = n$  ;
4    $HopCnt_{(i,k)} = HopCnt_{(i,k)} + 1$  ;
5   Update the existing entry in vector  $V$  for slot  $k$  with the received information but modified
      $HopCnt_{(i,k)}$  value and  $Sac_{(j,k)}$  node ;
6   Send(DynamicSlotAllocation( $Orig_{(i,k)}$ ,  $Sac_{(j,k)}$ ,  $HopCnt_{(i,k)}$ ,  $SlotReqTime_{(i,k)}$ )) to all
     neighbors ;

```

```

1 Procedure func3()
2 begin
3    $HopCnt_{(i,k)} = HopCnt_{(i,k)} + 1$  ;
4   Update the existing entry in vector  $V$  for slot  $k$  with the received information but modified
      $HopCnt_{(i,k)}$  value ;
5   Send(DynamicSlotAllocation( $Orig_{(i,k)}$ ,  $Sac_{(j,k)}$ ,  $HopCnt_{(i,k)}$ ,  $SlotReqTime_{(i,k)}$ )) to all
     neighbors ;

```

Algorithm 7: Verification of slot allotment status at node i

```

1 Procedure SLOT_ALLOTMENT_STATUS
2 begin
3   for  $j \leftarrow 0$  to  $S - 1$  do
4     if ( $v_{[j]}.HopCnt_{(i,k)} == 0$ ) then
5       Allot the slot  $k$  to node  $i$ ; Send(DynamicSlotAllotmentSucceeded( $i, k$ )) to all the nodes
         in the whole WSN ;
6       break;
7     if ( $j == S$ ) then
8       Call SLOT_ALLOTMENT_REQUEST procedure for the next slot  $k + 1$ ;

```

4. Proof of the proposed Algorithm. The correctness of the proposed algorithm is verified through various scenarios as described here.

The above steps for slot allotment are explained through an example given in Fig. 4.1. As per the scenario given in Fig. 4.1, node x_1 , x_2 , and x_7 have not been allotted to any feasible slot during the preparation of the initial schedule based on the average two-hop neighbors count. Here, x_1 and x_2 are the one-hop neighbors to each other whereas node x_7 is neither one-hop neighbor nor two-hop neighbor of node x_1 or x_2 . Let us assume that the node x_1 first starts the slot allotment procedure for the slot k at time t_1 and stores the information such as originating node, sacrificed intermediate node, hop count, requested slot, and request time stamp as $(x_1, x_1, 0, k, t_1)$ in the vector V and broadcast this information to its neighbors. When the node x_2 receives the message from node x_1 and by that time if node x_2 has already started the slot allotment procedure for the slot k at time t_2 which is latter than t_1 then it stores information such as the originating node, sacrificed intermediate node, hop count, requested slot, and request timestamp as $(x_2, x_2, 0, k, t_2)$ in the vector V at its own end. As per the proposed algorithm, since the received timestamp t_1 is less than the stored timestamp t_2 at node x_2 , so the stored information in the vector V at x_2 is updated to $(x_1, x_2, 1, k, t_1)$ and broadcast the updated information to it's neighbor. Here, the sacrificed intermediate node information remains same as x_2 as the node x_2 has sacrificed the slot k for node x_1 which starts the allocation procedure for slot k earlier than x_2 . Hence, x_2 can not be allotted to slot k and further it has to try with the next available slot.

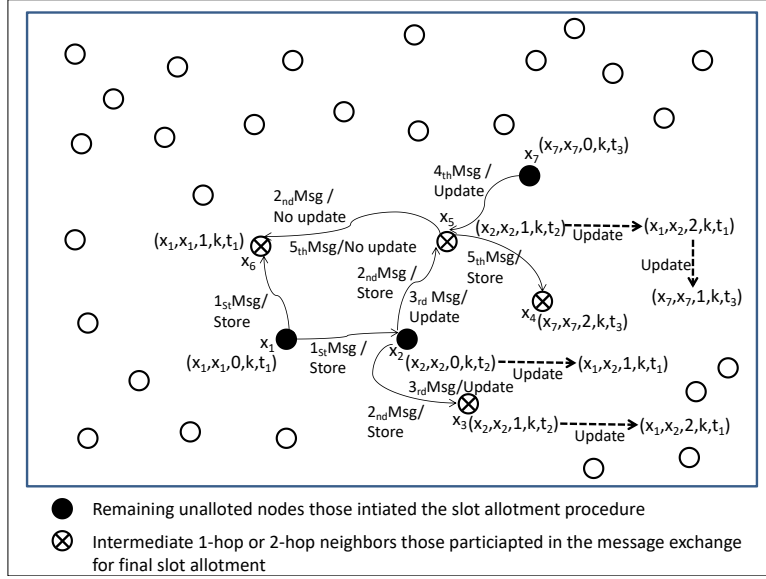


FIG. 4.1. Slot Not Allotted To One-hop Neighbors

At another instance of time t_3 , which is later than t_2 , node x_7 has started its slot allocation procedure for slot k by storing the information $(x_7, x_7, 0, k, t_3)$ into the vector V and broadcast this information to its neighbors. When a node x_5 received the message from node x_7 , by that time the node x_5 has already stored the information $(x_1, x_2, 2, k, t_1)$ in its vector V . According to the proposed algorithm, as the incremented received hop count is less than the stored hop count at node x_5 , the stored information in the vector V at node x_5 is updated to $(x_7, x_7, 1, k, t_3)$ and broadcast the updated information to its neighbor. Finally, node x_1 and x_7 are allotted to the same slot k as both these nodes are not neighbors to each other. Whereas, node x_2 is allotted to a slot other than k as it is a one-hop neighbor of x_1 .

According to the scenario given in figure 4.2, node x_1 and x_5 have not been allotted to any slot during the preparation of the initial schedule based on the average two-hop neighbors count. Here, x_1 and x_5 are two-hop neighbors to each other. Assume that the node x_1 starts the slot allotment procedure first for the slot k at time t_1 and stores the information such as originating node, sacrificed intermediate node, hop count, requested slot, and request time stamp information as $(x_1, x_1, 0, k, t_1)$ in the vector V and broadcast this information to its neighbors. Now, when the node x_2 receives the message from node x_1 and by that time node x_2 has no stored information regarding the slot k in its vector V then the node x_2 stores the received information as $(x_1, x_1, 0, k, t_2)$ in the vector V at its own end. In another instance at time t_2 , which is later than t_1 , the node x_5 has started its slot allocation procedure for slot k and stored the information $(x_5, x_5, 0, k, t_2)$ in vector V and broadcasted this information to its neighbors. Now, when the node x_2 receives the message from node x_5 , as per our algorithm, neither the received hop count nor the received timestamp is lesser than the stored one. Hence, there is no update made at the node x_2 . Now, when the node x_5 receives the message from node x_2 and as per our algorithm, since the received timestamp t_1 is less than the stored timestamp t_2 at the node x_5 , so, the stored information in the vector V at node x_5 is updated to $(x_1, x_5, 2, k, t_1)$ and broadcast the updated information to its neighbors. Here, the sacrificed intermediate node information became x_5 as the node x_5 has sacrificed the slot k for node x_1 which starts the allocation procedure for slot k earlier than x_5 . Hence, the node x_5 can not be allotted to slot k and try with the next slot. Now, when the node x_7 receives the message from node x_5 and as per our algorithm, if the received timestamp t_1 is less than the stored timestamp t_2 at node x_7 and the received sacrificed intermediate node information is same as stored originator node information then the stored information at node x_7 needs to be removed as node x_5 can no more be allotted to the slot k . Finally, based on our algorithm, the node x_1 is allotted to slot k and the node x_5 can not be allotted to the same slot k as both are the two-hop neighbors to each other. This proves the correctness of our algorithm.

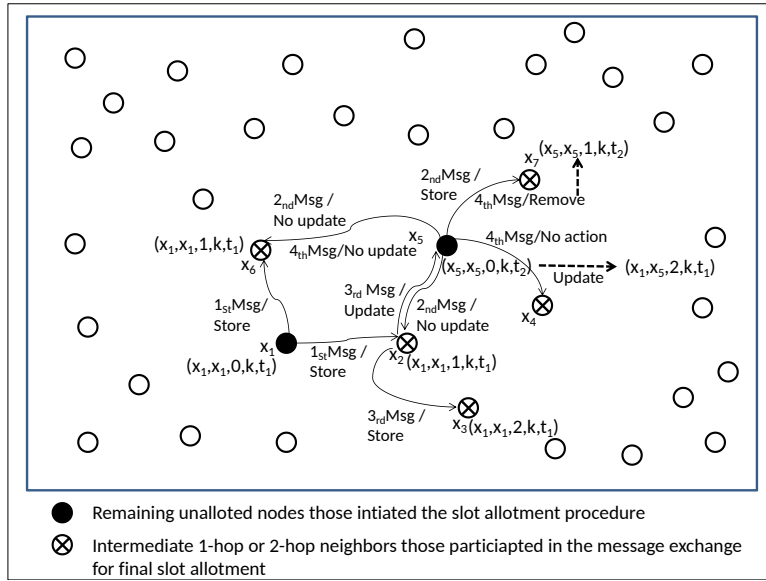


FIG. 4.2. Slot Not Allotted To Two-hop Neighbors

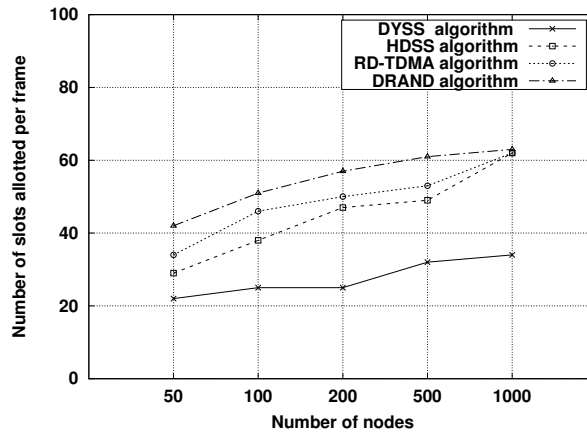


FIG. 5.1. Number of slots attached to the feasible schedule among various algorithms using uniform random distribution

5. Simulation Results. The simulation of the proposed dynamic slot allotment algorithm is carried out using the Castalia simulator[27]. The efficiency and correctness of the algorithm are tested in various deployment scenarios like uniform random distribution, grid, and randomised grid. The correctness of the algorithm is also carried out in fixed as well as varying node density with the number of sensor nodes varying from 50 to 1000. Important parameters like sensing range, data transmission range, etc. are taken into account based on the facts given in the datasheet of cc2420 [28] and TelosB [29].

Figure 5.1 shows the comparison of the efficiency of the proposed algorithm with other protocols with respect to the number of slots attached to the feasible schedule. This figure indicates that the proposed algorithm outperforms over others in terms of the number of slots allotted for preparing the feasible schedule. This result owes to the significant difference in the average two-hop neighbors count and the maximum two-hop neighbors count in real-time which is depicted in Fig. 5.2.

Figure 5.2 shows that the value of the average two-hop neighbors count and the maximum two-hop neighbors count differ from each other significantly. This happens because the deployment of sensor nodes in a WSN

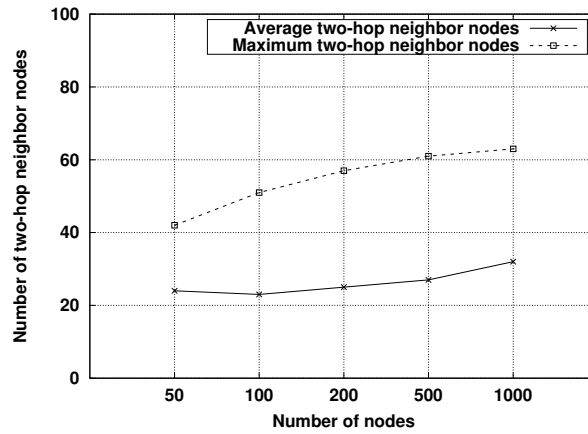


FIG. 5.2. Average two-hop neighbors vs maximum two-hop neighbors with variable number of nodes using uniform random distribution

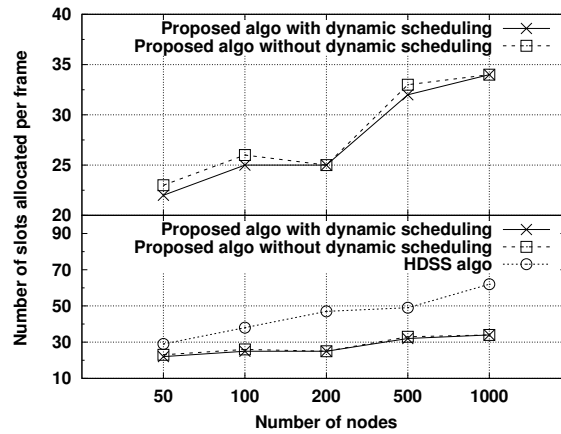


FIG. 5.3. Number of slots attached to a feasible schedule with and without applying dynamic slot scheduling algorithm for the un-allotted nodes

can not be 100% uniform. As there is a significant difference between the average two-hop neighbors and the maximum two-hop neighbors count, so, preparing the feasible schedule initially based on average two-hop neighbors will yield a better result in terms of the number of slots in the feasible schedule. This is already proved in the simulation result of the proposed algorithm as depicted in Fig. 5.1.

Figure 5.3 shows that the proposed algorithm prepares a feasible schedule with fewer slots as compared to the HDSS algorithm even if all the remaining un-allotted nodes are assumed to be allotted to separate slots. This graph also shows that the number of slots attached to the feasible schedule are further reduced when the dynamic slot scheduling is applied to allot the slot for the remaining un-allotted nodes. The further reduction in the number of slots is possible as some of the un-allotted nodes may not be the one-hop or two-hop neighbors to each other.

The time spent in the allocation of slots for the remaining un-allotted nodes in the proposed algorithm is compared with the reallocation of slots in the HDSS algorithm. This comparison is depicted in Fig. 5.4 which shows that our proposed algorithm takes very little time as compared to the HDSS algorithm in the allocation of slots to the remaining nodes. The initial slot allotment process for both the algorithms uses a similar approach except the proposed algorithm uses average two-hop neighbors count instead of maximum two-hop neighbors count as used by the HDSS algorithm. Accounting for the above fact, in this figure the initial slot allotment time is not depicted rather the time where both the algorithms significantly differ from each other is depicted

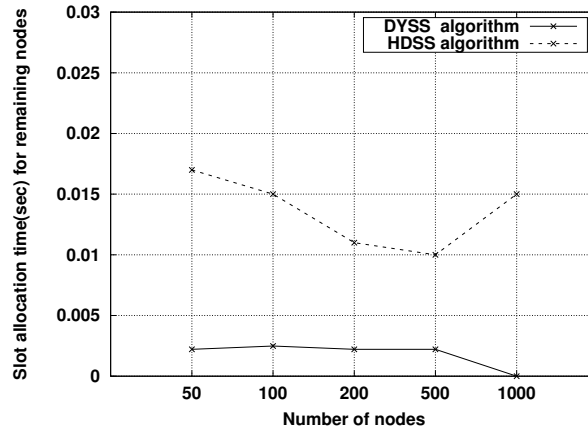


FIG. 5.4. Time spent in allocation of slots for the remaining nodes

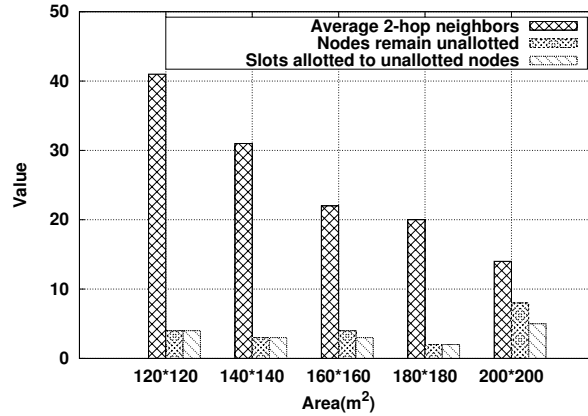


FIG. 5.5. Comparison of number of average two-hop neighbors, nodes remain un-allotted, slots allotted to un-allotted nodes with varying node density for 100 sensor nodes using uniform random distribution

to show a clear distinction between both the algorithms. As the time spent on the allocation of slots by the proposed algorithm is very less as compared to the HDSS algorithm. Hence, the energy consumption will also be proportionately less.

Figure 5.5 shows a comparison among the number average two-hop neighbors, nodes remain un-allotted, and slots allotted to un-allotted nodes with varying node density. Whereas Fig. 5.6 shows a similar comparison with varying the number of nodes using random uniform distribution. Figure 5.7 and 5.8 give a similar comparison using Randomised Grid. In all the cases, it is clearly shown that the nodes remain un-allotted are very less as compared to the average two-hop neighbor nodes as the distribution is randomly uniform. Since the number of remaining un-allotted nodes is very less, hence the reduction in the number of slots using the dynamic slot scheduling algorithm is also very less which is already depicted in Fig. 5.3.

6. Conclusion. In this paper, a dynamic slot scheduling algorithm for WSN has been proposed, which prepares a feasible schedule with less number of slots in a quick time. This ultimately helps to handle the collision due to correlated contention and at the same time minimizes latency during data transmission. Initially, slots are allotted to each node based on the average two-hop neighbors count as opposed to the maximum two-hop neighbors count as used in the earlier proposed HDSS algorithm. The use of average two-hop neighbors count reduces the number of slots in the schedule to a greater extent. Then the remaining un-allotted nodes are attached to slots in the best possible way using a novel dynamic slot allocation procedure to prepare the

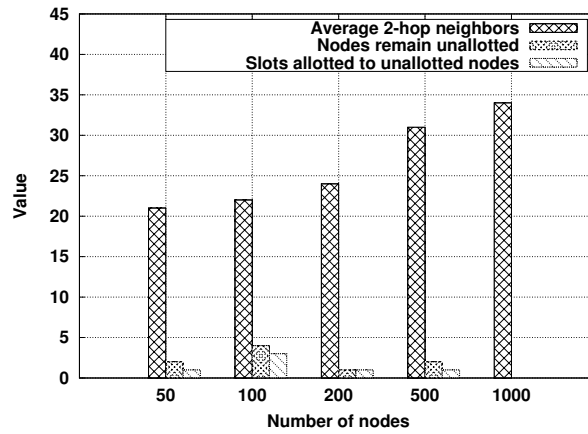


FIG. 5.6. Comparison of number of average two-hop neighbors, nodes remain un-allotted, slots allotted to un-allotted nodes with varying number of nodes using uniform random distribution

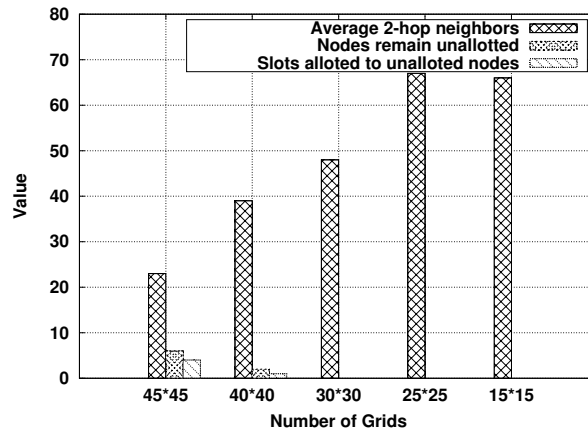


FIG. 5.7. Comparison of number of average two-hop neighbors, nodes remain un-allotted, slots allotted to un-allotted nodes with varying number of grids with Area 490*490 and number of nodes 1000 using Randomised Grid

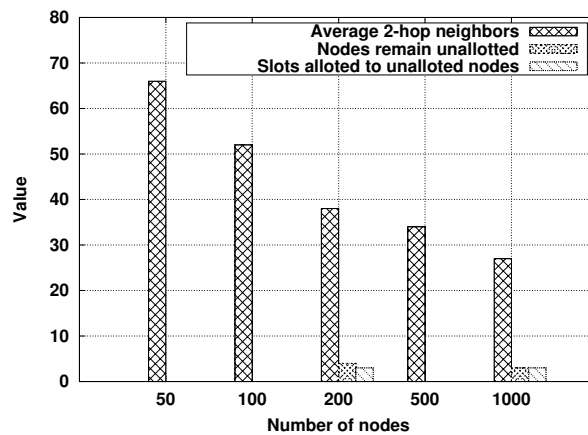


FIG. 5.8. Comparison of number of average two-hop neighbors, nodes remain un-allotted, slots allotted to un-allotted nodes with varying number of nodes using Randomised Grid

final schedule. The algorithm is tested in multiple environments like fixed and variable node density with uniform random distribution as well as the randomised grid to ensure it's correctness. The efficiency of the proposed algorithm has been compared with HDSS, DRAND, and RD-TDMA based on the number of slots. The comparison clearly shows that our algorithm outperforms others in terms of the number of slots attached to the final feasible schedule, which ultimately reduces the latency and also handles the collision during data transmission. The performance of the proposed algorithm is studied in an ideal scenario, i.e., a noiseless channel. In the future, we will further extend this algorithm to work in a real environment where every packet transmission is noisy in nature.

REFERENCES

- [1] S. C. ERGEN AND P. VARAIYA, *PEDAMACS: Power efficient and delay aware medium access protocol for sensor networks*, IEEE Transactions on Mobile Computing, vol. 5, no. 7, pp. 920-930, 2006.
- [2] W. YE, J. HEIDEMANN, AND D. ESTRIN, *An energy-efficient mac protocol for wireless sensor networks*, Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 1567-1576, 2002.
- [3] S. KUMAR AND H. KIM, *Energy efficient scheduling in wireless sensor networks for periodic data gathering*, IEEE Access, vol. 7, pp. 11410-11426, 2019.
- [4] K. MORIYAMA AND Y. ZHANG, *An efficient distributed tdma mac protocol for large-scale and high-data-rate wireless sensor networks*, IEEE 29th International Conference on Advanced Information Networking and Applications, pp. 84-91, 2015.
- [5] A. AHMAD AND Z. HANZLEK, *Distributed real time tdma scheduling algorithm for tree topology wsns*, 20th IFAC World Congress, vol. 50, no. 1, pp. 5926-5933, 2017.
- [6] A. AHMAD AND Z. HANZLEK, *An energy efficient schedule for ieee 802.15.4/zigbee cluster tree wsn with multiple collision domains and period crossing constraint*, IEEE Transactions on Industrial Informatics, vol. 14, no. 1, pp. 12-23, 2018.
- [7] M. R. LENKA, A. R. SWAIN, AND M. N. SAHOO, *Distributed slot scheduling algorithm for hybrid csma/tdma mac in wireless sensor networks*, IEEE International Conference on Networking, Architecture and Storage (NAS), pp. 1-4, 2016.
- [8] M. R. LENKA, A. R. SWAIN, AND B. P. NAYAK, *A hybrid based distributed slot scheduling approach for wsn mac*, Journal of Communications Software and Systems, vol. 15, no. 2, pp. 109-117, 2019.
- [9] A. BHATIA AND R. C. HANSDAH, *Rd-tdma: A randomized distributed tdma scheduling for correlated contention in wsns*, 28th International Conference on Advanced Information Networking and Applications Workshops, pp. 378-384, 2014.
- [10] A. AHMAD AND Z. HANZALEK, *An energy-efficient distributed tdma scheduling algorithm for zigbee-like cluster-tree wsns*, ACM Transactions on Sensor Networks, vol. 16, no. 1, pp. 3:1-3:41, 2019.
- [11] R. SEVERINO, N. PEREIRA, AND E. TOVAR, *Dynamic cluster scheduling for cluster-tree wsns*, 16th IEEE International Symposium on Object/component/service-oriented Real-time distributed Computing (ISORC), pp. 1-8, 2013.
- [12] J. LONG, M. DONG, K. OTA, AND A. LIU, *A green tdma scheduling algorithm for prolonging lifetime in wireless sensor networks*, IEEE Systems Journal, vol. 11, no. 2, pp. 868-877, 2017.
- [13] I. RHEE, A. WARRIER, J. MIN, AND L. XU, *Drand: Distributed randomized tdma scheduling for wireless ad hoc networks*, IEEE Transactions on Mobile Computing, vol. 8, no. 10, pp. 1384-1396, 2009.
- [14] W. YE, J. HEIDEMANN, AND D. ESTRIN, *An energy-efficient MAC protocol for wireless sensor networks*, IEEE Infocom, pp. 1567-1576, 2002.
- [15] S. LI, D. QIAN, Y. LIU, AND J. TONG, *Adaptive distributed randomized TDMA scheduling for clustered wireless sensor networks*, International Conference on Wireless Communications, Networking and Mobile Computing, pp. 2688-2691, 2007.
- [16] F. DOBSLAW, T. ZHANG, AND M. GIDLUND, *End-to-End reliability-aware scheduling for wireless sensor networks*, IEEE Transactions on Industrial Informatics, vol. 12, no. 2, pp. 758-767, 2016.
- [17] D. YANG, Y. XU, H. WANG, T. ZHENG, H. ZHANG, AND M. GIDLUND, *Assignment of segmented slots enabling reliable real-time transmission in industrial wireless sensor networks*, IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3966-3977, 2015.
- [18] Y. WANG AND I. HENNING, *A deterministic distributed TDMA scheduling algorithm for wireless sensor networks*, International Conference on Wireless Communications, Networking and Mobile Computing, pp. 2759-2762, 2007.
- [19] I. SLAMA, B. JOUABER, AND D. ZEGHLACHE, *Priority-based hybrid MAC for energy efficiency in wireless sensor networks*, Wireless Sensor Network, vol. 2, no. 10, pp. 755-767, 2010.
- [20] I. RHEE, A. WARRIER, M. AIA, J. MIN, AND M. SICHITIU, *Z-MAC: a hybrid MAC for wireless sensor networks*, IEEE/ACM Transactions on Networking, vol. 16, no. 3, pp. 511-524, 2008.
- [21] S. ZHUO, Y. SONG, AND Z. WANG, *Queue-length aware hybrid CSMA/TDMA MAC protocol for providing dynamic adaptation to traffic and duty-cycle variation in wireless sensor networks*, 9th IEEE International Workshop on Factory Communication Systems; pp. 105-114, 2012.
- [22] J. OLLER, I. DEMIRKOL, J. CASADEMONT, J. PARADELLS, G. GAMM, AND L. REINDL, *Has time come to switch from duty-cycled MAC protocols to wake-up radio for wireless sensor networks?*, IEEE/ACM Transactions on Networking, vol. 24, no. 2, pp. 674-687, 2016.
- [23] Y. LI, X. ZHANG, J. ZENG, Y. WAN, AND F. MA, *A Distributed TDMA Scheduling Algorithm Based on Energy-Topology Factor in Internet of Things*, IEEE Access, vol. 5, pp. 10757-10768, 2017.
- [24] I. SLAMA, B. SHRESTHA, B. JOUABER, D. ZEGHLACHE, AND T. ERKE, *DNIB: Distributed neighborhood information based*

- TDMA scheduling for wireless sensor networks*, 68th IEEE Vehicular Technology Conference, 2008.
- [25] J. LEE AND S. CHO, *Tree TDMA MAC Algorithm Using Time and Frequency Slot Allocations in Tree-Based WSNs*, *Wireless Personal Communications*, vol. 95, no. 3, pp. 2575-2597, 2017.
- [26] T. DANMANEE, K. NAKORN, AND K. ROJVIBOONCHAI, *CU-MAC: A Duty-Cycle MAC Protocol for Internet of Things in Wireless Sensor Networks*, *Transactions on Electrical Engineering, Electronics, and Communications*, vol. 16, no. 2, pp. 30-43, 2018.
- [27] *Castalia a simulator for wireless sensor networks*, [http://castalia.npc.nicta.com.au/pdfs/Castalia User Manual.pdf](http://castalia.npc.nicta.com.au/pdfs/Castalia%20User%20Manual.pdf).
- [28] *Cc2420 data sheet*, <http://www.stanford.edu/class/cs244e/papers/cc2420.pdf>.
- [29] *Telosb data sheet*, [http://www.xbow.com/Products/Product pdf files/Wireless pdf/TelosB Datasheet.pdf](http://www.xbow.com/Products/Product%20pdf%20files/Wireless%20pdf/TelosB%20Datasheet.pdf).

Edited by: P. Vijaya

Received: Dec 10, 2019

Accepted: May 21, 2020



ARTEFACTS REMOVAL FROM ECG SIGNAL: DRAGONFLY OPTIMIZATION-BASED LEARNING ALGORITHM FOR NEURAL NETWORK-ENHANCED ADAPTIVE FILTERING

TALABATTULA VISWANADHAM *AND P RAJESH KUMAR †

Abstract. Electrocardiogram (ECG) artefact removal is the major research topic as the pure ECG signals are an essential part of diagnosing heart-related problems. ECG signals are highly prominent to the interaction with the other signals like the Electromyography (EMG), Electroencephalography (EEG), and Electrooculography (EOG) signals and the interference mainly occurs at the time of recording. The removal of the artefacts from the ECG signal is a hectic challenge, for which, a novel algorithm is proposed in this work. The proposed method utilizes the adaptive filter termed as the (Dragonfly optimization + Levenberg Marquert learning algorithm) DLM-based Nonlinear Autoregressive with eXogenous input (NARX) neural network for the removal of the artefacts from the ECG signals. Once the artefact signal is identified using the adaptive filter, the identified signal is subtracted from the primary signal that is composed of the ECG signal and the artefacts through an adaptive subtraction procedure. The clean signal thus obtained is used for effective diagnosis purposes, and the experimentation performed to prove the effectiveness of the proposed method proves that the proposed method obtained a maximum Signal-to-noise ratio (SNR) of 52.8789 dB, a minimum error of 0.1832, and minimum error of 0.428.

Key words: ECG artefact removal, Dragonfly optimization, LM algorithm, NARX neural network, adaptive filter.

1. Introduction. With a lot of technical advances, the people of the world are afflicted with chronic diseases, among which the cardiovascular diseases are most commonly available diseases all over the world. The process of treatment and diagnosis of cardiovascular disease needs continuous monitoring with extensive care, but due to the increase in the number of patients, it is a tough task to undergo continuous monitoring of the patients that imposes the need for remote health monitoring. The remote health monitoring gains remarkable importance that involves the process of monitoring the ECG signals of the patient from a remote area by placing a mobile device in the patient's body during the normal day to day activity of the patient [17]. ECG is the graph that represents the electrical conducting system of the heart that lies in the value range of ± 2 mV and bandwidth range of 0.05 Hz to 125 Hz [1]. Thus, ECG consists of three waves, namely P wave, T wave, and U wave along with a QRS complex, and these are used for diagnosing the cardiac diseases [19]. The ECG has a lot of advantages not only in diagnosing the cardiac diseases but also in treating the obstructive sleep apnea or wearable physiological monitor and in checking the efficiency of the therapeutic drugs [19] [20]. The major issue is that the ECG signals are affected by the presence of noise, such as power line interference, motion artifacts, electromyogram effects, and baseline drift with respiration [3].

The ECG signal is affected by various kinds of artefacts at the time of the acquisition in the clinical atmosphere. The artefacts that affect the ECG in the clinical atmosphere include the baseline wander (BW), power-line interference (PLI), muscle artifacts (MA), and motion artifacts [4]. These artefacts are added in the ECG during the time of the recording [1] and the various noises present in the ECG affects the diagnosis procedures, which poses the necessity for a separation of the ECG signals from the midst of the artefact for the purpose of simple interpretation [18]. The artefacts have a lot of impact on the ST segment, decrease the quality of the ECG signal, and degrade the frequency resolution and, in turn, generate signals of large amplitude in ECG that appears like the PQRST waveforms. Moreover, the artefacts hide the tiny features that are essential for diagnosis and clinical monitoring. The main objective is regarding the removal of these artifacts that enable an artefact-free ECG for proper diagnosis. The main problem is regarding the separation of the high-resolution

*Aditya Institute of Technology and Management, Tekkali, K Kotturu, Andhra Pradesh 532201, India ([talabattula viswanadham@gmail.com](mailto:talabattula.viswanadham@gmail.com)).

†Andhra University college of engineering Visakhapatnam, Andhra Pradesh, 530003, India



ARTEFACTS REMOVAL FROM ECG SIGNAL: DRAGONFLY OPTIMIZATION-BASED LEARNING ALGORITHM FOR NEURAL NETWORK-ENHANCED ADAPTIVE FILTERING

TALABATTULA VISWANADHAM *AND P RAJESH KUMAR †

Abstract. Electrocardiogram (ECG) artefact removal is the major research topic as the pure ECG signals are an essential part of diagnosing heart-related problems. ECG signals are highly prominent to the interaction with the other signals like the Electromyography (EMG), Electroencephalography (EEG), and Electrooculography (EOG) signals and the interference mainly occurs at the time of recording. The removal of the artefacts from the ECG signal is a hectic challenge, for which, a novel algorithm is proposed in this work. The proposed method utilizes the adaptive filter termed as the (Dragonfly optimization + Levenberg Marquert learning algorithm) DLM-based Nonlinear Autoregressive with eXogenous input (NARX) neural network for the removal of the artefacts from the ECG signals. Once the artefact signal is identified using the adaptive filter, the identified signal is subtracted from the primary signal that is composed of the ECG signal and the artefacts through an adaptive subtraction procedure. The clean signal thus obtained is used for effective diagnosis purposes, and the experimentation performed to prove the effectiveness of the proposed method proves that the proposed method obtained a maximum Signal-to-noise ratio (SNR) of 52.8789 dB, a minimum error of 0.1832, and minimum error of 0.428.

Key words: ECG artefact removal, Dragonfly optimization, LM algorithm, NARX neural network, adaptive filter.

1. Introduction. With a lot of technical advances, the people of the world are afflicted with chronic diseases, among which the cardiovascular diseases are most commonly available diseases all over the world. The process of treatment and diagnosis of cardiovascular disease needs continuous monitoring with extensive care, but due to the increase in the number of patients, it is a tough task to undergo continuous monitoring of the patients that imposes the need for remote health monitoring. The remote health monitoring gains remarkable importance that involves the process of monitoring the ECG signals of the patient from a remote area by placing a mobile device in the patient's body during the normal day to day activity of the patient [17]. ECG is the graph that represents the electrical conducting system of the heart that lies in the value range of ± 2 mV and bandwidth range of 0.05 Hz to 125 Hz [1]. Thus, ECG consists of three waves, namely P wave, T wave, and U wave along with a QRS complex, and these are used for diagnosing the cardiac diseases [19]. The ECG has a lot of advantages not only in diagnosing the cardiac diseases but also in treating the obstructive sleep apnea or wearable physiological monitor and in checking the efficiency of the therapeutic drugs [19] [20]. The major issue is that the ECG signals are affected by the presence of noise, such as power line interference, motion artifacts, electromyogram effects, and baseline drift with respiration [3].

The ECG signal is affected by various kinds of artefacts at the time of the acquisition in the clinical atmosphere. The artefacts that affect the ECG in the clinical atmosphere include the baseline wander (BW), power-line interference (PLI), muscle artifacts (MA), and motion artifacts [4]. These artefacts are added in the ECG during the time of the recording [1] and the various noises present in the ECG affects the diagnosis procedures, which poses the necessity for a separation of the ECG signals from the midst of the artefact for the purpose of simple interpretation [18]. The artefacts have a lot of impact on the ST segment, decrease the quality of the ECG signal, and degrade the frequency resolution and, in turn, generate signals of large amplitude in ECG that appears like the PQRST waveforms. Moreover, the artefacts hide the tiny features that are essential for diagnosis and clinical monitoring. The main objective is regarding the removal of these artifacts that enable an artefact-free ECG for proper diagnosis. The main problem is regarding the separation of the high-resolution

*Aditya Institute of Technology and Management, Tekkali, K Kotturu, Andhra Pradesh 532201, India ([talabattula viswanadham@gmail.com](mailto:talabattula.viswanadham@gmail.com)).

†Andhra University college of engineering Visakhapatnam, Andhra Pradesh, 530003, India

ECG signals from the recorded ECG that is affected by the background noise [4]. The normal ECG signal has a predictable direction, duration, and amplitude of the characteristic waves, and an ECG signal is said to be normal or abnormal by assessing the ECG [1]. The ECG with the artefacts can be used for diagnosis by an experienced cardiologist, but the ECG analyzer can yield better accuracy. However, removal of the artefacts from the ECG signal enables accurate and simple interpretation [2].

The removal of noise from the ECG signal is performed using a number of the traditional algorithms that use any of the techniques, namely spatial or temporal averaging techniques. Initially, the noise is considered as random and stationary, and the noise reduction using the temporal averaging is the mean value of the frames or beats [9], and this method needs a number of the time frames for the reduction of the noise. At the same time, the spatial averaging method suffers from the problem of placing a number of electrodes in the same physical position. Along with the linear noise filtering method, the adaptive filtering methods have been used for separating and identifying the component waves of the ECG from the noisy ECG. The parameter of the filter is synchronized with the period of the signal for generating the quasi-periodic pattern of the cardiac signal. The other methods for ECG artefact removal are the subspace rotations, neural networks, and bi-spectral analysis [6]. Ensemble Averaging (EA) is the other familiar method that extracts the required components from the noisy ECG signal that averages the beats but loses the significant variations of the inter-beat in the cardiac cycle due to the averaging procedure [10] [14]. The optimization techniques [33-36] have applications in ECG artefact removal.

This paper proposes a novel algorithm for training the NARX neural network that is based on the proposed DLM optimization-based algorithm. The main intention of the paper is to remove the artefact present in the ECG signal for the effective diagnosis of cardiac diseases and other related problems. The paper presents the artefact cancelation strategy that is based on the adaptive subtraction procedure in which the artefact signal is filtered using the adaptive filter and subtracted from the artefact ECG signal. The adaptive filter uses the NARX neural network that uses the DLM optimization algorithm for tuning the weights of the network, and the effective tuning is brought about using the proposed algorithm. The proposed algorithm trains the network effectively based on the weights that correspond to the minimum value of the error. Thus, the proposed method of artefact removal stands as an effective procedure in eliminating the noise signals, such as ECG, EMG, and EOG.

The main contribution of the paper is the DLM optimization Algorithm, which determines the optimal weight for tuning the NARX neural network that serves as an adaptive filter in removing the artefacts from the ECG signal. DLM is the integration of the LM and Dragonfly optimization algorithm.

The organization of the paper is: Section 1 introduces the paper, and section 2 describes the literature works with the challenges. The proposed method is introduced and described clearly in section 3, section 4 presents the results and discussion of the paper, and section 5 concludes the paper.

2. Literature Review.

2.1. NARX Neural Network. Payam Amani et al. [29] introduced a multi-step ahead response time predictor for database queries depends on a nonlinear autoregressive neural network model with exogenous inputs. The experimentation was performed to analyze the performance of the predictor on a lab setup with a MySQL-server. Zina Boussaada et al. [30] developed a race sailboat using exclusively renewable sources. It predicted the direct solar radiation on a horizontal surface using a NARX neural network. The experimental results have shown that the prediction performance was best when the training phase of the neural network is performed at regular intervals.

Eugen Diaconescu [31] tested the performance of the prediction for diverse time series using a NARX dynamic recurrent neural network (RNN). The author utilized conventional statistical techniques to occur indications to make efficient the process of prediction chaotic time series with RNN. Hong He et al. [32] introduced an ECG measuring experiment at seven acupoints of the Pericardium Meridian of Hand-Jueyin to attain the meridian information transmission data. Here, a NARX network was used to model the meridian information transmission system.

2.2. Artefact removal in ECG signal. Here, eight research works in artefact removal in ECG signal are discussed. Syed Anas Imtiaz et al. The paper [1] proposed a method for the automatic artefact removal

using the Automatic artefact identification algorithm that is based on three parameters, namely the quality, interpretation quality, and computational complexity to find the absolute best use of the data and help the medical professionals in diagnosis. The method provides reliable data, but the presence of the low SNR artefacts that occur as a result of breathing interrupts the incoming waveform leading to the incapability of the filter to rectify the signal. Amit Kumar and Mandeep Singh [2] proposed a method Short Time Fourier Transform (STFT) for the decomposition of the artefacts from the ECG signal through the optimal selection of the wavelets thus, sustaining the diagnostic information. The method is highly robust and capable of removing the artefacts from the noisy physiological and non-stationary ECG Signals, but the method yielded Poor Percentage Root Mean Square Difference (PRD). Shing-Hong Liu et al. [3] proposed a method for the removal of the artefacts that initially computes the acceleration signal of the vibration using an accelerometer that is taken as a reference in the adaptive filter. Finally, the Least Mean Square (LMS) algorithm is employed for determining the optimal weight of the filter. The method possesses a stable convergence irrespective of the level of the noise, but the performance is found to degrade if the redundant signals and the reference signals are similar.

Muhammad Zia Ur Rahman et al. [4] used an efficient sign based normalized adaptive filter named Computationally efficient adaptive filtering technique, and it possesses weight update loops for removing the artefacts from the ECG. The method efficiently removes the non-stationary noise but used only for wireless biotelemetry ECG systems. Jinseok Lee et al. The paper [5] presented a real-time method for identifying and removing the artefacts from the ECG using the Empirical mode decomposition (EMD). There are two approaches, among which the first one uses first-order intrinsic mode function (F-IMF) of EMD, and the second approach uses the three statistical measures on the F-IMF time series for measuring the characteristics of randomness and variability. The method offers proper robustness, but it suffers from Segment disconnectivity if the detected corrupted segment is not utilized for the Atrial Fibrillation (AF) detection, which, in turn, results in performance degradation.

Taigang He et al. [6] use the Independent component analysis (ICA) for detecting and removing the artefacts from the ECG's, and the advantages of the method is that the method offers simplicity, efficiency, and hence potential for processing the ECG online using the ICA is better, but the ECG still contains the artefacts that provide the artefactual data. The method failed to remove the artefacts properly. Mbachu C.B et al. [7] proposed a method for filtering the artefacts from ECG using a method named as the Rectangular Window-Based Digital Filters. Initially, the digital finite impulse response (FIR) low pass, high pass, and notch filters are designed based on the rectangular window. The advantage is that the filters are able to remove the unwanted signals and thus, minimizes the power line interference, but distortions are present due to cascade filtering output signal that generates ripples due to the usage of rectangular windows. Guang Zhang et al. [8] designed an enhanced Least mean-square (LMS) method for degrading Cardiopulmonary resuscitation (CPR) artefacts causing reliable discovery of the VF rhythm during the uninterrupted chest compression (CC). This method reduces the CPR artefacts effectively from the corrupted ECG signal, but the performance of the enhanced LMS method is poor, and a large amount of real corrupted ECG records are required for enhancing the performance.

2.3. Challenges.

1. During the overlapping of the spectral content with the ECG, the SNR is enhanced using digital filtering, but this digital filtering injects little distortion in the ST-segment regions. In almost all the situations, the shape of the component wave that is present in the noise signal is well-known but the requirement is understand the time of occurrence and the exact shape of the signal [9].
2. The filters used for removal of the artefacts from the ECG signal cause a reduction in the amplitudes of the component waves [6], and they are not successful. Moreover, some of the noise and artefacts possess a wide range of the frequency, and they are random in nature; thus, filters are not successful in eliminating the interference while it lies in the same frequency range of the cardiac signal.
3. The major and significant feature regarding the ECG waveform is the QRS complex that lies in a particular frequency band. The relative power of the wave requires a specific examination that ensures the reliability of the ECG waveform. The value of the pSQI defines the presence, or the absence of the data of interest, which in other words, can be briefed as the presence of the higher value of pSQI indicates the required data is present or else the required data is missing. The strict monitoring should

be enabled as the artefacts caused as a result of breathing, muscle contractions, and general body motion possess a frequency of 5 Hz [1].

4. The LMS adaptive filtering is used in [4-8] for the dismissal of the artefacts from the ECG signal, and this method is the old algorithm that uses the predefined reference signal to remove the artefacts efficiently.
5. The monitoring tools namely, EOG, EMG, and EEG are involved in the process of continuous monitoring along with the ECG that causes the rhythmic artefacts to be included in the mechanical activity of the ECG. The continuous monitoring process causes the interference of the EOG, EMG, and EEG signals in the ECG such that the removal of artefacts remains the better way for proper diagnosis [8].

3. The proposed method for the ECG artefact removal. The reason to remove the artefacts from the ECG signals is performed using the newly proposed DLM optimization-based NARX neural network. The paper gives a brief discussion of the proposed method of artefact removal from the ECG signals that gains a lot of advantages like an effective diagnosis of cardiac diseases and helps the physician to take effective measures.

3.1. Removal of artefacts from the ECG signals. ECG is the record of the electrical activity of the heart, and during the recording process, the artefacts are included with the ECG signals due to the interference effects of the other signals, such as EEG, EOG, and EMG that may create adverse effects on the diagnosis of the cardiac-related diseases. In order to overcome the problems caused by the artefacts, it is essential to perform the adaptive noise cancellation strategy [24] for the removal of the artefacts, such as EEG, EOG, and EMG signals. Thus, to get rid of the artefacts using the adaptive noise cancellation, the cancellation framework requires two inputs. Among two of the inputs, one input comes from the ECG signal source, whereas the second input is from the artefacts. Thus, the primary input signal is the combination of the signal from the ECG source, which is the clean signal and the interference signal that is obtained by passing the artefact through the unidentified non-linear dynamics. Thus, the primary signal is represented as

$$(3.1) \quad B(i) = C(i) + I(i)$$

where $B(i)$ is the primary input signal, $C(i)$ refers to the clean ECG signal, and $I(i)$ stands for the interference that is generated using the unknown nonlinear dynamics or otherwise the signal obtained using the noise source. The noise signal is subjective to adaptive filtering to produce the filtered output, which is similar to that of the interference signal created as a result of the nonlinear dynamics. Thus, the noise cancellation is performed that extracts the clean signal through the subtraction of the filtered output from the primary input signal. Thus, the clean signal extracted using the noise cancellation strategy is given as

$$(3.2) \quad C^*(i) = B(i) - A(i)$$

where $C^*(i)$ indicates the clean signal obtained as a result of the adaptive noise cancellation, $B(i)$ stands for the primary input signal, and $A(i)$ refers to the adaptive filtered output. Figure 3.1 shows the proposed noise cancellation strategy.

3.2. NARX neural network for the enhanced adaptive filtering of the artefacts from the ECG signal. The proposed method of the artefact removal from the ECG signals using the NARX neural network is presented in this section. The main aim of the NARX neural network is to predict the artefact present in the ECG signal to generate the clean ECG signal sufficient for perfect diagnosis. The input to the neural network is the artefact and the signal is predicted for which the delays are used.

