

Recent Advances in Securing Medical Data

Lead Guest Editor: Thippa Reddy G
Guest Editors: Celestine Iwendi





Recent Advances in Securing Medical Data

Security and Communication Networks

Recent Advances in Securing Medical Data

Lead Guest Editor: Thippa Reddy G

Guest Editors: Celestine Iwendi






Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in "Security and Communication Networks." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

Roberto Di Pietro, Saudi Arabia

Associate Editors

Jiankun Hu , Australia
Emanuele Maiorana , Italy
David Megias , Spain
Zheng Yan , China

Academic Editors




Saed Saleh Al Rabae , United Arab Emirates
Shadab Alam, Saudi Arabia
Goutham Reddy Alavalapati , USA
Jehad Ali , Republic of Korea
Jehad Ali, Saint Vincent and the Grenadines
Benjamin Aziz , United Kingdom
Taimur Bakhshi , United Kingdom
Spiridon Bakiras , Qatar
Musa Balta, Turkey
Jin Wook Byun , Republic of Korea
Bruno Carpentieri , Italy
Luigi Catuogno , Italy
Ricardo Chaves , Portugal
Chien-Ming Chen , China
Tom Chen , United Kingdom
Stelvio Cimato , Italy
Vincenzo Conti , Italy
Luigi Coppolino , Italy
Salvatore D'Antonio , Italy
Juhriyansyah Dalle, Indonesia
Alfredo De Santis, Italy
Angel M. Del Rey , Spain
Roberto Di Pietro , France
Wenxiu Ding , China
Nicola Dragoni , Denmark
Wei Feng , China
Carmen Fernandez-Gago, Spain
AnMin Fu , China
Clemente Galdi , Italy
Dimitrios Geneiatakis , Italy
Muhammad A. Gondal , Oman
Francesco Gringoli , Italy
Biao Han , China
Jinguang Han , China
Khizar Hayat, Oman
Azeem Irshad, Pakistan

M.A. Jabbar , India
Minho Jo , Republic of Korea
Arijit Karati , Taiwan
ASM Kayes , Australia
Farrukh Aslam Khan , Saudi Arabia
Fazlullah Khan , Pakistan
Kiseon Kim , Republic of Korea
Mehmet Zeki Konyar, Turkey
Sanjeev Kumar, USA
Hyun Kwon, Republic of Korea
Maryline Laurent , France
Jegatha Deborah Lazarus , India
Huaizhi Li , USA
Jiguo Li , China
Xueqin Liang, Finland
Zhe Liu, Canada
Guangchi Liu , USA
Flavio Lombardi , Italy
Yang Lu, China
Vincente Martin, Spain
Weizhi Meng , Denmark
Andrea Michienzi , Italy
Laura Mongioi , Italy
Raul Monroy , Mexico
Naghme Moradpoor , United Kingdom
Leonardo Mostarda , Italy
Mohamed Nassar , Lebanon
Qiang Ni, United Kingdom
Mahmood Niazi , Saudi Arabia
Vincent O. Nyangaresi, Kenya
Lu Ou , China
Hyun-A Park, Republic of Korea
A. Peinado , Spain
Gerardo Pelosi , Italy
Gregorio Martinez Perez , Spain
Pedro Peris-Lopez , Spain
Carla Ràfols, Germany
Francesco Regazzoni, Switzerland
Abdalhossein Rezai , Iran
Helena Rifà-Pous , Spain
Arun Kumar Sangaiah, India
Nadeem Sarwar, Pakistan
Neetesh Saxena, United Kingdom
Savio Sciancalepore , The Netherlands