3.2.1. Solution Encoding. Figure 3.2 shows the solution of the proposed learning algorithm that yields an optimized solution using three weights, namely the weights of the exogenous input vector, weights of the regressed output vector, and the weights of the exogenous output vector. These three weights are combined to generate an optimized weight such that the optimized output trains the NARX neural network in canceling the artefacts and the size of the solution vector depends on the number of hidden neurons present in the network. Let us consider the weights of the exogenous input vector as L_1, L_2, L_{d1} , the weights of the regressed output is represented as R_1, R_2, R_{d2} , and the exogenous output vector as O_1, O_2, O_{d1} . Then, the solution generated using the LM algorithm, and the Dragonfly optimization algorithm is represented as X^{l1}, X^{l2}, X^{lf} and X^{d1}, X^{d2}, X^{df} respectively. Thus, the proposed DLM generated the optimal weights that is given as X^{z1}, X^{z2}, X^{zf} .

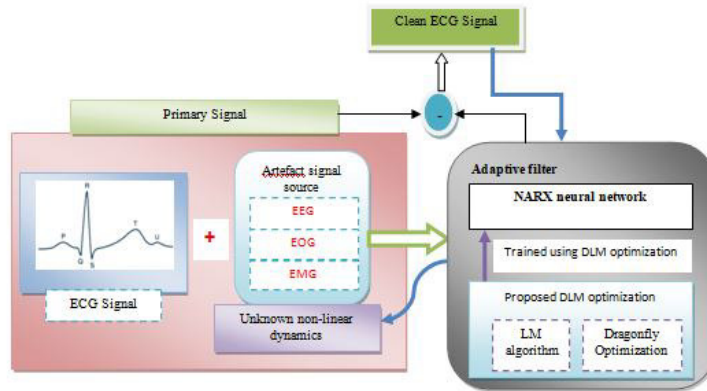


FIG. 3.1. ECG artefact removal

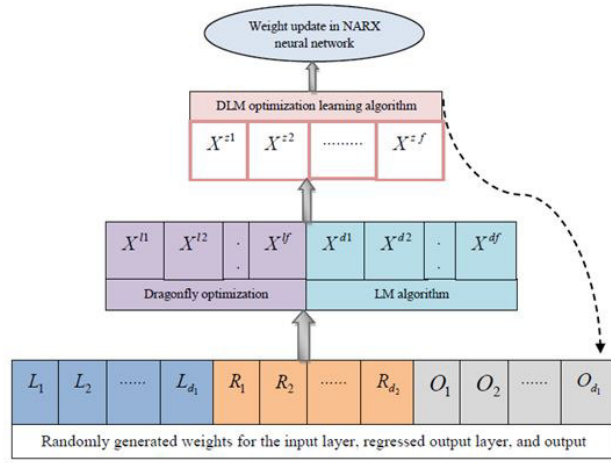


FIG. 3.2. Solution encoding of the DLM optimization-based NARX neural network

3.2.2. Architecture of the NARX neural network. NARX neural network [22] is a recurrent neural network that is used for the analysis and modelling of the nonlinear time series and holds a lot of merits when compared with the other classical prediction models. The NARX network possesses an effective learning rate, and in the proposed method of noise cancelation, the learning algorithm used is the dragonfly optimization and the LM algorithm. The NARX neural network is the collection of the multilayer fed forward network, recurrent loop, and the time delay. Figure 3.3 shows the architecture of the NARX neural network. The three layers include the input layer, the hidden layer, and the output layer. The network is subjected to the tapped delays both in the input layer and the output layer and the feedback flows in a single direction. The feedback follows the input layer, hidden layer, and provides the output in the output layer. There are three information vectors in the input layer namely, the exogenous input vector, delayed regressed output vector, and the delayed exogenous input vector. The output of the NARX neural network is given by

$$(3.3) \quad N(l + 1) = F[N(l), N(l - 1), N(l - 2) \dots, N(l - d_1); S(l), S(l - 1), S(l - 2) \dots, S(l - d_2)]$$

where $N(l)$ is the exogenous input vector, $N(l - 1), N(l - 2), \dots, N(l - d_1)$ are the delayed regressed output vector, and $S(l), S(l - 1), S(l - 2), \dots, S(l - d_2)$ are the delayed exogenous input vectors. At the beginning of the NARX network function, the weights are assigned between the input layer and the hidden layer and the regressed output vector and the hidden layer.

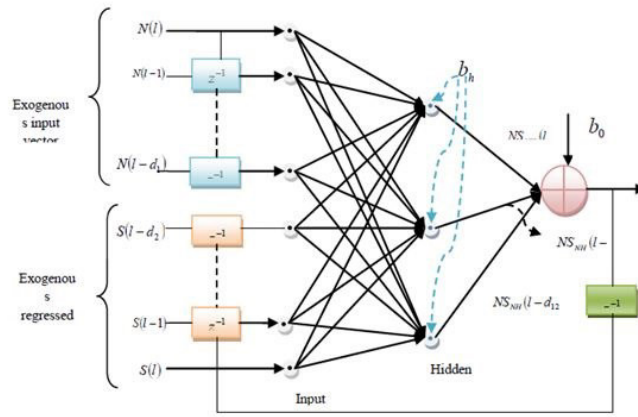


FIG. 3.3. The architecture of the NARX neural network used for adaptive filtering of the artefacts

3.2.3. A novel learning algorithm for the NARX neural network: - Proposed DLM-Dragonfly Levenberg Marquet optimization-based Learning algorithm. The paper proposes a new learning algorithm using dragonfly optimization [21] and the LM algorithm [23]. The dragonfly optimization is the meta-heuristic optimization algorithm that holds better among most of the evolutionary algorithms because of the following reasons. The main advantage is regarding the search space as all the information regarding the search space is restored, and there are only very few control parameters in the search space leading to the increase in the flexibility of the algorithm. The proposed algorithm is robust, and the problem of converging to the local minimum using the LM algorithm is solved using the Dragonfly that increases the convergence and converges to the global minimum. The proposed algorithm is advantageous as they utilize the advantages of both the LM and Dragonfly. The learning algorithms update the weights individually, and the error values are computed for both the algorithms. The learning algorithm with the minimum value of the error is used for updating the weights of the NARX neural network, and the following are the algorithmic steps of the proposed DLM optimization-based learning algorithm.

Step 1: Population Initialization. The initial step is the initialization that initializes the total swarm population that aims at the possibility of survival. The population initialization using the dragonfly optimization is represented as

$$(3.4) \quad D_d; (1 \leq d \leq n)$$

where D_d denotes the position of the d^{th} dragonfly and n is the total population of the dragonflies.

Step 2: Parameters influencing the position update. The dragon population aims at the possible location to live as they move in search of food, and they are distracted from the enemies that remain the two basic behaviours of the swarm population. The position update follows the initialization step, and the position of the dragonflies is updated based on five major factors, separation, alignment, cohesion, attraction, and distraction, and they are based on the swarm behaviours of survival. The swarm behaviour explains the factors as the separation is the factor that avoids collision among the individual dragonflies whereas, an alignment that corresponds to the velocity matching among the dragonflies. The parameter termed as collision aligns the dragonfly towards the center of mass. The five factors are modelled as

$$(3.5) \quad S_d = - \sum_{j=1}^N D - D_j$$

where D is the current position of the individual dragonfly, D_j is the position of the j^{th} neighbor, and N be number of neighboring dragonflies with respect to the reference dragonfly. S_d denote the separation factor of

the individual dragonflies.

$$(3.6) \quad G_d = \frac{\sum_{j=1}^N V_j}{N}$$

where G_d refers to the alignment of the d^{th} dragonfly and V_j is the velocity of the j^{th} neighbor.

$$(3.7) \quad C_d = \frac{\sum_{j=1}^N D_j}{N} - D$$

where D is the current position of the individual dragonfly and C_d refers to the cohesion of the d^{th} dragonfly.

Step 3: Define the step vectors. The position of the dragonfly is updated based on the step vector, and the main role of the step vector is to show the direction of the dragonfly, and it is employed for the larger dimensions.

$$(3.8) \quad \Delta D_{l+1} = (a_1 S_d + a_2 G_d + a_3 C_d + a_4 M_d + a_5 \varepsilon_d) + w \Delta D_l$$

where a_1, a_2, a_3, a_4 , and a_5 , are the weights of separation, alignment, cohesion, food factor, and enemy factor respectively. The values of the weights play a major role in the transition of the exploration phase and the exploitation phase of the dragonfly thus, the weights are tuned such that it ensures proper switch off between the two phases. W denotes the initial weight, represents the iteration number. s_d, G_d, C_d, M_d , and ε_d denotes the separation, alignment, cohesion, attraction, and distraction of the d^{th} dragonfly. The exploration phase of the dragonflies describes the hunting mechanism of the dragonfly for which it takes a back and forth movement. The exploration phase describes the grouping and the movement of the dragonflies in the same direction for a long distance.

Step 4: Compute the objective function of the Fireflies. The objective function is calculated to determine the optimal weight that is based on the minimum value of the error calculated using the LM algorithm and the Dragonfly optimization algorithm. The weight corresponding to the minimum error is selected as the weight vector to train the network.

Step 5: Update and determine the position of the food source and the enemy. The above discussion highlights the behaviour of the dragonflies and hence, it is essential to update the position of the enemy and the position of the food. The computation of the position of the food and the position of the enemy is performed as

$$(3.9) \quad M_d = D^* - D$$

where M_d represents the attraction of the dragonfly towards the food, D^* is the position of the food

$$(3.10) \quad D_l^d = D^- + D$$

D^- is the position of the enemy, and D_l^d is the distraction of the dragonfly away from the enemy.

Step 6: Determine the position of the dragonfly. The position update of the dragonfly using the step vector is formulated as

$$(3.11) \quad D_{l+1}^d = D_l^d + \Delta D_d^{l+1}$$

Moreover, the random walk is the search mechanism that is utilized to perform the search process in the absence of the neighbouring dragonfly. The position update of the dragonfly is computed as the below equation.

$$(3.12) \quad D_{l+1}^d = D_l^d + \text{levy}(H) * D_l^d$$

$$(3.13) \quad \text{Levy}(x) = 0.01 * \frac{r_1 * \sigma}{|r_2|^{\frac{1}{\alpha}}}$$

$$(3.14) \quad \sigma = \left(\frac{\Gamma(1 + \alpha) \times \text{Sin}(\frac{\pi\alpha}{2})}{\Gamma(\frac{1+\alpha}{2}) \times \alpha \times 2 \times \times (\frac{\alpha-1}{2})} \right)^{\frac{1}{\alpha}}$$

where $\Gamma(x) = (x-1)!$. T represents the current iteration, H is the dimension of the position vectors. r_1 and r_2 are the random vectors, α is the constant, and the value is 1.5. The optimal position of the dragonfly enables the optimal selection of the weights for training the NARX neural network.

Step 7: Weight update based on the dragonfly algorithm. The above steps are repeated for the maximum number of iterations, and upon the application of the firefly algorithm, the input vector and the updated weights are combined as represented below

$$(3.15) \quad X^d = (W, B^{new})$$

where W refers to the input vector and B^{new} stands for the weight updated using the dragonfly optimization algorithm.

Step 8: Calculation of the Mean Square Error. The MSE is computed between the target value and the current value computed using the dragonfly optimization.

$$(3.16) \quad e^d = \frac{1}{l} \sum_{j=1}^l (X_i^d - X_i^g)^2$$

where X_i^g represents the truth table of the original data and X_i^d denotes the output from the dragonfly algorithm.

Step 9: Update the weights using the LM algorithm. The weights of the NARX neural network are initialized that depends on the number of the hidden layers and the weights are denoted as below

$$(3.17) \quad X = X^{c1}, X^{c2}, \dots, X^{cf}$$

where f is the total number of weights initialized using the LM algorithm.

Step 10: The computation of the mean square error. It follows the following equation,

$$(3.18) \quad E(X) = e^T e$$

where $E(X)$ denotes the performance index, $e^T e$ indicates the target output, and e represents the expected output.

Step 11: Weight Update based on LM algorithm. The weight of the NARX neural network is updated using the LM algorithm as

$$(3.19) \quad \Delta X = [J^T J + \gamma k]^{-1} * J^T e$$

where J is the Jacobian matrix, J^T is the Jacobian transform matrix and γ is the learning rate. The update in the learning rate parameter depends on the decay function λ . For the greater values of $E(X)$, the learning rate is multiplied using the rate of the decay function. Once the learning rate is determined, then the value of $E(X)$ is recomputed using the weighted function $X = X + \Delta X$ as the trail weight. Similarly, the learning rate is divided by the decay rate, whenever the function $E(X)$ decreases. Then, the incremented values of the weights is found by the formula

$$(3.20) \quad X = X + \Delta X$$

Step 12: Formulation of the learning rate. Whenever the performance index $E(X)$ exceeds the trail weighted function, the learning rate is updated as

$$(3.21) \quad \gamma = \gamma * \lambda$$

The learning rate is multiplied with the decay value to obtain the new learning rate, and the weight update is performed based on the new learning rate that follows the steps from step 8.

Whenever the performance index $E(X)$ is less than the trail weighted function, the learning rate is updated by dividing the current learning rate by the decay function. Thus, the learning rate is represented as,

$$(3.22) \quad \gamma = \frac{\gamma}{\lambda}$$

Step 13: Random generation of the weights using the LM algorithm. The weight is updated using the following equation

$$(3.23) \quad X_{l+1}^{LM} = X_t - [H + \mu * T]^{-1} * q$$

where H denotes the Hessian matrix of the algorithm, and it is determined by multiplying the Jacobian matrix and the transverse of the Jacobian matrix.

Step 14: Computing the gradient matrix. The Jacobian matrix is employed for determining the gradient matrix that is denoted as

$$(3.24) \quad q = J^T e$$

where J represents the Jacobian matrix and e denotes the error value.

Step 15: Generated output based on the newly updated weights. The LM algorithm computes the output based on the original value of the input vector and the newly updated weights. The output is represented as

$$(3.25) \quad X_j^{LM} = (W, B^{new})$$

Step 16: Re-compute the error. The error is recomputed between the output based on the newly updated weight, and the ground value, and the following formula is used.

$$(3.26) \quad e^{LM} = \frac{1}{l} \sum_{j=1}^l (X_j^{LM} - X_i^g)$$

where l corresponds to the total number of the iterations, X_i^g denotes the truth value of the input vector, X_{LM}^j represents the output vector generated using the LM algorithm.

Step 17: Generation of the weight vector using the proposed DLM-training algorithm. The weight vectors are obtained using the proposed DLM algorithm for training the NARX neural network. The optimal generation of the weights are brought about through the proposed DLM algorithm. For the optimal selection of the weight vectors, the errors of the outputs using both the training algorithm are determined individually and the errors of the algorithms are compared. The weight vector corresponding to the low value of the error is considered for training the NARX network. When the error of the LM algorithm exceeds the error of the dragonfly optimization algorithm, then the weight vector obtained using the dragonfly optimization is selected as the optimal weights. When the error of the LM algorithm is lower than the error obtained using the dragonfly optimization, then the weight vector corresponding to the LM algorithm is used. Moreover, the error value has an impact on the damping factor, and the value of the damping factor reduces when the error of the current iteration is less than the error value of the previous iteration. Similarly, the value of the damping factor increases when the error of the present iteration is greater than the previous iteration. Thus, the error values of the dragonfly algorithm and the LM algorithm is employed for the optimal selection of the weights vectors in the DLM optimization algorithm.

$$(3.27) \quad W_{l+1} = \begin{cases} W_l^{LM} + 1 & \text{when } (e^{LM} < e^d) \\ D_{l+1}^d & \text{when } (e^d < e^{LM}) \end{cases}$$

The above equation represents the condition for the optimal selection of the weight vector.

3.2.4. Proposed method of the artefact removal using the DLM optimization-based NARX neural network. The main aim of the paper is depicted in Algorithm 11 that shows the steps involved in the elimination of the artefacts from the ECG signal. The ECG artefact signal comprises of the ECH signal and the artefacts, such as the EEG, EMG, and EOG signals. These artefacts are added with the ECG signal at the time of recording and monitoring and hence, removal of the artefacts is essential for the effective diagnosis. Thus, the adaptive filter is proposed that aims at the removal of the artefacts from the ECG signal, and the adaptive filter is made of NARX neural network that is trained using the DLM optimization algorithm. The proposed DLM algorithm determines the optimal weight for training the NARX neural network to perform the process of adaptive filtering. The proposed filter filters the artefact signal, and the filtered output is fed to the adaptive subtraction that causes the subtraction of the artefact from the ECG artefact signal such that a clean ECG signal is generated.

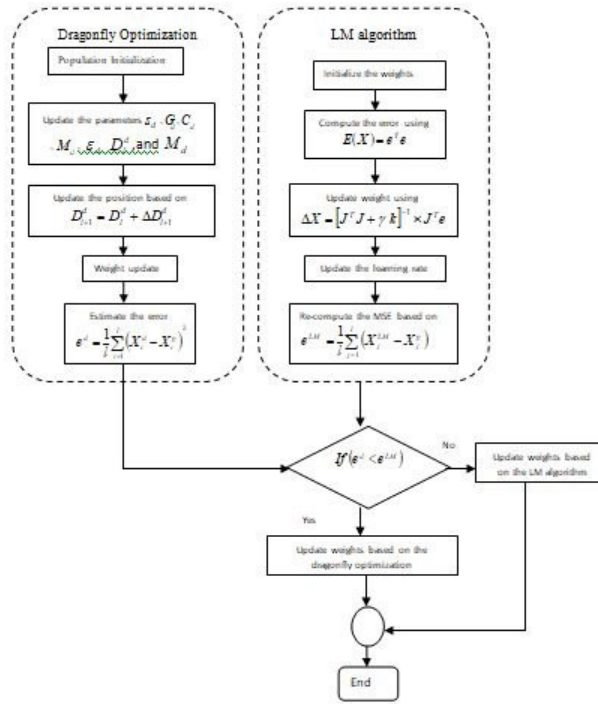


FIG. 3.4. Flowchart for the Proposed DLM optimization algorithm

Algorithm 1: The proposed method of artefact removal using the adaptive filter

- 1 # Proposed ECG artefact removal using the adaptive filter- DLM optimization-based NARX neural network **Data:** ECG Artefact signal
 - Result:** Clean ECG signal
 - 2 Read
 - 3 Compute the primary signal, $B(i) = C(i)+I(i)$
 - 4 # Perform adaptive filtering using the DLM-based NARX neural network
 - 5 {
 - 6 Read the artefact signal
 - 7 Update weights of NARX neural network using DLM
 - 8 **if** ($e^{lm} < e^d$) **then**
 - 9 | $W_{l+1} = W_{l+1}^{LM}$
 - 10 **else**
 - 11 | $W_{l+1} = W_{l=1}^{LM}$
 - 12 Estimate the filtered output $A(i)$
 - 13 }
 - 14 Calculate the clean signal, $C^*(i) = B(i) - A(i)$
-

4. Results and Discussion. In this section, the result of the proposed method is discussed in detail to elaborate the superior performance of the proposed method. The experimentation of the proposed technique of artefact removal from the ECG signals is done in a system with 2 GB RAM, Intel core processor, Windows 10 Operating System. The technique is implemented using the software tool MATLAB.

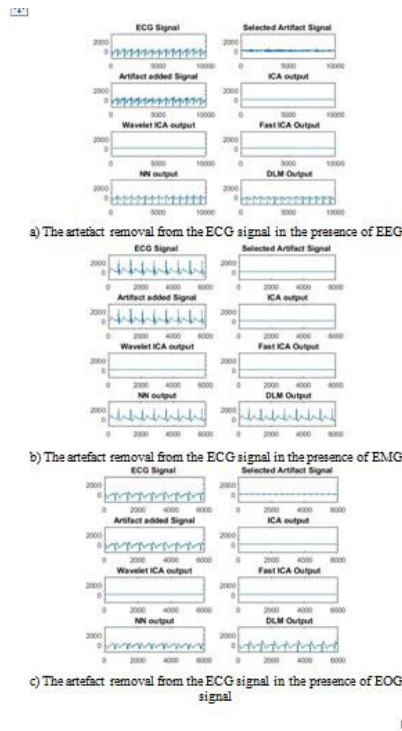


FIG. 4.1. *Experimental results of the proposed method of artefact removal using various artefact removal methods*

4.1. Database Description. Physionet database was created and contributed by Tatiana Lugovaya. The database has 310 ECG recordings, obtained from 90 persons. Each recording contains, ECG lead I, recorded for 20 seconds, digitized at 500 Hz with 12-bit resolution over a nominal ± 10 mV range, 10 annotated beats, information containing age, gender, and recording date. The raw ECG signals are rather noisy and contain both high and low frequency noise components. Each record includes both raw and filtered signals: Signal 0: ECG I (raw signal), Signal 1: ECG I filtered (filtered signal).

4.2. Experimental results. Figure 4.1 shows the experimental results of the proposed method of artefact removal using various artefact removal methods. The artefacts, such as EEG, EMG, and EOG are added with the pure ECG signal during the time of recording such that the presence of artefact affects the effective decision-making of the doctor. Thus, the various artefact removal methods concentrate on the removal of the artefacts as is depicted in figures 4.1 a, b, and c respectively.

4.3. Competing methods. The competing methods used are ICA [6], WICA [27], FICA, and NN [28] for comparing the results of the artefact removal with the proposed DLM to prove the superiority of the proposed method.

ICA: ICA is a source separation method, and its application to biomedical signals is rapidly expanding. ICA offers simplicity, efficiency, and hence the potential for processing the ECG online, but the ECG still contains the artefacts that provide the artefactual data. The method failed to remove the artefacts properly.

WICA [27]: The Wavelet-Independent Component Analysis (WICA) approach allows extending the removal of the artefacts in the clinical applications. WICA is the integration of DWT and ICA, which takes the advantages of both techniques.

FICA: Fixed point or FastICA algorithm of ICA is a technique for the removal of eye blink artifact from EEG and ECG signals. FastICA algorithm has been applied to synthetic signals prepared by adding random noise to the ECG signal. It divides the signal into two independent components, namely ECG pure and artifact signal. Similarly, it is applied to remove the artifacts from the EEG signal.

NN: In [28], an adaptive filtering approach based on a discrete wavelet transform and artificial neural network is developed for ECG signal noise reduction. This method integrates the multi-resolution property of wavelet decomposition and the adaptive learning ability of artificial neural networks, and fits well with ECG signal processing applications.

4.4. Comparative analysis. Figures 4.2-4.4 show the analysis of the proposed method of artefact removal in terms of the SNR, MSE, and RMSE in the presence of the EEG, EMG, and EOG signals. Figure 4.2.a) shows the analysis in terms of SNR in the presence of an EEG signal. For the first signal, the SNR values of the methods like the ICA, WICA, FICA, NN, and DLM are 48.4474 dB, 50.9866 dB, 44.4993 dB, 47.5152 dB, and 50.7498 dB respectively. The value of SNR is 44 dB for ICA, 44.10 dB for WICA, 44 for FICA, 50.7035 for NN, and 52.8753 for DLM respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of SNR is greater when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the SNR value of the proposed DLM is greater when compared with the other methods. Finally, for the fifth signal, the SNR dB's are 43.76, 44.3, 43.76, 49.83, and 51.92 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.2.b) shows the analysis in terms of MSE in the presence of an EEG signal. The effective method responds with the minimum value of the MSE error. For the first signal, the MSE values of the methods like the ICA, WICA, FICA, NN, and DLM are 20, 20, 17.5515, 0.3088, and 0.2099 respectively. The value of MSE is 15.6646 for ICA, 15.5487 for WICA, 14.3068 for FICA, 8.6930 for NN, and 4.5218 for DLM respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of MSE is minimum when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the MSE value of the proposed DLM is less when compared with the other methods. Finally, for the fifth signal, the MSE values are 20, 15.4641, 15.2445, 0.8452, and 0.38 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.2.c) shows the analysis in terms of RMSE in the presence of an EEG signal. The effective method responds with the minimum value of the RMSE error. For the first signal, the RMSE values of the methods like the ICA, WICA, FICA, NN, and DLM are 4.4721, 4.4721, 4.1894, 0.5557, and 0.4582, respectively. The value of RMSE is 3.9578 for ICA, 3.9431 for WICA, 3.7824 for FICA, 2.9483 for NN, and 2.1264 for DLM, respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of RMSE is minimum when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the RMSE value of the proposed DLM is less when compared with the other methods of artefact removal. Finally, for the fifth signal, the RMSE values are 4.4721, 3.9324, 3.9044, 0.9193, and 0.6165 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.3.a) shows the analysis in terms of SNR in the presence of the EMG signal. For the first signal, the SNR values of the methods like the ICA, WICA, FICA, NN, and DLM are 44.29 dB, 46.3359 dB, 44.2964 dB, 47.1836 dB, and 50.9222 dB respectively. The value of SNR is 37.9067 dB for ICA, 38.1276 dB for WICA, 37.9062 for FICA, 50.6968 for NN, and 52.5789 for DLM respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of SNR is greater when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the SNR value of the proposed DLM is greater when compared with the other methods. Finally, for the fifth signal, the SNR dB's are 43.0636, 44.8689, 43.0636, 49.5461, and 51.939 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.3.b) shows the analysis in terms of MSE in the presence of the EMG signal. The effective method responds with the minimum value of the MSE error. For the first signal, the MSE values of the methods like the ICA, WICA, FICA, NN, and DLM are 20, 20, 20, 84.98, and 0.192 respectively. The value of MSE is 20 for ICA, 20 for WICA, 20 for FICA, 100.8027 for NN, and 4.5206 for DLM respectively, for the second signal. It is clear that for the proposed DLM algorithm, the value of MSE is minimum when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the MSE value of the proposed DLM is less when compared with the other methods. Finally, for the fifth signal, the

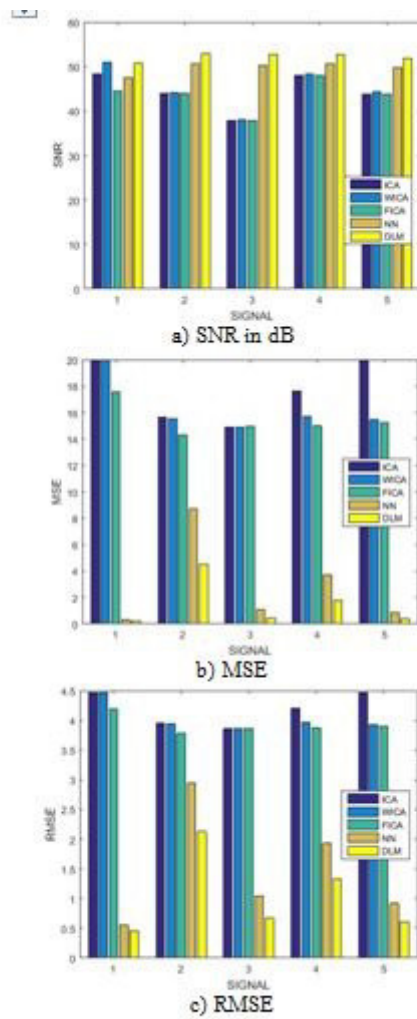


FIG. 4.2. Analysis of the artefact removal in the presence of the EEG signal

MSE values are 20, 20, 12.66, 12.2077, and 0.3839 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.3.c) shows the analysis in terms of RMSE in the presence of the EMG signal. The effective method responds with the minimum value of the RMSE error. For the first signal, the RMSE values of the methods like the ICA, WICA, FICA, NN, and DLM are 4.4721, 4.4721, 4.4721, 9.2184, and 0.4381 respectively. The value of RMSE is 4.4721 for ICA, 4.4721 for WICA, 4.4721 for FICA, 10.04 for NN, and 2.1264 for DLM, respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of RMSE is minimum when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the RMSE value of the proposed DLM is less when compared with the other methods of artefact removal. Finally, for the fifth signal, the RMSE values are 4.4721, 4.4721, 3.559, 3.4939, and 0.6196 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.4.a) shows the analysis in terms of SNR in the presence of the EOG signal. For the first signal, the SNR values of the methods like the ICA, WICA, FICA, NN, and DLM are 44.2953 dB, 46.4878 dB, 44.2953

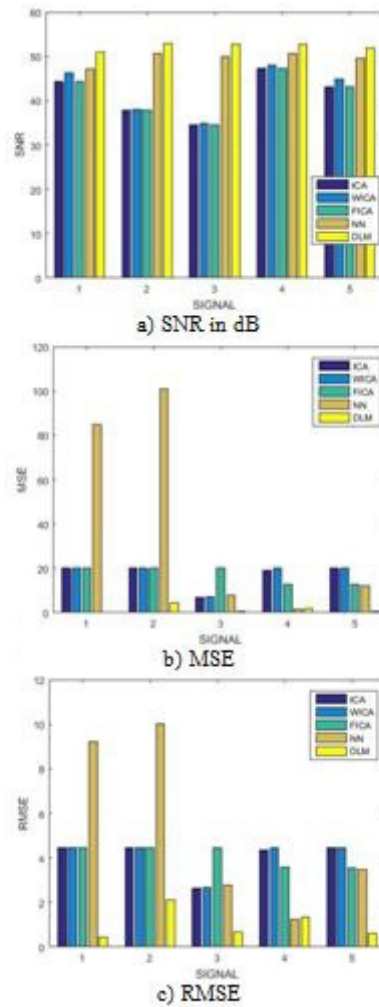


FIG. 4.3. Analysis of the artefact removal in the presence of the EMG signal

dB, 48.6889 dB, and 50.8412 dB respectively. The value of SNR is 37.9052 dB for ICA, 38.1418 dB for WICA, 37.9052 for FICA, 50.6948 for NN, and 52.8694 for DLM respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of SNR is greater when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the SNR value of the proposed DLM is greater when compared with the other methods. Finally, for the fifth signal, the SNR dB's are 42.2801, 43.8368, 42.5828, 49.4697, and 51.8589 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.4.b) shows the analysis in terms of MSE in the presence of the EOG signal. The effective method responds with the minimum value of the MSE error. For the first signal, the MSE values of the methods like the ICA, WICA, FICA, NN, and DLM are 20, 20, 20, 116.62, and 0.1832 respectively. The value of MSE is 20 for ICA, 20 for WICA, 16.3022 for FICA, 92.5998 for NN, and 4.5254 for DLM respectively, for the second signal. It is clear that for the proposed DLM algorithm, the value of MSE is minimum when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the MSE value of the proposed DLM is less when compared with the other methods. Finally, for the fifth signal, the MSE values are 20, 20, 7.6076, 11.8829, and 0.3772 for the artefact removal methods like the

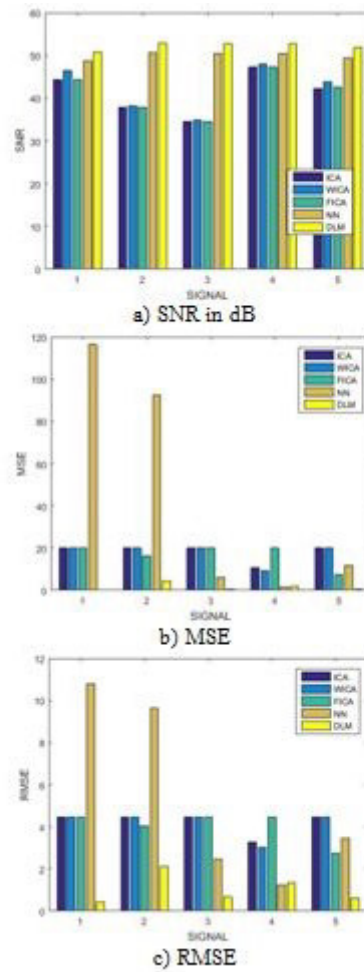


FIG. 4.4. Analysis of the artefact removal in the presence of the EOG signal

ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

Figure 4.4.c) shows the analysis in terms of RMSE in the presence of the EOG signal. The effective method responds with the minimum value of the RMSE error. For the first signal, the RMSE values of the methods like the ICA, WICA, FICA, NN, and DLM are 4.4721, 4.4721, 4.4721, 10.7991, and 0.4280 respectively. The value of RMSE is 4.4721 for ICA, 4.4721 for WICA, 4.0376 for FICA, 9.6228 for NN, and 2.1273 for DLM, respectively for the second signal. It is clear that for the proposed DLM algorithm, the value of RMSE is minimum when compared with the existing artefact removal methods. Similarly, the analysis using the third and the fourth signal indicate that the RMSE value of the proposed DLM is less when compared with the other methods of artefact removal. Finally, for the fifth signal, the RMSE values are 4.4721, 4.4721, 4.7582, 3.4471, and 0.6141 for the artefact removal methods like the ICA, WICA, FICA, NN, and DLM respectively that prove the proposed method is superior over the existing methods.

4.5. Comparative discussion. The comparison Table 4.5 presents the comparison of the artefact removal methods with respect to the SNR, MSE, and RMSE parameters. The maximum SNR of 52.8789 dB is obtained using the proposed DLM method, whereas the other methods like the ICA, WICA, FICA, NN, and DLM obtained an SNR value of 44.4474 dB, 50.9866 dB, 47.9796 dB, and 50.7035 dB respectively. Similarly, the MSE value for the proposed method is minimum as 0.1832, but for the methods like the ICA, WICA, FICA,

TABLE 4.1
Comparative discussion of the artefact removal methods

Methods	SNR (dB)	MSE	RMSE	Computational time (Sec)
ICA	44.4474	6.94	2.6345	12.5
WICA	50.9866	7.1040	2.6653	11
FICA	47.9796	7.60	2.75	9.5
NN	50.7035	0.3088	0.5557	7
DLM	52.8789	0.1832	0.428	6

NN, and DLM, the MSE error is 6.94, 7.1040, 7.60, and 0.3088 respectively. Likewise, the minimum RMSE of 0.428 is obtained by the proposed method when compared with the other existing artefact removal methods. Also, the proposed method has the computational time of 6 sec, which is minimum than the computational time of other comparative methods.

5. Conclusion. The paper concentrates on the proposed method of artefact removal using the DLM-based NARX neural network. The proposed algorithm uses both the dragonfly optimization and LM learning algorithm for framing the DLM algorithm that trains the NARX neural network. The artefact removal used a simple subtraction method that subtracts the artefact from the ECG signal such that the ECG signals obtained are clear and is suitable for diagnosing the cardiac-related diseases. The adaptive tuning of the artefact removal is carried out using the DLM-based NARX neural network such that the proposed methods stand as an effective approach for artefact removal. The experimentation performed using the artefact signals, such as EMG, EEG, and EOG proves that the proposed method is effective when compared with the existing methods. The maximum SNR of 52.8789 dB, a minimum error of 0.1832, and a minimum error of 0.428 is obtained using the proposed DLM-based NARX neural network that generates the clean ECG signal. The proposed method stood as an effective method in extracting the clean signal from the artefact ECG signal. The performance of the proposed method is further increased by using recent optimization algorithms. The training speed of the proposed method also needs further improvement.

REFERENCES

- [1] S. A. IMTIAZ, J. MARDELL, S. S. YARAHMADI, E. R.-VILLEGAS, *ECG artefact identification and removal in mHealth systems for continuous patient monitoring* vol.3, no.3, pp.171 - 176, 2016.
- [2] A. KUMAR AND M. SINGH, *Optimal Selection of Wavelet Function and Decomposition Level for Removal of ECG Signal Artifacts* Journal of Medical Imaging and Health Informatics, Vol. 5, 138–146, 2015.
- [3] S.-H. LIU, *Motion artifact reduction in Electrocardiogram using adaptive filter* Journal of Medical and Biological Engineering, 31(1): 67-72, 2010.
- [4] M. Z. U. RAHMAN, R. A. SHAIK, D. V. R. K. REDDY, *Efficient sign based normalized adaptive filtering techniques for cancelation of artifacts in ECG signals: Application to wireless biotelemetry* Signal Processing, Volume 91, Issue 2, February 2011, Pages 225-239.
- [5] J. LEE, D. D. MCMANUS, S. MERCHANT, AND K. H. CHON, *Automatic Motion and Noise Artifact Detection in Holter ECG Data Using Empirical Mode Decomposition and Statistical Approaches* IEEE Transactions On Biomedical Engineering, Vol. 59, No. 6, June 2012.
- [6] T. HE, G. CLIFFORD, AND L. TARASSENKO, *Application of independent component analysis in removing artefacts from the electrocardiogram* Neural Computing & Applications, April 2006, Volume 15, Issue 2, pp 105–116.
- [7] C. B. MBACHU, I. DIGO VICTOR, I. EMMANUEL, AND I. I. NSIONU, *Filtration Of Artifacts In ECG Signal Using Rectangular Window-Based Digital Filters* IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011.
- [8] G. ZHANG, T. WU, Z. WAN, Z. SONG, M. YU, D. WANG, L. LI, F. CHEN, *A new method to detect ventricular fibrillation from CPR artifact-corrupted ECG based on the ECG alone* Biomedical Signal Processing and Control, Volume 29, August 2016, Pages 67-75. J.
- [9] S. PAUL, M. R. REDDY, AND V. J. KUMAR, *A transform domain SVD filter for suppression of muscle noise artefacts in exercise ECG's* IEEE Trans. Biomed. Eng., vol. 47, no. 5, pp. 654–663, May 2000.
- [10] P. LANDER AND E. J. BERBARI, *Time-frequency plane wiener filtering of the high-resolution ECG: Development and application* IEEE Trans. Biomed. Eng., vol. 44, no. 4, pp. 256–265, Apr. 1997.
- [11] G. LU, J. S. BRITAIN, P. HOLLAND, J. YIANNI, A. L. GREEN, J. F. STEIN, T. Z. AZIZ, AND S. WANG, *Removing ECG noise from surface EMG signals using adaptive filtering* Neurosci. Lett., vol. 462, no. 1, pp. 14–19, Oct. 2009.

- [12] D. L. DONOHO, *De-noising by soft-thresholding* IEEE Trans. Inf. Theory, vol. 41, no. 1, pp. 613–627, May 1995.
- [13] O. SAYADI AND M. B. SHAMSOLLAHI, *ECG denoising with adaptive bionic wavelet transform* in Proc. IEEE Conf. Eng. Med. Biol. Soc., 2006, pp. 6597–6600.
- [14] R. SAMENI, M. B. SHAMSOLLAHI, C. JUTTEN, AND G. D. CLIFFORD, *A nonlinear Bayesian filtering framework for ECG denoising* IEEE Trans. Biomed. Eng., vol. 54, no. 12, pp. 2172–2185, Dec. 2007.
- [15] Y. KISHIMOTO, Y. KUTSUNA, AND K. OGURI, *Detecting motion artifact ECG noise during sleeping by means of a tri-axis accelerometer* in Proc. IEEE Conf. Eng. Med. Biol. Soc., 2007, pp. 2669–2672.
- [16] S. W. YOON, S. D. MIN, Y. H. YUN, S. LEE, AND M. LEE, *Adaptive motion artifacts reduction using 3-axis accelerometer in e-textile ECG measurement system* J. Med. Syst., vol. 32, no. 2, pp. 101–106, Apr. 2008.
- [17] A. AGARWAL, A. SINGH, A. ACHARYYA, R. A. SHAFIK, S. R. AHAMED, *Energy-Efficient and High-Speed Robust Channel Identification Methodology to Solve Permutation Indeterminacy in ICA for Artifacts Removal from ECG in Remote Healthcare* In Proceedings of the 2013 International Symposium on Electronic System Design, pp. 52 - 56, 2013.
- [18] S. A. ANAPAGAMINI, R. RAJAVEL, *Removal of artifacts in ECG using Empirical mode decomposition* In Proceedings of the 2013 International Conference on Communication and Signal Processing, pp.288 - 292, 2013.
- [19] J. S. LIN, S. Y. HUANG, K. W. PAN AND S. H. LIU, *A physiological signal monitoring system based on an SoC platform and wireless network technologies in homecare technology* J. Med. Bio. Eng., 29: 47-51, 2009.
- [20] S. KIM, H. NAKAMURA, T. YOSHIDA, M. KISHIMOTO, Y. IMAI, N. MATSUKI, T. ISHIKAWA AND T. YAMAGUCHI, *Development of a wearable system module for monitoring physical and mental workload* Telemed. J. E-Health, 14: 939-945, 2008.
- [21] S. MIRJALILI, *Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems* Neural Computing and Applications, Vol. 27, No. 4, pp. 1053–1073, May 2016.
- [22] J. M. P. MENEZES JR AND G. A. BARRETO, *Long-term time series prediction with the NARX network: An empirical evaluation* Neurocomputing, vol.71, no.16–18, pp.3335-3343, October 2008.
- [23] GAIDHANE, H. VILAS, V. SINGH, Y. V. HOTE, AND M. KUMAR, *New approaches for image compression using neural network* Journal of Intelligent Learning Systems and Applications, vol.3, no.04, pp.220-229, 2011.
- [24] A. JAFARIFARMAND, M. A. BADAMCHIZADEH, *Artifacts removal in EEG signal using a new neural network enhanced adaptive filter* Neurocomputing, vol.103, pp.222-231, 1 March 2013.
- [25] Y. LI, *Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach* Physiological Measurements, vol.27, pp.425–36, 2006.
- [26] P. MISHRA AND S. K. SINGLA, *Artifact Removal from Biosignal using Fixed Point ICA Algorithm for Pre-processing in Biometric Recognition* vol.13, no.1, Jan 2013.
- [27] B. AZZERBONI, M. CARPENTIER, E.L. FORESTA, AND E C. MORABITO, *Neural-ICA and Wavelet Transform for Artifacts Removal in surface EMG* In Proceedings of IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 2004.
- [28] S. POUNGONSRI AND X.-HUAYU, *An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks* Neurocomputing, vol. 117, pp. 206-213, 2013.
- [29] P. AMANI, M. KIHIL, AND A. ROBERTSSON, *NARX-based Multi-step Ahead Response Time Prediction for Database Servers* In proceedings of 11th International Conference on Intelligent Systems Design and Applications (ISDA), Cordoba, Spain, 2011.
- [30] Z. BOUSSAADA, O. CUREA, A. REMACI, H. CAMBLONG, AND N. M. BELLAAJ, *A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation Energies*, vol. 11, no. 3, 2018.
- [31] E. DIACONESCU, *The use of NARX Neural Networks to predict Chaotic Time Series* WSEAS Transactions on Computer Research, vol. 3, no. 3, pp. 182-191, March 2008.
- [32] H. HE, X. YAN AND W.WEI, *Meridian ECG Information Transmission System Modeling Using NARX Neural Network* In proceedings of IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 2016.
- [33] A. RATRE AND V. PANKAJAKSHAN *Tucker visual search-based hybrid tracking model and Fractional Kohonen Self-Organizing Map for anomaly localization and detection in surveillance videos* The Imaging Science Journal, pp. 1-16, 2017.
- [34] A. V. DHUMANE AND R.S. PRASAD *Multi-objective fractional gravitational search algorithm for energy efficient routing in IoT* Wireless Networks, pp. 1-15, 2017.
- [35] S. NIPANIKAR, V H. DEEPTHI, AND N. KULKARNI, *A sparse representation based image steganography using Particle Swarm Optimization and wavelet transform* 2017.
- [36] P.M. SHELKE AND R. S. PRASAD *An improved anti-forensics JPEG compression using Least Cuckoo Search algorithm*The Imaging Science Journal, vol. 66, no. 3, pp. 169-183, 2018.

Edited by: P. Vijaya

Received: Dec 7, 2019

Accepted: Jun 23, 2020



A COMPREHENSIVE REVIEW ON STATE-OF-THE-ART IMAGE INPAINTING TECHNIQUES

BALASAHEB H. PATIL *AND P.M. PATIL[†]

Abstract. Image inpainting is the process of restoring missing pixels in digital images in a plausible way. A study on image inpainting technique has acquired a significant consideration in various regions, i.e. restoring the damaged and old documents, elimination of unwanted objects, cinematography, retouch applications, etc. Even though, limitations exist in the recovery process due to the establishment of certain artifacts in the restored image areas. To rectify these issues, more and more techniques have been established by different authors. This survey makes a critical analysis of diverse techniques regarding various image inpainting schemes. This paper goes under (i) Analyzing various image inpainting techniques that are contributed in different papers; (ii) Makes the comprehensive study regarding the performance measures and the corresponding maximum achievements in each contribution; (iii) Analytical review concerning the chronological review and various tools exploited in each of the reviewed works. Finally, the survey extends with the determination of various research issues and gaps that might be useful for the researchers to promote improved future works on image inpainting schemes.

Key words: Image inpainting; Region Filling; Performance Measures; Chronological review; Tools; Research Gaps.

AMS subject classifications. 94A08

1. Introduction. “Inpainting refers to the art of restoring lost parts of the image and reconstructing them based on the background information, i.e. image inpainting is the process of reconstructing lost or deteriorated parts of images using information from surrounding areas” [66, 67, 68, 91]. It comprised of tasks like object disocclusion, filling holes and image restoration, etc. At first, the theory of digital inpainting was established by Bertalmio et al [90]. According to this technique, higher-order PDE was exploited for restoring purposes. Here, the areas to be filled are based on the assistance of gradient direction. The two most important classification of inpainting consists of textural and structural inpainting [69, 70, 71].

The regions outside the area to be inpainted are modelled by the texture inpainting approaches [72]. This was exploited to the textures with randomized 2D models. Consequently, the structural inpainting schemes attempt to rebuild the structures such as object and line contours. Usually, structural inpainting is deployed when the portion to be inpainted is small [73, 74, 92, 93]. It concerns on linear structures that could be considered as 1D pattern such as object and line contours. Moreover, image compression could be done effectively by neglecting certain portions at the encoder side and inpainting those portions by a similar technique at the decoder side [75, 94].

Also, morphological processes namely corrosion could be exploited for inpainting the smaller portions of missing values. In fine painting museums, inpainting of corrupted painting were usually done by skilled artists and generally, it is found to be much time-consuming. There were numerous techniques implemented for image inpainting [76, 77, 78]. “Microsoft Kinect sensor” is an inexpensive device, which has influenced a lot of analysts to handle with deep data. However, issues exist with this device in terms of its resolution and accuracy [79, 80, 81].

The main contribution of this paper is depicted below:

1. This work conducts a survey of diverse techniques related to various image inpainting schemes that are contributed to each paper.

*Research Scholar All India Shri Shivaji Memorial Society’s Institute of Information Technology, Pune and Assistant Professor Dept of E&TC Vidya Pratishthan’s Kamalnayan Bajaj Institute of Engineering & Technology, Baramati, India (balasahebhpati1144@gmail.com).

[†]Professor TSSM’S Bhivarabai Sawant College of Engineering and Research, Pune, India

2. Accordingly, the performance measures and the corresponding maximum achievements are also investigated.
3. Furthermore, various tools exploited in each work are reviewed and a chronological analysis is also made.
4. At last, the research challenges and gaps on image inpainting systems are also exhibited.

The paper is prearranged as follows. Section 2 portrays the related works and section 3 presents a comprehensive review of adopted algorithms, performance measures, as well as the maximum achievements and Section 4 describes the assessment on adopted tools and chronological review in each work. Moreover, Section 5 elaborates on the research gaps and challenges, and Section 6 concludes the paper.