De Rosal Ignatius Moses Setiadi ,
Indonesia
Wenbo Shi, China
Ghanshyam Singh , South Africa
Vasco Soares, Portugal
Salvatore Sorce , Italy
Abdulhamit Subasi, Saudi Arabia
Zhiyuan Tan , United Kingdom
Keke Tang , China
Je Sen Teh , Australia
Bohui Wang, China
Guojun Wang, China
Jinwei Wang , China
Qichun Wang , China
Hu Xiong , China
Chang Xu , China
Xuehu Yan , China
Anjia Yang , China
Jiachen Yang , China
Yu Yao , China
Yinghui Ye, China
Kuo-Hui Yeh , Taiwan
Yong Yu , China
Xiaohui Yuan , USA
Sherali Zeadally, USA
Leo Y. Zhang, Australia
Tao Zhang, China
Youwen Zhu , China
Zhengyu Zhu , China

Contents

Visual Attention and Motion Estimation-Based Video Retargeting for Medical Data Security

Qingfang Liu , Baosheng Kang , Qiaozhi Hua , Zheng Wen , and Haipeng Li 

Research Article (11 pages), Article ID 1343766, Volume 2022 (2022)

k Nearest Neighbor Similarity Join Algorithm on High-Dimensional Data Using Novel Partitioning Strategy

Youzhong Ma , Qiaozhi Hua , Zheng Wen , Ruiling Zhang , Yongxin Zhang , and Haipeng Li 


Research Article (16 pages), Article ID 1249393, Volume 2022 (2022)

Platform Firm's IT Capabilities, External Informal Knowledge Governance, and Green Knowledge Integration in Low-Carbon Economy

Guanghua Fu  and Bencheng Li

Research Article (11 pages), Article ID 3904413, Volume 2022 (2022)

Optimization of Urban Waste Transportation Route Based on Genetic Algorithm

Yanling Zhang, Xu Luo , Xiaoxuan Han, Yongxing Lu, Jiacheng Wei, and Chunyu Yu


Research Article (10 pages), Article ID 8337653, Volume 2022 (2022)

Risk Analysis of Distal Metastasis in Chondrosarcoma and the Development and Validation of a Novel Clinical Prediction Model: A Clinical Study Based on the SEER Database

Wenle Li, Rong Li, Wanying Li, Chan Xu, Minmin Ma, Haiwen Peng, Bing Wang , Qiang Liu , and Chengliang Yin 



Research Article (10 pages), Article ID 7542424, Volume 2022 (2022)

Agent-Based Data Extraction in Bioinformatics

Shakir Ullah Shah , Abdul Hameed, Abdulwahab Ali Almazroi, and Mohammed A. Alqarni



Research Article (11 pages), Article ID 4865209, Volume 2022 (2022)

Data Mining Method under Model-Driven Architecture (MDA)

Jiangning Xie , Feng Xu, Zhen Li, and Xueqing Li 




Research Article (10 pages), Article ID 5806829, Volume 2022 (2022)

Research Contribution and Comprehensive Review towards the Semantic Segmentation of Aerial Images Using Deep Learning Techniques

P. Anilkumar  and P. Venugopal 



Review Article (31 pages), Article ID 6010912, Volume 2022 (2022)

The Systematic Literature Review of Privacy-Preserving Solutions in Smart Healthcare Environment

Driss El Majdoubi , Hanan El Bakkali , Souad Sadki , Zaina Maqour, and Asmae Leghmid









Review Article (26 pages), Article ID 5642026, Volume 2022 (2022)

Creative Destruction Path Selection for Industrial Park Transformation and Upgrading under the Concept of Character Town in the Era of Big Data


Yue Bai  and Xuwen Li 

Research Article (11 pages), Article ID 2110039, Volume 2022 (2022)


Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis

V. Durga Prasad Jasti , Abu Sarwar Zamani , K. Arumugam , Mohd Naved , Harikumar Pallathadka , F. Sammy , Abhishek Raghuvanshi , and Karthikeyan Kaliyaperumal 
Research Article (7 pages), Article ID 1918379, Volume 2022 (2022)




An Efficient Blockchain Based Data Access with Modified Hierarchical Attribute Access Structure with CP-ABE Using ECC Scheme for Patient Health Record