2. Literature review.

2.1. Related works. In 2014, He et al. [1] have presented a novel wavelet frame oriented weight minimization approach for inpainting the image. Numerical analyses have revealed the significances of the presented scheme in conserving the image edges. In 2016, Chen et al. [2] have presented a scheme using PPA for resolving the issues in non-orthogonal wavelet inpainting. Mathematical analyses have shown the superiority of the presented approach over the compared models in terms of performance time. In 2017, Chen et al. [3] have proposed a novel scheme that exploited the Patch priority algorithm for covering up the required data that ensured the configuration stability. In the end, simulations were held and the outcomes have demonstrated the effectiveness of this method.

In 2017, Fei et al. [4] have implemented a paper depending on the ADMM) technique. In the end, mathematical analyses have shown that the presented scheme was capable of attaining considerable computational gain. In 2017, Xue et al. [5] have introduced a technique using the “Low gradient regularization” model, by which the penalty for small gradients was reduced. Also, the investigational results have shown the efficacy of the implemented model. Vahid et al. [6] have presented a novel technique that deployed the Orientation matrix model for non-textured inpainting of images. The presented approach was further deployed for better and faster inpainting.

In 2017, Wang et al. [7] have developed the space varying updating approach for enhancing the priority evaluation. Also, FFT was deployed that searched the whole image and provided quicker matching outcomes. Finally, the outcome shows the enhancements made by the implemented technique. In 2017, Ying et al. [8] have established an image inpainting scheme, which comprised of the PSNR values for enhancement. The analysis outcomes have shown that the presented method attains better PSNR value over other compared schemes. In 2007, Ubirat et al. [9] have introduced the concept for block-oriented image inpainting in the wavelet domain. In the end, simulations were held that proved the efficiency of the presented scheme in filling the areas with better visual quality.

In 2016, Barbu et al. [10] have exploited the variational model for image reconstruction, which depended on 2nd-order PDE. Certain successful image inpainting analysis and evaluation were also explained in this work. In 2016, Muddala et al. [11] have suggested an LDI, which aimed at enhancing the rendering quality of the images. As per the subjective and objective assessments, the developed technique outperformed the conventional schemes at the disocclusion. In 2012, Arnav et al. [12] have presented a scheme that prevailed over the drawbacks of low-resolution issues that resulted in poor occlusions. The experimental analysis illustrated that the presented model could efficiently reconstruct the images with reduced noise.

In 2012, Dhiyanesh and Sathiyapriya [13] have analyzed the active snake model for image segmentation. The experimentation illustrated that the presented method offers better outcomes over the other conventional methods. In 2013, Wang et al. [14] have introduced an effective transform-oriented approach for geometric techniques that tackled over the reduced efficiency. Further, the investigational analysis demonstrated that the adopted technique enhances and speeds up the performances along with improved restoration outcomes. In 2012, Li and Yan [15] have developed a novel approach, which was dependent on the TV approach. From the analysis, a reduced computational time was achieved by the presented scheme over the other evaluated schemes.

In 2017, Chunhong et al. [16] have modeled varied image inpainting schemes which classified the corrupted images using a tight frame model. Also, the betterment of the presented scheme was analyzed over other schemes in terms of improved quality. In 2019, Hu et al. [17] have developed a novel non-reference quality assessment scheme that solved the Thangka IIQA issues. Finally, a state-of-the-art index was generated by the

adopted technique for IQA that was associated with human vision. In 2019, Liu et al. [18] have dealt with the inverse issues of image inpainting model, for which multi-filters guided low-rank tensor coding was introduced. Besides, the presented scheme outperformed the traditional scheme concerning PSNR, and SSIM.

In 2019, Jiao et al. [19] have deployed the MLCN and encoder-decoder generator that have effectively restored the image. Moreover, the efficiency of the presented technique was illustrated via the simulation results. In 2019, Cheng and Li et al. [20] have developed a direction structure distribution analysis approach for MRF oriented inpainting schemes. Experimental outcomes have demonstrated that the proposed technique maintained better consistency with reduced cost on inpainting diverse types of corrupted images. In 2019, Tran and Hoang [21] have presented a novel digital inpainting technique, which considered the related data for restoring the intensities of pixels derived from the image dataset. Accordingly, for simulation purposes, 2D face images were assessed from public datasets.

In 2019, Anis et al. [22] have established method that considered the high-order PDE schemes for resolving the image inpainting and de-noising issues. Also, numerous arithmetical examples were provided that revealed the superiority of the presented model over the conventional methods. In 2019, Wali et al. [23] have developed the adaptive boosting method for TGV oriented image inpainting and denoising. Also, various investigational results have established that the presented model generates images with reduced artifacts. In 2018, Ding et al. [24] have adopted a novel exemplar-oriented image inpainting model that exploited the recently introduced scheme known as the PAMSE. Here, the adopted scheme proved the improved performance of the developed method in propagating texture and geometric structure in a simultaneous manner.

In 2018, Zhu et al. [25] have developed a new technique depending on CNN, which assisted in detecting the patch-oriented inpainting process. Finally, investigational outcomes demonstrated that the introduced scheme acquired better performance concerning running time, FPR, and TPR. In 2018, Karaca and Tunga [26] have offered a pattern and texture conserving interpolation-oriented approach for inpainting the lost regions in color images. Finally, the performance of the established scheme was evaluated on different color images with varied patterns and textures. In 2018, Yan et al. [27] have regarded the PSIS problem, and accordingly, a PSIS method was proposed depending on LC-based SIS. Finally, the analysis was carried out for computing the effectiveness of the developed model.

In 2018, Han et al. [28] have implemented a new “virtual view synthesis” scheme for a depth-image-oriented scheme that lessened the errors occurred during inpainting. At last, the investigational results have shown the superiority of the presented technique in achieving effective high-quality virtual view images. In 2018, Qin et al. [29] have established two approaches for reversible image recovery and visible-watermark elimination. Furthermore, the analysis outcomes have revealed the superiority and effectiveness of the presented scheme. In 2018, Lu et al. [30] have portrayed a novel MTC for restoring the images based on vectors. Further, experimentations on both real and synthetic images were carried out that proved the improved restoration performance of the developed approach.

In 2010, Xiong et al. [31] have established an efficient image compression scheme, which had combined the PAI for developing the visual redundancy in color images. Finally, the outcomes showed the betterment of the presented model in offering improved bit rate saving at better qualities. In 2018, Mariko et al. [32] have presented an image inpainting scheme that optimized the outline of the masked areas specified by users. The investigational outcomes have illustrated that the adopted technique performed higher results than traditional schemes without masked region restoration. In 2004, Criminisi et al. [33] have developed the approach for eliminating huge objects from digitalized images. Finally, investigational results have shown the enhancements made by the presented scheme.

In 2017, Li and Lv [34] have adopted a decoupled variational scheme for image inpainting in transform and image domain including Fourier and wavelet domain. The arithmetical experimentation and evaluations on diverse images have revealed the efficacy of the adopted scheme. In 2017, Wang et al. [35] have resolved the issues from the viewpoint of intensity function estimation. In the end, arithmetical experiments have confirmed the efficiency of the adopted technique, particularly in edge recovery. In 2017, Fuchs and Jan [36] have introduced an extensive analysis of higher-order issues that were simulated by numerical imaging applications. This work mainly concerned with higher-order bounded deviation along with dual solutions.

In 2019, Anh and Hoang [37] have established a new reconstruction approach that generated RGB images

without exploiting the descriptors. Finally, the experiments have shows the efficacy of the presented scheme. In 2016, Shen et al. [38] have developed a constrained inpainting scheme for recovering an image from its inaccurate or incomplete wavelet coefficients. From the analysis, the presented scheme recovered the images with better PSNR and visual quality over the existing methods. In 2016, Peter et al. [39] have established a novel approach which involves the comprehensive analysis of the weaknesses and strengths of the PDEs. Also, the analysis outcomes have offered simple harmonic diffusion and reduced compression rates when evaluated over the other existing schemes.

In 2017, Colomer et al. [40] have exploited the dictionary learning and sparse representation methods for inpainting the retinal vessels. Also, two diverse methods of evaluating the inpainting quality were offered that validated the non-artificial outcomes on inpainting. In 2015, Sarathi et al. [41] have presented approach for speedy automatic recognition of OD and its precise segmentation in digital eye images. Further, researches were performed on a labeled dataset with segmentation accuracy of 91%. In 2015, Zhang et al. [42] have presented an efficient and simple restoration approach based on image inpainting. The outcomes were found to have removed the noise efficiently while conserving the image details with improved PSNR.

In 2015, Li et al. [43] have established a fast local inpainting scheme depending on the “Allen–Cahn” approach. Also, numerous arithmetical outcomes were presented that portrayed the accuracy and robustness of the implemented scheme. In 2015, Jiao et al. [44] have presented a technique for restoring the highly corrupted digital “off-axis Fresnel holograms” depending on image inpainting. Besides, ABC was employed that increased the computational efficacy of the inpainting scheme. In 2015, Liang et al. [45] have introduced a forgery detection approach using exemplar-oriented inpainting. In the end, the simulation outcomes have illustrated the betterment of the presented scheme offers reduced processing time for varied images.

In 2014, Berntsson and George [46] have exploited the theory of parameter recognition as a method of inpainting. The arithmetical analysis and error analysis had shown the presented model’s betterment over the harmonic inpainting model. In 2014, Margarita et al. [47] have deployed the preliminary image segmentation method, and accordingly, SOMs were created for every homogeneous texture. Further, the outcomes were compared over the traditional SOM schemes and the desired inpainting agents were portrayed. In 2014, Chung and Yim [48] have introduced an effective error concealment technique for reconstructing the pixels, which were lost during video communication. The developed technique also offered considerable enhancement on PSNR over the existing schemes.

In 2014, Li et al. [49] have modeled the exemplar-oriented inpainting approach for maintaining better neighborhood consistency and structural coherence. Also, the investigational results have revealed the improvements of the adopted scheme for diverse tasks like text removal, scratch removal, object removal, and block removal. In 2014, Chen et al. [50] have introduced edge detection scheme, which enhanced the conventional depth map inpainting approach using extracted edges. The implemented scheme had predicted the lost depth values effectively and it had proved an enhanced performance over the traditional algorithms. In 2016, Kawai et al. [51] have combined the local planes for enhancing the background geometry, and accordingly, the inpainting quality was enhanced. From the analysis, the modeled technique was found to be better over the other compared approaches.

In 2013, Zhao et al. [52] have presented a new passive recognition technique for inpainting the corrupted JPEG images when they were stored in JPEG compressed format. Finally, the investigational results have detected and located the lost region accurately with higher quality. In 2013, Qin et al. [53] have developed the self-recovery method for tampered images with image inpainting and VQ indexing. Moreover, the efficiency of the implemented approach was demonstrated from the experimental results. In 2013, Chang et al. [54] have modeled an innovative forgery detection scheme for identifying the corrupted inpainting images, which was considered as an effective scheme for image processing. In the end, experimental outcomes have illustrated that the introduced scheme was faster with excellent performance over the compared schemes.

In 2012, Dong et al. [55] have presented a blind inpainting method that identified and recovered the damaged pixels of the provided image. Finally, the experimentations have illustrated that the presented scheme was better over several two-staged approaches. In 2012, Zhang and Dai [56] have implemented a wavelet decomposition scheme for filling the corrupted image with both texture information and missing structures. Also, the adopted scheme restored the images quickly and the PSNR was also better under the conventional methods. In 2012, Liu

et al. [57] have established a patch-oriented image compressing model motivated by the inpainting approaches. The analysis outcomes have achieved a better gain and it has also saved the bit-rate at better quality. In 2011, Li et al. [58] have exploited the Chambolle's dual schemes that resolved the TV colorization scheme and TV inpainting scheme. From the numerical analysis, the presented scheme was found to be easier and faster for implementation.

In 2011, Du et al. [59] have developed model for image inpainting, for which MS method was exploited and also, the level set technique was deployed for estimating the structure of the corrupted portions. In 2010, Cai et al. [60] have developed integrated frame-oriented method for recovering the missing coefficients or bits throughout the process of compression. Finally, arithmetical experimentation had demonstrated the efficacy of the presented approach to enhance the visual quality of images. In 2010, Ojeda et al. [61] have developed an approach for carrying out image segmentation. Initially, an image was designed locally via an autoregressive approach. Furthermore, the investigational outcomes were offered that confirmed the practical efficiency of the presented algorithm.

In 2009, Li et al. [62] have adopted a restoration scheme, where the unwanted distortions get reduced in imaged documents. The efficiency of the introduced scheme was demonstrated through real and synthetic documents. In 2008, Cai et al. [63] have adopted iterative tight frame method for inpainting the images. Here, the relationship of the presented technique was discussed with other wavelet-oriented schemes, and its effectiveness was proved. In 2008, Wang and Sung [64] have introduced a new approach to automatic image recovery and authentication, where the distorted area of the image was recovered and detected automatically. At the final point, experimental outcomes were provided, which proved the efficacy of the presented approach. In 2007, Celia et al. [65] have adopted the method for restoring the images. The presented scheme also filled the corrupted or missing information along with noise removal.

3. A comprehensive review of adopted algorithms, Performance measures and the maximum achievements.

3.1. Analysis of Adopted Methods. This area makes a survey of various strategies embraced in each work, which is given in a diagrammatic depiction by Fig. 3.1. In the review, it was observed that Bregman method was exploited by [1, 13, 55] and Patch Scale Optimization was utilized in [3, 28, 33, 37, 57]. MRF scheme was adopted in [18, 20]. In addition, Wavelet transform model were adopted in [9, 12, 34, 56, 60, 63], while PDE schemes were adopted in [10, 14, 15, 22, 31, 39, 40, 46, 65]. Moreover, the TV-based model like the LRTV model was adopted in [5] and the TGV scheme was adopted in [23, 36, 38, 53]. Likewise, Gaussian approach was deployed in [4, 21, 24, 50], whereas NN based models like CNN was adopted in [19, 25, 42] and NN was used in [47]. FFT was adopted in [2, 7, 30]. Furthermore, Local gradient scheme, Criminisi algorithm, View synthesis method, Tight frame algorithm, Cross-correlation, LC model, Huffman coding, SVM, kernel Hilbert space, Region Growing model, Allen-Cahn model, ABC, Fragment splicing detection, Spatial interpolation, CGPS, Delaunay triangulation, forgery detection model, Chambolle's dual method, MS model, AR, RBF-smoothing and Wong's watermarking scheme were adopted by [35, 37-38, 41, 46, 48-49, 51, 56, 58-61, 63-64].

3.2. Analysis of Performance Measures. Table 3.1 depicts the performance measures determined in diverse contributions concerning image inpainting. From Table 3.1, it is noticed that PSNR performance analysis is done in 38 papers which contributed about 58.46% of the reviewed works, and 21 papers was analyzed on Computation time which contributed on 32.31% of the total works. Also, the SSIM, SNR, Standard deviation and MSE have been contributed around 16.92% (11 papers), 4.62% (3 papers), 4.62% (3 papers), and 9.23% (6 papers). Further, mean opinion score, error, and converging time have been adopted in 6.15% (4 papers), 7.69% (5 papers), and 4.62% (3 papers). Moreover, the Bit saving, GWN, AP, and compression have contributed about 3.08% of the entire contribution. Over 1.54% of total contributions have analyzed the measures like MSSIM, frequency of success, gain, no. of masked pixels, damage ratio, TPR, FPR, recall, maximum size, variance, subjective scores, versatility, sparsity, VSI, overlapping score, segmentation accuracy, CPU time, accuracy, computational speed missing ratio, frame rate, detection overlap, detection error, and tampering percentage respectively.

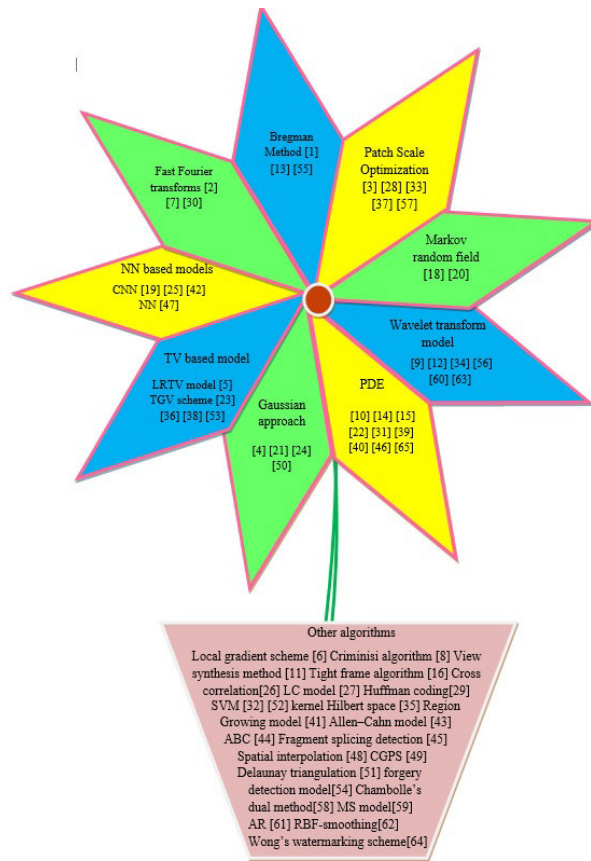


FIG. 3.1. Pictorial Representation of the Adopted Methods for image in painting models in the reviewed works

3.3. Maximum Performance. The most extreme presentation accomplished by all of the reviewed papers regarding the image inpainting system is shown in Table 3.2. In the review, PSNR presented in [29] has achieved a better value of 51.68dB, and computation time adopted in [6] has achieved a higher value of 0.44sec. Also, SSIM Value has achieved a better value of 0.8475 and it has been introduced by [19] and SNR Value has achieved a better value of 28.59 dB and it has been introduced by [16] respectively. Similarly, MSE, mean opinion score, error, and converging time have attained better values of 16.3269mm, 100, 0.043, and 1sec and it has been adopted by [4, 10, 17, 50]. The measures, bit saving, GWN, AP, compression and frame rate have attained higher values of 84.8%, 2%, 98.3, 7.751 and 29.94 fps and they have been adopted in [25, 29, 31, 35, 51]. Also, standard deviation, frequency of success, gain, number of masked pixels and damage ratio were introduced in [4, 6, 13-15], as well as they have obtained better values of 10, 1, 99.24%, 12469 and 7.5% correspondingly. Also, TPR, FPR, recall, Maximum Size, and Variance have attained higher values of 89.8, 1.4, 86.03%, 16 and 0.02 and they have been measured in [9, 10, 25, 54]. Also, tampering percentage, overlapping score, sparsity, segmentation accuracy, CPU time, accuracy and computational speed have attained higher values 33.9%, 0.91, 20%, 87%, 0.16s, 92% and 5,995ms correspondingly and they were presented in [34, 41, 43, 47, 53].

4. Assessment of adopted tools and chronological review in each work.

4.1. Review on Adopted Tools. The simulations of the reviewed works are made in diverse test systems such as MATLAB, C++, C, FreeFem++, Kakadu software, and so on. The pie chart representation of the adopted tools in the reviewed works is given by Fig. 4.1. In 29 papers MATLAB was implemented that have presented about 45% of the total contribution, and C++ was used in 4 papers which offered about 6% of the whole contribution. Also, the C language was offered in 1 paper and FreeFem++ have been presented in 1

TABLE 3.1
Review of various performance measures exploited for image inpainting schemes

Measures	Citations
PSNR	[2-6, 8-13, 16-22, 26-30, 34-35, 37-38, 42-43, 46, 48-49, 55-57, 60, 63-64]
Computation times	[2, 6-7, 14-15, 20, 23-24, 26, 30, 34, 40, 42, 44-45, 48-49, 51, 59, 62, 65]
SSIM value	[1, 3, 17-20, 22-23, 27-28, 37]
SNR value	[1, 16, 23]
MSE	[19, 21, 39, 48, 50, 61]
Mean opinion score	[17, 24, 40, 58]
Error	[2, 4, 43-44, 46]
Converging time	[10, 36, 38]
Bit saving	[31, 57]
GWN	[35, 61]
AP	[25, 54]
Compression	[29, 39]
MSSIM	[11]
Standard deviation	[13, 24, 58]
Frequency of success	[4]
Gain	[14]
No. of Masked pixels	[15]
Damage ratio	[6]
TPR	[25]
FPR	[25]
Recall	[54]
Maximum size	[9]
Variance	[10]
subjective scores	[32]
versatility	[33]
sparsity	[34]
VSI	[37]
Overlapping score	[41]
segmentation accuracy	[41]
CPU time	[43]
accuracy	[47]
computational speed	[47]
Missing ratio	[49]
Frame rate	[51]
Detection overlap	[52]
Detection error	[52]
Tampering percentage	[53]

paper, that has offered about 1% of the total contribution. Furthermore, Kakadu software has been exploited by 1 paper that contributes about 1% of the entire contribution. Accordingly, OpenCV library 2.3.1 was adopted by 2 papers that offer about 3.07% of the whole contribution.

4.2. Chronological Review. This review analyses various papers presented in different years. The percentage of contributions to corresponding years in pie chart format is illustrated in Fig. 4.2. Initially, 10.77% of papers are reviewed from the years, 2004-2009. Similarly, 18.46% and 21.54% of the total reviewed papers are existing in the year, 2010-2012 as well as 2013-2015. Moreover, 26.15% of contributions on image inpainting schemes are reviewed from the year 2016-2017. The papers reviewed for image inpainting in the years 2018-2019 is 23.07% of the whole contributions.

4.2.1. Research gaps and challenges. The research challenges and gaps on image inpainting systems are as follows:

- Image Inpainting [82, 83] is a technique that recovers the damaged images and it also concerns removing the undesirable portions from the image.
- This method eliminates the breaks in the image, eradicates the texts, and fills the missing part from the image.

TABLE 3.2
Review of various performance measures exploited for image inpainting schemes

Sl. no	Author [Citation]	Performance measures	Maximum performance
1	[29]	PSNR	51.68dB
2	[6]	Computation Times	0.44sec
3	[19]	SSIM Value	0.8475
4	[16]	SNR Value	28.59 dB
5	[50]	MSE	16.3269mm
6	[17]	Mean opinion score	100
7	[4]	Error	0.043
8	[10]	Converging Time	1sec
9	[31]	Bit Saving	84.8%
10	[35]	GWN	2%
11	[25]	AP	98.3
12	[29]	Compression	7.751
13	[51]	Frame Rate	29.94 fps
14	[13]	Standard Deviation	10
15	[4]	Frequency of Success	1
16	[14]	Gain	99.24%,
17	[15]	No. of Masked Pixels	12469
18	[6]	Damage Ratio	7.5%
19	[25]	TPR	89.8
20	[25]	FPR	1.4
21	[54]	Recall	86.03%
22	[9]	Maximum Size	16
23	[10]	Variance	0.02
24	[53]	Tampering Percentage	33.9%
25	[41]	Overlapping Score	0.91
26	[34]	Sparsity	20%
27	[41]	Segmentation accuracy	87%
28	[43]	CPU time	0.16s
29	[47]	Accuracy	92%
30	[47]	Computational speed	5,995ms

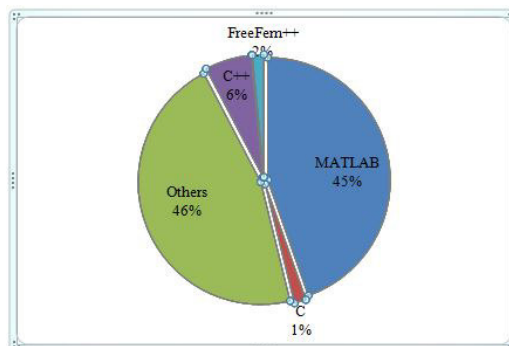


FIG. 4.1. Analysis on adopted tools in reviewed works

- Inpainting can be done by an individual, if he has more knowledge regarding this technique or if he is focused in that field [84, 85].
- However, owing to manual processes, it includes more time to offer essential results. For inpainting an ancient painting or to inpaint a scratched image with lost regions, it is required to estimate and fill up the missing image regions such that the painting or restored image appears as likely as its original version [86].
- Exactly, what formulates the inpainting issue so inspirational is the complexity of image benefits.
- Also, the image functions with multilevel complexities have forced the analysts to develop inpainting

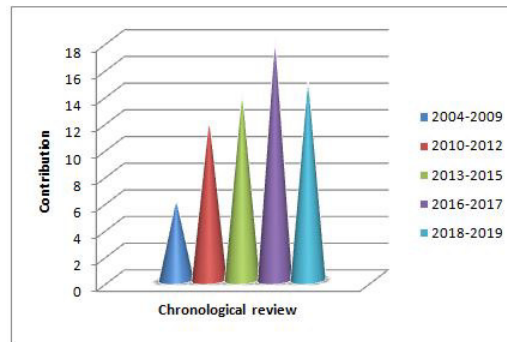


FIG. 4.2. Bar chart representing a chronological review

structures targeted at the real versions of images. Therefore, these inpainting approaches are at low stages.

- The significant challenging task in image inpainting method is the computation of the image quality, which should be the same before and after the inpainting process [87, 88].
- In recent days, various image inpainting schemes are available. Also, numerous attacks were deployed by the analysts for digital image inpainting that are categorized under the following classes such as, PDE based inpainting, hybrid inpainting, texture synthesis based inpainting, and exemplar-based inpainting. PDE oriented schemes can be well suitable for filling the little gaps, text overlay, and so on.
- However, the PDE method generally fails if exploited to the textured field or regions with regular patterns.
- The entire texture oriented schemes vary concerning their ability to produce texture with differed statistical features, gradient, and colour intensity.
- Also, the texture inpainting technique could not be adopted well for natural images and they do not include the capability to handle boundaries and edges effectively.
- At certain conditions, the user needs to decide which texture should be replaced by which one.

Thus, these schemes can be deployed for inpainting [89, 90] only the small regions, and more developments are required for inpainting the larger regions in an efficient manner.

5. Conclusion. This paper has introduced a survey of image inpainting models. In this, different methodologies adopted in the reviewed works were analyzed and exhibited. Moreover, this review analyzes the performance measures related with the reviewed image inpainting scheme. In conclusion,

- ▷ This paper looked around 65 research papers and stated a noteworthy investigation of different algorithms.
- ▷ The analysis has reviewed the performance measures and the corresponding maximum achievements contributed by different image inpainting schemes.
- ▷ Further, the various tools exploited in every reviewed works were also analyzed as well as specified diagrammatically.
- ▷ Also, the chronological review was done for the analyzed 65 works.
- ▷ Finally, this paper offered diverse research challenges that could be helpful for the specialists to accomplish more examination on the highlights of image inpainting schemes.

REFERENCES

- [1] L. HE, Y. WANG, *Iterative Support Detection-Based Split Bregman Method for Wavelet Frame-Based Image Inpainting*, IEEE Journals & Magazines, Vol. 23, no.12, pp. 5470 – 5485, 2014
- [2] D-Q. CHEN, Y. ZHOU, *Inexact alternating direction method based on proximity projection operator for image inpainting in wavelet domain*, Neurocomputing, Vol.189, pp.145-159, 12 May 2016
- [3] Z. CHEN, C. DAI, L. JIANG, B. SHENG, Y. YUAN, *Structure-aware image inpainting using patch scale optimization*, Journal of Visual Communication and Image Representation, Vol. 40, pp. 312-323, October 2016.

- [4] F. WEN, L. ADHIKARI, L. PEI, R.F. MARCIA, P. LIU, R.C. QIU, *The \LaTeX Nonconvex Regularization-Based Sparse Recovery and Demixing With Application to Color Image Inpainting*, IEEE Journals & Magazines, Vol.05, pp. 11513 - 11527, 2017.
- [5] H. XUE, S. ZHANG, D. CAI, *Depth Image Inpainting: Improving Low Rank Matrix Completion With Low Gradient Regularization*, IEEE Journals & Magazines, Vol. 26, no.09, pp.4311 - 4320, 2017.
- [6] V. K. ALILOU, F. YAGHMAEE, *Non-texture image inpainting using histogram of oriented gradients*, Journal of Visual Communication and Image Representation, Vol. 48, pp.43-53, October 2017.
- [7] H. WANG, L. JIANG, R. LIANG, X-X. L, *Exemplar-based image inpainting using structure consistent patch matching*, Neurocomputing, 31 May 2017.
- [8] H. YING, L. KAI, Y. MING, , *An Improved Image Inpainting Algorithm Based on Image Segmentation*, Procedia Computer Science, Vol. 107, pp.796-801, 2017.
- [9] U.A. IGNACIO, C.R. JUNG, *Block-based image inpainting in the wavelet domain*, Vol. 23, pp. 733-741, 2007.
- [10] T. BARBU, *Variational image inpainting technique based on nonlinear second-order diffusion* Computers & Electrical Engineering, Vol. 54, pp. 345-353, August 2016.
- [11] S.M. MUDDALA, M. SJÖSTRÖM, R. OLSSON, *Virtual view synthesis using layered depth image generation and depth-based inpainting for filling disocclusions and translucent disocclusions*, Journal of Visual Communication and Image Representation, Vol.38, pp. 351-366, July 2016.
- [12] A.V. BHAVSAR, A.N. RAJAGOPALAN, *Range map superresolution-inpainting, and reconstruction from sparse data*, Computer Vision and Image Understanding, vol. 116, no. 4, pp.572-591, April 2012.
- [13] B. DHYANESH AND K. S. SATHIYAPRIYA , *Image inpainting and image denoising in wavelet domain using fast curve evolution algorithm*, IEEE on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, pp. 166-169, 2012.
- [14] M. WANG, B. YAN, K.N. NGAN, *An efficient framework for image/video inpainting* Image Communication, Vol. 28, no.07, pp.753-762, August 2013.
- [15] M. LI, Y. WEN, *The A New Image Inpainting Method Based on TV Model*, Physics Procedia, vol. 33, pp. 712-717, 2012.
- [16] C. CHUNHONG, D. WEI, W. MINMIN AND H. KAI, *Inpainting of multiple blind motion-blurred images based on multi-scale tight wavelet frame*, 2017 3rd IEEE International Conference on Computer and Communications (ICCC), pp. 1841-1846, Chengdu, 2017.
- [17] W. HU, Y. YE, F. ZENG, J. MENG, *A new method of Thangka image inpainting quality assessment*, Journal of Visual Communication and Image Representation, vol. 59, pp. 292-299, February 2019.
- [18] Q. LIU, S. LI, J. XIAO, M. ZHANG, *Multi-filters guided low-rank tensor coding for image inpainting* Signal Processing: Image Communication, vol. 73, pp. 70-83, April 2019.
- [19] L. JIAO, H. WU, H. WANG, R. BIE , *Multi-scale semantic image inpainting with residual learning and GAN*, Neurocomputing, vol. 331, pp. 199-212, 28 February 2019..
- [20] J. CHENG, Z. LI, *Markov random field-based image inpainting with direction structure distribution analysis for maintaining structure coherence*, Signal Processing, vol.154, pp.182-197, January 2019.
- [21] A. SOBIECKI, G. A. GIRALDI, L. A. PEREIRA NEVES AND C. E. THOMAZ, *An Automatic Framework for Segmentation and Digital Inpainting of 2D Frontal Face Images*, IEEE Latin America Transactions, vol. 10, no. 6, pp. 2263-2272, Dec. 2012.
- [22] A. THELJANI, Z. BELHACHMI, M. MOAKHER, *High-order anisotropic diffusion operators in spaces of variable exponents and application to image inpainting and restoration problems* Nonlinear Analysis: Real World Applications, vol. 47, pp. 251-271, June 2019.
- [23] S. WALI, H. ZHANG, H. CHANG, C. WU, *A new adaptive boosting total generalized variation (TGV) technique for image denoising and inpainting*, Journal of Visual Communication and Image Representation, vol. 59, pp. 39-51, February 2019.
- [24] D. DING, S. RAM, J. J. RODRIGUEZ, *Perceptually aware image inpainting*, Pattern Recognition, vol. 83, pp.174-184, November 2018.
- [25] X. ZHU, Y. QIAN, X. ZHAO, B. SUN, Y. SUN, *A deep learning approach to patch-based image inpainting forensics*, Signal Processing: Image Communication, vol. 67, pp. 90-99, September 2018.
- [26] E. KARACA, M. A. TUNGA, *An interpolation-based texture and pattern preserving algorithm for inpainting color images* Expert Systems with Applications, vol. 91, pp. 223-234, January 2018.
- [27] X. YAN, Y. LU, L. LIU, S. WANG, *The \LaTeX Partial secret image sharing for (k,n) threshold based on image inpainting*, Journal of Visual Communication and Image Representation, vol. 50, pp. 135-144, January 2018.
- [28] D. HAN, H. CHEN, C. TU, Y. XU, *View synthesis using foreground object extraction for disparity control and image inpainting*, Journal of Visual Communication and Image Representation, vol. 56, pp. 287-295, October 2018.
- [29] C. QIN, Z. HE, H. YAO, F. CAO, L. GAO, *Visible watermark removal scheme based on reversible data hiding and image inpainting*, Signal Processing: Image Communication, vol. 60, pp. 160-172, February 2018.
- [30] L. TAN, W. LIU, Z. PAN, *Color image restoration and inpainting via multi-channel total curvature* Applied Mathematical Modelling, vol. 61, pp. 280-299, September 2018.
- [31] Z. XIONG, X. SUN AND F. WU, *Block-Based Image Compression With Parameter-Assistant Inpainting*, IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1651-1657, June 2010.
- [32] M. ISOGAWA, D. MIKAMI, D. IWAI, H. KIMATA AND K. SATO, *Mask Optimization for Image Inpainting*, IEEE Access, vol. 6, pp. 69728-69741, 2018.
- [33] A. CRIMINISI, P. PEREZ AND K. TOYAMA, *Region filling and object removal by exemplar-based image inpainting*, IEEE Transactions on Image Processing, vol. 13, no. 9, pp. 1200-1212, Sept. 2004.
- [34] F. LI, X. LV, *A Decoupled method for image inpainting with patch-based low rank regularization* Applied Mathematics and Computation, vol. 314, pp. 334-348, 1 December 2017.

- [35] S. WANG, W. GUO, T-Z. HUANG, G. RASKUTTI, *Image inpainting using reproducing kernel Hilbert space and Heaviside functions* Journal of Computational and Applied Mathematics, vol. 311, pp. 551-564, February 2017.
- [36] M. FUCHS, J. MÜLLER, *A higher order TV-type variational problem related to the denoising and inpainting of images* Nonlinear Analysis: Theory, Methods & Applications, vol. 154, pp. 122-147, May 2017.
- [37] A. TRAN, H. TRAN, *Data-driven high-fidelity 2D microstructure reconstruction via non-local patch-based image inpainting* Acta Materialia, vol. 178, pp. 207-218, 1 October 2019.
- [38] L. SHEN, Y. XU, X. ZENG, *Wavelet inpainting with the 0 sparse regularization* Applied and Computational Harmonic Analysis, vol. 41, no. 1, pp. 26-53, July 2016.
- [39] P. PETER, S. HOFFMANN, F. NEDWED, L. HOELTGEN, J. WEICKERT, *Evaluating the true potential of diffusion-based inpainting in a compression context* Signal Processing: Image Communication, vol. 46, pp. 40-53, August 2016.
- [40] A. COLOMER, V. NARANJO, K. ENGAN, K. SKRETTING, *Assessment of sparse-based inpainting for retinal vessel removal* Signal Processing: Image Communication, vol. 59, pp. 73-82, November 2017.
- [41] M. P. SARATHI, M. K. DUTTA, A. SINGH, C. M. TRAVIESO, *Blood vessel inpainting based technique for efficient localization and segmentation of optic disc in digital fundus images* Biomedical Signal Processing and Control, vol. 25, pp. 108-117, March 2016.
- [42] X. ZHANG, F. DING, Z. TANG, C. YU, *Salt and pepper noise removal with image inpainting* AEU - International Journal of Electronics and Communications, vol. 69, no. 1, pp. 307-313, January 2015.
- [43] Y. LI, D. JEONG, J. CHOI, S. LEE, J. KIM, *Fast local image inpainting based on the Allen-Cahn model* Digital Signal Processing, vol. 37, pp. 65-74, February 2015.
- [44] A. S. M. JIAO, P. W. M. TSANG, T. -C. POON, *Restoration of digital off-axis Fresnel hologram by exemplar and search based image inpainting with enhanced computing speed* Computer Physics Communications, vol. 193, pp. 30-37, August 2015.
- [45] Z. LIANG, G. YANG, X. DING, L. LI, *An efficient forgery detection algorithm for object removal by exemplar-based image inpainting* Journal of Visual Communication and Image Representation, vol. 30, pp. 75-85, July 2015.
- [46] F. BERNTSSON, G. BARAVDISH, *Coefficient identification in PDEs applied to image inpainting* Applied Mathematics and Computation, vol. 242, pp. 227-235, 1 September 2014.
- [47] M. FAVORSKAYA, L. C. JAIN, A. BOLGOV, *Image Inpainting based on Self-organizing Maps by Using Multi-agent Implementation* Procedia Computer Science, vol. 35, pp. 861-870, 2014.
- [48] B. CHUNG, C. YIM, *Hybrid error concealment method combining exemplar-based image inpainting and spatial interpolation* Signal Processing: Image Communication, vol. 29, no. 10, pp. 1121-1137, November 2014.
- [49] Z. LI, H. HE, Z. YIN, F. CHEN, *A color-gradient patch sparsity based image inpainting algorithm with structure coherence and neighborhood consistency* Signal Processing, vol. 99, pp. 116-128, June 2014.
- [50] W. CHEN, H. YUE, J. WANG, X. WU, *An improved edge detection algorithm for depth map inpainting* Optics and Lasers in Engineering, vol. 55, pp. 69-77, April 2014.
- [51] N. KAWAI, T. SATO AND N. YOKOYA, *Diminished Reality Based on Image Inpainting Considering Background Geometry*, IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 3, pp. 1236-1247, 1 March 2016.
- [52] Y. Q. ZHAO, M. LIAO, F. Y. SHIH, Y. Q. SHI, *Tampered region detection of inpainting JPEG images* Optik, vol. 124, no. 16, pp. 2487-2492, August 2013.
- [53] C. QIN, C-C. CHANG, K-N. CHEN, *Adaptive self-recovery for tampered images based on VQ indexing and inpainting* Signal Processing, vol. 93, no. 4, pp. 933-946, April 2013.
- [54] I-C. CHANG, J. C. YU, C-C. CHANG, *A forgery detection algorithm for exemplar-based inpainting images using multi-region relation* Image and Vision Computing, vol. 31, no.1, pp. 57-71, January 2013.
- [55] B. DONG, H. JI, J. LI, Z. SHEN, Y. XU, *Wavelet frame based blind image inpainting* Applied and Computational Harmonic Analysis, vol.32, no. 2, pp. 268-279, March 2012.
- [56] H. ZHANG, S. DAI, *Image Inpainting Based on Wavelet Decomposition* Procedia Engineering, vol. 29, pp. 3674-3678, 2012.
- [57] D. LIU, X. SUN, F. WU, *Inpainting with image patches for compression* Journal of Visual Communication and Image Representation, vol. 23, no. 1, pp. 100-113, January 2012.
- [58] F. LI, Z. BAO, R. LIU, G. ZHANG, *Fast image inpainting and colorization by Chambolle's dual method* Journal of Visual Communication and Image Representation, vol. 22, no. 6, pp. 529-542, August 2011.
- [59] X. DU, D. CHO, T. D. BUI, *Image segmentation and inpainting using hierarchical level set and texture mapping* Signal Processing, vol. 91, no. 4, pp. 852-863, April 2011.
- [60] J-F. CAI, H. JI, F. SHANG, Z. SHEN, *Inpainting for compressed images* Applied and Computational Harmonic Analysis, vol. 29, no. 3, pp.368-381, November 2010.
- [61] S. OJEDA, R. VALLEJOS, O. BUSTOS, *A new image segmentation algorithm with applications to image inpainting* Computational Statistics & Data Analysis, vol. 54, no. 9, pp. 2082-2093, 1 September 2010.
- [62] L. ZHANG, A. M. YIP, M. S. BROWN, C. L. TAN, *A unified framework for document restoration using inpainting and shape-from-shading* Pattern Recognition, vol. 42, no. 11, pp. 2961-2978, November 2009.
- [63] J-F. CAI, R. H. CHAN, Z. SHEN, *A framelet-based image inpainting algorithm* Applied and Computational Harmonic Analysis, vol. 24, no. 2, pp. 131-149, March 2008.
- [64] S-S. WANG, S-L. TSAI, *Automatic image authentication and recovery using fractal code embedding and image inpainting* Pattern Recognition, vol. 41, no. 2, pp. 701-712, February 2008.
- [65] C. A. Z. BARCELOS, M. A. BATISTA, *Image restoration using digital inpainting and noise removal* Image and Vision Computing, vol. 25, no. 1, pp. 61-69, January 2007.
- [66] K. SHI, Z. GUO, *On the existence of weak solutions for a curvature driven elliptic system applied to image inpainting* Applied Mathematics Letters, In press, journal pre-proof, Available online 16 August 2019, Article 106003.
- [67] X. YANG, B. GUO, Z. XIAO, W. LIANG, *Improved structure tensor for fine-grained texture inpainting* Signal Processing:

- Image Communication, vol. 73, pp. 84-95, April 2019.
- [68] F. WU, Y. KONG, W. DONG, Y. WU, *Gradient-aware blind face inpainting for deep face verification* Neurocomputing, vol. 331, pp. 301-311, 28 February 2019.
- [69] M. A. QURESHI, M. DERICHE, A. BEGHADI, A. AMIN, *A critical survey of state-of-the-art image inpainting quality assessment metrics* Journal of Visual Communication and Image Representation, vol. 49, pp. 177-191, November 2017.
- [70] H. WANG, Z. HE, Y. HE, D. CHEN, Y. HUANG, *Average-face-based virtual inpainting for severely damaged statues of Dazu Rock Carvings* Journal of Cultural Heritage, vol. 36, pp. 40-50, March–April 2019.
- [71] T. SANDERS, C. DWYER, *Subsampling and inpainting approaches for electron tomography* Ultramicroscopy, vol. 182, pp. 292-302, November 2017.
- [72] M. JAMPOUR, C. LI, L-F. YU, K. ZHOU, H. BISCHOF, *Face inpainting based on high-level facial attributes* Computer Vision and Image Understanding, vol. 161, pp. 29-41, August 2017.
- [73] P. TRAMPERT, W. WANG, D. CHEN, R. B. G. RAVELLI, P. SLUSALLEK, *Exemplar-based inpainting as a solution to the missing wedge problem in electron tomography* Ultramicroscopy, vol. 191, pp. 1-10, August 2018.
- [74] A. IGNAT, A. VASILIU, , *A study of some fast inpainting methods with application to iris reconstruction* Procedia Computer Science, vol. 126, pp. 616-625, 2018.
- [75] S. LI, M. ZHAO, *Image inpainting with salient structure completion and texture propagation* Pattern Recognition Letters, vol. 32, no. 9, pp. 1256-1266, 1 July 2011.
- [76] W. CASACA, M. BOAVENTURA, M. P. ALMEIDA, L. G. NONATO, *Combining anisotropic diffusion, transport equation and texture synthesis for inpainting textured images* Pattern Recognition Letters, vol. 36, pp. 36-45, 15 January 2014.
- [77] T-Y. KUO, P-C. SU, Y-P. KUAN, *SIFT-guided multi-resolution video inpainting with innovative scheduling mechanism and irregular patch matching* Information Sciences, vol. 373, pp. 95-109, 10 December 2016.
- [78] J. FAYER, B. DURIX, S. GASPARINI, G. MORIN, *Texturing and inpainting a complete tubular 3D object reconstructed from partial views* Computers & Graphics, vol. 74, pp. 126-136, August 2018.
- [79] A. BILESAN, M. OWLIA, S. BEHZADIPOUR, S. OGAWA, A. KONNO, *Marker-based motion tracking using Microsoft Kinect* IFAC-PapersOnLine, vol. 51, no. 22, pp. 399-404, 2018.
- [80] S. BADR, P. NOUFAL, *Integrated Data Hiding and Compression Scheme Based on SMVQ and FoE Inpainting* Procedia Technology, vol. 24, pp. 1008-1015, 2016.
- [81] X-H. YANG, B-L. GUO, X-X. WU, *Wavelet Inpainting Based on Tensor Diffusion* Acta Automatica Sinica, vol. 39, no. 7, pp. 1071-1079, July 2013.
- [82] A. CHEVEIGNÉ, D. ARZOUNIAN, *Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data* NeuroImage, vol. 172, pp. 903-912, 15 May 2018.
- [83] F. CHEN, T. HU, L. ZUO, Z. PENG, M. YU, *Depth map inpainting via sparse distortion model* Digital Signal Processing, vol. 58, pp. 93-101, November 2016.
- [84] X. YANG, P. XU, H. JIN, J. ZHANG, *Low-rank tensor completion with fractional-Jacobian-extended tensor regularization for multi-component visual data inpainting* Digital Signal Processing, In press, journal pre-proof, Available online 13 August 2019..
- [85] D. CALVETTI, F. SGALLARI, E. SOMERSALO, *Image inpainting with structural bootstrap priors* Image and Vision Computing, vol. 24, no. 7, pp. 782-793, 1 July 2006.
- [86] M. GHONIEM, Y. CHAHIR, A. ELMOATAZ, *Nonlocal video denoising, simplification and inpainting using discrete regularization on graphs* Signal Processing, vol. 90, no. 8, pp. 2445-2455, August 2010.
- [87] J. A. DOBROSOTSKAYA AND A. L. BERTOZZI, *A Wavelet-Laplace Variational Technique for Image Deconvolution and Inpainting* IEEE Transactions on Image Processing, vol. 17, no. 5, pp. 657-663, May 2008.
- [88] D. DING, S. RAM AND J. J. RODRÍGUEZ, *Image Inpainting Using Nonlocal Texture Matching and Nonlinear Filtering* IEEE Transactions on Image Processing, vol. 28, no. 4, pp. 1705-1719, April 2019.
- [89] H. LI, W. LUO AND J. HUANG, *Localization of Diffusion-Based Inpainting in Digital Images* IEEE Transactions on Information Forensics and Security, vol. 12, no. 12, pp. 3050-3064, Dec. 2017.
- [90] R. L. BIRADAR1, AND V.V KOHIR, *A novel image inpainting technique based on median diffusion* Indian Academy of Sciences, vol. 38, Part 4, pp. 621–644, August 2013.
- [91] M. GHORAI, S. MANDAL, AND B. CHANDA, *A group-based image inpainting using patch refinement in MRF framework* IEEE Transactions on Image Processing, 27(2), pp.556-567, 2017.
- [92] T. BARBU, *Variational image inpainting technique based on nonlinear second-order diffusions* Computers & Electrical Engineering, 54, pp.345-353, 2016.