F. Sammy  and S. Maria Celestin Vigila
Research Article (11 pages), Article ID 8685273, Volume 2022 (2022)


Cognitive-Behavioral Therapy (CBT) Is Applied in Post-Traumatic Stress Disorder (PTSD) of Chinese Shidu Parents Who Lost Their Only Child

Guilin Yu , Hongfeng Liu, Chiang-Hanisko Lenny, Daijun Chen, Yin Yu, and Chanyuan Sun
Research Article (6 pages), Article ID 8001358, Volume 2022 (2022)



Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal

M. M. Venkata Chalapathi , M. Rudra Kumar , Neeraj Sharma, and S. Shitharth 
Research Article (10 pages), Article ID 8777026, Volume 2022 (2022)

Exploration of Environmental Protection-Oriented Ecoenvironmental Performance Audit System

Jie Wang , Xiaomei Wang, and Na Li
Research Article (10 pages), Article ID 2657411, Volume 2022 (2022)

Somewhat Homomorphic Encryption: Ring Learning with Error Algorithm for Faster Encryption of IoT Sensor Signal-Based Edge Devices

V. Subramaniaswamy, V. Jagadeeswari, V. Indragandhi, Rutvij H. Jhaveri , V. Vijayakumar, Ketan Kotecha , and Logesh Ravi
Research Article (10 pages), Article ID 2793998, Volume 2022 (2022)





Ecological Welfare of China's Forest Towns: Concept, Formation Mechanism, and Evaluation Index System

Chen Chen, Haitao Sun, and Yingli Huang 
Research Article (6 pages), Article ID 8948709, Volume 2022 (2022)

Healthcare Security Incident Response Strategy - A Proactive Incident Response (IR) Procedure


Ying He , Leandros Maglaras , Aliyu Aliyu , and Cunjin Luo 
Research Article (10 pages), Article ID 2775249, Volume 2022 (2022)

A Novel One-Shot Object Detection via Multifeature Auxiliary Information

Yu Song , Min Li , Weidong Du, Yao Gou , Zhaoqing Wu , and Yujie He
Research Article (9 pages), Article ID 6805526, Volume 2022 (2022)

Contents



Factors Influencing Green Entrepreneurship of Returning Migrant Workers under the Dual-Carbon Background

Li Beiwei, Yue Zhengliang , and Liu Hongtao
Research Article (10 pages), Article ID 7611810, Volume 2022 (2022)









An Experimental and Modeling Study on the Combustion of Gasoline-Ethanol Surrogates for HCCI Engines

Peng Yin, Wenfu Liu, Yong Yang, Haining Gao, and Chunhua Zhang 
Research Article (10 pages), Article ID 5362928, Volume 2022 (2022)


A New V-Net Convolutional Neural Network Based on Four-Dimensional Hyperchaotic System for Medical Image Encryption

Xiaowei Wang, Shoulin Yin, Muhammad Shafiq , Asif Ali Laghari, Shahid Karim, Omar Cheikhrouhou , Wajdi Alhakami, and Habib Hamam
Research Article (14 pages), Article ID 4260804, Volume 2022 (2022)




Traditional and Hybrid Access Control Models: A Detailed Survey

Muhammad Umar Aftab , Ali Hamza , Ariyo Oluwasanmi , Xuyun Nie , Muhammad Shahzad Sarfraz , Danish Shehzad , Zhiguang Qin , and Ammar Rafiq 
Review Article (12 pages), Article ID 1560885, Volume 2022 (2022)





A Blockchain-Assisted Electronic Medical Records by Using Proxy Reencryption and Multisignature

Xiaoguang Liu , Jun Yan, Shuqiang Shan, and Rongjun Wu
Research Article (13 pages), Article ID 6737942, Volume 2022 (2022)


iReTADS: An Intelligent Real-Time Anomaly Detection System for Cloud Communications Using Temporal Data Summarization and Neural Network

Gotam Singh Lalotra , Vinod Kumar , Abhishek Bhatt, Tianhua Chen, and Mufti Mahmud 
Research Article (15 pages), Article ID 9149164, Volume 2022 (2022)


Early Detection of Cognitive Decline Using Machine Learning Algorithm and Cognitive Ability Test

A. Revathi , R. Kaladevi , Kadiyala Ramana , Rutvij H. Jhaveri , Madapuri Rudra Kumar , and M. Sankara Prasanna Kumar 
Research Article (13 pages), Article ID 4190023, Volume 2022 (2022)

Analysis and Improvement of Blockchain-Based Multilevel Privacy-Preserving Location Sharing Scheme for Telecare Medical Information Systems

Zhenjie Huang , Yafeng Guo , Hui Huang , Runlong Duan , and Xiaolong Zhao 
Research Article (15 pages), Article ID 1926902, Volume 2022 (2022)

Differential Evolution and Multiclass Support Vector Machine for Alzheimer's Classification




Jhansi Rani Kaka  and K. Satya Prasad
Research Article (13 pages), Article ID 7275433, Volume 2022 (2022)

Research on Privacy Protection Technology of Mobile Social Network Based on Data Mining under Big Data

Jiawen Du  and Yong Pi 

Research Article (9 pages), Article ID 3826126, Volume 2022 (2022)

Cloud-Assisted Privacy-Preserving Method for Healthcare Using Adaptive Fractional Brain Storm Integrated Whale Optimization Algorithm

S. Thanga Revathi , A. Gayathri, J. Kalaivani, Mary Subaja Christo , Danilo Pelusi, and M. Azees 



Research Article (10 pages), Article ID 6210054, Volume 2021 (2021)

A Personalized Eccentric Cyber-Physical System Architecture for Smart Healthcare

Amutha Balakrishnan, Ramana Kadiyala , Gaurav Dhiman , Gokul Ashok, Sandeep Kautish , Kusum Yadav, and J. Maruthi Nagendra Prasad 

Research Article (36 pages), Article ID 1747077, Volume 2021 (2021)

A Novel Random Error Approximate Adder-Based Lightweight Medical Image Encryption Scheme for Secure Remote Monitoring of Health Data

Nagarajan Manikandan, Rajappa Muthaiah, Yuvaraja Teekaraman , Ramya Kuppusamy, and Arun Radhakrishnan 

Research Article (14 pages), Article ID 3570904, Volume 2021 (2021)

Blockchain and Business Process Management in Health Care, Especially for COVID-19 Cases

Ibrahim Abunadi  and R. Lakshmana Kumar 

Research Article (16 pages), Article ID 2245808, Volume 2021 (2021)

Research Article

Visual Attention and Motion Estimation-Based Video Retargeting for Medical Data Security

Qingfang Liu ^{1,2}, Baosheng Kang ¹, Qiaozhi Hua ³, Zheng Wen ⁴ and Haipeng Li ⁵

¹School of Information Science and Technology, Northwest University, Xi'an 710127, China

²Network Center, Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China

³Computer School, Hubei University of Arts and Science, Xiangyang 441000, China

⁴School of Fundamental Science and Engineering, Waseda University, Tokyo 1698050, Japan

⁵Capinfo Company Ltd, Beijing 100010, China

Correspondence should be addressed to Qiaozhi Hua; 11722@hbuas.edu.cn

Received 15 January 2022; Accepted 12 July 2022; Published 9 August 2022

Academic Editor: G. Thippa Reddy

Copyright © 2022 Qingfang Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medical data security is an important guarantee for intelligent medical system. Medical video data can help doctors understand the patients' condition. Medical video retargeting can greatly reduce the storage capacity of data on the premise of preserving the original content information as much as possible. The smaller volume of medical data can reduce the execution time of data encryption and threat detection algorithm and improve the performance of medical data security methods. The existing methods mainly focus on the temporal pixel relationship and foreground motion between adjacent frames, but these methods ignore the user's attention to the video content and the impact of background movement on retargeting, resulting in serious deformation of important content and area. To solve the above problems, this paper proposes an innovative video retargeting method, which is based on visual attention and motion estimation. Firstly, the visual attention map is obtained from eye tracking data, by K-means clustering method and Euclidean distance factor equation. Secondly, the motion estimation map is generated from both the foreground and background displacements, which are calculated based on the feature points and salient object positions between adjacent frames. Then, the visual attention map, the motion estimation map, and gradient map are fused to the importance map. Finally, video retargeting is performed by mesh deformation based on the importance map. Experiment on open datasets shows that the proposed method can protect important area and has a better effect on salient object flutter suppression.