- [93] V. KUMAR, J. MUKHERJEE, AND S.K.D. MANDAL, *Image inpainting through metric labeling via guided patch mixing* IEEE Transactions on Image Processing, 25(11), pp.5212-5226, 2016.
- [94] N. CAI, Z. SU, Z. LIN, H. WANG, Z. YANG, AND B.W.K. LING, *Blind inpainting using the fully convolutional neural network* The Visual Computer, 33(2), pp.249-261, 2017.

Edited by: P. Vijaya

Received: Dec 7, 2019

Accepted: Jun 22, 2020



AN EFFICIENT WAY OF FINDING POLARITY OF ROMAN URDU REVIEWS BY USING BOOLEAN RULES

HALIMA SADIA, MOHIB ULLAH, TARIQ HUSSAIN*, NIDA GUL, MUHAMMAD FAROOQ HUSSAIN, NAUMAN UL
HAQ, AND ABU BAKAR

Abstract. Opinion mining is the technique of analyzing the sentiment, behavior, feelings, emotions, and attitudes of customers about a product, topic, comments on social media, etc. Online shopping has revolutionized the way customers do shopping. The customer likes to visit the online store to find their product of interest. It is becoming more difficult for customers to make purchasing decisions solely based on photos and product descriptions. Customer reviews provides a rich source of information to compare products and make purchasing decisions commonly on the basis of other customer reviews. Clients provide comments in the language of their choice, e.g. the people of Pakistan use Roman script based the Urdu language. Normally such comments are free from scripting rules. Hundreds of comments are given on a single product, which may contain noisy comments. Identifying noisy comments and finding the polarity of these comments is an active area of research. Limited research is being carried out on roman Urdu sentiment analysis. In this research paper, we propose a novel approach by using Boolean rules for the identification of the related and non-related comments. Related reviews are those which show the behavior of a customer about a particular product. Lexicons are built for the identification of noise, positive and negative reviews. The precision of the evaluation results is 68%, recall is also 68% and F-measure is 68% The accuracy of the whole evaluation is 60%.

Key words: Roman Urdu, Polarity, Boolean rules, Opinion Mining, Noise, Reviews polynomial

AMS subject classifications. 90C09

1. Introduction. Sentiment analysis methodologies are proposed for the classification of opinions in multiple languages. Opinion mining also known as sentiment mining is a process of identifying opinions from any language. Opinion mining is a combination of data mining and natural language processing techniques [1]. In [2] define opinion mining as the process that searches results of a given product and produce a list of their reviews attribute. Customer opinion mining is the process of checking customer interest in a particular product [3]. Opinion mining is a technique of analyzing reviews of customers which helps both the customer and retailer. The retailer easily identifies customer interest and demand for products from these reviews. Another angle of opinion is customer convincing customers [4]. Most of the customers can make their purchasing decision based on customer opinions present on the retailer website [5]. It is time-consuming to extract important information from multiple reviews over one product. The Internet is an important part of our daily life. Internet brings us a lot of advantages; online shopping is one of them, it easier for the customers to buy goods or services from a vendor. Consumer finds their product of interest by visiting the vendor's website[6]. It is difficult for customers to use Urdu based keyboards as most handheld devices need software updates and installations packages for using Urdu languages [7]. The customers mostly rely on the Roman script for Urdu which is easily available but unfortunately with no scripting structure. Normally Urdu language written in Roman script is called Roman Urdu. Let us take an example of Roman Urdu phrase "ye mobile bht mehnga he" or "ye mobile mene istemal kia he bht acha h". On shopping sites over internet, customers use their native languages for posting their reviews and these reviews are understood by few people, mining useful information from these reviews become a difficult task [8][4]. Especially, it creates a problem for non-native users to understand these reviews and to obtain useful information. Another main issue is syntax and the changing meaning of words in different situations. Some words show positive meaning in one situation and negative in another[9]. For example, "ye mobile acha he" in this sentence the word "acha" shows that mobile is good. If we say ye mobile "acha nahi" he

*Institute of Computer Science and Information Technology, The University of Agriculture Peshawar 25130, Pakistan.
(uom.tariq@gmail.com)

in this sentence the word "acha" is the same but the word "nahi" changing the semantics change and that means mobile is not good. Another important issue is the use of several syntax for the same word by the different users e.g. multiple users will use roman urdu word "acha" meaning "good" in English with different syntax like "icha" or "echa" etc. So, it is difficult to handle these reviews according to its orientation [10] [11]. Also, if the sentence contains both positive and negative opinion words like "ye mobile acha nahi he" here the word "acha" is positive, and "nahi" is negative so the opinion is negative, but they consider it as positive only because of the word "acha". Customer reviews are posted in different native languages. NLP techniques which enable computers to understand human natural languages. Linguistics tasks are performed through NLP such as translation, summarization, analyses, Part-of-Speech (POS) tagging, information retrieval, speech recognition [12][13]. In opinion mining, POS tagging techniques plays a vital role as it is used to identify different features word and opinion sentence in opinion repository. Many types of lexicons are available for word matching and feature identifications. Many researchers have worked on opinion mining and proposed different models. But most of them worked over English, Chinese, Arabic, Urdu, Pashto and Sindhi, etc [14][17]. Urdu is the national language of Pakistan and is like the Hindi language. Both the languages have different scripts but having some pronunciations. They are mostly used in South East Asia with nearly 1.5 billion populations [10][18]. Users also use Roman-Urdu on social media when communicating with their friends and posting their comments on different blogs and websites. This proposed model is used to identify the polarity of customer reviews written in Roman Urdu by applying proposed Boolean rules in form of precision, recall, F-measure and accuracy.

This research paper has several contributions:

- With the emerging power of internet online shopping has become more attractive to customers. Customer likes to buy products online, but before buying the products they visit reviews given by other customers. Reviews play an important role in buying decisions of a customer. Customer post their reviews in the language of their choice, in Pakistan people normally post their reviews in Roman Urdu. Roman Urdu is unstructured free text. It is observed from the customer reviews that there are many noisy comments in the review list Identifying customer reviews as noise or related comments and finding the polarity of Roman Urdu comments is a research issue. Another problem is that two negative words change the polarity of a review, this is an ultimate need to identify and categorize those comments. In this research is going to identify noisy and related comments and find the polarity of those comments.
- The paper provides how to identify noise in user reviews? " In this research discusses how to find polarity of roman Urdu reviews?
- In this research proposed to identify related and noisy comments from Roman Urdu reviews "
- Introduces state of the art model to find the polarity of comments using proposed Boolean Rules.

In the next section 2, the literature review of the Roman Urdu and boolean rules. In section 3, boolean rules Principles and Categories is explained. Section 4 explains the results evaluation. Section 5 provides the experimental results of the roman urdu language by using boolean rules. And finally, conclusions, recommendation and future work is explained in section 6 and 7.

2. Related works. The authors in [19] have worked on opinion mining. They tried to mine all opinions of customers and summarize them. Their summarization task was different from traditional summarization because they only mine the features of a product on which the customer has expressed their opinions. First, they mine the product features which are commented by the users, secondly identify the opinion sentence and its polarity and third summarizing the results. The experimental results show that techniques are more effective. The research conducted in [12] has presented their work in the Sindhi language. They proposed a technique for part of speech tagging. And develop a linguistic rule for the Sindhi language. The orthography of Sindhi language is difficult due to the absence of diacritic symbols. For the development of the Sindhi Part of Speech (SPOS) tagging system, they used a supervised approach. They set 186 disambiguation rules for POS tagging. The POS tagging algorithm take words and check into the lexicon if the word were present in the lexicon, the associated tag was assigned and if the word was not presented in the lexicon then tag set was assigned according to linguistic rules. The output of the system was Sindhi word, in English, and the Sindhi language. In [20] present models for finding the polarity of tweets. They build a two-way task for classifying the polarity of tweets as positive and negative and three-way tasks for finding polarity as positive, negative,

and neutral. They utilize three types of models for their experiment i.e. unigram, feature-based and tree kernel model. As a baseline, they used the n-gram model, for the feature-based model they used for tree kernel base model they design a tree of tweets. Their experimental results for a two-way task shows that unigram model is a hard baseline which only achieve 20% in both tasks and the feature-based model achieve similar accuracy as the unigram model, but the tree kernel model achieves the accuracy of 2.58% and 2.66% over these two models. They also experimented with these models in combination. Combining unigrams with feature models achieved 0.78% over the combined model of the feature model with tree kernel. For all their experiment they used SVM (support vector machine). Experimental results for a three-way task shows that tree kernel model achieved more accuracy form unigrams and baseline model. It achieved 4.02% accuracy over unigram model and 4.29% over Senti-feature model. In, [21] proposed a system for mining opinion written in Arabic language. The proposed method was the combination of three methods i.e. lexicon-based method, maximum entropy, and KNN method. Set of documents were classified by using a lexicon-based method for training machine. Then for the maximum entropy method, these classified documents act as a training set. Further other documents are classified as maximum entropy. For classification of the rest documents that output of the two methods was used as a training set for the k-nearest method. Experimental results showed that 50% when only the lexicon-based method was applied to the dataset it gives 50% accuracy which exceeded to 60% when lexicon and entropy-based methods were applied in the combination and 80% when three methods were applied together. In, [22] worked on Punjabi text classification. They proposed a hybrid system by blending Naïve Bayesian and N-grams. They extract the features of N-grams and used it to train Naïve Bayesian. They then validated the trained model using testing data. Experimental results were compared with existing models and results from comparison show a better efficiency of the proposed method. In [23] proposed for text classification techniques for Roman-Urdu reviews by using the Waikato Environment for Knowledge Analysis (WEKA). With the emerging use of the internet and e-commerce, opinion mining and sentiment analysis become a very important field for both researchers and retailers. They create a data set of 150 positive and 150 negative reviews for training machine. Naïve Bayesian, Decision Tree and KNN classification models were designed to analyze the polarity of new customer reviews based on trained data set. Their result shows that and F-measure and Naïve Bayesian performed better than Decision Tree and KNN classification models in terms of accuracy, precision, recall. In, [4] proposed a method in which they provide a facility to the non-Urdu speaking customers to get benefit from the comment posted in Roman Urdu. They took the data from a website called WHATMOBILE where people give there-reviews over a phone they want to buy or been using. They define their work in to 4 main steps. They 1st made a Crawler which take the comments from the site which contain both useful and noisy comments, Then they use BING Translator to translate the comments from roman Urdu to English to make the computer understand, they then extract their opinion the relevant comments by using POS and preprocessing and removing the noise and use their local database for the Identification of the opinion polarity and at last show the user reviews and rating in graphical form. The experimental results recorded 27% precision and categorized 21.1% reviews falsely due to noisy data.

3. Proposed method. This model consists of five steps. First, reviews written in roman Urdu are extracted from the mobile website whatmobile.com written in Roman-Urdu and are collected in a word document. These reviews are positive, negative, and noisy. Different lexicons for positive, negative, and for noise identifying are created for testing and results. After the data processing, the stopping words are removed from the reviews. The reviews are then parsed according to the noise dictionary and noise is removed. After removing noise, the reviews are parsed for identifying the polarity of reviews through Boolean Rules.

Rule no 1 According to AND gate, when there is one positive and one negative word exist in a review then the polarity of the review will be negative.

Rule no 2 According to the XNOR gate, when two negatives exist in a review then the polarity of the review will be positive.

Rule no 3 According to the XNOR gate, when two positive words are existing in a review then the polarity of the review will be positive.

Rule no 4 According to AND gate, when one negative and one positive word exists in a review then the polarity of the review will be negative.

Rule no 5 If there is a single positive in a review then the review will be positive.

Rule no 6 If there is a single negative exists in a review then the review will be negative.

3.1. Preprocessing. In this step, noisy data present in the reviews is eliminated. Irrelevant views like comments which do not have any positive or negative opinion about products will be categorized as noisy data. Lexicon of noise is created to identify noise from reviews.

Algorithm 1 The proposed algorithm

```

1: Input  $R = R_1, R_2, R_3, R_4, \dots, R_n$ ,
2: Output related, non-related, positive & negative
3: Step 1: Preprocessing
4: For  $C_i$  to the Size of Array
5:  $C_i \leftarrow \dots$  Remove stop words  $C_i$ .
6:  $TC \leftarrow \dots$  Tokensize  $C_i$ .
7: Step 2: Identifying related and Noisy Components
8: For  $j = 0$  to  $TC_j$  length by 1.
9: IF ( $TC_j \in$  noisy Dictionary).
10: noisycomponent.add( $C_i$ ).
11: else.
12: Relatedcomponent.add( $C_i$ ).
13: End of Loop.
14: Step 3: Boolean Rules
15: For  $k=0$  to related comment. Size by 1
16: ( $C_{swr} \in$  Remove stop words  $C_k$ .
17:  $TC_{idl} \leftarrow \dots$  Tokensize  $C_{swr}$ .
18: for  $W=0$  to  $TC_{idl}$  by 1.
19: if ( $TC_w \in$  positive dictionary.
20: ( $TP$ ) $\leftarrow \dots$ 1.
21: else ( $TN$ ) $\leftarrow \dots$ 1.
22: if  $TP = 1$  and  $TN = 1$ .
23:  $CM.TP$  AND  $TN$ .
24: Else if ( $TP = 1$  and  $TP = 1$ ).
25:  $CM.TP$  XNOR  $TP$ .
26: Else if ( $TN = 1$  and  $TN = 1$ ).
27:  $CM.TN$  XNOT  $TN$ .
28: Else if ( $TN = 1$  and  $TP = 1$ ).
29:  $CM.TP$  AND  $TN$ .
30: Step 4: Results
31: For  $z = 0$  to  $CM.size$  by 1.
32: IF ( $CM_i == CMM$  .
33:  $TP$ 
34: Else IF ( $CM_i \neq CMM$  .
35:  $TN$ 
36: Step 5: Display Results
37: For  $x = 0$  to  $TC_{idl}$  size by 1 .
38: IF ( $TC_{idl} =$  positive).
39: write  $TC_{idl}$ .positive).
40: Else
41: write  $TC_{idl}$ .negative).
42: End of Loop.
43: End of Loop.

```

Algorithm 2 Steps

- 1: **Step 1:** Data is extracted.
- 2: **Step 2:** Preprocessing is performed to identify related and non-related reviews
- 3: **Step 3:** Remove noisy data from the user reviews.
- 4: **Step 4:** Opinion words are extracted
- 5: **Step 5:** Boolean rules are applied.
- 6: **Step 6:** Polarity of reviews are identified.
- 7: **Step 7:** The results are displayed.

TABLE 3.1
Original Data Set 1st January 2016 to 31st December 2017

Total Reviews	454	Positive	Negative
Samsung mobile	82	51	29
Nokia mobile	96	61	36
Q mobile	170	75	93
Noise	106		

Data set. Reviews in roman Urdu are crawled from website whatmobile.com in the time span of 1st January 2016 to 31st December 2017. Three types of mobiles are selected. Samsung, Nokia, and Qmobile. A total of 454 reviews are crawled. These reviews are consisting of positive, negative, and noisy. Positive reviews are those which show some positive opinion about a product e.g. "ye mobile bht acha h". Negative reviews are those which shows negative opinion about a product e.g. "bht bakwaas mobile he ye". Noisy reviews are those reviews which are not related e.g. "me ye mobile bechna chahta hun agr koi lena chahy to". All the product reviews are manually evaluated, and the result is calculated. This research aims to find the polarity of reviews about products and not to criticize or demotivate any product.

3.2. Noise Lexicon. Table 3.2 shows noise lexicon and noisy comments. The lexicon consists of terms that are used to identify or extract the noise from user opinions. The lexicon is generated by identifying the words from the noisy reviews. As roman Urdu has no rules for writing reviews so the users have used different writing styles and spellings for the same word to express their opinion e.g. warranty, woranti, warenti etc.

In this step, the opinion words are extracted from the refined reviews. The opinions are identified through these words. The opinion words are stored in the opinion word lexicon.

Positive lexicon. Table 3.3 consists of positive lexicon and positive words. The lexicon consists of words that are used to identify positive reviews. Similarly like in noisy reviews users also have used different writing styles in positive reviews e.g. zabardast, zberdast, zbardast etc.

Negative lexicon. Table 3.4 consists of negative lexicon and negative reviews. A negative lexicon consists of words that are used to identify negative reviews. Like positive lexicon, negative lexicon has words with different writing styles and spelling e.g. bekaar, bekar, bkwaas, bakis etc.

3.3. Applying Boolean Rules. In this step, the Boolean rules are applied to the reviews to identify the polarity. Figure 3.1 shows Boolean rules of all required gates.

3.4. Identification. After applying the Boolean rules, the opinion sentence is identified.

4. Result evaluation. The results of the proposed model are evaluated by using standard methodologies of information retrieval i.e. precision, recall, F-measure, and accuracy. Following 2 by 2 confusion matrix which is also called contingency, the matrix is used to evaluate measurements. A confusion matrix [24] includes information on current and anticipated classifications made through a rating model. The operation of such a model is usually measured using the matrix data. The confusion matrix is also called a contingency table. The entries within the confusion matrix cover the following meaning in the background of our study shown in table 4.1.

5. Experimental Results. This section contains the experimental results of the model. For experimental reasons, users' opinions are collected on the "whatmobile.com" website. More than 400 reviews of NOKIA,

TABLE 3.2
Noise Lexicon

Noise lexicon words	Reviews
Warranty	mery pas lumia 530 hy warranty nhi hy full box hy kisi kochahye to msg kry no call only sms plzz final 4200 iam frommultan 03012939265
Sale	Me j5 mobile sale krna chata ho. Condition9/10 white colorprice in 16000 location Faisalabad sell no03117215820

TABLE 3.3
Positive lexicon

Positive lexicon words	Positive Reviews
Zabardast	Zabardast mobile habattery timing ny to tamam smart phones kopechy chor dia
Sale	Outclass mobile he. aik dinbatrery backup. camera front and backbht acha h

SAMSUNG, and Qmobile have been posted from the website in spam from January 1, 2016, to December 31, 2017, to identify polarity. The data has been collected into an MSWord document. Dictionaries for positive, negative, and noise identification have been created for testing and results. For evaluation purposes, each review is manually analyzed and checked if it is positive, negative, or noise. Manually observed results are then compared with practical results. A total of 104 non-related or noisy reviews were crawled and 103 are identified successfully. The percentage of noisy reviews was 99.30%. Lexicon of noise is created to extract noisy comments from the reviews. There were a total of 454 reviews of three products, total positive reviews were 186 and practically 163 were evaluated positive, total negative were 164 and practically 153 were identified as negative, and total 104 were non-related comments and 138 were identified correctly. The results show that 35% of reviews are identified positively, 33% as negative, and 30% noise of total reviews. Table 5.1 shows the noisy comments observed during reviews crawling.

While the TP is true positive comments i.e. number of all actual positive. TN is true negative is the total number of negative comments. FP is the total number of negative comments that are retrieved as positive. While FN is a total number of comments which were positive but practically, they are identified as negative. Precision recall and F-measure and accuracy have been calculated by using the contingency Table 5.4.

Boolean rules. Table 5.5 shows Boolean rules were created to identify the polarity of reviews on the basis of positive and negative words that exist in the reviews.

Lexicons. Three types of lexicons will be created to identify noise and to find the polarity of reviews. Table 5.6 shows no of words of all lexicons. All of the lexicons are consist of following no of words.

5.0.1. Noise Lexicon. Table 5.7 shows noise lexicon and noisy comments.

Three types of lexicons are created to identify noise and to find the polarity of reviews.

Positive lexicon. Table 5.8 consists of positive lexicon and positive words.

Negative lexicon. Table 5.9 consists of negative lexicon and negative reviews.

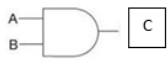
5.1. Individual Products Result. In the case of the first product, out of 82, there were 18 true positives, 19 true negatives, 14 false positives, and 31 false negatives. Table 5.10 shows the contingency matrix of Samsung mobile. Figure 5.2 shows the graph of Samsung mobile. Table 5.10 contingency matrix of Samsung mobile

In the case of second product total, 96 comments were crawled, out of 96 there were 33 true positives, 23 true negatives, false-negative 27, and 15 were false positive. Table 5.12 shows the contingency matrix of the Nokia mobile. Figure 5.2. shows the contingency matrix graph of the Nokia mobile.

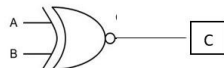
In the case of the third product total, 170 reviews were crawled. Out of total 170, the true positive rate was 45, the true negative was 29, false-negative 32 and false-positive were 64. Table 5.14shows the contingency matrix of Qmobile. Figure 5.3 shows the contingency matrix graph of Qmobile.

TABLE 3.4
Negative lexicon

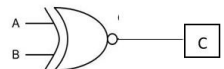
Negative lexicon words	Negative Reviews
bura	yar ya mobile kisi kamka nai hay bht hi bura hay
Bekar	full time bekar set ha yah 3.30ghantylaita ha charging ma ... Paisy brbad krny ha tw lylo

Rule no 1	Inputs		Output (AND)	Equation 
	A	B	C	
	Positive	Negative	Negative	

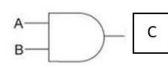
(a) Boolean Rule 1 for (AND Gate) Equation

Rule no 2	Inputs		Output (XNOR)	Equation 
	A	B	C	
	Positive	Positive	Positive	

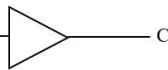
(b) Boolean Rule 2 for (XNOR Gate) Equation

Rule no 3	Inputs		Output (XNOR)	Equation 
	A	B	C	
	Negative	Negative	Positive	

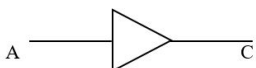
(c) Boolean Rule 3 for (XNOR gate) Equation

Rule no 4	Inputs		Output (AND)	Equation 
	A	B	C	
	Negative	Positive	Negative	

(d) Boolean Rule 4 for (AND gate) Equation

Rule no 5	Inputs	Output	Equation 
	A	C	
	Positive	Positive	

(e) Boolean Rule 5 for single positive Equation

Rule no 6	Inputs	Output	Equation 
	A	C	
	Negative	Negative	

(f) Boolean Rule 1 for single Negative Equation

FIG. 3.1. Rules for Logic Gates

TABLE 4.1
Result evaluation

		Prediction	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	FN	

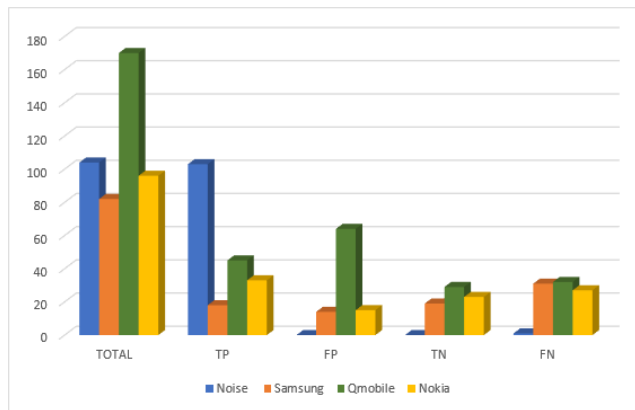


FIG. 5.1. Contingency matrix graph of All Reviews

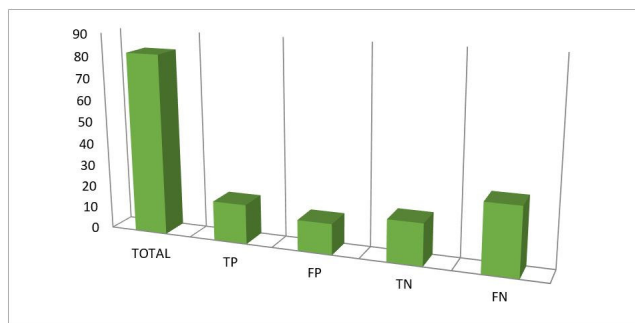


FIG. 5.2. Contingency matrix graph of Samsung mobile

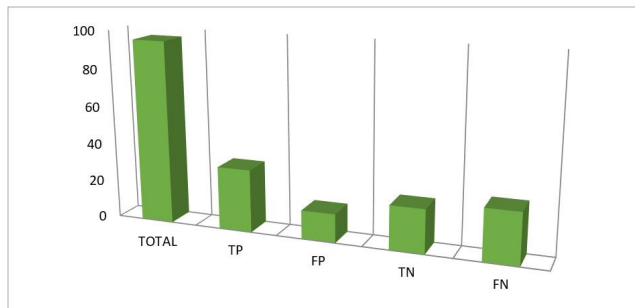


FIG. 5.3. Contingency Matrix Graph of Nokia Mobile

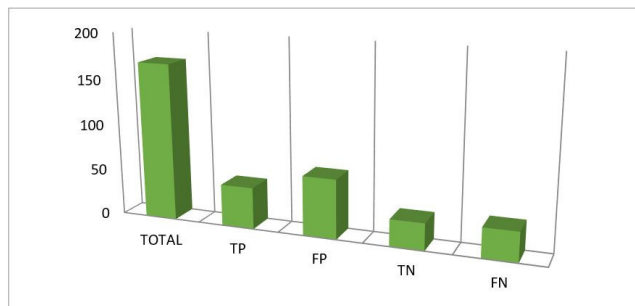


FIG. 5.4. Contingency matrix graph of Qmobile

TABLE 5.1
Related and non-related comments

Related comments	Nonrelated Comments
Bakwass tareen cell phone hai guys, ye acha nahih, Aj mujhy S8 purchase key howy 2 din howyJaib sy giraor glass breake ho gaya yani Gorillaglass ki wat lag gai or warranty green touch waly1 mah bad clame ker k daingy or 10% amount bhily gy	Main apna phone J7 sale kr rha hun koi falt nhihai condition 9.5/10 haii only 18000 SeriesPerson contact me at this \\number. 03558153892 03438848790 from rawalpindi
Yar ya mobile kisi kam ka nai hay bht hi bura hay	Agar kisi k pass J5 (15) dead halatmain hu jiskiLCD thik hu grwo sale krna चाहy tu plzRabtakrain... 0300-4746675

TABLE 5.2
Total results of all products

	Noise		Positive		Negative	
	Reviews	Percentage	Reviews	Percentage	Reviews	Percentage
Practical Evaluation	138	30.0%	163	35.0%	153	33.0%
Manually	140	22.90%	186	40.97%	164	36.12%

5.2. Combine Products Result. TP is true positive comments i.e. number of all actual positive. TN is true negative is the total number of negative comments. FP is the total number of negative comments that are retrieved as positive. While FN is a total number of comments which were actually positive but practically, they are identified as negative. Table 5.16 shows the contingency table results of all products.

6. Discussion. The experimental results show that the model identified the related and non-related comments successfully. Lexicons of noise, positive and negative are generated for identification of noise and finding polarity of reviews. Boolean rules are generated to identify the polarity of reviews. The work has been evaluated in two steps. In the first step, the noisy comments are identified with the help of noise lexicon while in the second step the Boolean rules are applied on reviews, and by the lexicon of positive and negative the reviews are identified successfully. The true positive rate was 199, false-positive was 93, the true negative was 70 and false negative was 91. The precision of the whole evaluation is 0.68, recall is 0.68, F-measure is 0.67, and accuracy is 0.60. All products are also evaluated individually and this time the true positive 96, false-positive 93, true negative 70, and false negative is 90. The precision is 0.507, recall is 0.516, F-measure is 0.511, and accuracy is 0.407.

7. Conclusion. Opinion mining is a vast and challenging area of research as it deals with natural language. Due to the complexity of human language, it is very difficult to obtain useful information. Human language has no such rules of typing that's why users can type the same word with a different style. Roman Urdu is also a simple way of conveying a message. The majority of people in Pakistan convey their message in roman Urdu, but not enough work has been done on roman Urdu. The paper proposed state of the art opinion mining parser is generated by using Boolean rules model to find the polarity of customer reviews that are posted on the different retailer websites. Different lexicon of for identification of noise, positive and negative reviews. In this research the experimental results show that practical evaluation was mainly based on lexicons. The precision of practical evaluation was 68%, the recall was 68% and F-measure was 67%, the accuracy of the whole evaluation was 60%. There were a total of 454 reviews. Positive reviews were 186 and 163 are identified correctly. Total negative were 164 and 153 are identified correctly and total non-related reviews were 104 and 103 are identified successfully. There were different causes for the deviation of results from the original results. One main reason is no predefined rules for posting reviews in roman Urdu. The user has not used proper rules and spelling for posting reviews. So some of the words in negative lexicon also match the words present in positive reviews. For example, the word "kam" is present in a negative lexicon also present in positive review "Ye mobile acha hai or price bhi bohat kam aur affordable he very nice" so because of this reason many positive reviews are identified

TABLE 5.3
Contingency table results

Products	Total	TP	FP	TN	FN
Noise	104	103	0	0	1
Samsung	82	18	14	19	31
Qmobile	170	45	64	29	32
Nokia	96	33	15	23	27
Total	452	199	93	71	

TABLE 5.4
Final Results

Total Reviews	Precision	Recall	F-measure	Accuracy
199	0.68	0.68	.68	0.60

TABLE 5.5
Boolean rules

Rule no	Input		Output	Output Gate
1	True	False	False	AND
2	True	True	True	EXNOR
3	False	False	True	EXNOR
4	False	True	False	AND
5	Positive		Positive	Single gate
6	Negative		Negative	Single gate

TABLE 5.6
No of words of all lexicon

Lexicon	No of word
Noise lexicon	45
Positive lexicon	50
Negative lexicon	55

TABLE 5.7
Shows the noise lexicon

Noise lexicon words	Reviews
Warranty	mery pas lumia 530 hy warranty nhihy full box hy kisi ko chahye tomsg kry no call only sms plzzfinal 4200 iam from multan 0301-*****
Sale	Me j5 mobile sale krna chata ho. Condition 9/10 white colorprice in16000 location Faisalabad sell no 0321-*****

TABLE 5.8
Positive lexicon

Positive lexicon words	Positive Reviews
Zabardast	Zabardast mobile ha batterytiming ny to tamam smart phones kopechy chor dia
Outclass	Outclass mobile he. aik din batrery backup. camera front and back bhtacha h

TABLE 5.9
Negative lexicon

Negative lexicon words	Negative Reviews
Bura	yar ya mobile kisi kam ka naihay bht hi bura hay
Bekar	full time bekar set ha yah 3.30ghanty laita ha charging ma ... Paisybrbad krny ha tw ly lo

TABLE 5.10
Contingency matrix of Samsung mobile

		Prediction	
		Negative	Positive
Actual	Negative	19	14
	Positive	31	18

TABLE 5.11
Total Result of Samsung mobile

Total Reviews	Precision	Recall	F-measure	Accuracy
96	0.56	0.36	0.43	0.45

TABLE 5.12
Contingency table of Nokia mobile

		Prediction	
		Negative	Positive
Actual	Negative	23	15
	Positive	27	33

TABLE 5.13
Total Result of Nokia mobile

Total Reviews	Precision	Recall	F-measure	Accuracy
96	0.56	0.36	0.43	0.45

TABLE 5.14
Contingency table of Q mobile

		Prediction	
		Negative	Positive
Actual	Negative	29	64
	Positive	32	45

TABLE 5.15
Total Result of Q mobile

Total Reviews	Precision	Recall	F-measure	Accuracy
170	0.41	0.58	0.48	0.43

TABLE 5.16
Total Result of all products

Total Reviews	Precision	Recall	F-measure	Accuracy
170	0.50	0.56	0.51	0.47

TABLE 5.17
Contingency table results of all products

Orientation	Number of comments
True Positive (TP)	96
False positive (FN)	93
True Negative (TN)	70
False Negative (FN)	

TABLE 5.18
Final results of all products

Serial no	Products	Total	TP	FP	TN	FN
1	Samsung	81	18	14	19	31
2	Qmobile	170	45	64	29	32
3	Nokia	96	33	15	23	27
4	Total	348	96	93	71	90

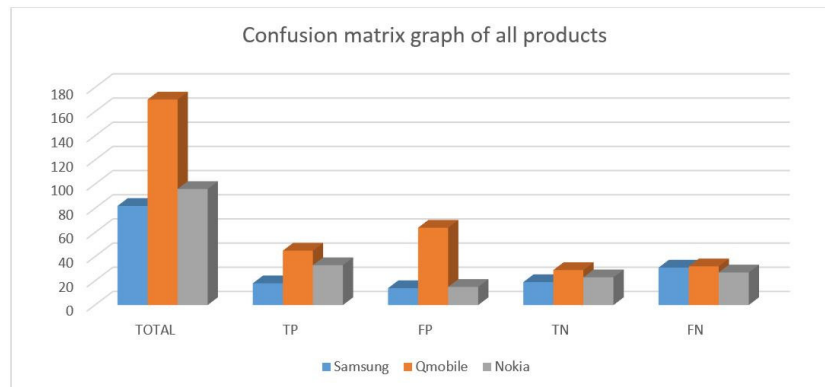


FIG. 5.5. Contingency matrix graph of n Reviews

as negative. The second main reason for result deviation was Boolean rules that were created for identification of review that consist of double negative words i.e. rules no 3. Rule no 3 was made for identification of reviews that consist of two negative words for example "bura" and "nahi". And the reviews were identified as positive, but there were some other reviews which are consist of some more negative words "3rdclass" and "bekar". So, these two words were also present in different reviews and that are identified as positive because of rule no 3.

7.1. Recommendations. The model gave satisfactory results up to some extent. Precision was 50% and more enhancement could be achieved by implementing new methods for handling different errors. Errors and inconsistencies can be handled in different ways. Improving Boolean rules can increase the model accuracy. All the lexicons are designed manually which did not cover the whole lexicons and could not achieve the perfect result, instead of manually if semantic lexicon will use such as wordnet it will give a much better result.

7.2. Future work. In the future we suggested the accuracy of the model can be increased by using a wordnet lexicon generator. Secondly, the accuracy can be increased by identifying the word that changes the polarity of the words from negative to positive and vice versa. From the research, it is found that the word "nahi" changed the polarity from negative to positive and vice versa. Improvement can be done by improving the word "nahi" and making a separate code for this word to evaluate it much better.

Authors' contributions:. All the authors contributed to this research. The order of authors in this manuscript is maintained depending on the level of contributions they made in this research.

REFERENCES

- [1] HALL, MARK AND FRANK, EIBE AND HOLMES, GEOFFREY AND PFAHRINGER, BERNHARD AND REUTEMANN, PETER AND WITTEN, IAN H, *The WEKA data mining software: an update*, ACM SIGKDD explorations newsletter, 11, 1, (2009), pp. 10–18.
- [2] DAVE, KUSHAL AND LAWRENCE, STEVE AND PENNOCK, DAVID M, *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews*, Proceedings of the 12th international conference on World Wide Web, (2003), pp. 519–528.
- [3] HU, MINQING AND LIU, BING, *Mining and summarizing customer reviews*, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.
- [4] DAUD, MISBAH AND KHAN, RAFIULLAH AND DAUD, AITAZAZ AND OTHERS, *Roman Urdu opinion mining system (RUOMiS)*, arXiv preprint arXiv:1501.01386, 2015
- [5] ZNHANG, KUNPENG AND NARAYANAN, RAMANATHAN AND CHOUDHARY, ALOK N, *Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking.*, WOSN, 10, pp. 11–11, 2010
- [6] ABBASI, AHMED AND CHEN, HSINCHUN AND SALEM, ARAB *Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums*, ACM Transactions on Information Systems (TOIS), 26, 3, pp. 1–34, 2008, ACM New York, NY, USA
- [7] HAN, JIAWEI AND PEI, JIAN AND KAMBER, MICHELINE, *Data mining: concepts and techniques*, 2011, Elsevier
- [8] WILSON, THERESA AND WIEBE, JANYCE AND HOFFMANN, PAUL, *Recognizing contextual polarity in phrase-level sentiment analysis*, Proceedings of human language technology conference and conference on empirical methods in natural language processing, pp. 347–354, 2005

- [9] CERCEL, DUMITRU-CLEMENTIN AND TRĂUŞAN, ŞTEFAN, *Research Challenges in Opinion Mining From A Natural Language Processing Perspective*, University Politehnica of Bucharest Scientific Bulletin Series C-Electrical Engineering and Computer Science, 78, 3, pp. 157–168, 2016
- [10] RASHID, AYESHA AND ANWER, NAVEED AND IQBAL, MUDDASER AND SHER, MUHAMMAD, *A survey paper: areas, techniques and challenges of opinion mining*, International Journal of Computer Science Issues (IJCSI), 10, 6, pp. 18, 2013, Citeseer
- [11] KOULOUMPIS, EFTHYMIOS AND WILSON, THERESA AND MOORE, JOHANNA, *Twitter sentiment analysis: The good the bad and the omg!*, Fifth International AAAI conference on weblogs and social media, 2011
- [12] MAHAR, JAVED AHMED AND MEMON, GHULAM QADIR, *Rule based part of speech tagging of sindhi language*, 2010 International Conference on Signal Acquisition and Processing, pp. 101–106, 2010, IEEE
- [13] PAK, ALEXANDER AND PAROUBEK, PATRICK, *Twitter as a corpus for sentiment analysis and opinion mining.*, LREc, 10, 2010, pp. 1320–1326, 2010
- [14] WANG, DINGDING AND ZHU, SHENGHUO AND LI, TAO, *SumView: A Web-based engine for summarizing product reviews and customer opinions*, Expert Systems with Applications, 40, 1, pp. 27–33, 2013, Elsevier
- [15] YASSINE, MOHAMED AND HAJJ, HAZEM, *A framework for emotion mining from text in online social networks*, 2010 IEEE International Conference on Data Mining Workshops, pp. 1136–1142, 2010, IEEE
- [16] HA JMOHAMMADI, MOHAMMAD SADEGH AND IBRAHIM, ROLIANA AND OTHMAN, ZULAIHA ALI, *Opinion mining and sentiment analysis: A survey*, International Journal of Computers & Technology, 2, 3, pp. 171–178, 2012
- [17] VON HELVERSEN, BETTINA AND ABRAMCZUK, KATARZYNA AND KOPEĆ, WIESŁAW AND NIELEK, RADOSŁAW, *Influence of consumer reviews on online purchasing decisions in older and younger adults*, Decision Support Systems, 113, pp. 1–10, 2018, Elsevier
- [18] DU, RUNDONG AND LU, ZHONGMING AND PANDIT, ARKA AND KUANG, DA AND CRITTENDEN, JOHN AND PARK, HAESUN, *Toward Social Media Opinion Mining for Sustainability Research*, Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015
- [19] LIU, BING, *Sentiment analysis and opinion mining*, Synthesis lectures on human language technologies, 5, 1, pp. 1–167, 2012, Morgan & Claypool Publishers
- [20] AGARWAL, APOORV AND XIE, BOYI AND VOVSHA, ILIA AND RAMBOW, OWEN AND PASSONNEAU, REBECCA J, *Sentiment analysis of twitter data*, Proceedings of the workshop on language in social media (LSM 2011), pp. 30–38, 2011
- [21] EL-HALEES, ALAA M, *Arabic opinion mining using combined classification approach*, Arabic opinion mining using combined classification approach, 2011, Naif Arab University for Security Sciences
- [22] *N-gram based approach for opinion mining of Punjabi text*, KAUR, AMANDEEP AND GUPTA, VISHAL, International Workshop on Multi-disciplinary Trends in Artificial Intelligence, pp. 81–88, 2014, Springer
- [23] BILAL, MUHAMMAD AND ISRAR, HUMA AND SHAHID, MUHAMMAD AND KHAN, AMIN, *Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques*, Journal of King Saud University-Computer and Information Sciences, 28, 3, pp. 330–344, 2016, Elsevier
- [24] LI, NAN AND WU, DESHENG DASH, *Using text mining and sentiment analysis for online forums hotspot detection and forecast*, Decision support systems, 48, 2, pp. 354–368, 2010, Elsevier

Edited by: P. Vijaya

Received: Nov 26, 2019

Accepted: Jun 26, 2020



FORECASTING THE IMPACT OF SOCIAL MEDIA ADVERTISING AMONG COLLEGE STUDENTS USING HIGHER ORDER STATISTICAL FUNCTIONS

MEENA ZENITH. N* AND RADHIKA. R[†]

Abstract. Nowadays, social media has emerged as one of the activities among the users in their day-to-day life activity. This research work plans to develop a statistical review that concerns on social media advertising among college students from diverse universities. The review analysis on social media advertising is given under six sections such as (i) Personal Profile; (ii) Usage; (iii) Assessment; (iv) Higher Order statistics like Community, Connectedness, Openness, Dependence, and Participation; (v) Trustworthiness such as Trust, Perceived value and Perceived risk; and (vi) Towards advertisement which involves attitude towards advertisement, response towards advertisement and purchase intension. The initial stage is on questionnaire preparation based on social media networking. The records stated in the questionnaire are intimately taken about the usage of social media sites and the advertisement on networking. In the second stage, the planned questionnaire is distributed over college students from diverse universities. The entire questions are made mandatory in this questionnaire and after this, the students from various universities are demanded to fill up their responses to this questionnaire. These responses from the students are then taken for analysis purposes. In this research work, the analysis is performed based on higher-order statistical analysis that favorably concerned with correlation coefficients and entropy. This in turn helps to determine the correlation and entropy among the response towards the social media network.

Key words: Social media; Advertising; Universities; Correlation analysis; Entropy analysis

AMS subject classifications. 62L10

1. Introduction. Nowadays, Online Social Networks (OSNs) [9-12], which involves Facebook, Sharechat, Digg, Twitter, and Instagram have become more popular. The users tend to post the photos, videos, news, and so on in OSN, and such users have some followers whose views and comments on that posted information. So far, many OSNs [13-15] host are available on online applications. By using this hosted application, the advertiser can make post job information, and the users can invite associates for online games. The information offered by the users is termed as information producers, and one who views this information is referred to as information consumers. The most recent successful targeted information on advertising systems is provided to the producers and consumers, which facilitates the producers to target users based on profile information, online activities, and user demographics. The advertised information is then clicked on by the targeted users, as the personal interest of the user matched with the ad. The potential benefits have been attained by information producers, because of the clicks and e-commerce activities performance of the customers.

Currently, in our complicated society, both the consumers and businesses need the advertisement factor because it evolves as a core communications system [24-25]. Further, it acts as the main factor in most of the organization's marketing programs as of its tendency to delivering cautiously prepared messages to destination audiences. Based on the suitable advertising plan, the entire advertising activities [16-19] are made sure, which can effectively lead the organization's advertising programs to be efficient and cost-effective. The suitable messages are delivered by an advertising plan via suitable media or vehicles to suitable audiences. Therefore, in any of the advertising plan, media planning is concerned as the major component.

The classification of conventional media selection models [20-23] is given in the following: (1) models that are based on experience and judgment since are not capable on considering the huge count of media combinations, (2) clean quantitative models that might not integrate qualitative criteria like experience and knowledge on the decision. In contrast, the selection process of advertising media has incorporated uncertain, inaccurate,

*Research scholar Noorul Islam centre for higher education (zenithmeena@gmail.com).

[†]Associate professor Noorul Islam centre for higher education (rradhika.19@yahoo.com).

and information like experience and judgment of individuals. Therefore the media selection decision causes a dilemma, where its solution depends on human judgment and thus becomes more complicated to resolve by only using human judgment. Still, various case studies are evolved in this social media advertising concept, yet the researchers are not concentrated over the prediction phase.

The core contribution of this framework is delineated as follows:

- This paper mainly attempts to introduce a novel tactic over social media advertising among college students from diverse universities.
- In this, the analysis is handled under six factors called (i) Personal Profile (ii) Usage (iii) Assessment (iv) Higher Order statistics (v) Trustworthiness and (vi) Towards advertisement.
- The first stage includes questionnaire preparation on the topic of social media networking, which concerns more on social media sites and advertising.
- The second stage is about the distribution of organized questionnaires among college students of diverse universities and is then subjected to analysis.
- To the end, based on higher-order statistical analysis, the analysis is carried out that focused on correlation coefficient and entropy function, which leads to finding the correlation and entropy among the response towards the social media network.

The organization of the work is as follows: Section 2 expresses the related works on social media advertisements. Section 3 depicts the review of social media networking. Section 4 explains the perspectives of social media usage and advertisement. Section 5 delineates the enquired conclusion on the impact of social media advertising. Section 6 ends the paper.

2. Literature Survey.

2.1. Related Works. In 2020, Sydney Chinchanchokchai, and Federico de Gregorio [1] have developed a rapid growth of advertising on Social Media Platforms (SMPs). The consumer socialization framework adopts to investigate predictors of advertising avoidance on SMPs such as Facebook, Twitter, and Instagram via an online survey. Results show that the effects of SMP usage, susceptibility to social media influence, and susceptibility to peer influence on SMP ad avoidance are all mediated by attitude toward social media advertising in general. Greater SMP usage and higher susceptibility to social media influence are positively related to SMP advertising attitudes, while greater peer influence susceptibility is negatively related.

In 2017, Gupta et al. [2] have made an effort for computing the importance of multimedia tool named YouTube. The crucial success factors were determined based on the Content analysis of hundred YouTube advertisements. Some of the crucial success factors were Visual Category, Message Appeals, Audio content, Content category, and the response of viewers via the number of views and likes. In accordance with this, a methodology has been implemented, which might aid the managers who have improved the promotional strategies for the association. This research work has been verified by deploying the Attention, Interest, Desire, and Action (AIDA) model.

In 2020 Sreejesh S et al. [3] have used social media platforms as a promotion channel and allows consumers to socialize and network better. In this media, attention is often restricted towards primary purpose and it affects consumer response towards the advertisement. The media interactivity can affect the reaction of customers towards social media. Since high interactivity in this medium directs the users to involve more into the primary purpose of socialization, this feature of the media adversely affects the advert and its effectiveness.

In 2017, Richard et al. [4] have introduced a novel method for questioning the correlation among monopolistic behavior and the social optimum, while admitting the advertisement. Further, a common outcome has been obtained that creates the dozen special cases of interest to users, by featuring conditions regarding consumer preferences and using the uncommon method over comparative static analysis. Further, the reasonable preference specification has shown enough performance by generating this case. The derivations of outcomes were made, which pursues the advertisement complementary other than influential advertising paradigm that has possessed stable quasilinear preferences by the consumers.

In 2018, Lee [5] has scrutinized the efficiency of Social-Local-Mobile (SoLoMo) advertising and Location-Based Advertising (LBA). The outcome has depicted the effectiveness of SoLoMo advertising than LBA. In this research work, the differences in the reaction of customers to the ads on diverse situational platforms and contexts were determined. In the literature, it has lacked in the straight distinguishing of efficiency of LBA

and SoLoMo advertising, which has contributed to the main academic of this framework. This was the initial step that was taken in comparing among the ads over LBA and SoLoMo. This framework further investigated the factors that connected with the advancement of brand interaction, modern smartphone functionalities, sociability, perceived location awareness, and influences the attitudes towards ads. Moreover, the telecommunication promotion effectiveness was exhibited by investigating the customers that react following ads in diverse situational contexts on diverse platforms. To the end, the practical suggestions were also defined in this paper.