1. Introduction

With the rapid development of high-tech medical imaging [1–3], blockchain technology [4], artificial intelligence [5], Internet of Things (IoT) [6], and 5G network [7], intelligent medical system [8] and intelligent diagnosis [9] are becoming more and more popular. However, data security threats [10] make protecting the security of medical data an urgent problem. The volume of medical video data is greater than typical data, which makes the execution of medical data security methods, such as data encryption [11] and integrity detection [12], long. Video retargeting [13] can greatly reduce the storage capacity of video data on the premise of preserving the original content information as much as possible. Medical video retargeting can obtain smaller volume

of medical data, then reduce the execution time of data encryption and threat detection algorithm, and improve the performance of medical data security methods.

Traditional image and video retargeting methods, mainly including uniform scaling and direct cropping, only consider the original size and the target size of images, without considering image content. Their effects are unsatisfactory. To improve image and video retargeting performance, researchers proposed content-aware retargeting techniques, which are mainly classified into three types: discrete retargeting [14, 15], continuous retargeting [16–19], and multi-operator retargeting [20–23].

Video retargeting has one more time dimension than image retargeting. It needs to take consideration of the

correlation between the contents of adjacent frames. By regarding video as a three-dimensional pixel space-time matrix, Rubinstein et al. proposed FSC [15], looking for and deleting common pixel seams between adjacent frames to eliminate content jitter. NCV [17] proposed by Wolf et al. combines the gradient map, face detection, and foreground motion to produce importance map and then uses mesh deformation to realize video retargeting. Nam et al. [24] proposed a video retargeting method based on Kalman filter and saliency fusion to reduce video content jitter, so as to enhance the robustness of video retargeting. Wang et al. [25] proposed a multi-operator method based on improved seam carving to realize video retargeting. Cho and Kang [26] proposed an interpolation video retargeting method based on image deformation vector network, which uses the displacement vector generated by a convolutional neural network to perform interpolation. Kaur et al. [27] proposed a spatiotemporal seam carving video retargeting method based on Kalman filter.

The existing video retargeting methods mainly focus on the pixel relationship and foreground motion between adjacent frames. These methods aim to ensure the shape of important content in the process of retargeting. However, the above methods do not consider the attention of users to the video content, nor the impact of background movement on retargeting, resulting in serious deformation of the important content or poor quality of retargeting results. Furthermore, the human visual system can quickly find the required information from the visual scene and locate the visual attention to the focus in the scene [28]. Consequently, besides moving objects and important targets, the attention focus also includes the areas where change is about to happen next moment, such as the place where the sun will rise before sunrise, the place where actors will appear on the stage before the performance, and the direction where the ball is moving to.

This paper makes full use of the user's eye tracking data and the motion information of both the background and foreground in the video and proposes a video retargeting method based on visual attention and motion estimation to reduce the deformation of the important area. Firstly, clustering is carried out according to the eye tracking data to generate the visual attention energy map. Then, the motion estimation map is obtained according to the corresponding feature points of the foreground and background between adjacent frames. Thirdly, importance map is generated by composing visual attention energy map, motion estimation energy map, and gradient map. Finally, video retargeting is performed by mesh deformation.

The proposed method utilizes the attention attribute of the human visual system and the movement factor of content in video, so the retargeting result is more in line with people's visual requirements. The experimental results on public datasets show that the method in this paper is better

than the compared method in protecting important area and reducing salient object jitter.