In 2016, Lin and Kim [6] have presented the study on the impacts of advertising the consumer attitudes, such as perceptions of privacy risk, and purchase intent. The Technology Acceptance Model (TAM) derives the testing model. This learning further determined the intrusiveness and privacy focus were both suitable antecedent variables for perceiving the usefulness, yet not apparent accessibility of sponsored advertising. The privacy concerns can purchase the product when both antecedent variables as well influenced the consumer attitudes over sponsored advertising. The theoretical correlations among perceived usefulness, attitudes, ease of use, and purchase intentions have also been verified.

In 2018, Javan et al. [7] have introduced a two-phase methodology to advertise the media selection, which integrates the human-based information by incorporating the quantitative and qualitative models, when the associated complexity with media selection decision was reacted. The initial stage has determined the top media for advertising the hierarchy process based on four assumptions of AIDA. The second stage determines the optimum media mix by incorporating the fuzzy linguistic decision approach. Finally, implemented approach has been evolved empirically over the real-world case with acceptable outcomes.

In 2020, Fevzi Bitiktas, and Okan Tuna [8] have stated business-to-business (B2B) markets which rapidly increase the business world as in the lives of individuals. It can evaluate the current social media behaviors of container shipping companies to help practitioners to increase their relation to the algorithms of these platforms. After identifying container shipping companies that use social media most actively, to analyze their Facebook messages in terms of branding, message appeals, direct-sales, and information cues.

2.2. Review. Table 2.1 determines the pros and cons of the conventional model cases on social media networking. SMPs [1] can able to loosely target people in a particular field or internet area and ability to reach a younger age demographic, however, it is not right for all research projects and very little control over the message once posted. AIDA [2] poses better efficiency and an understanding of factors that attract the customers. Yet, it cannot catch much attention. Social media platforms [3] provide real time feedback and cost-effective with consumers, however, it can be dangerous if it is misused. The monotone comparative static approach [4] has a better capturing of the problematic condition and permits a wide variety of testing of the alternate hypothesis. Though, it requires similar identification potential and further needs research on empirical implementation. SoLoMo advertising [5] is considered to be more effective and discovers the difference between the customer's review of ads. But has privacy issues and difficulty in creating one campaign for all of them. TAM [6] has a successful explanation of the relation between interactive platforms and is more useful and easier to be used by the customers. But, there is a need for defining the threshold and needs help advertisers to formulate the social media advertising strategy. Genetic algorithm [7] poses a better determination of optimum media mix and further declines the complexity related to the decision process. However, it needs to more improved models on optimization. B2B [8] have less inventory and reduced transaction cost and it requires prequalification Low order conversion rates are the main drawbacks that considered in this methodology.

2.3. Problem Statement. Social Media sites like Facebook, Instagram is probably popular among varied age groups. More research works are in rush to analyze the impact of social media advertising, however, some sort of analysis are still under a crisis that should be dealt with effective manner. Here, some of the problems that to be rectified in the future are explained in this section in brief. Not many research works were under went for the usage constraint of the social media site. Usage constraints include visiting, usage time, and access time. In the literature, there is no record on usage analysis of social media sites over how much visitors visiting the social media, how much time they are spending averagely in a day, and which time they are accessing these sites. Similarly, for which purpose the social media sites are mostly used among the users is also still not determined. If having this information in advance, then it will lead the researchers to develop an improvement over the advertisement field in social media. Owing to the amount of time spending on these sites for the present time with last year helps in the precise prediction of user assessment for the following year. As the

TABLE 2.1
Features and Challenges of traditional models regarding the Social Media Advertising Factor

Author[citation]	Methodology	Features	Challenges
Sydney Chinchachokchai, and Federico de Gregorio [1]	SMPs	Able to loosely target people in a particular field or internet area Ability to reach a younger age demographic	Not right for all research projects very little control over the message once posted
Gupta et al. [2]	AIDA	Better efficiency A better understanding of factors that attracts the customers	Cannot catch much attention
Sreejesh S et al. [3]	social media platforms	Real time feedback cost-effective with consumers	Awareness of discussions and interactions are needed among consumers It can be dangerous if it is misused
Richard et al. [4]	Monotone comparative static approach	Better capturing of problematic condition Permit a wide variety of testing of the alternate hypothesis	Requires similar identification potential Needs further research on empirical implementation
Lee [5]	SoLoMo advertising	More effective Discovers the difference among the customer's review on ads	Issues on privacy Difficulty in creating one campaign for all of them
Lin and Kim [6]	TAM	Successful explanation on the relation among interactive platforms More useful and easier to be used by the customers	Need for defining the threshold Help advertisers to formulate the social media advertising strategy
Javan et al. [7]	Genetic algorithm	Better determination of optimum media mix Further declines the complexity related to the decision process	Needs more improved models for optimization
Fevzi Bitiktas, and Okan Tuna [8]	B2B	Less inventory Reduced transaction cost More efficient pricing	Requires prequalification Low order conversion rates

same, analyzing the community in terms of finding people, cultivating, relationships, and sharing the feeling aids the researchers to find on more intimate feelings of the users about the advertisement in social media sites. On checking the connectedness issues of the users over the social media sites will aid to resolve the critical issues like content sharing, editing and communicating information, etc. On finding the priority of users over social media in terms of product, information gathering, and sharing the needs, it also paves the way for crucial development over the business field, as most of the users intently depend on these social media sites for their everyday works. Further, analyzing the trustworthiness of the users over these network sites will help the future work to enhance the trust problems. Notably, the person's attitude towards the advertisement is also the major course for business development. On confining the awareness and positive feeling towards the advertisement on products or brands, reaches the superior performance over the social media advertising field for the researchers. The response towards advertisement by the clients or users to purchase, recommendation, and sharing of the product is also considered the most viable research area, by using their response more intend work can be made in the future work over advertising. Moreover, the risk prediction is the main task that needs more attention in this following field, as the user's privacy issues, money values everything depends on this. Hence, more intimate research will be needed for these risk issues. After this, the users feeling or intention towards the advertisement has also not yet analyzed, this lacks the research many times. Therefore more advancement is in need over the social media advertising that concerns the researchers to do well for future work.

3. Review Towards Social Media Networking.

3.1. Organized Questionnaire. The constructed questionnaire is comprised of 54 questions and are classified as, the usage of social media sites with 3 questions, Assessments of social media sites having 4 questions, Higher-order statics analysis that involves the Community, Connectedness, Openness, Dependence, and Participation with 4 questions each, Trustworthiness over social media sites that consists of Trust, Perceived

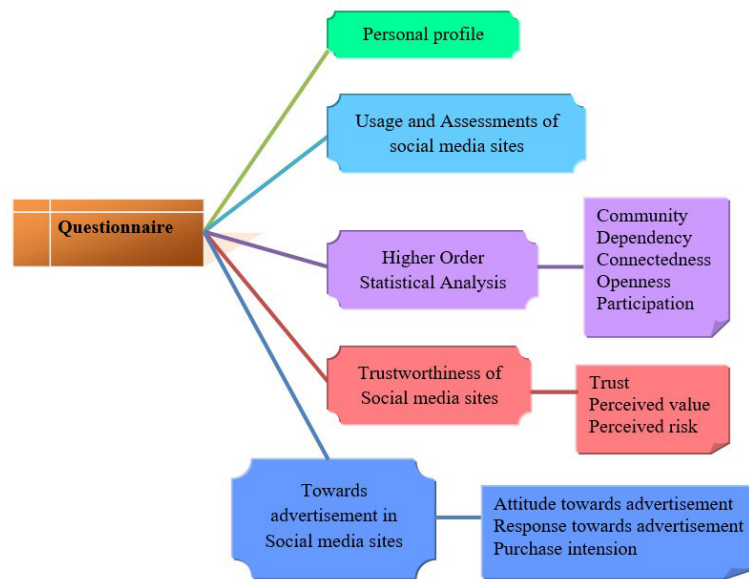


FIG. 3.1. Art on Constructed Questionnaire

TABLE 3.1
Personal Profile of college students

Personal Profile	
1	Name
2	Gender
3	Age group
4	Education level
5	Educational institution

TABLE 3.2
Usage of Social Media Sites

1.	Since when you have been visiting these social media sites? Facebook
2.	Since when you have been visiting these social media sites? Instagram
3.	How much time you spend on social media site in a day? Facebook
4.	How much time you spend on social media site in a day? Instagram
5.	Between what times do you access social media sites mostly? Facebook
6.	Between what times do you access social media sites mostly? Instagram

TABLE 3.3
Assessments of Social Media Sites

1.	Which of the following are you connected with on social media networks?
2.	Comparing last year, have you increased or decreased or spent same amount of time using social networking sites
3.	Looking at the next twelve months, compared to the last year for product information search do you think you will be increasing, decreasing or spending the same amount of time using social networking sites?
4.	What type of products you like to stay connected through social media sites?

value and Perceived risk with 4 questions each, and Towards advertisement over social media sites which includes Attitudes towards advertisement, Response towards advertisement and Purchase intention with 4 questions each. The overall artistic representation of the constructed questionnaire is illustrated in Fig. 3.1, and the corresponding categories of the questionnaire are symbolized in Tables 3.1-3.6.

TABLE 3.4
Higher Order Statics over Social Media Sites

Community	
1.	Finding people of the same interest or background on social media sites is easy
2.	Cultivating more intimate relationship with others on social media is easy
3.	Sharing emotions and communicating feelings with friends on social media sites is easy
4.	It is easy to be part of the community or interest groups on social media sites
Connectedness	
5.	Same social media identity (login ID) is used to login different social media sites
6.	Sharing contents from one social media site and posting it in other social media by sharing or through links is easy
7.	Special advanced skills are not required to use social media sites
8.	Editing and communicating information on the social media sites in the form of text, picture, video is easy
Openness	
9.	Joining social media sites is easy
10.	It is easy to join the groups and communities that I am interested in social media sites
11.	Information can be acquired on social media platform freely
12.	Publishing posts on social media sites can be done freely
Dependence	
13.	When choosing products, social media is my first priority for gathering information
14.	Searching information about products through social media sites is easy
15.	More time is being spent on social media than other online media such as company websites, online shopping website
16.	Making comments or sharing experience with my friends about the products through social media sites is done frequently
Participation	
17.	I am willing to help friends who have problems regarding the use of social media
18.	I often participate in the discussion about products proposed by my friend on social media site
19.	I am subscribed to updates and alerts regarding a brand or product through social media site
20.	Product information is searched through social media sites often

TABLE 3.5
Trustworthiness over Social Media Sites

Trust	
1.	Information on social media is trustworthy
2.	I will share my experience with my friends about buying products or acquiring information on social media
3.	I trust the opinion on social media when considering the product
4.	The probability of getting poor quality products through social media platforms is low
Perceived value	
5.	It is possible to find products that are more suitable for my personal quality and styles on social media site
6.	It is possible to save a lot of money acquiring information about product on social media site
7.	The probability of leaking my privacy in purchasing products through social media platforms is low
8.	After I acquire information about products from social media, I know their quality and function
Perceived risk	
9.	The financial risk in buying products through social media sites is low
10.	The probability of wasting time on obtaining information about products through social media sites is low
11.	The probability of harming my physical health by purchasing products (long exposure to mobile or computer screen) is low
12.	The probability of getting me under social pressure in purchasing products through social media is low

3.2. Data Acquisition. On considering the data acquisition, 100 samples are gathered using Google form from various college students under different universities. The questionnaires are filled with proper responses by college students from diverse universities. The appropriate response can be given by them when the questionnaire was sent in advance. Additionally, the data can be collected via another method called personal interview. In reality, the doubt from the college students associated with the questionnaire can be made clear through this personal interview, and hence the honesty of the responses can be enhanced.

4. Perspectives Of Social Media Usage And Advertisement.

4.1. Personal profile of College students. Fig. 4.1 signifies the personal profile of various college students regarding their gender, age group, educational level, and educational institution. From Fig. 4.1.(a), among the total gender group, 65% of them are female students and the rest 35% are male students. On

TABLE 3.6
Towards Advertisement in Social Media Sites

Attitude towards advertisement	
1.	Social media advertisements make me more aware about various brands and new products
2.	The advertisements displayed on social media gives a positive feeling
3.	The social networking sites are targeting the advertisements to specific audience based on their interest
4.	Brands that use social media for marketing purpose are more innovative than others who are not using it
5.	I prefer brand that is advertised on social media sites
Response towards advertisement	
6.	I talk about the product to my friends
7.	I purchase the product
8.	Just view the produc
9.	Recommend the products to others
10.	Take part in events posted by the company
Purchase intension	
11.	I like to try a product recommended on social media
12.	Seeing social media advertisements increases my interest in buying the same product
13.	I am very likely to buy products shared by my friends on social media platform
14.	Using social media platform help me make decisions better before purchasing products
15.	Social media advertisements greatly influence the purchase choice

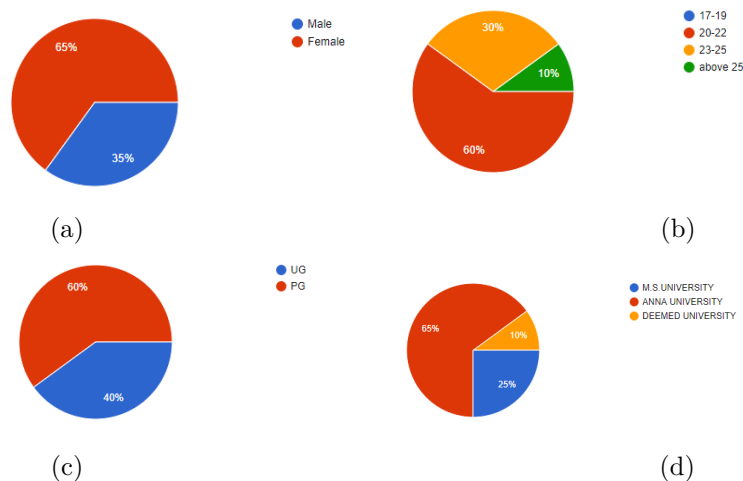


FIG. 4.1. Personal Profile of College students (a) Gender (b) Age group (c) Education level and (d) Educational Institution

considering Fig. 4.1.(b), most of the students belong to the age group among 20-22 and is given as 60%. Other than that, 30% of students are among the age group 23-25 and 10% of them are above 25 age. Similarly, while taking their educational qualification, the majority of the students are from PG degree i.e. 60%, the rest 40% of them are undertaking their UG degree. Finally, the educational institution is discussed in brief as follows: 65% of the students are doing their major from Anna University and only 25% and 10% of the students are from M.S. University and Deemed University, respectively.

4.2. Usage of Social Media Sites. Fig. 4.2 depicts the usability of social media sites by college students under three sets of questions. The first question explains on the year since the students visiting social media sites like Facebook and Instagram. For this respective question, 40% and 60% of the students are using Facebook and Instagram for less than a year, 5% and 30% are using them for 1-2 years, only 20% and 50% of them are using these social media for 3-4 years and 35% and 5% of the students are interested in this Facebook and Instagram for above 5 years. The second question is about the average time that spends by the students on social media sites within a day. Most of them use social media sites for less than an hour and that is exemplified as 45% and 70% for Facebook and Instagram, respectively. 40% and 20% of the students utilize them for 1-3 hrs. 15%

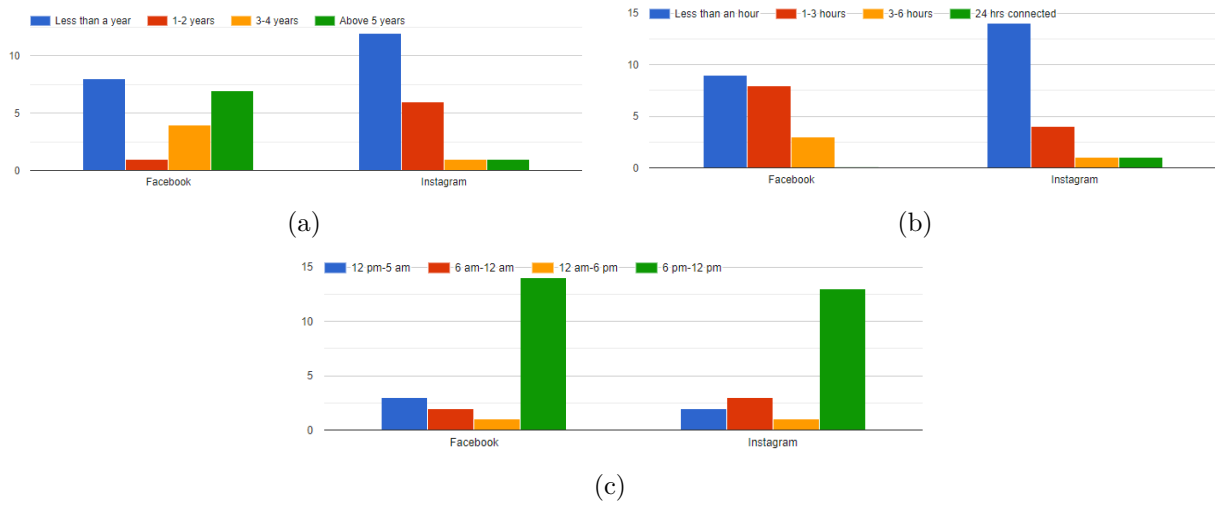


FIG. 4.2. Usability of Social Media sites (a) year since the students visiting the social media sites (b) average time spend by the students on social media sites in a day and (c) access time of the students for a day

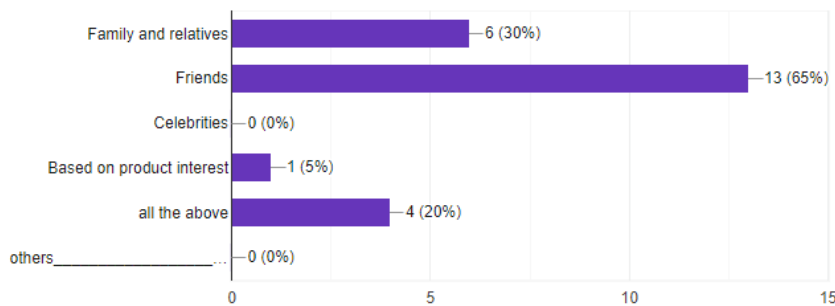


FIG. 4.3. Assessment of social media sites based on their interest

and 5% of them use social media sites for 3-6 hrs. Only 5% of the students are using the Instagram site for 24 hrs. The last and final question on this usage is among which times, most of the students are accessing social media sites. Between the times 12 pm to 5 am, only 15% and 10% of the students mostly prefer social media sites. 10% and 15% of them are using these Facebook and Instagram between 6 am to 12 am, only 5% of the overall students are accessing the social media sites among the time 12 am-6 am. Largely, these Facebook and Instagram are preferred at the time between 6 pm and 12 pm and that is 70% and 65%, respectively.

4.3. Assessment of Social Media Sites. Figs. 4.3 and 4.4 delineates the assessment of social networking sites by college student from various universities. Fig. 4.3 expresses the interest of the students on social media networks. In which, 30% are giving preference to family and relatives, the majority of 65% prefer friends, only 5% decide based on the product interest and 20% of the students relies on social media for all these above-mentioned tasks.

Fig. 4.4 exacts the assessment on social media sites under three common questionnaires. The initial question in Fig. 4.4.(a) is about the comparison with last year about the time spending on social networking sites. 40% of the students respond to this question as increasing, 35% as decreasing and 25% says nearly the same as last year. Next one in Fig. 4.4.(b) is regarding the product information search when comparing the last year with the upcoming year, what will you suggest about the time spent on social media sites. For this, only 25% of students suggest as increasing, 40% think on as decreasing and the rest stay about the same. The

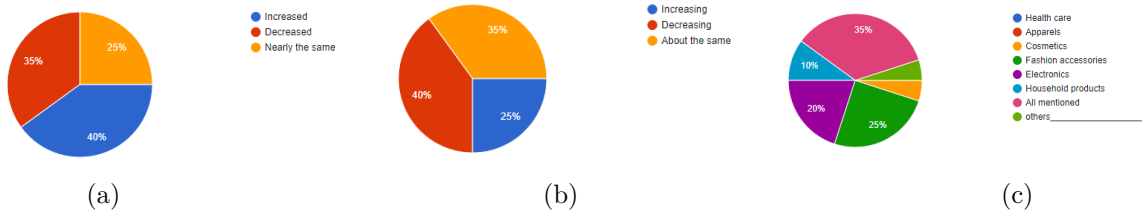


FIG. 4.4. Assessment of Social Media Sites

final question in Fig. 4.4.(c) is on the type of product that the students like to stay connected via social media sites. 10% of the students choose the household products, 5% selects on the cosmetics, 25% prefer on fashion accessories, 20% of students on Electronic products, 35% mentioned on all the above products and only 5% choose the other.

4.4. Higher Order Statistical Analysis of Social Media Sites. Table 4.1 explains the higher-order statistical analysis of social media sites with the input drivers Community, Connectedness, Openness, Dependence, and Participation. Considering the community, while finding the people with the same interest or background is easier in social media site is the first statement. To this, the response among the student is: only 10% strongly agreed on that fact, 50% only agreed, 10% of them neither agree nor disagree with this statement and 30% disagreed on this report. Next statement says the cultivation of more intimate relation with each other is easier, the level of agreement for this is 25% strongly agree, 45% agreed to this report, 10% neither agree nor disagree on that fact and only 5% disagree on this statement. It is easy to share the emotions and communicate the feelings with friends in social media sites, is the next question. For this, 15% of the students strongly agree, 30% only agreed with this fact, 25% of them neither agree nor disagree, 20% of students disagreed with this statement and 5% of them strongly disagree. To be a part of community or interest group is easier within the social media sites, for this final question on the community, the response of the students are strongly agreed by 25% of them, only agreed by the students of 35%, 30% neither agree nor disagree with this report and 15% of the students only disagreed on this work. Next is the connectedness, for diverse social media sites, 15% strongly agree for the same login ID or identity, 60% only agreed for this identical ID, 5% recommended neither agree nor disagree with this statement, 15% are not interested in using the same ID so disagree for that and 10% strongly disagree on this argument. The subsequent statement is about the sharing and posting of content over the social media site via links and sharing is an easier task, for this, most of them agreed and is 60%, 15% strongly agreed on this and 15% of students neither agree nor disagree for this argument. For using the social media sites, 40% of them strongly agreed on the non-requirement of special advanced skills, where 40% of them also agreed on this report, 15% neither agree nor disagree for this and 5% of them only disagreed on this statement. The next argument is: Editing and communicating information through social media site is easier, and for this 25% of them strongly agreed, 50% only agreed on this statement, 20% of them neither agreed nor disagreed and 5% only disagreed with this argument. The subsequent phase is Openness that includes four questions. The first is joining social media site is easier, the response among the student for this is 25% strongly agreed on this statement, 55% agreed with this argument, 15% of them neither agree nor disagree and 5% disagreed on this report. The next one is on social media sites, joining the interested group and community is easier, where 20% strongly agreed for this, 50% agreed on this statement, 25% neither agree nor disagree on that and 5% disagreed with this report. Information acquisition on social media platforms can be done freely, for this 15% strongly agreed, 60% of students only agreed, 20% of them neither agreed nor disagreed on this report, and 5% on the overall students disagreed and strongly disagreed on this statement. Publishing posts on social media sites can be done freely is the last statement. Owing to this, 30% of the students strongly agreed on this fact, 45% of them only agreed for this and 15% of the students neither agreed nor disagreed as well disagree on this argument. The consequent phase is dependence, where the initial statement is while gathering information on products, 15% strongly agreed that social media site is their first priority, 30% only agreed as their first priority, 10% of the students neither agree nor disagree for this, 35% disagree as that is not their first priority and 10% strongly disagreed on this report. Searching the product information

via social media site is 10% strongly recommended as easier, 40% also agreed as easier, 50% neither agreed or nor disagreed as easier, 5% disagreed on this easier report. The next is about the time spent on social media sites than others media like online shopping websites, company websites, etc, for this 25% strongly agreed on using more time, 40% as well agreed with this long time utilizing, 20% of them neither agree nor disagree, 10% disagreed on this long time usage and 5% strongly disagreed. Making comments or sharing the experience on products among friends on social media site is 20% agreed by them as done frequently, 45% neither agreed nor disagreed on this statement, 15% disagreed and 20% strongly disagreed as they do not do frequently. While taking the participation, helping friends over the problems in social media site is the first statement, and for this 5% are strongly willing to do, 40% agreed to help them, 30% neither agree nor disagree over this, 20% disagreed to help and 5% strongly disagreed on this thinking. Discussion on product proposal of the friend on the social media site, 25% strongly agreed to participate, 20% agreed on to participate in this discussion, 15% of them neither agree nor disagree, 10% disagree on this participation and 25% strongly disagree to participate. Updating and alerting on products or brands on social media site, 5% strongly agreed on subscribing, 30% agreed to subscribe for this, 15% neither agreed nor disagreed on this report, 30% disagreed for subscribing and 20% strongly disagreed to subscribe for this. The searching on product information through social media sites is made often, for this 5% strongly agreed, 45% only agreed for this, 25% of them neither agreed nor disagreed, 10% disagreed on this argument and 20% strongly disagreed for that statement.

4.5. Trustworthiness of Social Media Sites. Table 4.2 reveals the trustworthiness of social media sites with trust, perceived value, and perceived risk that contains four questions in each group. In Trust, information on social media is strongly agreed as trustworthy by 10% of students, 15% only agreed, 25% neither agreed nor disagreed as trustworthy, 40% disagreed and 30% strongly disagreed on their trustworthiness. Sharing the experience on product and information within friends on social media site is 5% strongly made among the overall students, 40% of them only agreed for this, 20% neither agreed nor disagreed, 20% only disagreed on this statement and 15% strongly disagreed with that. The expert opinion on product among social media site is 5% strongly agreed as trustworthy, 15% only agreed as trusty, 45% neither agreed nor disagreed, 20% only disagreed on their trustworthiness and 15% strongly disagreed. The probability of acquiring poor quality product through social media platform is 5% strongly agreed as low by the students, 25% of them only agreed as low, 25% neither agreed nor disagreed about this report, 15% only disagreed as they think the probability may be high and 30% strongly disagreed on this argument. On considering the Perceived value, finding the suitable product for customer's in accordance to their personal quality and style on social media site is 15% strongly agreed as possible, 45% of them only agreed as possible, 5%, neither agreed nor disagreed as having possibility, 10% disagreed on their possibility and 20% strongly disagreed as they think it is not possible. Saving the money by acquiring product information on social media sites is 15% strongly agreed by them as possible, 45% mostly agreed as possible, 5% neither agreed nor disagreed on this, 10% disagreed as possible and 20% strongly disagreed on their possibility. Leakage of privacy information in purchasing a product via social media platform is low, 10% strongly agreed for this argument, 20% only agreed as having probability, 25% of the students neither agree nor disagree, 30% of them disagree with this probability and 15% strongly disagreed as the probability of this statement is considered by them as low. The quality and function of the product are known only after getting the product information from social media site, for this 10% strongly agreed, 35% only agreed for this, 25% of the students neither agree nor disagree, 20% disagreed with this argument and 10% of them strongly disagreed. In the Perceived risk, about the financial risk on buying a product via social media site is 15% strongly agreed as low, most among the overall students with 35% agreed as low, 10% neither agreed nor disagreed, 30% disagreed on this low value and 10% strongly disagreed on this statement. The possibility of wasting time for gaining the product information from social media site is 15% strongly agreed as low, 30% only agreed as low, 20% both agreed nor disagreed on this low argument, 20% of them disagreed and 15% strongly disagreed with this low statement. The probability on harming the health by buying a product is 5% strongly agreed by the students as low, 20% only agreed as low, 30% of them neither agreed nor disagreed, 35% disagreed for this statement as it may be high, and 10% strongly disagreed. The probability of getting under social pressure in buying a product through social media platform is low, for this statement, 15% strongly agreed, 25% only agreed, 25% of them neither agreed nor disagreed, 15% of the students disagreed on this report and 20% of them strongly disagreed with this argument.

TABLE 4.1
Analysis on Level of Agreement over Higher Order Statistics

Level of Agreement	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Community					
Finding people who have same interest or background on social media sites is easy	10%	50%	10%	30%	-
Cultivating more intimate relationship with others on social media is easy	25%	45%	10%	5%	-
Sharing emotions and communicating feelings with friends on social media sites is easy	15%	30%	25%	20%	5%
It is easy to be part of the community or interest groups on social media sites	25%	35%	30%	15%	-
Connectedness					
Same social media identity (login ID) is used to login different social media sites	15%	60%	5%	15%	10%
Sharing contents from one social media site and posting it in other social media by sharing or through links is easy	20%	65%	15%	-	-
Special advanced skills are not required to use social media sites	40%	40%	15%	5%	-
Editing and communicating information on the social media sites in the form of text, picture, video is easy	25%	50%	20%	5%	-
Openness					
Joining social media sites is easy	25%	55%	15%	5%	-
It is easy to join the groups and communities that I am interested in social media sites	20%	50%	25%	5%	-
Information can be acquired on social media platform freely	15%	60%	20%	5%	5%
Publishing posts on social media sites can be done freely	30%	45%	15%*	15%	-
Dependence					
When choosing products, social media is my first priority for gathering information	15%	30%	10%	35%	10%
Searching information about products through social media sites is easy	10%	40%	50%	5%	-
More time is being spent on social media than other online media such as company websites, online shopping website	25%	40%	20%	10%	5%
Making comments or sharing experience with my friends about the products through social media sites is done frequently	-	20%	45%	15%	20%
Participation					
I am willing to help friends who have problems regarding the use of social media	5%	40%	30%	20%	5%
I often participate in the discussion about products proposed by my friend on social media site	25%	20%	15%	15%	25%
The brand on social media site can be subscribed to updates and alerts regarding a product.	5%	30%	15%	30%	20%
Product information is searched through social media sites often	5%	45%	25%	10%	20%

4.6. Towards Advertisement in Social Media Sites. Table 4.3 portrays the level of agreement over the advertisement in social media sites, which comprised of three main factors like Attitude towards advertisement, Response towards advertisement, and Purchase intention. In Attitude towards advertisement, social media advertisement on awareness probability over the new product and various brands is 20% strongly agreed as best, 40% only agreed as good, 10% neither agreed nor disagreed, 15% of them disagreed as bad and rest 15% strongly disagreed as worst. The advertisement displayed on social media is 5% strongly agreed as providing positive feeling, 35% of them only agreed on the same feeling, 30% of the students neither agreed nor disagreed on this positive feeling concept, 20% disagreed on this positive feeling and 10% of them strongly disagreed. The

TABLE 4.2
Analysis on Level of Agreement over Trustworthiness

Level of Agreement	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Trust					
Information on social media is trustworthy	10%	15%	25%	40%	30%
I will share my experience with my friends about buying products or acquiring information on social media	5%	40%	20%	20%	15%
I trust the opinion of experts on social media while considering any product	5%	15%	45%	20%	15%
The probability of getting poor quality products through social media platforms is low	5%	25%	25%	15%	30%
Perceived value					
It is possible to find products that are more suitable for my personal quality and styles on social media site	25%	25%	15%	25%	10%
It is possible to save a lot of money acquiring information about product on social media site	15%	45%	5%	10%	20%
The probability of leaking my privacy in purchasing products through social media platforms is low	10%	20%	25%	30%	15%
After I acquire information about products from social media, I know their quality and function	10%	35%	25%	20%	10%
Perceived risk					
The financial risk in buying products through social media sites is low	15%	35%	10%	30%	10%
The probability of wasting time on obtaining information about products through social media sites is low	15%	30%	20%	20%	15%
The probability of harming my physical health by purchasing products (long exposure to mobile or computer screen) is low	5%	20%	30%	35%	10%
The probability of getting me under social pressure in purchasing products through social media is low	15%	25%	25%	15%	20%

advertisement is mainly targeting the particular audience based on their interest in social networking sites, for this argument, 15% strongly agreed, 50% of the students only agreed with the most count, 10% of them neither agreed nor disagreed, 10% disagreed and strongly disagreed on this statement. Brands that use social media for advertising purpose are more original than others who are not using it, on considering the report, 10% of them strongly agreed, 30% of the students agreed on this report, 25% neither agreed nor disagreed for this statement, 20% disagree for this and 15% strongly disagreed on this argument. The brand that advertised in social media site is agreed as preferable by 15% of the students, 30% of them neither agreed nor disagreed, 30% disagreed on this preference and 20% strongly disagreed as they do not prefer the brand. The subsequent one is the Response towards the advertisement. In this, about the product to a friend, 20% strongly agreed to convey, 25% only agreed to talk on the product, 15% neither agreed nor disagreed on talking, 15% disagreed with this conveying concept, and 15% strongly disagreed to talk about the product. On product purchase, 15% strongly agreed to purchase, 15% agreed to buy, 35% of the students neither agreed nor disagreed, 20% disagreed to buy that product, and 15% of them strongly disagreed on purchasing. On viewing the product, 10% strongly agreed to view, 40% of them agreed on viewing over the product, 20% of the students neither agreed nor disagreed on this report, 10% disagreed on this viewing concept, and 15% strongly disagreed. Product recommendation to others, 15% only agreed to recommend on that, 45% neither agreed nor disagreed, 20% of them disagreed to suggest the product among others and 20% strongly disagreed over this suggesting concept. Take part over the events posted by the company, 5% only strongly agreed for this, 10% only agreed, 15% of the students neither agreed nor disagreed, 35% disagreed on this fact and 35% strongly disagreed for this take part intention. Finally, the purchase intention, the four sets of questions in this is explained with their responses as follows. Try a product suggested over social media, 10% of the students strongly agreed to try that, 40% of them agreed on trying the product, 15% neither agreed nor disagreed, 15% disagreed on try over statement, and 20% of the students disagreed. Social media advertising increases the interest on buying the advertised product, 10% strongly agreed to buy the same product, 30% agreed over this buying concept, 20% of them neither agreed

TABLE 4.3
 Analysis on Level of Agreement over Advertisement

Level of Agreement	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Attitude towards advertisement					
Social media advertisements make me more aware about various brands and new products	20%	40%	10%	15%	15%
The advertisements displayed on social media gives a positive feeling	5%	35%	30%	20%	10%
The social networking sites are targeting the advertisements to specific audience based on their interest	15%	50%	10%	10%	10%
Brands that use social media for marketing purpose are more innovative than others who are not using it	10%	30%	25%	20%	15%
I prefer brand that is advertised on social media sites	-	15%	30%	30%	20%
Response towards advertisement					
I talk about the product to my friends	20%	25%	15%	15%	25%
I purchase the product	15%	15%	35%	20%	15%
Just view the product	10%	40%	20%	10%	15%
Recommend the products to others	-	15%	45%	20%	20%
Take part in events posted by the company	5%	10%	15%	35%	35%
Purchase intension					
I like to try a product recommended on social media	10%	40%	15%	15%	20%
Seeing social media advertisements increases my interest in buying the same product	10%	30%	20%	20%	20%
I am very likely to buy products shared by my friends on social media platform	10%	25%	25%	25%	15%
Using social media platform help me make decisions better before purchasing products	10%	45%	20%	10%	10%
Social media advertisements greatly influence the purchase choice	20%	45%	-	20%	15%

nor disagreed, 20% disagreed on this same buying intention and 20% strongly disagreed to purchase the same product. Purchase the product that shared by the friend on social media, 10% of them strongly agreed to do this, 25% agreed to purchase as per their friend’s suggestion, 25% of the students neither agreed nor disagreed on this fact, 25% disagreed to buy the product as per their friends intention and 15% strongly disagreed on this concept. Social media platforms aided on making a better decision before purchasing product, 10% strongly agreed as better decision, 45% agreed this as a good concept, 20% of them neither agreed nor disagreed over this, 10% disagreed on this statement and 10% strongly disagreed. Social media advertisements greatly influence the purchase choice, 20% strongly agreed that as true, 45% of them agreed on this fact, only 20% disagreed over this argument and 15% of the students strongly disagreed over this influence concept.

5. Enquired Conclusion On Impact Of Social Media Advertising.

5.1. Simulation Procedure. The implemented analysis was evaluated using MATLAB 2018a after entering the gained raw data from the students. The result of this analysis was mainly concerned over the social media sites under six constraints like a personal profile, usage, assessment, higher order statistics, trustworthiness, and towards advertisement and each constraint was comprised of 4, 6, 4, 20, 12 and 15 questions, correspondingly. In this, the correlation analysis was made by correlating the entire questions in this questionnaire, and the high correlation (rank) was obtained. The mathematical formula for the correlation coefficient for A and B matrix was given by Eq. (1). In this, σ_A and μ_A exemplifies the standard deviation and mean of A. σ_B and μ_B signifies the standard deviation and mean of B. The Entropy is also computed in this analysis and the mathematical expression is stated as per Eq. (2). In this, Boltzmann constant is given as k_b that is equivalent to $1.38065 \times 10^{23} \text{J/K}$ and p_i refers to the probability.

$$(5.1) \quad \rho(A, B) = \frac{1}{N - 1} \sum_{i=1}^N \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right)$$

TABLE 5.1
4×4 matrix representation on result of correlation analysis under Personal Profile

1	0.2344	0.2568	0.2899
0.2344	1	-0.6086	-0.1671
0.2568	-0.6086	1	0.3662
0.2899	-0.1671	0.3662	1

TABLE 5.2
6×6 matrix representation on result of correlation analysis under Usage

1	0.3352	-0.4254	0.007	-0.0586	0.0433
0.3352	1	-0.2215	-0.5737	-0.1485	-0.2163
-0.4254	-0.2215	1	0.2349	0.0365	-0.1616
0.007	-0.5737	0.2349	1	0.3155	0.2152
-0.0586	-0.1485	0.0365	0.3155	1	0.7371
0.0433	-0.2163	-0.1616	0.2152	0.7371	1

TABLE 5.3
4×4 matrix representation on result of correlation analysis under Assessment

1	0.2349	0.0365	-0.1616
0.2349	1	0.3155	0.2152
0.0365	0.3155	1	0.7371
-0.1616	0.2152	0.7371	1

$$(5.2) \quad S = -k_B \sum_i p_i \log p_i$$

5.2. Correlation Analysis under Personal Profile. Table 5.1 expresses the correlation analysis under the personal profile of the students for the regarded 4 questions. In this, a 4×4 matrix is constructed on the result of correlation analysis, as four questions are analyzed. The highest correlation obtained by the question is given in Table 3.1, “2. Students Gender”.

5.3. Correlation Analysis under Usage. Table 5.2 describes the correlation analysis under the usage of social media sites by students for accounted 6 questions. Owing to this, a 6×6 matrix is constructed as the result of correlation analysis, as six questions are analyzed. The highest correlation gained in the questionnaire is, “5. Between what times do you access social media sites mostly? Facebook” in Table 3.2.

5.4. Correlation Analysis under Assessment. Table 5.3 exemplifies the correlation analysis under assessment of social media sites among students for the constrained 4 questions. In this, a 4×4 matrix is constructed on the result of correlation analysis, as four questions are analyzed. The highest correlation which is attained by the questionnaire in Table 3.3 is, “3. Looking at the next twelve months, comparing to the last year for product information search do you think you will be increasing, decreasing or spending the same amount of time using social networking sites?”.

5.5. Correlation Analysis under Higher Order Statistics. Table 5.4 symbolizes the correlation analysis under higher order statistics of students regarding the social media sites for the considered 20 questions. In this, 20×20 matrix is modeled as the result of correlation analysis, as twenty questions are analyzed. The highest correlation accomplished by the question is in Table 3.4 under connectedness, “6. Sharing contents from one social media site and posting it in other social media by sharing or through links is easy”.

5.6. Correlation Analysis under Trustworthiness. Table 5.5 portrays the correlation analysis under the trustworthiness of social media sites for the accounted 12 questions. In this, the 12×12 matrix is formed as the result of correlation analysis, as twelve questions are analyzed. The highest correlation that is gained from the questionnaire is in Table 3.5 under perceived value, “7. The probability of leaking my privacy in purchasing products through social media platforms is low”.

TABLE 5.4
20x20 matrix representation on result of correlation analysis under Assessment

1	0.3352	-0.4254	0.007	-0.0586	0.0433	-0.0421	0.3764	0.2783	0.0828	-0.1493	0.2837	0.2596	0.0452	-0.2219	0.3437	-0.0295	0.2594	-0.2377	0.1135
0.3352	1	-0.2215	-0.5737	-0.1485	-0.2163	0.255	0.411	0.4837	0.2219	-0.2328	-0.1363	-0.0988	-0.4284	-0.4164	0.1493	-0.4082	-0.4718	-0.2176	0.2056
-0.4254	-0.2215	1	0.2349	0.0365	-0.1616	0.0918	-0.2563	-0.3099	0.0618	-0.0143	-0.3421	-0.0689	-0.2812	-0.1921	0.0059	0.1713	-0.0623	-0.0592	-0.2065
0.007	-0.5737	0.2349	1	0.3155	0.2152	-0.2095	-0.2863	-0.3964	-0.2651	0.3678	0.1126	-0.2374	0.1248	0.3623	-0.0858	0.342	0.4349	0.2101	-0.0386
-0.0586	-0.1485	0.0365	0.3155	1	0.7371	0.3536	-0.3534	-0.4068	-0.0958	0.2392	0.1334	0.0936	0	0.2009	-0.0163	-0.1365	-0.0369	0.3851	-0.0303
0.0433	-0.2163	-0.1616	0.2152	0.7371	1	0.0927	-0.1422	-0.093	-0.0404	0.3142	0.3076	0.2973	0.2428	0.1225	0.1817	-0.1873	0.0912	0.2787	-0.2816
-0.0421	0.255	0.0918	-0.2095	0.3536	0.0927	1	-0.2246	-0.1097	-0.2756	-0.3537	-0.1993	0.0142	-0.1505	-0.2168	-0.2665	-0.4255	-0.1212	-0.1583	0.0145
0.3764	0.411	-0.2563	-0.2863	-0.3534	-0.1422	-0.2246	1	0.4355	0.2879	-0.1031	0.1443	0.0645	-0.3295	-0.3031	0.3989	-0.0607	0.023	-0.2667	0.0294
0.2783	0.4837	-0.3099	-0.3964	-0.4068	-0.093	-0.1097	0.4355	1	0.3256	-0.1727	0.2257	0.1527	0.1569	-0.2009	0.1798	0.0796	-0.0685	-0.2751	0.1314
0.0828	0.2219	0.0618	-0.2651	-0.0958	-0.0404	-0.2756	0.2879	0.3256	1	0.2703	0.3826	0.4619	-0.2718	-0.3532	0.0185	-0.2185	-0.1607	-0.2238	-0.0228
-0.1493	-0.2328	-0.0143	0.3678	0.2392	0.3142	-0.3537	-0.1031	-0.1727	0.2703	1	0.1287	0.0309	-0.082	0.0088	0.1025	-0.1694	-0.0165	0.1725	-0.3644
0.2837	-0.1363	-0.3421	0.1126	0.1334	0.3076	-0.1993	0.1443	0.2257	0.3826	0.1287	1	0.158	0	0.0068	-0.1022	0.1102	0.2901	-0.0666	-0.0979
0.2596	-0.0988	-0.0689	-0.2374	0.0936	0.2973	0.0142	0.0645	0.1527	0.4619	0.0309	0.158	1	0.1367	-0.3633	0.1377	-0.2545	-0.0413	0	-0.1409
0.0452	-0.4284	-0.2812	0.1248	0	0.2428	-0.1505	-0.3295	0.1569	-0.2718	-0.082	0	0.1367	1	0.3444	0.2185	0.386	0.4387	0.2715	0.1559
-0.2219	-0.4164	-0.1921	0.3623	0.2009	0.1225	-0.2168	-0.3031	-0.2009	-0.3532	0.0088	0.0068	-0.3633	0.3444	1	-0.3445	0.3821	0.0902	-0.0725	0.4592
0.3437	0.1493	0.0059	-0.0858	-0.0163	0.1817	-0.2665	0.3989	0.1798	0.0185	0.1025	-0.1022	0.1377	0.2185	-0.3445	1	0.2523	0.1168	0.1768	-0.3378
-0.0295	-0.4082	-0.1713	0.342	-0.1365	-0.1873	-0.4255	-0.0607	0.0796	-0.2185	-0.1694	-0.1022	-0.2545	0.386	0.3821	0.2523	1	0.1449	0.0369	0.1628
0.2594	-0.4718	-0.0623	0.4349	-0.0369	0.0912	-0.1212	0.023	-0.0685	-0.1607	-0.0165	0.2901	-0.0413	0.4387	0.0902	0.1168	0.1449	1	0.171	-0.1444
-0.2377	-0.2176	-0.0592	0.2101	0.3851	0.2787	-0.1583	-0.2667	-0.2751	-0.2238	0.1725	-0.0666	0	0.2715	-0.0725	0.1768	0.0369	0.171	1	-0.1312
0.1135	0.2056	-0.2065	-0.0386	-0.0303	-0.2816	0.0145	0.0294	0.1314	-0.0228	-0.3644	-0.0979	-0.1409	0.1559	0.4592	-0.3378	0.1628	-0.1444	-0.1312	1

TABLE 5.5
12x12 matrix representation on result of correlation analysis under Trustworthiness

1	0.1449	0.0369	0.1628	-0.0485	0	-0.0034	0.3262	0.2848	-0.2346	-0.2948	-0.411
0.1449	1	0.171	-0.1444	0.3946	-0.3693	-0.0333	-0.0232	0.1165	0.0028	-0.1801	-0.1763
0.0369	0.171	1	-0.1312	0.4357	0.1929	-0.3975	-0.182	-0.032	-0.4364	-0.091	-0.1754
0.1628	-0.1444	-0.1312	1	-0.2647	0.3897	0.0487	-0.234	0.4236	-0.0053	-0.574	0.0966
-0.0485	0.3946	0.4357	-0.2647	1	0.0362	-0.3201	0.0911	-0.2766	-0.5296	-0.2334	-0.1042
0	-0.3693	0.1929	0.3897	0.0362	1	-0.4651	-0.2949	-0.2422	-0.2828	-0.3605	-0.2842
-0.0034	-0.0333	-0.3975	0.0487	-0.3201	-0.4651	1	0.3602	0.4516	0.51	0.5121	0.469
0.3262	-0.0232	-0.182	-0.234	0.0911	-0.2949	0.3602	1	0.1415	0.0593	0.1598	-0.0795
0.2848	0.1165	-0.032	0.4236	-0.2766	-0.2422	0.4516	0.1415	1	0.0678	-0.2504	0.1101
-0.2346	0.0028	-0.4364	-0.0053	-0.5296	-0.2828	0.51	0.0593	0.0678	1	0.4597	0.3429
-0.2948	-0.1801	-0.091	-0.574	-0.2334	-0.3605	0.5121	0.1598	-0.2504	0.4597	1	0.3278
-0.411	-0.1763	-0.1754	0.0966	-0.1042	-0.2842	0.469	-0.0795	0.1101	0.3429	0.3278	1

TABLE 5.6
15x15 matrix representation on result of correlation analysis Towards Advertisement

1	0.3256	-0.1727	0.2257	0.1527	0.1569	-0.2009	0.1798	0.0796	-0.0685	-0.2751	0.1314	-0.351	-0.2377	0.0663
0.3256	1	0.2703	0.3826	0.4619	-0.2718	-0.35332	0.0185	-0.2185	-0.1607	-0.2238	-0.0228	0.0817	0.1074	0.0058
-0.1727	0.2703	1	0.1287	0.0309	-0.0802	0.0088	0.1025	-0.1694	-0.0165	0.1725	-0.3644	0.2671	-0.0466	-0.224
0.2257	0.3826	0.1287	1	0.158	0	0.0068	-0.1022	0.1102	0.2901	-0.0666	-0.0979	0.3313	-0.3957	-0.0031
0.1527	0.4619	0.0309	0.158	1	0.1367	-0.3633	0.1377	-0.2545	-0.0413	0	-0.1409	0.117	0.2417	-0.2416
0.1569	-0.2718	-0.082	0	0.1367	1	0.3444	0.2185	0.386	0.4387	0.2715	0.1559	0.2866	0.11	-0.3935
-0.2009	-0.3532	0.0088	0.0068	-0.3633	0.3444	1	-0.3445	0.3821	0.0902	-0.0725	0.4592	-0.1496	0	0.0134
0.1798	0.0185	0.1025	-0.1022	0.1377	0.2185	-0.3445	1	0.2523	0.1168	0.1768	-0.3378	0.1228	0.3056	-0.3264
0.0796	-0.2185	-0.1694	0.1102	-0.2545	0.386	0.3821	0.2523	1	0.1449	0.0369	0.1628	-0.0485	0	-0.0034
-0.0685	-0.1607	-0.0165	0.2901	-0.0413	0.4387	0.0902	0.1168	0.1449	1	0.171	-0.1444	0.3946	-0.3693	-0.0333
-0.2751	-0.2238	0.1725	-0.0666	0	0.2715	-0.0725	0.1768	0.0369	0.171	1	-0.1312	0.4357	0.1929	-0.3975
0.1314	-0.0228	-0.3644	-0.0979	-0.1409	0.1559	0.4592	-0.3378	0.1628	-0.1444	-0.1312	1	-0.2647	0.3897	0.0487
-0.351	0.0817	0.2671	0.3313	0.117	0.2866	-0.1496	0.1228	-0.0485	0.3946	0.4357	-0.2647	1	0.0362	-0.3201
-0.2377	0.1074	-0.0466	-0.3957	0.2417	0.11	0	0.3056	0	-0.3693	0.1929	0.3897	0.0362	1	-0.4651
0.0663	0.0058	-0.224	-0.0031	-0.2416	-0.3935	0.0134	-0.3264	-0.0034	-0.0333	-0.3975	0.0487	-0.3201	-0.4651	1

5.7. Correlation Analysis Towards Advertisement. Table 5.6 explains the correlation analysis towards the advertisement of social media sites for the constrained 15 questions. In this, a 15x15 matrix is designed as the result of correlation analysis, as fifteen questions are analyzed. The highest correlation gained in this questionnaire section is from Table 3.6 under Response towards advertisement, “6. I talk about the product to my friends”.

5.8. Overall Correlation Analysis. Table 5.7 demonstrates the overall correlation analysis of the defined topic of social media sites. In the view of the attained correlation of the above six questions, the overall correlation analysis is obtained. Here, a 6x6 matrix is developed as the result of the aforesaid correlation analysis. The finally attained best correlation among the entire questionnaire is from the usage section in Table

TABLE 5.7
6×6 matrix representation on result of correlation analysis under Usage

1	0.3352	-0.4254	0.007	-0.0586	0.0433
0.3352	1	-0.2215	-0.5737	-0.1485	-0.2163
-0.4254	-0.2215	1	0.2349	0.0365	-0.1616
0.007	-0.5737	0.2349	1	0.3155	0.2152
-0.0586	-0.1485	0.0365	0.3155	1	0.7371
0.0433	-0.2163	-0.1616	0.2152	0.7371	1

TABLE 5.8
Entropy Analysis of Overall Questionnaire

																	Age group		
0.9341	1.2955	0.971	1.2362	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Since when have you been visiting these social media sites, Facebook																			
1.7394	1.3955	1.4577	1.2568	1.319	1.4166	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Which of the following are you connected with on social media networks																			
1.4577	1.2568	1.319	1.4166	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
It is easy to join the groups and communities that I am interested in social media sites																			
1.7394	1.3955	1.4577	1.2568	1.319	1.4166	1.9955	1.5589	1.5589	2.2589	1.6855	1.7345	2.1211	1.8834	1.6955	1.2789	1.6842	1.6805	1.601	1.6805
The probability of getting me under social pressure in purchasing products through social media is low																			
1.6841	1.6805	1.601	1.6805	1.2789	1.6855	2.1261	1.5955	2.0414	1.8577	1.9464	2.2855	-	-	-	-	-	-	-	-
The advertisements displayed on social media gives a positive feeling																			
1.5589	2.2589	1.6855	1.7345	2.1211	1.8834	1.6955	1.2789	1.6842	1.6805	1.601	1.6805	1.2789	1.6855	2.1261	-	-	-	-	-

3.2, “5. Between what time do you access the social media sites mostly? Facebook”.

5.9. Overall Entropy Analysis. Table 5.8 describes the entropy analysis of the overall contributed questionnaire. On considering the personal profile of the students, the maximum entropy is gained by the question from Table 3.1 i.e. “3. Age group”. Similarly, for the usage section, the obtained entropy is maximum for the question from Table 3.2, “1. Since when have you been visiting these social media sites, Facebook”. Subsequently, an assessment section, from Table 3.3, question “1. Which of the following are you connected with on social media networks?” has obtained the maximum entropy of all other questions in that section. While taking the higher order statistics drives like community, connectedness, openness, dependence, and participation, the maximum attained entropy is for Table 3.4 and the question is “10. It is easy to join the groups and communities that I am interested in social media sites”. The next section is trustworthiness, where the question that achieved maximum entropy is from Table 3.5, “12. The probability of getting me under social pressure in purchasing products through social media is low”. To the end, among towards advertisement section, the maximum entropy gained question is in Table 3.6, “2. The advertisements displayed on social media give a positive feeling”. Among these obtained maximum entropies, the overall entropy that gained the maximum value of 2.2516 is for the question in the personal profile section “Age group”.

6. Conclusion. In this paper, a new strategy was introduced to make a statistical review on social media advertisements. This analysis was mainly involved with six stages like (i) Personal Profile; (ii) Usage (iii) Assessment (iv) Higher Order statistics (v) Trustworthiness and (vi) Towards advertisement. In the first stage, the questionnaire was prepared by concerning the intimate usage of social media sites and advertisements over the social media network. Then on the next stage, the prepared questionnaire was distributed among the various universitycollege students and suggested to fill up according to their thoughts or responses. These gathered responses were then taken for analysis. Further, the analysis has been made based on the higher-order statistical analysis that auspiciously focused on correlation coefficients and entropy function, which were aided on determining the correlation and entropy between the responses towards the social media network. In the view of the attained correlation of the above six questions, the overall correlation analysis was obtained. Here, the 6×6 matrix was developed as the result of the aforesaid correlation analysis. The finally attained best correlation among the entire questionnaire was from the usage section in Table 3.2, “5. Between what time do you access social media sites mostly? Facebook”. Moreover, among these obtained maximum entropies, the overall entropy that gained the maximum value of 2.2516 was for the question in the personal profile section “Age group”. In the future we will provide advertising industry, to make the advertising more and stronger.

REFERENCES

- [1] S. CHINCHANACHOKCHAI, AND F. GREGORIO, *A consumer socialization approach to understanding advertising avoidance on social media* Journal of Business Research, vol.110, pp.474-483, March 2020.
- [2] H. GUPTA, S. SINGH, P. SINHA, *Multimedia tool as a predictor for social media advertising- a YouTube way* Multimedia Tools and Applications, vol.76, no.18, pp 18557-18568, September 2017.
- [3] S. SREEJESH, J. PAUL, C. STRONG, AND J. PIUS, *Consumer response towards social media advertising: Effect of media interactivity, its conditions and the underlying mechanism* International Journal of Information Management, vol.54, 2020.
- [4] R. E. JUST, R. D. POPE, *The many conditions under which monopolistic advertising can differ from the social optimum* Journal of Economics and Finance, vol.41, no.3, pp 421-440, July 2017.
- [5] Y-C. LEE, *Comparing factors affecting attitudes toward LBA and SoLoMo advertising* Information Systems and e-Business Management, vol.16, no.2, pp 357-381, May 2018.
- [6] C. A. LIN, TONGHOONKIM, *Predicting user response to sponsored advertising on social media via the technology acceptance model* Computers in Human Behavior, vol.64, pp.710-718, November 2016
- [7] H.T. JAVAN, A. KHANLARI, O. MOTAMEDI, H. MOKHTARI, *A hybrid advertising media selection model using AHP and fuzzy-based GA decision making* Neural Computing and Applications, vol.29, no.4, pp 1153-1167, February 2018.
- [8] F.BITIKTAS, AND O. TUNA, *Social media usage in container shipping companies: Analysis of Facebook messages* Research in Transportation Business & Management, vol. 34, 2020
- [9] T. SILAWAN AND C. ASWAKUL, *SybilVote: Formulas to Quantify the Success Probability of Sybil Attack in Online Social Network Voting* IEEE Communications Letters, vol. 21, no. 7, pp. 1553-1556, July 2017.
- [10] H. KO, S. PACK AND W. LEE, *Timer-Based Push Scheme for Online Social Networking Services in Wireless Networks* IEEE Communications Letters, vol. 16, no. 12, pp. 2095-2098, December 2012.
- [11] K. LIANG, J. K. LIU, R. LU AND D. S. WONG, *Privacy Concerns for Photo Sharing in Online Social Networks* IEEE Internet Computing, vol. 19, no. 2, pp. 58-63, Mar.-Apr. 2015.
- [12] H. QINLONG, M. ZHAOFENG, Y. YIXIAN, N. XINXIN AND F. JINGYI, *Improving security and efficiency for encrypted data sharing in online social networks* China Communications, vol. 11, no. 3, pp. 104-117, March 2014.
- [13] A. THAPA, M. LI, S. SALINAS AND P. LI, *Asymmetric Social Proximity Based Private Matching Protocols for Online Social Networks* IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 6, pp. 1547-1559, 1 June 2015.
- [14] J. CHEN, J. W. PING, Y. XU AND B. C. Y. TAN, *Information Privacy Concern About Peer Disclosure in Online Social Networks* IEEE Transactions on Engineering Management, vol. 62, no. 3, pp. 311-324, Aug. 2015.
- [15] K. WONG, A. WONG, A. YEUNG, W. FAN AND S. TANG, *Trust and Privacy Exploitation in Online Social Networks* IT Professional, vol. 16, no. 5, pp. 28-33, Sept.-Oct. 2014.
- [16] K. YADATI, H. KATTI AND M. KANKANHALLI, *CAVVA: Computational Affective Video-in-Video Advertising* IEEE Transactions on Multimedia, vol. 16, no. 1, pp. 15-23, Jan. 2014.
- [17] J. QIN, H. ZHU, Y. ZHU, L. LU, G. XUE AND M. LI, *POST: Exploiting Dynamic Sociality for Mobile Advertising in Vehicular Networks* IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 6, pp. 1770-1782, 1 June 2016.
- [18] Z. CHENG, X. WU, Y. LIU AND X. HUA, *Video eCommerce++: Toward Large Scale Online Video Advertising* IEEE Transactions on Multimedia, vol. 19, no. 6, pp. 1170-1183, June 2017.
- [19] K. REN, W. ZHANG, K. CHANG, Y. RONG, Y. YU AND J. WANG, *Bidding Machine: Learning to Bid for Directly Optimizing Profits in Display Advertising* IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 4, pp. 645-659, 1 April 2018.
- [20] M. TAVANA, E. MOMENI, N. REZAEINIYA, S. M. MIRHEDAYATIAN, H. REZAEINIYA, *A novel hybrid social media platform selection model using fuzzy ANP and COPRAS-G* Expert Systems with Applications, vol.40, no.14, pp.5694-5702, 15 October 2013
- [21] D. NOHARA, Y. SENG M. MATSUBARA, T. SAKAI, *Media selection for refolding of thermolysin by use of immobilized preparation* Journal of Bioscience and Bioengineering, vol.89, no.2, pp.188-192, 2000.
- [22] Y. HUANG, C.-G. YANG, H. BAEK, S-G LEE, *Erratum to: Revisiting media selection in the digital era: adoption and usage* Service Business, vol.10, no.1, pp 261-261, March 2016
- [23] D. T. TOSTI, J. R. BALL, *A behavioral approach to instructional design and media selection* AV communication review, vol.17, no.1, pp 5-25, March 1969
- [24] Z. XUE, J. WANG, G. DING, Q. WU, Y. LIN AND T. A. TSIFTSIS, *Device-to-Device Communications Underlying UAV-Supported Social Networking* IEEE Access, vol. 6, pp. 34488-34502, 2018.
- [25] J-P. HUANG, B. HEIDERGOTT, I. LINDNER, *Naïve learning in social networks with random communication* Social Networks, vol.58, pp.1-11, 2019

Edited by: P. Vijaya

Received: Dec 5, 2019

Accepted: Jun 22, 2020



NOVEL METRIC FOR LOAD BALANCE AND CONGESTION REDUCING IN NETWORK ON-CHIP

ABDELKADER AROUI*, ABOU ELHASSAN BENYAMINA†, PIERRE BOULET‡, KAMEL BENHAOUA§ AND AMIT KUMAR SINGH¶

Abstract. The Network-on-Chip (NoC) is an alternative pattern that is considered as an emerging technology for distributed embedded systems. The traditional use of multi-cores in computing increase the calculation performance; but affect the network communication causing congestion on nodes which therefore decrease the global performance of the NoC. To alleviate this problematic phenomenon, several strategies were implemented, to reduce or prevent the occurrence of congestion, such as network status metrics, new routing algorithm, packets injection control, and switching strategies. In this paper, we carried out a study on congestion in a 2D mesh network, through various detailed simulations. Our focus was on the most used congestion metrics in NoC. According to our experiments and performed simulations under different traffic scenarios, we found that these metrics are less representative, less significant and yet they do not give a true overview of reading within the NoC nodes at a given cycle. Our study shows that the use of other complementary information regarding the state of nodes and network traffic flow in the design of a novel metric, can really improve the results. In this paper, we put forward a novel metric that takes into account the overall operating state of a router in the design of adaptive XY routing algorithm, aiming to improve routing decisions and network performance. We compare the throughput, latency, resource utilization, and congestion occurrence of proposed metric to three published metrics on two specific traffic patterns in a varied packets injection rate. Our results indicate that our novel metric-based adaptive XY routing has overcome congestion and significantly improve resource utilization through load balancing; achieving an average improvement rate up to 40 % compared to adaptive XY routing based on the previous congestion metrics.

Key words: Embedded Systems, Network on-Chip, Routing, Load balancing, Congestion reducing.

AMS subject classifications. 68M10, 68M12, 68M14

1. Introduction. The Network-on-Chip (NoC) has proved to be more reliable, modular and reusable [4] solutions than shared buses. NoC is proposed as an alternative interconnection solution in mutli-processors on chip (MPSoC).As it handles the high throughput and latency communication demands amongst various on-chip tiles, it connects cores, caches and memory controllers using packet switching routers, covering both regular and irregular topologies [17, 10].

Routing algorithm effectively facilitate communication and routing packets between the cores. Although this algorithm decides how to route incoming packets to various destinations, network congestion may occur due to low bisection bandwidth and poor routing. In fact, because of non-uniformity traffic, some network nodes must send more packets than others causing a rise in congestion and affecting network performance significantly i.e. network performance is highly dependent on the routing algorithm.

Routing in network on chip can be generally classified into oblivious and adaptive [24]. In oblivious routing, the packets are routed without any information about the traffic levels of the network, whereas in adaptive routing algorithm, the routing decision is made by taking into consideration the current congestion status of the network.

The process of an adaptive routing algorithm is also divided into two phases: the routing function and the selection function. The routing function defines the set of available output channels for a packet dependent on the source and destination positions; whereas the selection function selects an output channel from the set of

*Computer Science Department, University of Oran 1 Ahmed Ben Bella, Algeria (aroui.abdelkader@edu.univ-oran1.dz).

†Computer Science Department, University of Oran 1 Ahmed Ben Bella, Algeria. (benyamina.lahssan@univ-oran1.dz).

‡CNRS, University Lille, Centrale Lille, IMT Lille Douai, UMR 9189-CRISTAL, Lille, France (pierre.boulet@univ-lille1.fr).

§Computer Science Department, University of Mustapha Stambouli, Mascara, Algeria. (kbenhaoua@gmail.com).

¶School of Computer Science and Electronic Engineering, University of Essex, UK. (a.k.singh@essex.ac.uk).

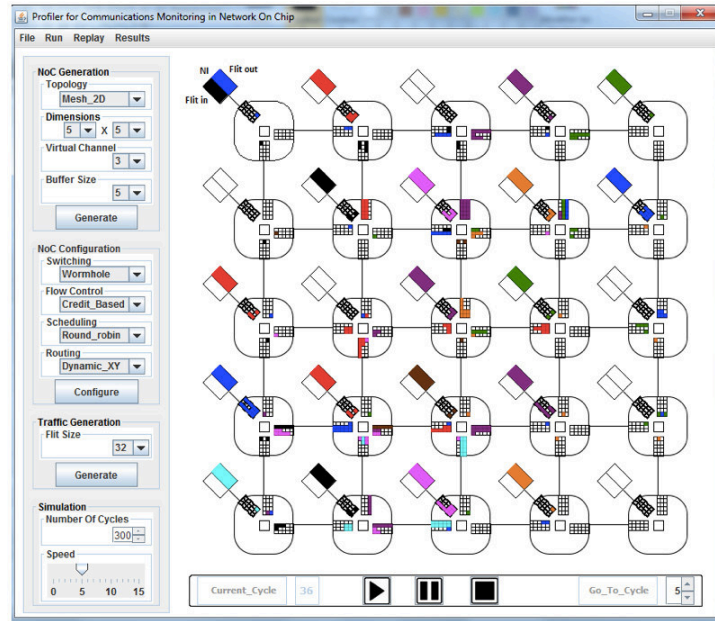


FIG. 1.1. *Our Graphical simulator for Data-Flow Monitoring on Network on-Chip [23]*

available output channels.

Adaptive routing algorithms benefit from reachable metrics of a router to evaluate congestion level of a network. Various metrics have been considered in estimating the congestion level in NoC.

In this paper, we study a subset of the most commonly used metrics in the NoCs congestion domain. After a thorough graphic analysis at each cycle of the execution of some small scenarios on our simulator cf. Fig. 1.1, it appeared that the underestimation of congestion information in some scenarios force routing to congested areas, and random routing decisions are taking place when the candidate output ports have the same congestion value. The indeterminism of some congestion-based routing algorithms and the representativeness of existing congestion metrics are the two issues that motivated our study in this paper.

A novel metric-based approach is proposed and evaluated for 2D mesh topology through simulations using various workloads. The experimental results reveal that the proposed mechanism results in better performances compared to existing techniques.

The rest of this paper is structured as follows: Section 2 describes existing work on congestion treatment. Section 3 gives a general overview about the set-up of our Network-on-Chip architecture. Our proposed technique is presented and illustrated in Sect. 4. Some experiences and analyses are found in Sect. 5 and finally in Sect. 6 we recapitulate and propose some of the future works and perspectives.

2. Related Work. Multi-processor system-on-Chip (MPSoC) has emerged as a positive solution to address the increased computational requirements of modern and future applications [3]. This recent breakthrough of MPSoC which consist multiple processing elements (PEs) in the same chip, thanks to the evolution of semiconductor technology, allow us to integrate several elements in the same chip. MPSoC also provide increased parallelism towards achieving high performance [2]. The NoC has been introduced as a power efficient and scalable communication infrastructure between processors.

Although a large variety of NoC topologies have been presented in the literature, 2D mesh is the most widely studied. Due to spatio-temporal irregularity of network traffic, particularly in 2D mesh topology, some network nodes have to dispatch much more packets than others. Congestion will take place at those nodes and significantly affects network performance [32].

If the congestion arises several times, the concerned resources will be getting additional higher temperature (hotspot formation). This phenomenon reduces the effective rate of the NoC, or in the worst case, can cause

blocking of traffic and failure of certain network resources. For this purpose, a huge number of approaches were proposed to surmount the negative effect of congestion and improve the network performance: reconfigurable routers and novel NoC architectures were designed to avoid router and network congestion; congestion metrics were included to complete the design of adaptive routing algorithms intended for congestion reduction [14, 6, 28]; End-to-end flow control technique also was suggested to adjust injecting packets into any emerging network congestion [30, 26, 27]; thus selection strategies were developed to reduce congestion and improve network performance with regards to latency [1, 8, 15].

Routing algorithm [18, 20, 21] is an important part in NoC design, that has a significant impact on the traffic flow and performance improvement of NoC. In adaptive routing, a proper selection strategy is needed to choose the appropriate path for packets to reach their destinations as fast as possible. Several efforts have shown that adaptive routing can reduce network congestion by conducting traffic away from the congested regions. Congestion-aware adaptive routing algorithms were merely designed to select the least congested route to produce the load balance in the network by considering the current congestion status of the network in their routing decisions.

A specific network for congestion data transmission has been added to the proposed NoC architecture [22]. The mechanism that aggregates and transmits metrics of congestion beyond direct neighbors throughout the monitoring network which has been proposed by Gratz and al, it was labelled as Regional Congestion Awareness algorithm (RCA) [9]. It was proved in [29], that destination based adaptive routing algorithms (DAR and DBAR) had provided better performances than regional adaptive routing techniques. The Gratz's work was improved, not long ago, by using a Global Congestion Awareness (GCA) mechanism [28].

Adaptive routing based on Ant Colony Optimization has been presented [13] to exploit traffic historical information to reach load balancing. The inconvenient of this technique was that the pheromone table grows fast with the scaling of NoC and thus the storage cost was too high. This problem was solved and technique was improved by proposing a Regional ACO-Based Cascaded Adaptive Routing [5] with static and dynamic regional table forming technique to decrease the cost of the table storage. In order to reduce NoC area, power and overall packet latency, two routing mechanisms for congestion handling have been presented in [11]. The mechanisms use σ bits to capture the congestion information of the links of the network for the node with n-hop visibility at each node. A new methodology of Congestion Aware Adaptive Routing (CAAR) was designed [16], to improve NoC latency and throughput during high congestion by prioritizing packets that suffer higher latency while travelling long distances. A congestion detection mechanism has been proposed in [12], which is capable of locating where the congestion is in the network within several cycles and to change the routing algorithm of congested nodes in the cycle following the congestion detection.

By monitoring and exchanging some network resources status information, we can, therefore, predict the occurrence of congestion in the network. According to [31], congestion may occur at a network node great possibility when more than a quarter of its buffer space has been occupied. The information exchanged in the network and between neighboring nodes is called metric. Various metrics have been used in the literature for estimating congestion level of network resources. The number of free VCs [16], the number of free buffer slots in the downstream router [33], and the active demand that each output port experiences (i.e., crossbar demand) [9, 7] are among such congestion metrics. These metrics were very interested in evaluating the state of the desired downstream router input port and simultaneously they neglected the state of the entire entity (router) that contains this resource and other resources. However, taking into account the downstream router input port state as congestion metric to decide on the next hop is not an effective method, since the router state is the sum of states of resource components that router i.e. all ports and all virtual Channels state. Based on this assumption and through the simulations we have discovered that routing algorithms listed above have made some improvement in overcoming the negative effect of congestion and improving network performance.

Consequently, we managed to propose firstly, a novel congestion metric that estimates the overall state of the router and reflects exactly what is happening within the router. Secondly, a congestion-aware routing algorithm based on this metric was designed to significantly reduce the negative impact of congestion and its occurrences as well as improving NoC's performance.

3. NoC Architecture Overview. In this work, 2D Mesh topology was chosen. It is the most used in the literature and the easiest to implement. Our platform is a set of processors elements (PEs) interconnected by

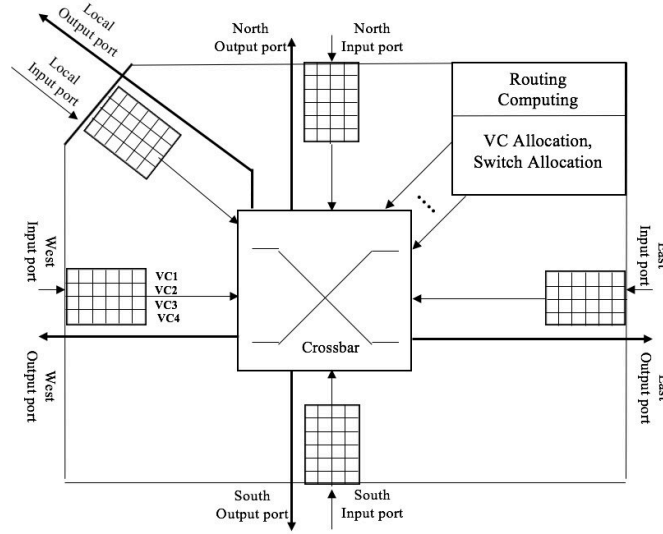


FIG. 3.1. Architecture of mesh Router

routers. The data flow exchanged between PEs passes by routers. A typical wormhole NoC router is shown in Fig. 3.1. Every router in our topology has five I/O ports to connect with its adjacent router and IP blocks: four connected to the neighboring routers and one for the local core (PE) through the Network Interface (NI). The input-ports contain the Virtual Channel (VC) buffers, and the output-ports are simple data buses.

The router architecture includes FIFOs (Buffers) for each input port, route computation unit, Virtual Channel (VC) allocation unit, crossbar control logic, and the crossbar. NI transforms data exchanged between cores on many packets then in to smaller data called flits. The flit is the smallest unit on which flow control can perform. A packet in this case is split into a header flit, body flit(s), and a tail flit. A flit enters into the router through one of the ports and got stored in its FIFO. The header flit indicates each start of a new packet, as soon as it is stored in its FIFO, linked to the routing unit which establishes the output port that the packet should follow. Once, this is done, the header flit tries to allocate a virtual channel for the next hop. If that is the case, it enters the switch arbitration stage, where it rivals for the output port with other flits from the other router input ports. Once the crossbar passage is granted, the flit traverses the switch and enters the channel. Following flits belonging to the same packet can proceed directly to the crossbar and lead to the output port [19]. The NoC routers use Round-Robin (RR) scheme for both VC and switch allocation due to its simplicity.

Many routing algorithms dealing with the mesh architecture networks have been proposed, XY routing algorithm is most suitable for networks which are implemented in this architecture. Two variants of this algorithm have been proposed, static and dynamic. In this paper, we opted for the dynamic

For a better distribution of the traffic on the network, the researchers proposed dynamic XY routing methods [18] which provide adaptiveness and ensure deadlock-free and live-lock-free routing at the same time. The adaptiveness lies in making routing decisions by monitoring network status in the proximity, and the deadlock-free and live lock-free features are incorporated by limiting a packet to traverse the network by using the shortest paths between the source and the destination [18].

Different coefficients have been used to complete the design of dynamic XY routing for congestion reduction in NoC, like the number of available VCs, the crossbar demand, and the number of free buffer slots at the port of corresponding entry in the downstream router.

Our approach in this paper is aimed to introduce a novel congestion metric into the design of our dynamic XY routing technique. More details of our contribution will be presented and illustrated in the next section.

4. Proposed technique. Starting with the assumption that load imbalance is one of the main critical issues in NoC, and that good network performance can be achieved by equally distributing traffic between

nodes in the network. Since congestion occurs on loaded nodes which affect significantly network performance, it is viable to find the right way to spread the excessive load of a node to its neighbors and reduce the number of congestion in the network. In this context, congestion control scheme which consists of dynamic adaptive routing output port selection is aimed to balance global traffic distribution as well as to alleviate congestion caused by heavy network traffics.

As we argued in the previous section, several proposed works were addressed the problem of network congestion either to reduce or to predict it by proposing routing and selection strategies that were based on most used and approved metrics in literature. Many solutions and methods have been presented as a form of techniques on new architectures that facilitate network monitoring, exchange and exploitation of congestion information in traffic routing decisions.

We have observed that techniques have changed from one article to another however the metrics remained unchanged. A few techniques have been proposed against some metrics to alleviate the congestion phenomenon. This track has motivated us to make an analysis study through simulations on the three most commonly used metrics: Available VC, Demand Crossbar and Buffer occupancy.

By running some small scenarios on our graphical simulator cf. Fig. 1.1 and through step by step debugging option that offer our simulator, two important findings came in sight: 1) the underestimation of node congestion information in certain scenarios forces routing traffic to congested areas, 2) Random routing decisions are made when the candidate exit ports have the same congestion value. The indeterminism of some congestion-based routing algorithms and the non-representativeness of congestion information are the cause of our motivation to our study.

In our bi-objective approach, a novel mechanism for controlling and evaluating node congestion is integrated in the design of the routing algorithm, which allows efficient distribution of heavy packet traffics on the chip and minimizes the congestion occurrence. We introduced a novel metric which is composed of two modules, the first one is original and designed to estimate information and state of node congestion, and to take good decisions of routing; whereas the second one measures the load applied on nodes to better balance the traffic on networks.

We note: *OutFlits*: the number of Flits leaving router at given cycle correspond to the number of active output port; *Candidate VCs*: the number of active virtual channels which contains data ready to cross the router by one of its Output ports; the maximum number of output ports in our router architecture is equal to 5. Router congestion status and information evaluation is given by Eq (4.1):

$$Router\ status = \underbrace{\frac{OutFlits}{CandidateVCs}}_{(a)} \times \underbrace{\frac{OutFlits}{Outputports}}_{(b)} \quad (4.1)$$

The sub-equation (a) Shows the extent of contention between VCs. The sub-equation (b) gives an idea about what is exactly happening within router's output ports when many are being requested at once, at a given cycle. (b) completes (a) to form router congestion information. In order to make the best use of network resources, we need to properly distribute the traffic and to avoid any uncontrolled routing decision. In the case of the equalization of congestion information between several VCs candidates in contention, the router occupancy rate information is added to the congestion information to form our congestion metric which is given by Eq (4.2):

$$Our\ CM = Router\ status \times Router\ occupancy\ rate \quad (4.2)$$

where *Router occupancy rate* detail is given by Eq. (4.3):

$$Router\ occupancy\ rate = \frac{1}{V} \sum_{i=1}^V Buffer\ occupancy\ rate_i \quad (4.3)$$

where V represents the router virtual channels number and $Buffer\ occupancy\ rate_i$ is the occupancy rate of i -th buffer of router.

The division by 0 leads to some complicated situations for our CM. In order to avoid such situation and to keep the efficiency of our metric we have proposed algorithm 1, The CM computation algorithm.

Algorithm 1 Congestion Metric Computation.

```

1: for (Each RouterIndex) do
2:   if (Candidat_VCs[RouterIndex]==0) then
3:     NovelCongestionMetric[RouterIndex]= 1.0
4:   else
5:     Router status [RouterIndex]= OutFlits2 / (Candidat VCs × Output ports)
6:     NovelCongestionMetric[routerindex] = Routerstatus[RouterIndex] × Occupancy_rate[RouterIndex]
7:   end if
8: end for

```

As discussed previously, to validate our proposal, we have opted for the XY dynamic algorithm as a routing technique for its simplicity and because it is the most implemented for 2D Mesh topology.

Our congestion control for routing decision making as detailed in Algorithm 2, depends on neighbor nodes congestion condition. it uses Eq. (4.2) for calculating congestion value in each node. To route packets, each node in network is responsible to get and compare the neighbors updated congestion values to select the output port that its congestion value is minimum.

Algorithm 2 Our congestion-based approach

```

1: for (Each winner Flit of node 's CandidateVCs and each output port) do
2:   Read the destination of Flit ready to leave
3:   Compare addresses of the destination and the current router.
4:   if (the destination is the local core of the current router) then
5:     Prepare to send the Flit to the local core
6:   else
7:     if the destination has the same x or y address as the current router then
8:       Prepare to send the Flit to the neighboring router on the y-axis or x-axis towards the destination
9:     else
10:      Get and Compare the congestion values of neighboring ports leading to destination
11:      Prepare to send the Flit to the port with the least congestion value according to the destination
12:    end if
13:  end if
14:  Increment the number of leaving flits
15:  Update CandidateVCs value
16:  Compute and update node congestion value according to algorithm (1)
17:  Send the Flit
18: end for

```

Our approach was designed to select the least congested route to minimize congestion occurrence and to produce the load balancing in the network by taking into account our proposed congestion metric to decide the next hop.

To demonstrate the most factors that adduce our approach in comparison to previous techniques, we put forward the following simulations to clarify and analyze the results and the performances of the approach.

5. Experiment. A NoC java-based simulator [23] is used in order to evaluate our strategy performance. Its graphical presentation offers the possibility to follow the execution details in real time, to know the real state of routers and to locate easily those who suffer from congestion. This simulator is accurate cycle, that means that one cycle is required to transmit one flit to the next router. For each cycle in application lifetime, our simulator starts by identifying candidate flits that request output ports, then running routing technique to determine the path to take, followed by an arbitration technique to obtain winning flits that cross the router continuing their route in network.

In this paper, we simulate 2D-mesh NoC platform that use a wormhole switching mechanism, round-robin

TABLE 5.1
Simulation setup.

Simulator	[23]
NoC Topology	7×7 2D Mesh
Virtual channel	Yes, 3 VCs per port
Buffers / packet size	5 / 5
Switching	Wormhole
Routing	Congestion-based adaptive routing
Arbitration	Round-robin
Congestion metrics	Available VCs, Buffer Occupancy Crossbar Demand, Our metric
Traffic distribution	Uniform, random (50 xml files)
Simulation Time	300;500;1000 clock cycles per xml file

arbitration and our proposed method as a routing technique. The inter-router transmission is based on the Credit-based flow. Each physical channel in router has a buffer of four flits where packet length is equal to buffer size. The simulation setup is summarized in Table 5.1.

In addition to our novel congestion metric we have chosen the three most common used metrics. We attentively study congestion by focusing on its occurrence in the network for each metric under different types of traffic. We examined uniform and random traffic scenarios. In uniform traffic, node (i, j) only generates packets to node $(2i, 2j)$. In random traffic, each node randomly generates packets to every other node with the same probability where a node (i, j) represents the core/router placed at a row of i and column j of a 2D mesh.

The injection of traffic into the network is regulated by a packet injection mechanism called PIR (Packets Injection Rate), i.e. a node that injects on average a packet into the network every 10 cycles, it has a PIR=0.1.

In our experiment, several runs were performed for 7×7 2D-Mesh. Simulations are performed under each traffic and for each congestion metric in varied PIR, to observe how congestion propagates in the network and to quantify the number of its occurrence. For each metric, different values of congestion occurrence are evaluated at different traffic scenarios and different injection rates.

In order to properly conduct the evaluation of our solution, calculated parameters were introduced, and monitored during our simulations i.e. **LU** that is the rate of usage and exploitation of the links in the network. This parameter gives us an indication about the number of resources involved in the distribution of the network traffic. A high rate of LU generates a good distribution load and viable network topology exploitation. To measure and to track the congestion evolution in the network, the number of congested nodes rate **C.N** and the congestion occurrence rate **C.O** were introduced.

To compare the network performance in implemented solutions, we also considered the average packet delay **Lat.** and the average throughput **Th.** as performance metrics, which are defined as follows:

$$Avg\ Packet\ Delay = \frac{1}{N} \sum_{i=1}^N latency_i \quad (5.1)$$

where N refers to the total number of received packets and $latency_i$ is the delay of the i -th packet.

$$Throughput = \frac{Total\ received\ flits}{number\ of\ nodes \times Total\ cycles} \quad (5.2)$$

where *Total received flits* represent the total number of received flits in network, number of nodes is the number of network nodes, and total cycles is the simulation length in clock cycles.

The average throughput **Th** is usually represented as a measured rate by calculating the number of successfully delivered Flits during the simulation period. This rate is affected by network congestion and packet loss, in other word, reducing congestion and its occurrence in the network nodes is translated by a high throughput rate and a good traffic fluidity.

TABLE 5.2
Simulation results for simulation time = 300 clock cycles

a) Results for Random and Uniform traffic, for PIR=0.3

Metric	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Available_VCs	0.25	0.94	44%	92%	17%	0.24	0.90	38%	92%	9%
Crossbar Demand	0.32	1.85	50%	92%	15%	0.29	1.47	43%	92%	8%
Buffer_Occupancy	0.21	0.95	39%	92%	17%	0.20	0.85	32%	92%	9%
OurMetric	0.38	0.86	59%	73%	14%	0.38	0.81	59%	73%	8%

b) Results for Random and Uniform traffic, for PIR=0.5

Metric	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Availavle_VCs	0.27	1.03	47%	92%	17%	0.26	0.91	43%	92%	10%
Crossbar Demand	0.28	1.25	45%	90%	16%	0.25	1.39	37%	90%	9%
Buffer_Occupancy	0.23	0.89	42%	92%	17%	0.22	0.81	36%	92%	10%
OurMetric	0.38	0.79	59%	73%	14%	0.38	0.77	59%	73%	8%

c) Results for Random and Uniform traffic, for PIR=1.0

Metric	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Availavle_VCs	0.21	0.75	39%	92%	17%	0.20	0.64	32%	92%	10%
Crossbar Demand	0.29	1.27	47%	90%	15%	0.24	1.21	38%	90%	9%
Buffer_Occupancy	0.20	0.77	37%	92%	17%	0.20	0.66	31%	92%	10%
OurMetric	0.38	0.90	59%	73%	14%	0.38	0.86	59%	73%	8%

Th: Average Throughput, **L.U:** Average Links Usage rate, **Lat:** Average Latency value divided per 10 **C.N:** Average Congested Nodes rate, **C.O:** Average Congestion Occurrence rate

In our paper, we aim to maximize **L.U** and reduce **C.O** and **C.N** while reducing the congestion negative effect and improving further the network performance with respect to throughput (**Th**) and latency (**Lat**).

A performance comparison between our congestion metric and the three standard metrics is performed. The simulation results for implemented strategies are summarized in Tables 5.2 and 5.3. Under each metric, PIR and traffic scenario, the five parameters are depicted.

By analyzing the results of the tables, it is quite clear that our metric solution achieved very good results compared to other metrics results. The positive outcome of our studies is due to better traffic distribution and optimal use of the network resources throughout our process. In this framework, our results proved to be more efficient in terms of the overall performance. As a result, we managed to achieve a better traffic fluidity including lower latency and a high rate of successful data delivery towards different network nodes.

In order to confirm this significant difference in the results; As it is shown, Table 5.4 clearly demonstrates the difference in performance between the four solutions. The traffic flow got far better in our solution, simply because of the existing gap in the network traffic load and the number of successfully delivered flits. What is also interesting in our approach is that despite all this doubled data load in the network, the average latency is improved and remained better than those of the other solutions.

The normalized load was introduced as parameter for evaluation. It is defined as the rate of the distribution

TABLE 5.3
Simulation results for simulation time = 500 clock cycles

a) Results for Random and Uniform traffic, for PIR=0.3

Metric	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Available_VCs	0.22	1.22	42%	98%	16%	0.21	1.12	35%	98%	29%
Crossbar Demand	0.25	1.42	41%	90%	15%	0.23	1.34	35%	98%	26%
Buffer_Occupancy	0.20	1.21	37%	98%	16%	0.18	1.11	30%	98%	29%
OurMetric	0.31	0.96	52%	82%	15%	0.31	0.95	51%	82%	26%

b) Results for Random and Uniform traffic, for PIR=0.5

Metric	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Available_VCs	0.22	1.14	42%	98%	16%	0.21	1.04	35%	98%	29%
Crossbar Demand	0.24	1.43	41%	90%	15%	0.23	1.33	35%	90%	26%
Buffer_Occupancy	0.20	1.1	37%	98%	16%	0.18	1.02	31%	98%	30%
OurMetric	0.31	0.93	52%	82%	15%	0.31	0.92	51%	82%	26%

c) Results for Random and Uniform traffic, for PIR=1.0

Metric	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Availavle_VCs	0.22	1.04	40%	98%	17%	0.20	0.98	34%	98%	30%
Crossbar Demand	0.24	1.32	41%	90%	15%	0.23	1.23	34%	90%	27%
Buffer_Occupancy	0.19	1.01	36%	98%	17%	0.18	0.91	30%	98%	30%
OurMetric	0.31	1.00	52%	82%	15%	0.31	0.92	51%	82%	26%

TABLE 5.4
Traffic Load and Latency results for Random traffic, PIR= 1.0 and Simulation time = 1000 clock cycles

Metric	Sent Flits	Received Flits	Latency
Available VCs	11820	9905	8,60
Crossbar Demand	11796	10518	10,70
Buffer Occupancy	10580	8646	7,90
Our Metric	16360	15424	6,90

of all the packets that have circulated in the network over the number of buffers existing in this network [25].

We compare the normalized load for the four solutions as illustrated in Figure 5.1. The load is almost identical in the case of reduced traffic. The higher the load, the greater the difference between the normalized load. The load results in our solution explain the gap of results and performance and confirm the uniform distribution of traffic across all network resources.

As shown in Figs. (5.2), (5.3) and (5.4), our approach significantly reduces the number of occurrences of congestion. It has better average latency and better performances, more particularly, in terms of network resource utilization and load balancing.

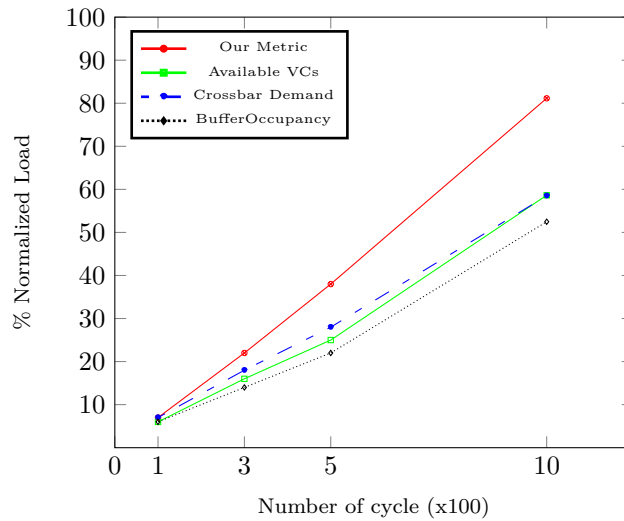


FIG. 5.1. Network load comparison

TABLE 5.5
Our Approach performances improving rates

Simulation Cycles	Random Traffic					Uniform Traffic				
	Th	Lat	L.U	C.N	C.O	Th	Lat	L.U	C.N	C.O
Nbr_Cycle = 300	24%-29%	21%	20%-30%	10%	7%	18%-32%	11%	20%-30%	20%	22%
Nbr_Cycle = 500	38%	15%-20%	44%-49%	10%	11%	31%-57%	5%-13	38%-57%	18%	26%
Nbr_Cycle = 1000	38%-86%	12%	44%-98%	20%	29%	45%-48%	1%-14%	62%-67%	11%	15%

Our approach gives advantages even in the case of the uniform traffic cf. Figs 5.2.b, 5.3.b and 5.4.b of which the majority of the packets are supposed to cross small distances to reach their destinations and few packets are supposed to send over longer distances in the NoC.

The benefit of our strategy appeared clearly in the random traffic scenarios, as it is illustrated in Figs. 5.2.a, 5.3.a and 5.4.a. that's where the performance gaps between the strategies are huge.

The experimental results summarized in Table 5.5 show that performances are dramatically improved by our congestion control mechanism and this for both types of traffic.

Compared to the three-standard metrics, our metric shows an improvement rate up to 86% in throughput, up to 21% in latency, up to 98% in use of network resources, of up to -20% in number of congested nodes and up to -29% in terms of reduction of occurrence of congestion in the network, those for the uniform and random traffic and for different configurations of simulations. The improvement rate of our congestion-control-based approach has reached its top for random traffic, PIR=1.0 and simulation time = 1000 clock cycles.

Compared to a existing congestion-based strategies and metrics, our proposed congestion control mechanism requires additional arithmetic and negligible extra cost, but offers better performance for both types of traffic: uniform and random, on 2D NoC mesh platforms.

6. Conclusion. In this paper we carried out studies in order to reduce the occurrence and effect of congestion in NoC. The graphical aspect and the step by step debug of the NoC behavior in simulator [23], has motivated us to propose a new control mechanism to overcome congestion and to compensate the performance degradation caused by the traditional indetermination of congestion. An efficient congestion metric is proposed while designing the adaptive XY routing algorithm in so as to overcome congestion and improve resource



FIG. 5.2. Average performances under implemented metrics, for simulation time =1000 clock cycles and PIR = 0.3.



FIG. 5.3. Average performances under implemented metrics, for simulation time =1000 clock cycles and PIR = 0.5.



FIG. 5.4. Average performances under implemented metrics, for simulation time =1000 clock cycles and PIR = 1.0.

Th.: Average Throughput, **C.O.:** Average Congestion Occurrence rate, **L.U:** Average Links Usage rate, **Lat.:** Average Latency value divided per 10 and **C.N:** Average Congested Nodes rate

utilization through load balancing.

In addition, experiments show that our newly introduced method has higher performances, besides our results have reached an average rate of improvement ranging up to 40%, compared to past techniques.

In the short term, we intend to study the impact of this novel metric on other routing algorithms and other selection strategies.

Acknowledgments. This work has been supported in part by the PHC TASSILI 14MDU917 project of the University of Oran 1, and by IRCICA, Université de Lille, USR 3380, F-59650 Villeneuve d’Ascq, France.

REFERENCES

- [1] G. ASCIA, V. CATANIA, M. PALESI, AND D. PATTI, *Implementation and Analysis of a New Selection Strategy for Adaptive Routing in Networks-on-Chip*, IEEE Transactions on Computers, 57 (2008), pp. 809–820.
- [2] M. K. BENHAOUA, A. SINGH, A. E. H. BENYAMINA, AND P. BOULET, *DynMapNoCSIM : A Dynamic Mapping SIMULATOR for Network on Chip based MPSoC*, Journal of Digital Information Management, 13 (2015), pp. 45–54.
- [3] M. K. BENHAOUA AND A. K. SINGH, *Dynamic Communications Mapping in Multi-tasks NoC-based Heterogeneous MPSoCs Platform*, Int. J. High Perform. Syst. Archit., 5 (2015), pp. 240–251.
- [4] L. BENINI AND G. DE MICHELI, *Networks on chips: a new SoC paradigm*, Computer, 35 (2002), pp. 70–78.
- [5] E. J. CHANG, H. K. HSIN, C. H. CHAO, S. Y. LIN, AND A. Y. WU, *Regional ACO-Based Cascaded Adaptive Routing for Traffic Balancing in Mesh-Based Network-on-Chip Systems*, IEEE Transactions on Computers, 64 (2015), pp. 868–875.
- [6] E. J. CHANG, H. K. HSIN, S. Y. LIN, AND A. Y. WU, *Path-Congestion-Aware Adaptive Routing With a Contention Prediction Scheme for Network-on-Chip Systems*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 33 (2014), pp. 113–126.
- [7] A. DANA AND N. SALEHI, *Congestion Aware Routing Algorithm for Mesh Network-on-chip Platform*, Indian Journal of Science and Technology, 5 (2012), pp. 2822–2830.
- [8] M. EBRAHIMI, M. DANESHTALAB, P. LILJEBERG, J. PLOSILA, AND H. TENHUNEN, *Agent-based on-chip network using efficient selection method*, in 2011 IEEE/IFIP 19th International Conference on VLSI and System-on-Chip, Oct. 2011, pp. 284–289.
- [9] P. GRATZ, B. GROT, AND S. W. KECKLER, *Regional congestion awareness for load balance in networks-on-chip*, in 2008 IEEE 14th International Symposium on High Performance Computer Architecture, Feb. 2008, pp. 203–214.
- [10] B. GROT, J. HESTNESS, S. W. KECKLER, AND O. MUTLU, *Express Cube Topologies for on-Chip Interconnects*, in 2009 IEEE 15th International Symposium on High Performance Computer Architecture, Feb. 2009, pp. 163–174.
- [11] N. GUPTA, A. SHARMA, V. LAXMI, M. S. GAUR, M. ZWOLINSKI, AND R. BISHNOI, *nLBDR: generic congestion handling routing implementation for two-dimensional mesh network-on-chip*, IET Computers Digital Techniques, 10 (2016), pp. 226–232.
- [12] Z. HAN, M. C. MEYER, X. JIANG, AND T. WATANABE, *Low-Cost Congestion Detection Mechanism for Networks-on-Chip*, in 2019 IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc), Oct. 2019, pp. 157–163.
- [13] H. K. HSIN, E. J. CHANG, C. H. CHAO, AND A. Y. WU, *Regional ACO-based routing for load-balancing in NoC systems*, in 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC), Dec. 2010, pp. 370–376.
- [14] N. JIANG, D. U. BECKER, G. MICHELOGIANNAKIS, AND W. J. DALLY, *Network congestion avoidance through Speculative Reservation*, in IEEE International Symposium on High-Performance Comp Architecture, Feb. 2012, pp. 1–12.
- [15] K. JIN, C. LI, D. DONG, AND B. FU, *HARE: History-Aware Adaptive Routing Algorithm for Endpoint Congestion in Networks-on-Chip*, International Journal of Parallel Programming, 47 (2019), pp. 433–450.
- [16] G. N. KHAN AND S. CHUI, *Congestion Aware Routing for On-Chip Communication in NoC Systems*, in Complex, Intelligent, and Software Intensive Systems, Advances in Intelligent Systems and Computing, Springer, Cham, July 2017, pp. 547–556.
- [17] J. KIM, J. BALFOUR, AND W. DALLY, *Flattened Butterfly Topology for On-Chip Networks*, in 40th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 2007), Dec. 2007, pp. 172–182.
- [18] M. LI, Q.-A. ZENG, AND W.-B. JONE, *DyXY - a proximity congestion-aware deadlock-free dynamic routing method for network on chip*, in 2006 43rd ACM/IEEE Design Automation Conference, 2006, pp. 849–852.
- [19] P. LOTFI-KAMRAN, *Per-packet global congestion estimation for fast packet delivery in networks-on-chip*, The Journal of Supercomputing, 71 (2015), pp. 3419–3439.
- [20] P. LOTFI-KAMRAN, M. DANESHTALAB, C. LUCAS, AND Z. NAVABI, *BARP-a Dynamic Routing Protocol for Balanced Distribution of Traffic in NoCs*, in Proceedings of the Conference on Design, Automation and Test in Europe, DATE '08, New York, NY, USA, 2008, ACM, pp. 1408–1413.
- [21] P. LOTFI-KAMRAN, A. M. RAHMANI, M. DANESHTALAB, A. AFZALI-KUSHA, AND Z. NAVABI, *EDXY A low cost congestion-aware routing algorithm for network-on-chips*, Journal of Systems Architecture, 56 (2010), pp. 256–264.
- [22] S. MA, N. E. JERGER, AND Z. WANG, *DBAR: An efficient routing algorithm to support multiple concurrent applications in networks-on-chip*, in 2011 38th Annual International Symposium on Computer Architecture (ISCA), June 2011, pp. 413–424.
- [23] M. A. MEGHABBER, A. AROUI, L. LOUKIL, A. E. H. BENYAMINA, K. BENHAOUA, AND T. DJERADI, *A flexible network on-chip router for data-flow monitoring*, in 2017 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), Oct. 2017, pp. 1–6.
- [24] L. NI AND P. MCKINLEY, *A survey of wormhole routing techniques in direct networks*, Computer, 26 (1993), pp. 62–76. Conference Name: Computer.
- [25] M. NICKRAY, M. DEHYADGARI, AND A. AFZALI-KUSHA, *Adaptive routing using context-aware agents for networks on chips*, in 2009 4th International Design and Test Workshop (IDT), Nov. 2009, pp. 1–6.
- [26] G. NYCHIS, C. FALLIN, T. MOSCIBRODA, AND O. MUTLU, *Next Generation On-chip Networks: What Kind of Congestion Control Do We Need?*, in Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks, Hotnets-IX, New York, NY, USA, 2010, ACM, pp. 12:1–12:6.
- [27] G. P. NYCHIS, C. FALLIN, T. MOSCIBRODA, O. MUTLU, AND S. SESHAN, *On-chip Networks from a Networking Perspective: Congestion and Scalability in Many-core Interconnects*, in Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, SIGCOMM '12, New York, NY, USA, 2012, ACM, pp. 407–418.

- [28] M. RAMAKRISHNA, P. V. GRATZ, AND A. SPRINTSON, *GCA: Global congestion awareness for load balance in Networks-on-Chip*, in 2013 Seventh IEEE/ACM International Symposium on Networks-on-Chip (NoCS), Apr. 2013, pp. 1–8.
- [29] R. S. RAMANUJAM AND B. LIN, *Destination-based adaptive routing on 2D mesh networks*, in 2010 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS), Oct. 2010, pp. 1–12.
- [30] M. S. TALEBI, F. JAFARI, A. KHONSARI, AND M. H. YAGHMAE, *A Novel Congestion Control Scheme for Elastic Flows in Network-on-Chip Based on Sum-Rate Optimization*, in Computational Science and Its Applications – ICCSA 2007, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Aug. 2007, pp. 398–409.
- [31] M. TANG AND X. LIN, *Quarter Load Threshold (QLT) flow control for wormhole switching in mesh-based Network-on-Chip*, Journal of Systems Architecture, 56 (2010), pp. 452–462.
- [32] M. TANG, X. LIN, AND M. PALESI, *Local Congestion Avoidance in Network-on-Chip*, IEEE Transactions on Parallel and Distributed Systems, 27 (2016), pp. 2062–2073.
- [33] H. C. TOUATI AND F. BOUTEKKOUK, *FACARS: A novel fully adaptive congestion aware routing scheme for network on chip*, in 2018 7th Mediterranean Conference on Embedded Computing (MECO), June 2018, pp. 1–6.

Edited by: Dana Petcu

Received: Jan 8, 2020

Accepted: May 7, 2020



A DISTRIBUTED NEURAL NETWORK TRAINING METHOD BASED ON HYBRID GRADIENT COMPUTING

ZHEN LU^{*}, MENG LU[†] AND YAN LIANG[‡]

Abstract. The application of deep learning in industry often needs to train large-scale neural networks and use large-scale data sets. However, larger networks and larger data sets lead to longer training time, which hinders the research of algorithms and the progress of actual engineering development. Data-parallel distributed training is a commonly used solution, but it is still in the stage of technical exploration. In this paper, we study how to improve the training accuracy and speed of distributed training, and propose a distributed training strategy based on hybrid gradient computing. Specifically, in the gradient descent stage, we propose a hybrid method, which combines a new warmup scheme with the linear-scaling stochastic gradient descent (SGD) algorithm to effectively improve the training accuracy and convergence rate. At the same time, we adopt the mixed precision gradient computing. In the single-GPU gradient computing and inter-GPU gradient synchronization, we use the mixed numerical precision of single precision(FP32) and half precision(FP16), which not only improves the training speed of single-GPU, but also improves the speed of inter-GPU communication. Through the integration of various training strategies and system engineering implementation, we finished ResNet-50 training in 20 minutes on a cluster of 24 V100 GPUs, with 75.6% Top-1 accuracy, and 97.5% GPU scaling efficiency. In addition, this paper proposes a new criterion for the evaluation of the distributed training efficiency, that is, the actual average single-GPU training time, which can evaluate the improvement of training methods in a more reasonable manner than just the improved performance due to the increased number of GPUs. In terms of this criterion, our method outperforms those existing methods.

Key words: Deep learning, gradient descent, mixed precision computing, distributed training

AMS subject classifications. 97R40, 97P50, 97P10

1. Introduction. In recent years, deep learning technology has a far-reaching impact on the fields such as computer vision [1, 2, 3, 4, 5], speech recognition [6, 7], and natural language processing [8, 9]. The rapid development of deep learning is not only due to the innovation of algorithms, but also benefited from the substantial improvement of computing power, which has become the core competitiveness of deep learning technology. In the field of computer vision, for example, several major technological breakthroughs are driven by larger neural networks and larger datasets. Neural networks have increased from 8 layers [1] and 22 layers [4] to hundreds of layers [5]. There are academic datasets of more than one million samples [10], and the industry has used datasets of billion samples [11]. Only with the support of computing power, researchers can complete such scale training tasks in a reasonable time. At present, the commonly used computing power promotion method in the industry is based on distributed data-parallel training [12, 13, 14, 15, 16]. Due to the limited computing resources of a single GPU, data is allocated to multiple GPUs, and the data of each GPU is combined to form a larger batch, that is, it has a larger batch size, which enables training across GPUs in an inter-computer manner. By training at the same time, the throughput of computation can be increased and the training speed can be improved. However, distributed data-parallel training is still an open research field, which mainly faces three problems.

Problem 1: Due to the optimization [12] and generalization [17, 18] difficulty of large batch training, the training accuracy will be lower than that of small batch. Gradient estimation on larger batch is more accurate, which allows a larger step size in the gradient descent, making the process of the optimization algorithm faster. However, as described in [19], when the batch size increases to 64K, ResNet-50 validation accuracy drops from 75.4% to 73.2%.

^{*}Guangzhou College of South China University of Technology, Guangzhou 510800, China (luzhenaaa@126.com).

[†]China mobile (Suzhou) software technology co., Suzhou 21500, China (lumeng@cmss.chinamobile.com).

[‡]Corresponding author, Alibaba Group, Guangzhou 510335, China(liangyan_709@163.com).

Problem 2: Distributed data-parallel training divides the gradient computing into multiple GPUs, and gradient need to be synchronized and aggregated. With the increase of the number of GPUs, the number of communications between multiple GPUs and the total amount of communication data increase, making it difficult to maintain the linear throughput scaling. Especially for the model with a large amount of parameters, there are too many gradient data to be synchronized, making the communication a bottleneck and affecting the training speed. If the GPUs are distributed on multiple machines, inter-machine communication will be more difficult than the intra-machine communication, which will result in lower scaling efficiency.

Problem 3: We define the convergence rate of training as the reciprocal of the number of epochs completed when converging. That is to say, the less epochs are used when convergence, the faster the convergence rate will be. Because the current evaluation criterion of distributed training speed is the total time consumed for fixed epochs (the default is 90 epochs), there are few researches and reports on convergence rate in the industry.

In this paper, an efficient distributed neural network training method is proposed for addressing the above three problems. Our contribution includes:

1. A hybrid gradient descent method is proposed, which is a new warmup scheme on the AdaBound algorithm [20] combined with the linear scaling SGD algorithm [12]. This method can effectively improve the convergence rate without losing the training accuracy, and fill the blank in the research area of improving the convergence rate in the distributed system.
2. In this paper, the hybrid gradient precision computing is used, and the FP16 numerical precision is used in single-GPU gradient computing and multi-GPU gradient synchronization. At the same time, many strategies to avoid numerical overflow are proposed. Without sacrificing the training accuracy, the training speed of single-GPU is improved, and the speed of multi-GPU synchronous aggregation is also improved because the communication volume is compressed.
3. We have constructed a real distributed training system by carefully integrating various training strategies, such as batch normalization (BN) [21], layer-wise adaptive rate scaling (LARS) [16] and so on. In our training system, we have accomplished the ResNet-50 training in 20 minutes, with the Top-1 accuracy being 75.6%, and the scaling efficiency more than 97%.

In order to compare with the existing distributed training methods in the industry, we propose a new performance evaluation criterion. At present, the commonly used evaluation criterion is the training time of training the classic ResNet-50 model, to run 90 epochs on imageNet dataset, and to achieve the Top-1 validation accuracy more than 75%. The prerequisite of this evaluation criterion is to achieve the specified accuracy within 90 epochs. We think that there are two shortcomings in this criterion. Firstly, the requirement of 90 epochs limits the real training speed. Due to the influence of convergence rate, the actual training speed and training time are as follows:

$$\begin{aligned} \text{Actual training speed} &= \text{convergence rate} \times \text{single-GPU speed} \times \text{total GPUs} \times \text{GPU scaling efficiency.} \\ \text{Actual training time} &= 1 / \text{actual training speed} \end{aligned}$$

Accordingly, DOWNBench [22] is used to evaluate the actual training time, which is the training time of the ResNet-50 model when the number of epochs is not limited, and the Top-5 verification accuracy is over 93%.

The second disadvantage is that due to the influence of the number of GPUs, the comparison of training time has become a competition of affordable GPU resources, which does not reflect the advantage of training methods. Some reports compare the GPU scaling efficiency, but this criterion only represents the relative speed of multi-GPU to single-GPU training, and does not represent the absolute speed. Therefore, this paper proposes a new criterion-actual average single-GPU training time: Actual average single-GPU training speed = convergence rate * single-GPU speed * GPU scaling efficiency Actual average single-GPU training time = 1 / actual single-GPU training speed.

The rest of this paper is organized as follows. We will discuss the existing work related to this paper in Section 2.1. The proposed training methods and engineering implementation details of the training system will be discussed in Section 2.2. Section 3 is about the experiments and result analysis. Section 4 concludes this paper.

2. Materials and Methods.

2.1. Related materials.

2.1.1. Gradient descent algorithm. At present, the optimization method for training deep neural network is the first-order optimization algorithm, that is, gradient descent method. The most commonly used one is the stochastic gradient descent method (SGD) [23], which is simple, and has good performance in many applications. However, SGD has a disadvantage, that is, it adjusts the gradient uniformly in all directions. This may cause the convergence rate to change when the training data is sparse. In recent years, many adaptive methods have been proposed to adjust the current gradient direction and magnitude through the direction and magnitude of the historical gradient. These methods include ADAM [24], ADAGRAD [25] and RMSprop [26]. Especially for ADAM, because of its fast convergence rate and stable effect, it is often used in the industry. However, the generalization ability of these adaptive methods is worse than SGD [27]. Recently, Luo et al. proposed a variant of ADAM [20], which is called ADABOUND. By limiting the learning rate in a dynamic upper and lower bound, we can smoothly transform ADAM to SGD. ADABOUND not only retains the fast convergence characteristic of ADAM, but also has the similar generalization ability as SGD.

The general process of gradient descent method is as follows [20]:

1. Calculate the gradient g_t of objective function with respect to current parameter x_t .
2. Calculate current gradient decay: $\eta_t = \alpha_t \cdot m_t / \sqrt{V_t}$.
3. Update parameters: $x_{t+1} = x_t - \alpha_t \cdot m_t / \sqrt{V_t}$.

where, t represents the iterative steps, α_t is the current learning rate of decay. For SGD, $m_t = g_t, V_t = I$. For ADAM, $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t, v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2, V_t = \text{diag}(v_t)$. For ADABOUND, the meanings of the symbols are the same as ADAM, with the value of $\alpha_t \cdot \sqrt{V_t}$ limited within the upper and lower boundaries. If $\alpha_t \cdot \sqrt{V_t}$ exceeds the boundaries, it will be truncated, and the upper and lower bounds will be adjusted dynamically. Finally, it will converge to the learning rate α of SGD, to finish the smooth transition from ADABOUND to SGD. β_1 and β_2 are the default hyperparameters.

2.1.2. Distributed training optimization method.

1) *Training accuracy optimization.* In the era of small batch training, there appears many training tricks for improving the training accuracy, which can still be used for large batch training. Ioffe et al. introduced the BatchNorm technology [21], which normalizes the features in the hidden layers to avoid the gradient vanishing. At present, it has become the standard training module for almost every deep learning framework. The initialization of weights will also affect the final result. Early, Gaussian distribution [1] is used for weight initialization. Later, Xavier initialization [28], KaiMing initialization [29] and other more advanced initialization technologies are proposed. In addition, weight decay is used for weight regularization by $l_1 - norm$, which can improve the generalization ability of the model.

In order to overcome the inherent optimization difficulty [12] and generalization problem of large batch training [17, 18], more specific training skills are required. Goyal et al. proposed an improved version of SGD [12], which has linear scaling on the learning rate. At the same time, the warmup scheme is set up before the linear scaling, that is, a small learning rate is used at the beginning and then it is switched back to the large learning rate when the training process is stable. Google [30] and Sony [15] adopt the strategy of gradually increasing the batch size to ensure the stability of the large batch training process. Because the difference between the weight gradient norms of each layer in neural network leads to the instability of training, [16] proposed a layer-wise adaptive learning rate scaling strategy called LARS, which adjusts learning rate of each layer by multiplying by $\|w\|_2 / \|\nabla w\|_2$, where w is the weight of each layer. LARS achieved excellent results in the super large batch training tasks.

2) *Training speed optimization.* The precondition of training speed optimization is that there is no loss of the training accuracy, so training speed optimization needs to cooperate with the above training accuracy optimization techniques. On this basis, optimization and improvement can be made in the aspects of single-GPU speed, convergence rate and multi-GPU communication. Akiba et al. [13] finished ResNet-50 training in 15 minutes through RMSprop warmup scheme and the adoption of BN without sliding average. ResNet-50 training has been completed in 6.6 minutes through mixed precision computing, tensor fusion [32], hierarchical and hybrid gradient collective communication by Jia et al. [31]. In addition, [12] and [15] have improved the

collective communication part, among which [12] has adopted recursive halving [33] and doubling algorithm [34] and [15] has adopted 2D-Torus all-reduce.

2.2. Methods in this paper.

2.2.1. Hybrid gradient descent. Our goal is to propose a gradient descent method, which can not only ensure fast convergence, but also maintain the advantages of linear scaling SGD in large batch training. Inspired by ADABOUND, we propose an improved version, called linear scaling ADABOUND, which specifies the following restrictions:

– upper bound

$$B_u = lr \cdot N + \frac{1}{(1 - \beta_2) \cdot t} - \frac{1}{(1 - \beta_2) \cdot T}$$

– lower bound

$$B_l = \frac{t}{T} \cdot lr \cdot N$$

$$\alpha \cdot \sqrt{V_t} = \text{Clip}(\alpha \cdot \sqrt{V_t}, B_l, B_u)$$

where, lr represents the learning rate of SGD in single-GPU batch-size. N is the number of GPUs, which meanwhile serves as the linear scaling rate. By comparing with the original ADABOUND, we scale the learning rate to $lr \cdot N$. t is the current epoch. T is the preset ADABOUND maximal epoch. α is the initial learning rate. Clip indicates that the value is limited to the upper and lower bounds, with the exceeding part being truncated. Therefore, when t increases from 0 to T , B_u decreases from $+\infty$ to $lr \cdot N$, while B_l increases from 0 to $lr \cdot N$. When $B_u = +\infty, B_l = 0$, the gradient descent method is ADAM; When $B_u = B_l = lr \cdot N$, the gradient descent method is linear scaling SGD. By this, inside each epoch, the smooth transition from ADAM to linear scaling SGD is completed.

In order to further ensure the stability and generalization of the training, we only take the linear scaling ADABOUND as the warmup scheme. After T epochs, the learning rate decay process of SGD starts from the learning rate of $lr \times N$ until the training convergence. In addition to the above, we adopt LARS in SGD stage to adaptively adjust the learning rate of each layer.

To sum up, the complete ADABOUND warmup + linear scaling SGD method is as follows:

For epoch = 1 : T

1. Calculate the gradient g_t of the objective function with respect to the current parameter x_t .
2. Average the gradient by sliding:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$

- 3.

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2, \quad V_t = \text{diag}(v_t)$$

4. Scale and limit the learning rate to the set range:

$$\alpha / \sqrt{V_t} = \text{Clip}(\alpha / \sqrt{V_t}, B_l, B_u)$$

5. Calculate the current gradient decay:

$$\eta_t = \alpha \cdot m_t / \sqrt{V_t}$$

6. Update parameters:

$$x_{t+1} = x_t - \alpha \cdot m_t / \sqrt{V_t}$$

For epoch = $T + 1 : 90$

1. Calculate the gradient g_t of the objective function with respect to the current parameter x_t .
2. Learning rate decreases exponentially:

$$\alpha_t = lr \cdot N \cdot \kappa^{epoch}$$

3. Layer-wise adaptive learning rate scaling:

$$\alpha_t = LARS(\alpha_t)$$

4. Update parameters:

$$x_{t+1} = x_t - \alpha_t \cdot g_t$$

where κ is the default hyperparameter.

The experiments in Section 3 show that compared with using only the ADABOUND or the linear scaling SGD, ADABOUND warmup + linear scaling SGD will produce higher training accuracy.

2.2.2. Gradient computing with mixed precision. The gradient computing with mixed precision scheme in this paper is similar to that in [31]. In the process of forward and back propagation, FP16 precision is used for computing; in the weight updating process, FP32 precision is used for updating. This paper makes the following improvements to this scheme:

1. Since the numerical representation range of FP16 is far smaller than FP32, the conversion from FP32 to FP16 may cause serious optimization problems such as too small weight and gradient being truncated to 0, resulting in gradient vanishing. After the forward propagation is completed, we multiply the loss by a scaling factor, and then continue with backpropagation, which is equivalent to the same scaling when calculating the gradient. This improvement avoids too small gradient value. Meanwhile, it is also easy for implementation.
2. In the process of forward propagation and backpropagation, the operator f with $|f(x)| \gg x$ should be avoided to be calculated under FP16, because the computing results may overflow the representation range of FP16, which makes the numerical accuracy deviate greatly. The specific operators include: exp, square, log and cross entropy loss.
3. In the collective communication process, the gradient of FP16 instead of FP32 is transmitted during gradient aggregation, which reduces the traffic and saves the bandwidth.

2.2.3. Strategy fusion and system engineering implementation. This section lists the problems that need to be noticed in the process of completing the actual distributed training system, as well as the solutions we proposed.

First, based on the improvement proposed earlier in this paper, we adopt a fusion scheme of variety of training strategies.

1) *Multi-GPU BN*. Compared with the original BN, the following constructions need to be applied in the mean and variance statistics of BN on multi-GPU:

In each iteration, the batch size allocated to each GPU remains unchanged. For neural networks with BN layer, forward loss computing depends on the mean and variance statistics of each single-GPU batch. If the single-GPU batch size changes, the loss function changes, invalidating the premise of linear scaling [12].

In each iteration, the mean and variance of BN are calculated in the batch of each GPU respectively, and the statistics between multiple GPUs are not carried out. On the one hand, the communication bandwidth between multiple GPUs can be saved. On the other hand, the loss computing of each GPU is independent and does not affect each other, meeting the premise assumption of the linear scaling.

The original BN mean and variance calculating method is moving average of the statistical value of each iteration. When the batch is large enough, the computing result of the moving average is relatively inaccurate [13]. Therefore, we only consider the last iteration, and use collective communication to calculate the statistical average over all GPUs. Because large-scale accumulation operation is required, in order to avoid the number value range overflow, we employ FP32 precision.

2) *Weight initialization.* This paper adopts the following initialization methods:

1. For the convolution layers, the KaiMing Gaussian initialization [28] is applied, that is, the initial weights of convolution layers follow Gaussian distribution $N(0, std)$.

$$std = \sqrt{\frac{2}{fan_{in}}}$$

where, fan_{in} is the input dimension of the three-dimensional convolution kernel, that is, the length · width · height of the convolution kernel.

2. For the residual block in ResNet, the last layer is BN layer, which performs a linear transformation: $\gamma\hat{x} + \beta$, where the input of BN layer is denoted by \hat{x} . The scale factor γ of $\gamma\hat{x} + \beta$ is initialized to 0. Given the input x , assume $block(x)$ is the output of the last layer in the residual block. With combination of the output of the last layer and the input of the residual block, the final output will be $block(x)+x$. If the scale factor γ of the last BN layer is initialized to 0, then $block(x)=0$, and only the input is returned finally, which is equivalent to changing ResNet into fewer layers, so that it is easier to train in the initial stage.

3) *Multi-machine multi-GPU communication.* Because the multi-machine and multi-GPU communication strategy is not the focus of this paper, the collective communication algorithm we adopted is the ring all-reduce [34] based on NCCL. We broadcast from a master node to ensure that the initial weights between multi-machine and multi-GPU are consistent, and distribute data evenly, so as to keep the load balancing of multiple GPUs.

4) *Single-GPU computing.* In order to improve the GPU scaling efficiency, we let the single GPU complete more computing tasks, and increase the proportion between single-GPU computing time and multi-GPU communication time.

The batch size on a single GPU is made as large as possible until it approaches the upper limit of the GPU memory.

3. Experiments and result analysis.

3.1. Experiment setup.

Hardware. Two training clusters are used in these experiments:

1. Cluster 1 includes 3 machines, with each machine equipped with 8 Nvidia Tesla V100 GPUs with 16G GPU memory, and the GPUs are interconnected through NVlink. 1 NVME 3.2T SSD hard disk is used as the storage. The Connectx-4 Lx EN Cards 40G is employed for the inter-machine communication, with RoCE (RDMA over Converted Ethernet) adopted as the communication mode, which is an RDMA communication mode used under Ethernet, and can be directly connected through memory for quick access to the remote data. We further use GPUDirect RDMA, which allows the direct connection of GPU memory between multiple GPUs to further improve the inter-GPU communication speed.
2. Cluster 2 consists of six machines, each of which is equipped with four Nvidia Tesla P40 GPUs with 24G GPU memory. The GPUs are interconnected by PCIe. One NVME 3.2T SSD hard disk is still employed as the storage. The ConnectX-4 Lx EN Cards 40G is used for inter-machine communication and the communication mode is RoCE. Due to the limitation of GPU type, GPUDirect RDMA cannot be applied in Cluster 2, and the mixed precision cannot be applied neither.

Software. We use horovod [31] as the training framework, which provides APIs to make it easier to write a distributed training script than previous training frameworks such as Distributed TensorFlow. NVIDIA’s collective communication library NCCL is used for multi-machine and multi-GPU communication.

Data. The training data is ImageNet database [10], with a total of 1.28 million training pictures and 50000 verification pictures, with the number of categories being 1000. The data augmentation technology in [1] is used in the experiment.

Model. Based on the existing relevant technical reports and the authoritative distributed training test standard DAWNBench [22], ResNet-50 [5] is used as our training neural network model. The basic information of the model is shown in Table 3.1.

TABLE 3.1
Basic information of ResNet-50 model.

Input Size	Parameter Size	FLOPs	Top-1 Accuracy	Top-5 Accuracy
224*224	25M	4G	75%	93%

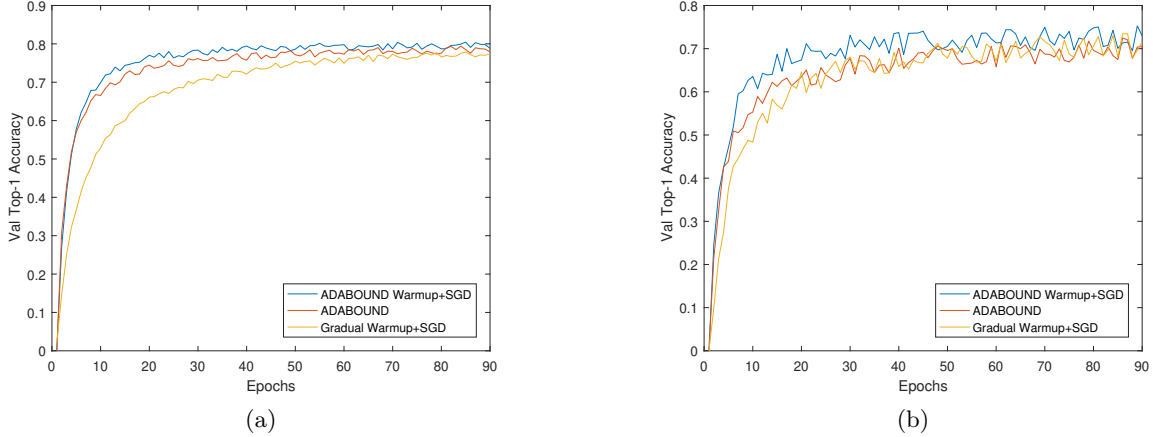


FIG. 3.1. Top-1 training and validation accuracy curves with different gradient descent algorithms.

Experimental details. Under the FP32, the batch of size 128 can be loaded on a single V100 GPU, and the total batch-size of cluster 1 is 3072.

Under the mixed precision, we doubled the batch size of V100, because compared with FP32, FP16 only occupies half of the FP32 memory, and the total batch size of cluster 1 is 6144.

Under the FP32, a single P40 GPU can be loaded with batch of size 128, and the total batch size of cluster 2 is 3072. Hyperparameters applied in the experiment: the number of iteration epochs of ADABOUND warmup stage is $T = 10$, which is selected according to the experiments shown in section 3.2.1. We directly apply the default hyperparameters for ADABOUND: $\beta_1 = 0.9, \beta_2 = 0.999$. All the other hyperparameters are the same as linear scaling [12]: the single-GPU learning rate $lr = 0.1$. The learning rate of SGD decays exponentially: $lr \cdot \kappa^{epoch}$, where, $\kappa = 0.9$. In SGD, weight decay [1] is adopted, with weight decay is 0.0001.

3.2. Experimental results.

3.2.1. Comparative experiment of gradient descent methods. Listed in this section are the experimental results on cluster 1. The results on cluster 2 are similar to those on cluster 1, therefore will not be discussed again. The gradient descent methods in the comparison include the proposed ADABOUND warmup + SGD, ADABOUND, and gradual warmup + SGD [12] proposed previously. The Top-1 training and validation accuracy curve are shown in Fig. 3.1. This shows that the ADABOUND warmup + SGD method is the most effective, with the convergence rate similar as that of the ADABOUND, and faster than warmup +SGD method. Convergence can be achieved within 50 epochs only. At the same time, on the validation set, the accuracy is the best, Top-1 accuracy = 75.6%, which is about 0.5% higher than that of the second best one, that is, the gradual warmup +SGD method. Because of the fast convergence rate, the accuracy is over 75.0% in the 32nd epoch.

We also tried different epoch T for warmup stopping. The comparison result is shown in Fig. 3.2, which shows that T has little effect on the result. With the increase of T , the convergence rate will be slightly improved, but the training accuracy will also be slightly reduced.

3.2.2. Comparative experiment of gradient computing precisions. Two aspects are compared in this experiment: the training accuracy of mixed precision computing and that of FP32 precision computing, as well as the training speed of mixed precision computing and that of FP32 precision computing. The experi-

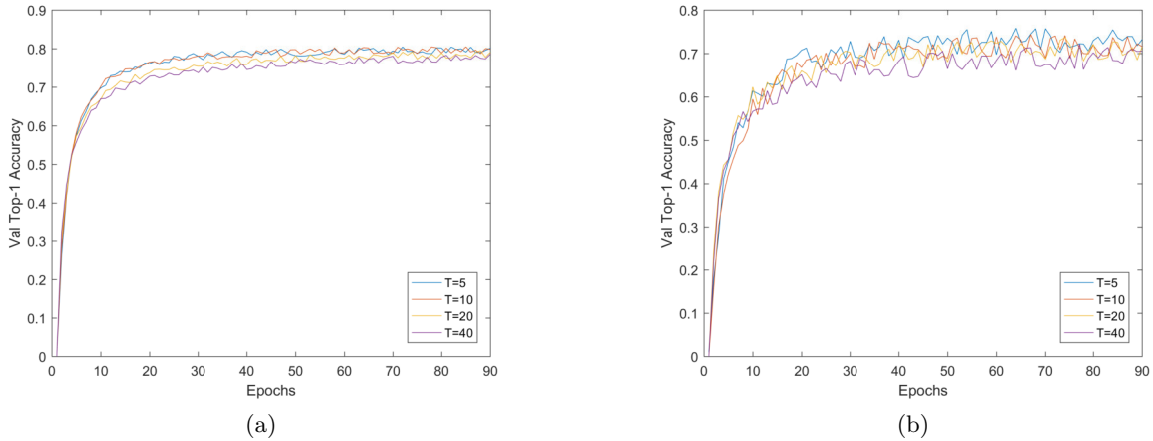


FIG. 3.2. Top-1 training and validation accuracy curves with different warm-up stopping epoch.

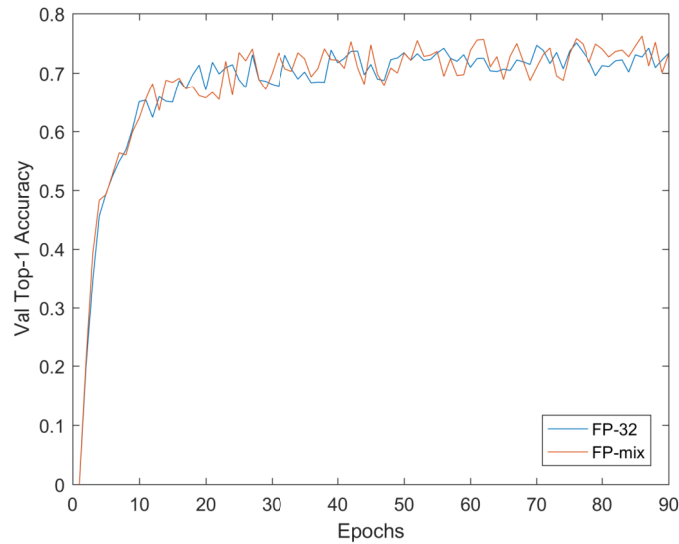


FIG. 3.3. Top-1 validation accuracy curves with different computing precision.

ment uses the gradient descent method of ADABOUND warmup + SGD, with other training strategies being consistent with Section 2.2.3.

1) *Training accuracy.* Because the P40 GPU of cluster 2 does not support FP16 computing perfectly enough, this experiment is only conducted on cluster 1. Figure 3.3 shows the accuracy results of validation set under two precision strategies. The training curve of mixed precision and that of FP32 precision almost coincides, except that the curve of mixed precision is more oscillating. Therefore, by using our methods to avoid the problem of too small numerical range, the mixed precision gradient computing has no effect on the training accuracy.

2) *Training speed.* First, we test the training speed of single GPU on cluster 1. As shown in Fig. 3.4, the training speed of mixed precision on V100 is much higher than that of FP32. Specifically, when batch size = 128, the training speed of mixed precision is close to twice that of FP32 (Table 3.2). With the maximum batch size 256 under mixed precision, the training speed is 2.03 times faster than that of FP32.

We continue to test the GPU scaling efficiency with mixed precision.

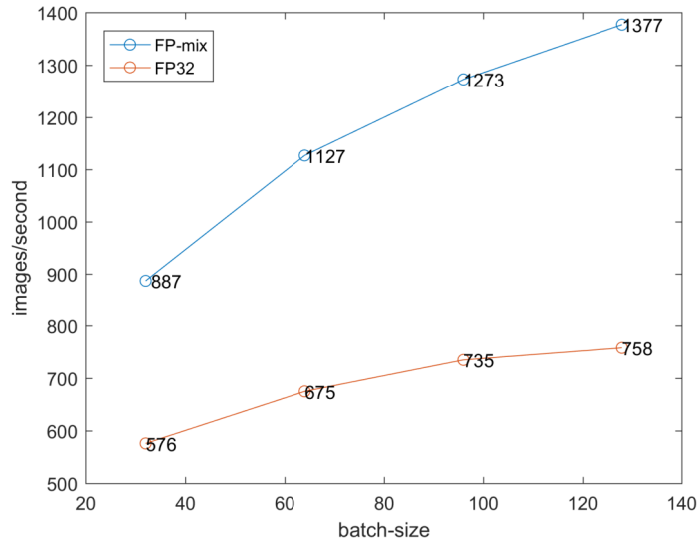


FIG. 3.4. Training throughput per V100 GPU with different computing precision and batch-size.

TABLE 3.2

Training throughput per V100 GPU with different computing precision and batch-size 128/256.

Data Type	Single-GPU Batch-size	Images per Second
FP-mix	128	1377
FP-mix	256	1542
FP32	128	758

The GPU scaling efficiency can be measured by the ratio of actual speed to ideal speed, and can also be measured as the ratio of ideal time-consuming to actual time-consuming. The computing method is as follows:

$$\begin{aligned} \text{GPU Scaling Efficiency} &= \text{Actual Speed} / \text{Ideal Speed} = \text{Ideal Time} / \text{Actual Time} = \\ &= 1 / (1 + \text{per GPU Communication Time} / \text{per GPU Computing Time}), \end{aligned}$$

where

$$\begin{aligned} \text{ideal time} &= \text{per GPU computing time} / \text{GPU number}, \\ \text{actual time} &= (\text{per GPU computing time} + \text{communication time}) / \text{GPU number}. \end{aligned}$$

The computing of mixed precision gradient can reduce the computing time and the communication time of single GPU at the same time. Therefore, when the reduced computing time of single GPU is less than the reduced communication time of single GPU, the scaling efficiency will be improved. On the other hand, when the reduced computing time of single GPU is greater than the reduced communication time of single GPU, the scaling efficiency will be weakened. See Fig 3.5 for the detailed experimental results. When the 24 V100 GPUs of cluster 1 are in full use, the GPU scaling efficiency with mixed precision = actual speed/ideal speed = $36082/37008=97.5\%$. The GPU scaling efficiency of FP32 accuracy = actual speed / ideal speed = $17773/18192=97.7\%$. The mixed precision is slightly lower than that of FP32. However, the speed of single GPU with mixed precision is much faster than that of single GPU with FP32 (see Table 3.2, 1542 images/s:758 images/s), which makes up for the slight deficiency of the GPU scaling efficiency. The total processing speed of 24 GPUs with mixed precision is also much higher than that of FP32, which is 2.03 times (36082 images/s: 17773 images/s). And under different GPU numbers, the linear scaling efficiency is achieved.

In addition, this paper uses the strategy of maximizing the batch size of a single GPU. If the batch size is less than 256 under mixed precision, not only the speed of a single GPU (Table 3.2), but also the GPU scaling efficiency will decrease (Table 3.3), thus affecting the total processing speed of multiple GPUs. Figure 3.6 shows

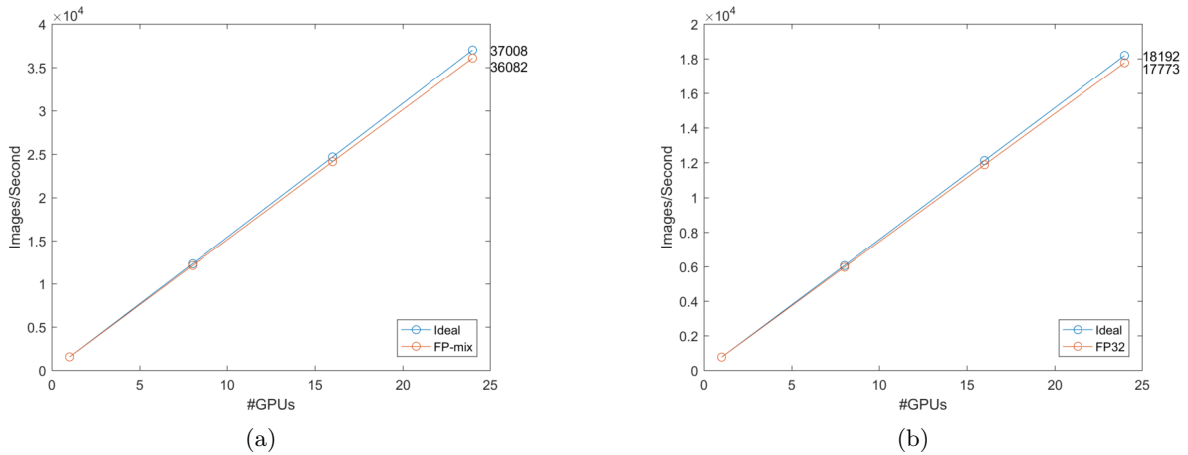


FIG. 3.5. Multi-V100-GPU training throughput with different computing precision.

TABLE 3.3
Multi-V100-GPU scaling efficiency with FP16 and different batch-size.

Batch-Size	64	128	192	256
GPU Scaling Efficiency	95.8%	96.4%	97.1%	97.5%

that the total processing speed and the GPU scaling efficiency of 24 GPUs will decrease with the decrease of the batch size.

Next, we test the GPU scaling efficiency improvement only brought by the FP16 collective communication of cluster 2 when the mixed precision computing is disabled. Under the FP32 precision gradient computing, the single-GPU processing speed of P40 is about 295 images/second. As shown in Fig. 3.7, when all 24 P40 GPUs of 6 machines in cluster 2 are in full use, the GPU scaling efficiency of FP16 communication = actual speed/ideal speed = $6903/7080=94\%$. The GPU scaling efficiency of FP32 accuracy communication = actual speed/ideal speed = $6456/7080=91.2\%$. Although the number of GPUs is the same as that of the cluster 1, due to the large number of machines, multi-GPU communication between different machines is more difficult than multi-GPU communication in the same machine, and the communication efficiency is lower. Moreover, it is unable to use GPUDirect RDMA for memory direct connection, so the GPU scaling efficiency of cluster 2 is inherently worse than that of cluster 1. However, through FP16 communication, the GPU scaling efficiency of cluster 2 is greatly improved compared with that of FP32 communication, and the linear scaling is almost achieved under different GPU numbers.

3.2.3. Experiment on the effectiveness of other training strategies. Based on hybrid gradient descent method and the mixed precision computing, we conduct the comparative experiments of the following influencing factors. The experiments are all completed in the cluster 1.

1) *Multi-GPU BN.* We compare the effects of several BN strategies on training accuracy and training speed. Table 3.4 shows that the Top-1 accuracy is 70.34%, when we randomly change the batch size of single GPU, do BN statistics cross GPUs, and do moving average for each iteration, the Top-1 accuracy is 70.34%.

By using the strategy of this paper, namely, fixed single-GPU batch size, independent BN statistics inside the GPU, and one-time statistics of the last iteration, the Top-1 accuracy can reach 75.6%, which is the best in the experiment.

In addition, cross-GPU BN statistics will have a great impact on communication, and the GPU scaling efficiency will be reduced from the highest value of 97.5% to 96.5%.

2) *Weight initialization.* We compare three common initialization methods of convolution weight: KaiMing initialization, Xavier initialization and $N(0, 1)$ Gaussian initialization. Table 3.5 shows that KaiMing initialization performs best on Top-1 accuracy, in addition, it will improve the convergence rate.

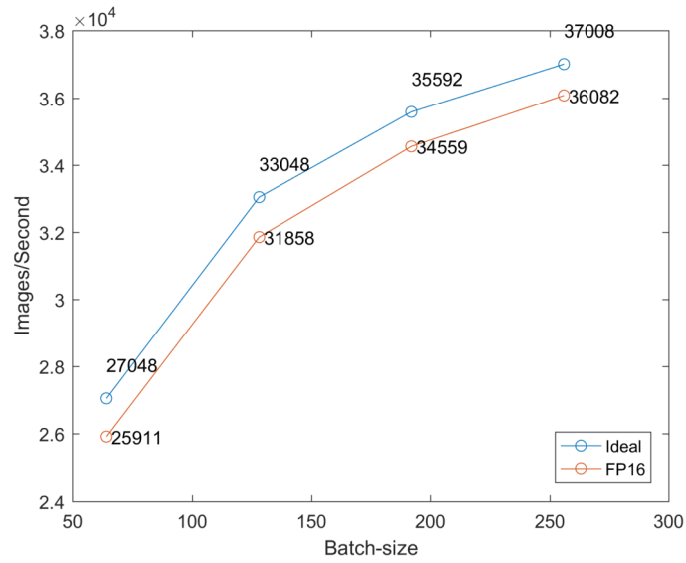


FIG. 3.6. Multi-V100-GPU training throughput with FP16 and different batch-size.

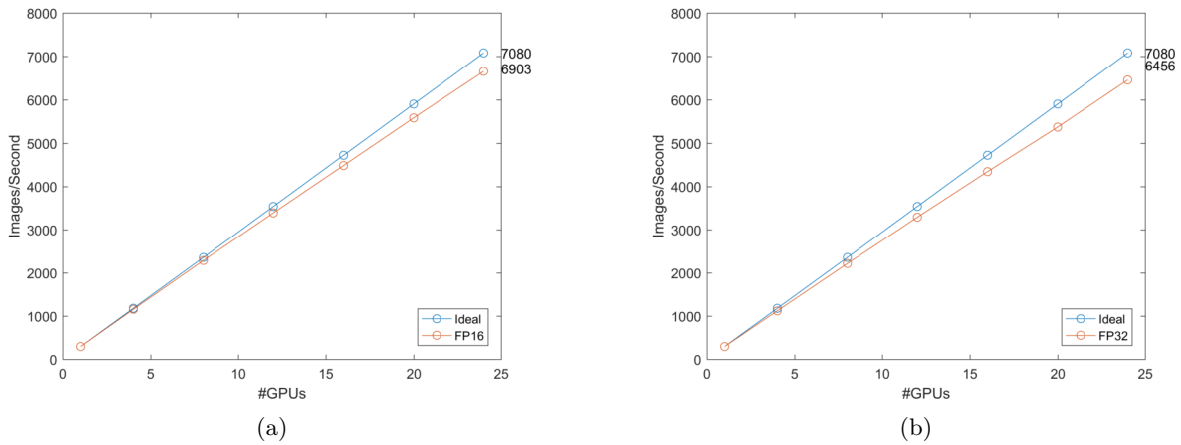


FIG. 3.7. Multi-P40-GPU training throughput with different collective communication precision.

TABLE 3.4

Top-1 validation accuracy with various strategies.

Fixed Batch Size	Independent Statistics	Without Moving Average	Top-1 Accuracy
×	×	×	70.34%
√	×	×	73.35%
√	√	×	75.05%
√	√	√	75.60%

TABLE 3.5

Performance with different weight initialization methods.

Initialization for Convolutional Layers	KaiMing Initialization	Xavier Initialization	Gauss Initialization
Top-1 accuracy	75.6%	75.23%	74.38%
Epoch when Top-1 acc=75.0%	32	35	36

TABLE 3.6
Performance with different γ initialization.

	$\gamma = 0$	$\gamma = 1$
Top-1 accuracy	75.6%	75.32%
Epoch when Top-1 acc= $\geq 75.0\%$	32	42

TABLE 3.7
90-epoch training time and accuracy results for ResNet-50 on ImageNet.

	Time	Top-1 Accuracy	Top-5 Accuracy
Cluster 1	53min	75.6%	93.21%
Cluster 2	108min	75.58%	93.22%

TABLE 3.8
Actual training time and actual average single-GPU training time for ResNet-50 on ImageNet with Top-1 validation accuracy of 75%

	Actual Training Time	Actual Average Single-GPU Training Time	Convergence Rate	Top-1 Accuracy
Cluster 1	18min	432min	32epoch	$>75\%$
Cluster 2	39min	936min	33epoch	$>75\%$

TABLE 3.9
Actual training time and actual average single-GPU training time for ResNet-50 on ImageNet with Top-5 validation accuracy of 93%.

	Actual Training Time	Actual Average Single-GPU Training Time	Convergence Rate	Top-5 Accuracy
Cluster 1	23min	552min	39epoch	$>93\%$
Cluster 2	50min	1200min	42epoch	$>93\%$

In addition, we also have conducted comparative experiment on whether to initiate the scale factor γ of the last BN layer of the residual block to 0. The experimental result (Table 3.6) shows that, when γ is initialized to 0, the convergence rate and the training accuracy can both be improved.

3.2.4. Comparison of existing methods. We build a complete distributed training system through the system design of software and hardware integration and high-quality engineering implementation. The absolute time-consuming and final training accuracy of 90 epochs trained on two clusters are listed in Table 3.7 by using the methods and strategies described in Section 2.2.

Taking the convergence rate into account, and by dividing with the number of GPUs, we obtain the actual training time and the actual average single-GPU training time under different accuracy criteria, as shown in Table 3.8 and Table 3.9. Among them, according to the actual training time as the evaluation criterion, our performance under DOWNBench is 23 minutes.

According to the actual average training time of single GPU as the evaluation criterion, the result of cluster 1 can be compared with that of [15, 35] and the result of cluster 2 can be compared with that of [31] since the GPU types are the same. The GPU scaling efficiency of [15, 35, 31] under 24 GPUs has not been published. As the number of GPUs increases, the GPU scaling efficiency generally decreases. For the sake of fairness, we use the published linear scaling efficiency of 99.2% in [31], and set the GPU scaling efficiency of [15, 35] 97.5% as same as ours. From Table 3.10 and Table 3.11, we can see that the actual average single-GPU training time of our method is far less than that of other methods, or even less by more than an order of magnitude. With little difference in the GPU scaling efficiency, the effectiveness of our method is mainly due to fast convergence rate and the super-high acceleration of the single GPU.

4. Conclusion. The main contribution of this paper comes from two aspects: algorithm and system construction. At the algorithm level, this paper proposes a distributed neural network training method based on hybrid gradient computing, which not only improves the convergence rate by hybrid gradient descent while maintaining the generalization ability of SGD, but also improves the speed of single-GPU computing and multi-GPU communication by mixed precision computing. At the system construction level, for achieving

TABLE 3.10
Actual average single V100-GPU training time with different methods.

	Actual Average Single-GPU Training Time
This work	432min
[15]	5251min
[35]	1941min

TABLE 3.11
Actual average single P40-GPU training time with different methods.

	Actual Average Single-GPU Training Time
This work	936min
[31]	8908min

excellent training accuracy and speed, we carefully integrate various training strategies and construct a complete distributed training system, which is demonstrated to be effective through a large number of experiments.

In the future, we will continue to improve the training speed and training accuracy. In terms of training speed, the model of ResNet-50, which belongs to the computing intensive model, has a larger ratio of computing volume and parameter volume than other models such as AlexNet and VGGNet, and is easy to achieve higher GPU scaling efficiency. In the future, we will try a variety of all-reduce communication algorithms to achieve 95% or higher GPU scaling efficiency with different types of models (such as IO intensive type). Moreover, we will also try the gradient accumulation on single GPU, which is mainly to increase the upper limit of single GPU's batch size. Before multi-GPU gradient aggregation, we can try to run more batches on the single GPU, and sum the gradients from the backpropagation of each batch cumulatively. In the aspect of training accuracy, we will use the heuristic approach to automatically adjust the hyperparameters that are manually set in the current training process. We believe there will be great possibility to further improve the training accuracy.

Acknowledgments. This paper would not have been possible without the consistent and valuable reference materials that we received from our teachers and friends, whose insightful guidance and enthusiastic encouragement in the course of our shaping this paper gain our deepest gratitude. We shall extend our thanks to editors and specialist reviewer for your instructions and helps to the article.

REFERENCES

- [1] A. KRIZHEVSKY, I. SUTSKEVER, AND G. HINTON, *ImageNet classification with deep convolutional neural networks*, NIPS, Lake Tahoe, Nevada, USA, 3-6 Dec 2012.
- [2] P. SERMANET, D. EIGEN, X. ZHANG, ET AL., *Overfeat: Integrated recognition, localization and detection using convolutional networks*, Arxiv preprint arxiv:1312.6229, 2013.
- [3] K. SIMONYAN, A. ZISSERMAN, *Very deep convolutional networks for large-scale image recognition*, ICLR, Banff, Canada, 14-16 April 2014.
- [4] C. SZEGEDY, W. LIU, Y. JIA, ET AL., *Going deeper with convolutions*, CVPR, Boston, MA, 7-12 Jun 2015.
- [5] K. HE, X. ZHANG, S. REN, J. SUN, *Deep residual learning for image recognition*, CVPR, Las Vegas, NV, USA, 27-30 Jun 2016.
- [6] G. HINTON, L. DENG, D. YU, *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups*, IEEE Signal Processing Magazine 2012, 29(6), 82-97.
- [7] W. XIONG, J. DROPPA, X. HUANG, ET AL., *The Microsoft 2016 Conversational Speech Recognition System*, ICASSP, Shanghai, China, 21-25 Mar 2017.
- [8] R. COLLOBERT, J. WESTON, L. BOTTOU, ET AL., *Natural language processing (almost) from scratch*, JMLR 2011, 12(1), 2493-2537.
- [9] Y. WU, M. SCHUSTER, Z. CHEN, ET AL., *Google's neural machine translation system: Bridging the gap between human and machine translation*, Arxiv preprint arxiv:1609.08144, 2016.
- [10] O. RUSSAKOVSKY, J. DENG, H. SU, ET AL., *ImageNet Large Scale Visual Recognition Challenge*, IJCV 2015, 115(3), 211-252.
- [11] D. MAHAJAN, R. GIRSHICK, V. RAMANATHAN, ET AL., *Exploring the Limits of Weakly Supervised Pretraining*, ECCV, Munich, Germany, 8-14 Sep 2018.
- [12] P. GOYAL, P. DOLLAR, R. GIRSHICK, ET AL., *Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour*, Arxiv preprint arxiv:1706.02677, 2017.
- [13] T. AKIBA, S. SUZUKI, K. FUKUDA, *Extremely Large Minibatch SGD: Training ResNet-50 on ImageNet in 15 Minutes*, Arxiv preprint arxiv:1711.04325, 2017.

- [14] C. YING, S. KUMAR, D. CHEN, ET AL., *Image Classification at Supercomputer Scale*, Arxiv preprint arxiv:1811.06992 v2, 2018.
- [15] H. MIKAMI, H. SUGANUMA, P. U-CHUPALA, ET AL., *Massively Distributed SGD: ImageNet/ResNet -50 raining in a Flash*, Arxiv preprint arxiv:1811.05233v2, 2019.
- [16] Y. YOU, I. GITMAN, B. GINSBURG, *Large Batch Training Of Convolutional Networks*, Arxiv preprint arxiv:1708.03888, 2017.
- [17] N. S. KESKAR, D. MUDIGERE, J. NOCEDAL, ET AL., *On large-batch training for deep learning: Generalization gap and sharp minima*, ICLR, Toulon, France, 24-26 April 2017.
- [18] P. CHAUDHARI, A. CHOROMANSKA, S. SOATTO, ET AL., *Entropy-SGD: Biasing Gradient Descent Into Wide Valleys*, ICLR, Toulon, France, 24-26 April 2017.
- [19] Y. YANG, Z. ZHANG, C. HSIEH, ET AL., *ImageNet training in 24 minutes*, Arxiv preprint arxiv:1709.05011, 2017.
- [20] L. C. LUO, Y. H. XIONG, Y. LIU, X. SUN, X., *Adaptive Gradient Methods with Dynamic Bound of Learning Rate*, ICLR, New Orleans, LA, USA, 6-9 May 2019.
- [21] S. IOFFE, C. SZEGEDY, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, Arxiv preprint arxiv:1502.03167v3, 2015.
- [22] C. A. COLEMAN, D. NARAYANAN, D. KANG, *DAWN Bench: An End-to-End Deep Learning Benchmark and Competition*, NIPS ML Systems Workshop, Long Beach, USA, 4-9 Dec 2017.
- [23] H. ROBBINS, S. MONRO, *A stochastic approximation method*, The Annals of Mathematical Statistics, 1951, 22(3), 400-407.
- [24] D. P. KINGMA, J. L. BA, *Adam: A method for stochastic optimization*. ICLR, San Diego, CA, USA, 7-9 May 2015.
- [25] J. DUCHI, E. HAZAN, Y. SINGER, *Adaptive subgradient methods for online learning and stochastic optimization*, JMLR, 2011, 12(7), 2121-2159.
- [26] T. TIELEMAN, G. HINTON, *RMSprop: Divide the gradient by a running average of its recent magnitude*, COURSERA: Neural networks for machine learning, 2012, 4(2), 26-31.
- [27] A. C. WILSON, R. ROELOFS, M. STERN, ET AL., *The marginal value of adaptive gradient methods in machine learning*, NIPS, Long Beach, USA, 4-10 Dec 2017.
- [28] X. GLOROT, Y. BENGIO, *Understanding the difficulty of training deep feedforward neural networks*, AISTATS, Chia Laguna Resort, Sardinia, Italy, 13-15 May 2010.
- [29] K. HE, X. ZHANG, S. REN, ET AL., *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, ICCV, Santiago, Chile, 13-16 Dec 2015.
- [30] S. L. SMITH, P. J. KINDERMANS, C. YING, ET AL., *Don't Decay the Learning Rate, Increase the Batch Size*, NIPS, Long Beach, USA, 4-9 Dec 2017.
- [31] X. JIA, S. SONG, W. HE, ET AL., *Highly Scalable Deep Learning Training System with Mixed-Precision: Training ImageNet in Four Minutes*, ArXiv preprint arXiv:1807.11205, 2018.
- [32] A. SERGEEV, M. D. BALSIO, *Horovod: fast and easy distributed deep learning in TensorFlow*, ArXiv preprint arXiv:1802.05799 (2018).
- [33] R. RABENSEIFNER, *Optimization of collective reduction operations*, ICCS, Krakow, Poland, 6-9 June 2004.
- [34] R. THAKUR, R. RABENSEIFNER, W. GROPP, *Optimization of collective communication operations in MPICH*, IJHPCA 2005, 19(1), 49-66.
- [35] M. YAMAZAKI, S. KASAGI, A. TABUCHI, *Yet Another Accelerated SGD: ResNet-50 Training on ImageNet in 74.7 seconds*, ArXiv preprint arXiv:1903.12650.

Edited by: Dana Petcu

Received: Apr 2, 2020

Accepted: May 18, 2020



AN ANALYTICAL MODEL OF A CORPORATE SOFTWARE-CONTROLLED NETWORK SWITCH

VALERY P. MOCHALOV, GENNADY I. LINEST, NATALYA YU. BRATCHENKO, AND SVETLANA V. GOVOROVA*

Abstract. Implementing the almost limitless possibilities of a software-defined network requires additional study of its infrastructure level and assessment of the telecommunications aspect. The aim of this study is to develop an analytical model for analyzing the main quality indicators of modern network switches. Based on the general theory of queuing systems and networks, generated functions and Laplace-Stieltjes transforms, a three-phase model of a network switch was developed. Given that, in this case, the relationship between processing steps is not significant, quality indicators were obtained by taking into account the parameters of single-phase networks. This research identified the dependencies of service latency and service time of incoming network packets on load, as well as equations for finding the volume of a switch's buffer memory with an acceptable probability for message loss.

Key words: switch architecture; thread table; service time; wait time.

AMS subject classifications. 68M10

1. Introduction. Problems and limitations in modern computer networks have led to the development and construction of software-defined networks (SDN - Software Defined Networks). The main approaches of the SDN concept are presented in the guidelines of the International Telecommunication Union - ITU-T Y.3000 series [19], which require the separation of data transmission and management processes, a logically centralized level of control, the use of a unified OpenFlow interface, and virtualization of physical resources. Unlike traditional switching and routing methods based on IP and MAC addresses, the OpenFlow protocol is able to implement more than forty criteria for selecting transmission routes for network packets [11, 18]. In [13], SDN and network functions virtualization technologies (NFV) were investigated with the aim of developing faster and more flexible fault-tolerant solutions. In addition, there is also a wide variety of implementations of the OpenFlow protocol from switch providers [5] mainly related to the construction of additional software layers on existing products. This is of interest for the future development of a general performance analysis and for comparing various SDN solutions (consisting of controllers, switches, and application modules) for more complex scenarios, such as data centers and cloud computing, distributed across wide area networks. [1] describes the task of setting up the manager to work as an SDN controller. The article includes a brief overview of the various OpenFlow SDN-based controllers available in various programmable languages. It focuses on two controllers that support OpenFlow, namely the POX, a Python-based controller, and the Java-based Floodlight controller. The performance comparison of both controllers is tested on various network topologies by analyzing network bandwidth and round-trip latency using an effective network simulator called Mininet. Single, linear, tree-like and user (user-defined) topologies are developed in Mininet by activating external controllers. It should be noted that SDN is a new network architecture that is adaptive, dynamic, efficient, and manageable. The typical architecture of an SDN network [8, 19] is shown in cf. Fig. 1.1. The SDN controller and network operating system periodically update their internal data about the status of network elements, topology, data transfer routes, threads, and resources. Upon receiving a service request, the controller processes the first network packet of the corresponding thread and sets the control and forwarding rules for all subsequent packets, i.e. data management occurs at the thread level. The first packet of each new thread is sent to the controller,

*Department of Information Communications Institute of Information Technologies and Telecommunications North Caucasian Federal University 1 Pushkin Str., Stavropol, 355017, Russia. Corresponding author: Natalya Yu. Bratchenko (nbratchenko@ncfu.ru). This study was carried out with the financial support of the Russian Foundation for Basic Research (RFBR) as part of research project No. 19-07-00856\19.

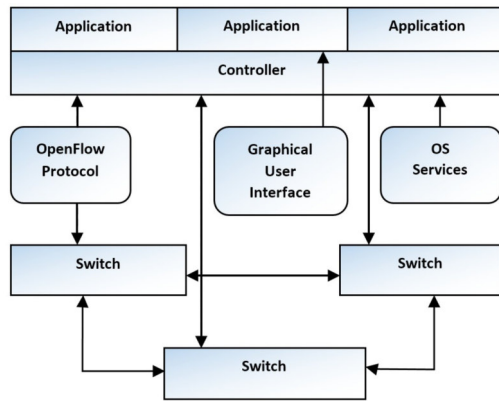


FIG. 1.1. Typical SDN Network Architecture

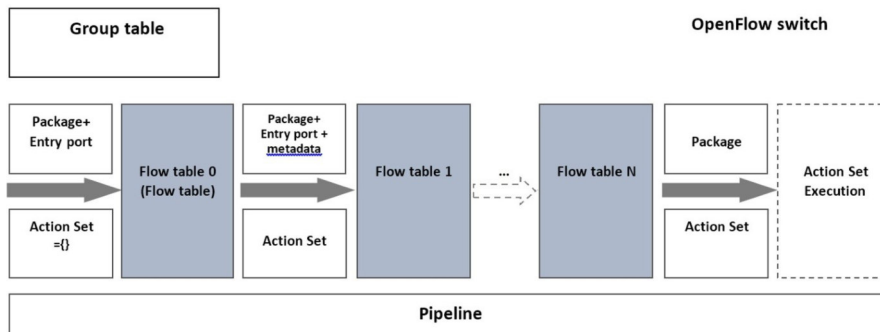


FIG. 1.2. Contents of SDN Switch Address Tables

which creates the corresponding entry in the transfer table in the switch. Each switch fills its addressing tables only according to the controller. If we assume that there is no queue of serviced network packets at the controller input, we can describe it as a queuing system (QS) with an infinite number of serviced devices, and the SDN switch model, which includes the phases of receiving and processing packet headers, thread control, and making changes to addressing tables, can be represented by multiphase QS with unlimited memory. The network's operation in stationary mode, as well as its transitions to various states, can be described by the Markov chain [4, 8, 12].

The switch management port is connected to the controller processor by a secure OpenFlow messaging channel. In this case, both a special control network and the existing transport network can be used. Each switch includes a chain of tables connected in series for addressing packet threads (cf. Fig. 1.2), which contain algorithms and instructions for redistributing packets: forwarding to the next table number, to one of the output ports, or to the control input of the controller. Upon receipt of input packets, the address of a received packet is checked against the entries in the thread tables. If an address match is not established, the packet is sent to the controller, which determines the rules for processing it and sets them in the switch's addressing tables. At the same time, the controller, depending on the state of the network (topology, load, deadlocks and packet thread blocks), can change the contents of address tables, and also compares information about the condition of elements.

2. Methods. Each switch contains a set of thread entries that include match fields, counters, and instructions for defined actions related to standard pre-processing and packet forwarding. The structure of the thread entries is shown in cf. Fig. 2.1.

Operations on packet threads can be divided into stages: the switch receives packets from subscribers,

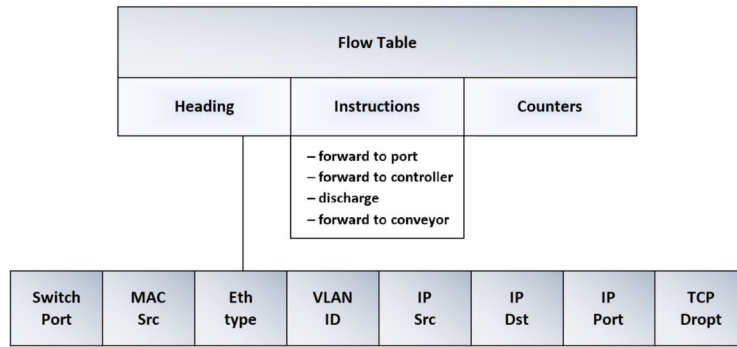


FIG. 2.1. Switch Thread Entries Structure

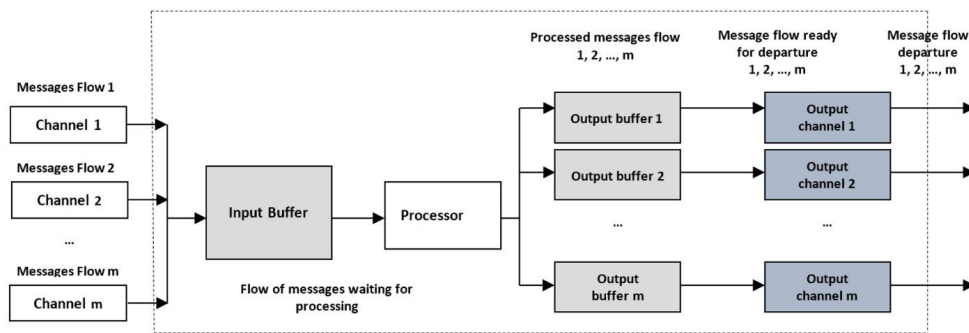


FIG. 2.2. Switch Architecture Variant

checks for records of incoming packet threads in the address tables, and forwards requests to the controller, which makes the decision of how to process the packets. A variant of switch architecture is shown in cf. Fig. 2.2.

Packet processing includes the following actions [8, 15, 17]:

1. The packet of the newly formed thread arrives with the speed of the communication channel at the input port of the switch and is placed in its buffer memory.
2. The formation of packet threads. Checking that the address of the received packet matches the entries in the thread tables. If a match is found, then step 5 is executed.
3. If no match is found, the packet is sent to the SDN controller.
4. According to the routing algorithm, the SDN controller adds the corresponding record to the switch and other switches along the transmission path of the thread.
5. Waiting for the release of the channel in the direction of the outgoing port. Packet encapsulation and transmission to the specified address.

Each fragment of the SDN network contains several switches interconnected by high-speed duplex communication channels. We believe that asynchronous sealing should be used on channels connecting subscribers to the switch, as it does not significantly affect overall network performance. Meanwhile, if a packet is sent to the output port of a switch's high-speed channel immediately after arrival, we assume that its service time is zero. If the packet remains in the switch's memory and its header is sent to the controller, then we describe the service time by an arbitrary distribution function. In real network elements, the amount of memory is always limited, so when the established buffer size is exceeded, there is a high probability of packet loss. Buffer memory is a shared resource for all communication channels. The required amount of memory can be estimated based on the given probability of packet failures entering the node. For approximate estimates of the memory size of a switch, we will use functions that account for the volume of requests, the message service time, and the probability of losses. In this case, the number of servicing devices is equal to the number of places in

the drive m and there is no queue. The memory can be represented by the total of m independent single-line QS and failures, service blocking and discipline, distributed according to exponential law. The number of service elements of such a multi-line center is equal to the number of packet threads received at the switch input. The OF-CONFIG protocol used in SDN allows switch resources to be allocated, forming several virtual switches from one physical switch, while distributing physical memory between threads. Message volumes are distributed according to the exponential law $L(x) = 1 - e^{-fx}$. Messages are received on m channels at the intensities $a_i (i = \overline{1, m})$. The thread of received messages is simplest at the intensity $\lambda = \sum_{i=1}^m a_i$. The paper [10] reflects clear expressions for determining memory volume with a reasonable likelihood for message refusal. For LMS $M|G|1|\infty$ the Laplace-Stieltjes transform (LST) was found for the distribution function (DF) of the volume of the serviced message $R(x)$:

$$i(s) = 1 - \frac{a_i}{g_i} \left[\frac{1}{f} + \frac{f}{(s+f)^2} \right] \quad (2.1)$$

where i is the number of the receiving channel; a_i is the intensity of the input thread ($i = \overline{1, m}$); $g_i > 0$ is the message receipt time on channel i ; f is the DF parameter of the message volume.

From here, the first two instances of request volume will be equal to:

$$l_{1i} = \frac{2a_i}{g_i f^2} = \frac{2\rho_i}{f}, l_{2i} = \frac{2a_i}{g_i f^3} = \frac{6\rho_i}{f^2} \quad (2.2)$$

where ρ_i – is the load of channel i .

$$(2i - l_{1i}^2) = \frac{2\rho_i}{f^2} (3 - 2\rho_i)$$

Then the LST of the stationary total message volume will be equal to:

$$\delta(s) = \prod_{i=1}^m \left\{ 1 - \frac{a_i}{g_i} \left[\frac{1}{f} + \frac{f}{(s+f)^2} \right] \right\} \quad (2.3)$$

From here it follows that the first and second instances of total message volume will be equal to:

$$\delta_1 = \frac{2}{f} \sum_{i=1}^m \rho_i, (\delta_2 - \delta_1^2) = \frac{2}{f^2} \sum_{i=1}^m \rho_i (3 - 2\rho_i) \quad (2.4)$$

The calculation of the memory volume V is carried out taking into account the probability of message loss:

$$p_{\Pi} = 1 - R(V) \quad (2.5)$$

where $R(V) = \int_0^V D(V-x)dL(x)$; $L(x) = 1 - e^{-fx}$ - DF message volume; $D(x) = \rho(\delta < x)$ - DF stationary total message volume δ

The first two instances of DF $R(x)$ are equal to:

$$r_1 = \delta_1 + \phi_1, r_2 = \phi_2 + 2\phi_1\delta_1 + \delta_2 \quad (2.6)$$

where $\phi_1 = \frac{1}{f}$ - average message size; δ_1, δ_2 - instances of total message volume.

To solve cf. Eq. 2.5 we can use formula:

$$D(V) = p_0 + (1 - p_0) \frac{\gamma(p, gx)}{\Gamma(p)} \quad (2.7)$$

where p_0 is the probability of not receiving requests; $\gamma(p, gx)$ is the gamma distributions are expressed by $\gamma(p, gx) = \int_0^{gx} t^{p-1} e^{-t} dt$, $\Gamma(p) = \gamma(p, \infty)$, and their parameters p and g are defined as

$$p = \frac{r_1^2}{r_2 - r_1^2}, g = \frac{r_1}{r_2 - r_1^2} \quad (2.8)$$

The next stage of request processing is analyzing the service part of the packets and forming network packet threads. The following actions are performed with each packet [3, 10]:

1. Identifying the package and implementing the procedure for determining its place in the thread.
2. Perform a search on all types of threads.
3. If the thread is not found, then the corresponding packet is transmitted to the controller and a new thread is formed with information about the current packet.

We believe that the request is instantly serviced upon identification of the packet, otherwise, i.e. in the absence of identification, the latency is described by an exponential distribution with parameter p .

The DF of service time for the input packet thread looks like this: $B(t) = p + (1-p)(1 - e^{pt})$, and its LST looks like:

$$\beta(q) = p + \frac{(1-p)p}{p+q} = \frac{p(1+q)}{(p+q)} \quad (2.9)$$

from which we get the average service time value:

$$\beta_1 = -\beta'(0) = \frac{1-p}{p} \quad (2.10)$$

If the system load is determined as $\rho = \alpha\beta_1 = \alpha(1-p)/p$, then the LST service timeout is defined as:

$$W(q) = \frac{(1-\rho)(p+q)}{p+q-\alpha(1-p)} = \frac{(1-p)(p+q)}{q+p(1-\rho)} \quad (2.11)$$

The average timeout value is $W_1 = -W'(0) = \frac{p}{p(1-\rho)}$.

If the image takes the form of the rational fraction $\frac{A_n(p)}{B_m(p)}$, and P_1, P_2, \dots, P_n - are roots of multiplicity r_1, r_2, \dots, r_n , so that

$$r_1 + r_2 + \dots + r_n = m, \quad (2.12)$$

$$B_m(p) = \beta_0(p - P_1)^{r_1}(p - P_2)^{r_2} \dots (p - P_n)^{r_n}. \quad (2.13)$$

Then the original can be found with the formula:

$$f(t) = \sum Res \left[\frac{A_n(p)e^{pt}}{B_m(p)} \right] \quad (2.14)$$

If the roots of the denominator P_1, P_2, \dots, P_m are simple, then

$$f(t) = \sum \frac{A_n(P_k)}{B_m(P_k)} e^{P_k t} \quad (2.15)$$

Then, the inverse of the LST function $W(q)$ will be determined by the relation:

$$W(t) = \sum Res \left[\frac{(1-\rho)(\rho+q)}{q(q+p(1-\rho))} e^{qt} \right] \quad (2.16)$$

The distribution function of the random variable (RV) U takes the form of:

$$U(t) = p \{U < t\} = \int_0^t W(t-u) dB(u) = \int_0^t e^{-\mu u} W(t-u) dU \quad (2.17)$$

where $B(t) = 1 - e^{-\mu t}$

For the case $p\{W > 0\} = 1 - W(0) = \frac{(n\rho)^n p_0}{n!(1-\rho)}$ the average value of stationary latency will be equal to:

$$W_1 = EW = \int_0^\infty dW(t) = \frac{n^{n-2} \rho^n p_0}{\mu(1-\rho)^2 (n-1)!} \quad (2.18)$$

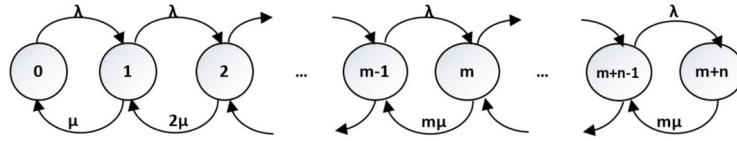


FIG. 2.3. State and Transition Graph of QS type $M|M|m|n$

The average value of the stationary service time will be equal to:

$$U_1 = EU = \int_0^\infty t dU(t) = \beta_1 + T_1 = \frac{1}{\mu} + \frac{n^{n-2} \rho^n p_0}{\mu(1-\rho)^2(n-1)!} \tag{2.19}$$

where EW, EU are the mathematical expectations of RV W and U .

To simplify the calculations, the obtained random variables were approximated by the DF $Z(x) = p_0 + (1-p_0) \frac{\gamma(p, gx)}{\Gamma(p)}$, where p_0 is the stationary probability of the absence of requests

$$p = \frac{\delta_1^2}{(1-p_0)\delta_2 - \delta_1^2}, g = \frac{(1-p_0)\delta_1}{(1-p_0)\delta_2 - \delta_1^2} \tag{2.20}$$

where p and g are the gamma distribution parameters, and $\delta_1, \delta_2 - \delta_1^2$, are the first instance and the variance of the total message volume.

At the data transfer stage, the switch executes the instructions associated with a packet, adds routing information received from the SDN controller, and transfers the generated packet thread to the outgoing port to be sent to the communication channel [15, 19]. The methods of transmitting network packets, which determine both the speed of data transmission over communication channels and the storage time of copies of messages in buffer memory, significantly depend on the linear protocols used, the characteristics of acknowledgment stages and *time-out*. We assume that copies of the transmitted packets are stored after the end of their transmission in the buffer memory for some time-out period until confirmation of the delivery of the packets is received. No confirmation necessitates retransmission of the packet. This eliminates possible errors and packet loss due to insufficient buffer memory in the receiving switch. Obviously, the transmission phase of network packets leaving the switch can be described by a system in the form $M|M|m|n$ that is m single-line QS with n place buffers, the simplest incoming thread and exponential distribution of service time. A diagram of the intensity of QS type $M|M|m|n$ transitions is shown in cf. Fig. 2.3.

We know [6, 17] that the probability of finding an $M|M|m|n$ system in a p_k state is:

$$p_k = \frac{\frac{\rho^k}{k!}}{\sum_{k=0}^m \frac{\rho^k}{k!} + \frac{\rho^{m+1}}{m \cdot m!} - \frac{1 - (\frac{\rho}{m})^n}{1 - \frac{\rho}{m}}}, 0 \leq k \leq m \tag{2.21}$$

where $\rho = \frac{\lambda}{\mu}$ is the load.

Accordingly, the probability of a p_{m+s} state is:

$$p_{m+s} = \frac{\frac{\rho^m}{m!} (\frac{\rho}{m})^s}{\sum_{k=0}^m \frac{\rho^k}{k!} + \frac{\rho^{m+1}}{m \cdot m!} - \frac{1 - (\frac{\rho}{m})^n}{1 - \frac{\rho}{m}}}, 1 \leq \rho \leq n \tag{2.22}$$

We believe that due to the lack of a free buffer in the receiving switch, a transmitted packet has probability p of being lost. Then the intensity of the output thread is determined as $y = \lambda - \lambda p_{m+n} = \lambda \sum_{i=0}^{m+n-1} p_i$.

The corresponding probabilities of the states of the system in question will be equal to:

$$p_0 = \frac{\rho}{\sum_{k=0}^m \frac{\rho^k}{k!} + \frac{\rho^{m+1}}{m \cdot m!} - \frac{1 - (\frac{\rho}{m})^n}{1 - \frac{\rho}{m}}} \tag{2.23}$$

TABLE 3.1
Simulation Results

V	p_{loss}		
	$\rho = 0.4$	$\rho = 0.5$	$\rho = 0.7$
20	0.093	0.171	0.279
30	0.017	0.083	0.073
40	0.0063	0.0038	0.017

$$p_{m+n} = \frac{\frac{\rho^m}{m!} \left(\frac{\rho}{m}\right)^n}{\sum_{k=0}^m \frac{\rho^k}{k!} + \frac{\rho^{m+1}}{m \cdot m!} - \frac{1 - \left(\frac{\rho}{m}\right)^n}{1 - \frac{\rho}{m}}} = \frac{\frac{\rho^m}{m!} \left(\frac{\rho}{m}\right)^n}{\rho} p_0 \quad (2.24)$$

Taking into account lost packets, the output thread intensity will be equal to:

$$y = \lambda(1 - p_{m+n}) = \lambda \left(1 - \frac{\frac{\rho^m}{m!} \left(\frac{\rho}{m}\right)^n}{\rho} p_0\right) = \lambda \frac{\rho - \frac{\rho^m}{m!} \left(\frac{\rho}{m}\right)^n}{\rho} p_0 \quad (2.25)$$

In addition to these characteristics, one can evaluate the quality indicators of the transmitting part of the switch by such parameters as:

- the time the packet stays in the system

$$V = \frac{N}{m\mu(1 - p_0)} = \frac{\sum_{k=0}^{m+n} kp_k}{m\mu(1 - p_0)}; \quad (2.26)$$

- latency for packets in the queue

$$W = \frac{N_0}{m\mu(1 - p_0)} = \frac{\sum_{k=m+1}^{m+n} (k - m)p_k}{m\mu(1 - p_0)}; \quad (2.27)$$

- service time

$$\begin{aligned} T_s = V - W &= \frac{\sum_{k=0}^{m+n} kp_k}{m\mu(1 - p_0)} - \frac{\sum_{k=m+1}^{m+n} (k - m)p_k}{m\mu(1 - p_0)} \\ &= \frac{1}{m\mu} + \frac{(m - 1) \sum_{k=m}^{m+n} p_k + \sum_{k=2}^{m-1} (k - 1)p_k}{m\mu(1 - p_0)} \end{aligned} \quad (2.28)$$

3. Results. Simulation results of the required memory volume V of the switch for a given probability of message loss (p_{loss}) are presented in cf. Table 3.

Cf. Table 3 shows that with an increase in buffer memory, the probability of message loss decreases. With a load of $\rho = 0.4$, an increase in memory volume by a factor of two decreases the probability of information loss by about 14 times, at $\rho = 0.5$ it is 45 times, and at $\rho = 0.7$ it is 16 times. Thus, we can conclude that there is an optimal load for the network, which will minimize losses.

The results of analytical modeling of the formation of packet threads using cf. Eqs. 2.18 and 2.19 are shown in cf. Figs. 3.1 and 3.2.

The presented graphs determine the dependence of the dynamic characteristics of the switch on the load for two message volumes: $\delta_1 = 31.33 \cdot 10^3$ ch, (curve 1) and $\delta_1 = 53.31 \cdot 10^3$ ch (curve 2), with variance $\delta_2 - \delta_1^2 = 105.90 \cdot 10^6$ ch². cf. Fig. 3.1 shows that the latency for the message volumes being studied does not significantly differ from load values of 0.6 to 0.7. Then, with an increase in load, latency increases inversely

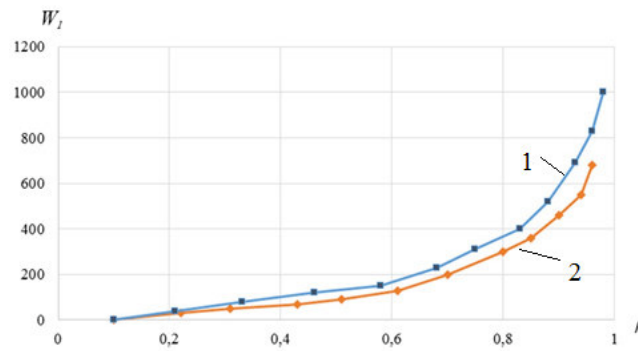


FIG. 3.1. Average Load Latency

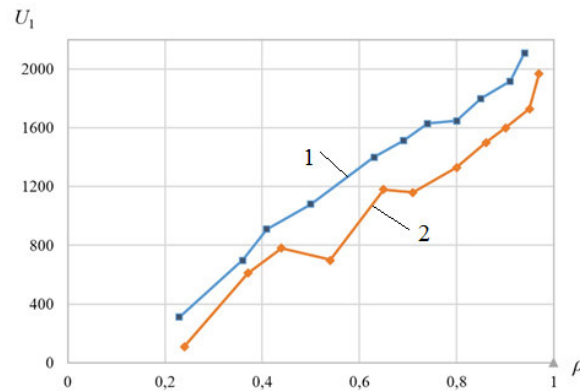


FIG. 3.2. Relation Between Average Service Time and Load

with an increase in message size. In general, the functions have a power-law nature, starting from a certain load value (in our case 0.7), latency sharply increases. A longer average latency is observed for a smaller message size. This is done to equalize the processing time of the message packet as a whole. More latency means less time for processing data and, conversely, less latency means more processing time.

Fig. 3.2 shows that the average service time is also inversely related to message size. In contrast to cf. Fig. 3.1, a direct proportional dependence on the load is observed here. In the case of $\delta_1 = 31.33 \cdot 10^3$ ch (curve 1) a certain failure is observed, i.e. a reduction in service time with a load value of 0.5. For the case $\delta_1 = 53.31 \cdot 10^3$ ch (curve 2), the failure is less pronounced and corresponds to a load value of 0.7. Therefore, we can conclude that an optimal relationship exists between message size and service time that can be calculated using cf. Eqs. 2.18 and 2.19.

4. Discussion. Article [14] presents an analytical model of a software-controlled switch on an SDN network and focuses on a study of its stability in the process of changing information interactions. The complexity of the mathematical apparatus, as well as a large number of limitations, are the constraining factors of its application in practice. Evaluations of the switch operation process presented here are time dependent. We know that, when considering random stationary processes, when the characteristics of the system do not change over time, the system can be considered stationary in its established mode, with parameters that do not depend on time. This factor can greatly simplify the model.

Article [16] establishes a connection between message delays in a network switch and the amount of buffer memory. It explores the effect of the internal buffer in software and hardware SDN switches on the system's quality indicators. The disadvantages of this model are the assumptions about the discrete nature of message processing, as well as the possibility of conducting research only at the level of private distributions, which

complicates its application in the analysis of real information systems.

The analytical model of a network switch proposed in [9] provides an efficient enough study of its latencies with various methods of information exposure but does not provide an assessment of quality indicators.

A similar mathematical model is presented in [2]. Limitations on the degree of load used on network devices here do not facilitate conducting research with a fairly wide range of load changes and nonstationary threads.

The analytical model developed for studying the quality indicators of a program-controlled network SDN switch is formally presented as a mass service network (CeMO) with a Poisson thread of incoming packets, exponential service, failures, and locks. We know that when studying this CeMO, it is possible to independently investigate its constituent nodes, which are an $M|M|m|n$ QS type. Therefore, the operations performed by the switch are conventionally divided into three groups: receiving and placing network packets in the buffer memory, processing and generating packet threads, and transmitting network packets to communication channels. A function generator and Laplace-Stieltjes transforms were used for this. The first stage of servicing requests includes the process of writing packets to the switch's multi-line buffer memory. The number of single-line memory elements here corresponds to the number of buffer memory locations; there is no queue. In the next center, the addresses of the received packets are checked against the entries in the thread tables. If a match is found, then a packet thread is generated; if no match is found, the packet is sent to the SDN controller. The third single-line center, which implements the process of transmitting packets leaving the switch, is described as an $M|M|m|n$ system i.e. m independent single-line QS with n place buffers. Considering that, in this case, the relationship between the processing steps is not significant, the packet processing steps being discussed can be considered independent, which is why the quality indicators for this CeMO are obtained by taking into account the parameters of single-phase QS.

5. Conclusion. The new dynamic networking architecture of SDNs has successfully transformed traditional networks into diverse application platforms. However, current SDN technology must be perfected, including with the standardization and unification of various interfaces, ensuring heterogeneous network compatibility, coordinating several controllers, and fixing some security issues. Extending SDN applications also faces major challenges. The use of SDN is no longer limited to campus networks and WANs between data centers but can also be used in wider spaces, including not only terrestrial networks, but also in space. This study developed and investigated a mathematical model for the operation of a three-phase network switch based on the general theory of queuing networks using the Laplace-Stieltjes transform. A number of assumptions were made that made it possible to consider a three-phase switch as single-phase with three independent threads. Using the SDN model presented in this article, network administrators and schedulers can better predict likely performance changes resulting from traffic changes. This will allow them to make operational decisions to prevent small problems from turning into major bottlenecks. The results of the study can be used in the design and operation of communication networks that implement the concept of SDN.

REFERENCES

- [1] I. Z. BHOLEBAWA AND U. D. DALAL, *Performance analysis of SDN/OpenFlow controllers: POX versus floodlight.*, *Wireless Personal Communications*, 98(2), 2018, 1679-1699. doi:10.1007/s11277-017-4939-z
- [2] Y. GOTO, B. NG, W. K. G. SEAH AND Y. TAKAHASHI, *Queueing analysis of software defined network with realistic OpenFlow-based switch model*, *Computer Networks*, 164, 2019, 106892. doi:10.1016/j.comnet.2019.106892.
- [3] D. GROSS, J. F. SHORTIE, J. M. THOMPSON AND C. M. HARRIS, *Fundamentals of queueing theory (1st ed.)*, Hoboken, NJ, USA: Wiley, 2008, doi:10.1002/9781118625651
- [4] L. HANHUA, W. YASHI, M. LIJUAN AND H. ZHENQI, *OSS/BSS Framework based on NGOSS.*, 2009 International Forum on Computer Science-Technology and Applications (pp. 466-471), 2009,. doi:10.1109/IFCSTA.2009.120
- [5] D. KREUTZ, F. M. V. RAMOS, P. ESTEVES VERISSIMO, C. ESTEVE ROTHENBERG, S. AZODOLMOLKY, AND S. UHLIG, *Software-defined networking: a comprehensive survey*, *Proceedings of the IEEE*, 103(1), 14-76, 2015. doi:10.1109/JPROC.2014.2371999
- [6] T. LECHLER, B. J. TAYLOR, AND B. KLINGENBERG, *The telecommunications carriers' dilemma: Innovation vs. Network Operation.*, *PICMET '07 - 2007 Portland International Conference on Management of Engineering & Technology* (pp. 2940-2947), 2007. doi:10.1109/PICMET.2007.4349638
- [7] P. N. BROWN AND Y. SAAD, *The problem with threads*, 39(5), 33-42, 2006. doi:10.1109/MC.2006.180
- [8] T. LI, J. CHEN, AND H. FU, *Application scenarios based on SDN: an overview*, *Journal of Physics: Conference Series*, 1187(5), 2019, 052067. doi:10.1088/1742-6596/1187/5/052067

- [9] S. V. MALAKHOV, V. N. TARASOV, AND I. V. KARTASHEVSKIY, *Teoreticheskoye i eksperimental'noye issledovaniye zaderzhki v programmno-konfiguriruyemykh setyakh. infokommunikatsionnyye tekhnologii [Theoretical and experimental study of delay in software-configurable networks]*, Infokommunikatsionnyye Tekhnologii [Infocommunication Technologies], 13(4), 2015, 409-413. doi:10.18469/ikt.2015.13.4.08 (in Russian)
- [10] V. P. MOCHALOV, N. YU. BRATCHENKO, AND S. V. YAKOVLEV, *Analytical model of integration system for program components of distributed object applications*, 2018 International Russian Automation Conference (RusAutoCon) (pp. 1-4), 2018. doi:10.1109/RUSAUTOCON.2018.8501806 (in Russian)
- [11] B. A. A. NUNES, M. MENDONCA, X.-N. NGUYEN, K. OBRACZKA, AND T. TURLETTI, *A survey of software-defined networking: past, present, and future of programmable networks*, IEEE Communications Surveys & Tutorials, 16(3), 1617-1634, 2014. doi:10.1109/SURV.2014.012214.00180
- [12] M. OLSZEWSKI, J. ANSEL, AND S. AMARASINGHE, *Kendo: Efficient deterministic multithreading in software*, ACM SIGPLAN Notices, 44(3), 97, 2014. doi:10.1145/1508284.1508256
- [13] N. S. RAO, *Performance comparison of SDN Solutions for Switching Dedicated Long-Haul Connections*, Retrieved from Oak Ridge National Lab. (ORNL), Oak Ridge 2016, TN (United States) website: <https://www.osti.gov/biblio/1267045-performance-comparison-sdn-solutions-switching-dedicated-long-haul-connections>
- [14] K. E. SAMOUYLOV, I. A. SHALIMOV, I. G. BUZHIN, AND Y. B. MIRONOV, *Model' funktsionirovaniya telekommunikatsionnogo oborudovaniya programmno-konfiguriruyemykh setey [Model of functioning of telecommunication equipment for software-configured networks]*, Sovremennyye Informatsionnyye Tekhnologii i IT-Obrazovaniye [Modern Information Technologies and IT-Education], 14(1), 2018. Retrieved from <https://cyberleninka.ru/article/n/model-funktsionirovaniya-telekommunikatsionnogo-oborudovaniya-programmno-konfiguriruyemykh-setey> (in Russian)
- [15] J. SIMOES, AND S. WAHLE, *The future of services in next generation networks*. IEEE Potentials, 30(1), 24-29, 2011. doi:10.1109/MPOT.2010.939761
- [16] D. SINGH, B. NG, Y.-C. LAI, Y.-D. LIN, AND W. K. G. SEAH, *Analytical modelling of software and hardware switches with internal buffer in software-defined networks*, Journal of Network and Computer Applications, 136, 22-37, 2019. doi:10.1016/j.jnca.2019.03.006
- [17] H. SUTTER, AND J. LARUS, *Software and the concurrency revolution*, Queue, 3(7), 54, 2005. doi:10.1145/1095408.1095421
- [18] A. VISHNU PRIYA, AND N. RADHIKA, *Performance comparison of SDN OpenFlow controllers*, International Journal of Computer Aided Engineering and Technology, 11(4/5), 2019, 467. doi:10.1504/IJCAET.2019.10020284
- [19] Y.3300, *Framework of software-defined networking. (n.d.)*, Retrieved from <https://www.itu.int/rec/T-REC-Y.3300-201406-1/en>

Edited by: Dana Petcu

Received: Feb 16, 2020

Accepted: May 7, 2020

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 $\text{\LaTeX} 2_{\epsilon}$ 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.