2. Proposed Method

As shown in Figure 1, the framework of the proposed VAMEVR (visual attention and motion estimation-based video retargeting) method mainly includes visual attention data clustering, saliency detection, SIFT feature detection, motion estimation, mesh deformation, and so on.

2.1. Visual Attention. In a video, the areas concerned by the human visual system are usually regarded as important areas. These areas should be of increased energy to reduce deformation in the retargeting process. In this paper, the eye tracking data will be utilized as the basis of visual attention, and it will be abstracted as visual focus. Then, visual attention energy will be generated according to the visual focus.

2.1.1. Visual Attention Focus. This paper takes the eyeball tracking data of DAVSOD [29] dataset as demonstration. As shown in Figure 2, the eyeball tracking data exist in the form of discrete points. Through observation, it is found that most eyeball tracking data points are presented as two clusters.

In this paper, the K-means method [30] is utilized to cluster the eyeball tracking data points into 2 groups. The center of each group is just the visual focus. Firstly, we randomly select 2 data points as the initial cluster centroid. Secondly, we divide the data points into 2 mutually exclusive clusters according to the Euclidean distance from each point to the initial selected data points. Thirdly, the average positions of each cluster are obtained as the new cluster centroid. Then, repeat steps 2 and 3 until the centroid position does not vary.

The example of the focusing result is presented in Figure 3. Figure 3(a) shows the original frame. Figure 3(b) shows the eye tracking data and focusing result. The white point is regarded as the eye tracking data, and the two red points are the centers of two clusters. Figure 3(c) shows the visual attention energy map.

2.1.2. Visual Attention Energy. Visual attention energy indicates the attention of the human visual system to important position in the image. The greater the energy is, the higher the attention is, and vice versa.

Two cluster centroids described in Section 2.1.1 are denoted as $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$. The distances from each pixel of the frame to P_1 and P_2 are separately set as r_1 and r_2 . Then, visual attention energy $e(x_i, y_j)$ of each pixel position in the frame is defined as

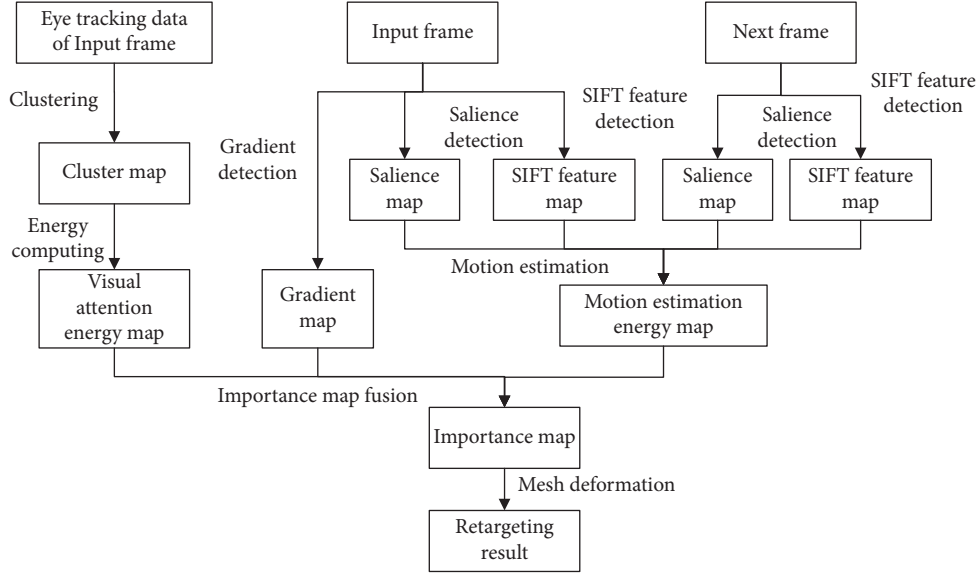


FIGURE 1: The framework of the proposed VAMEVR method.

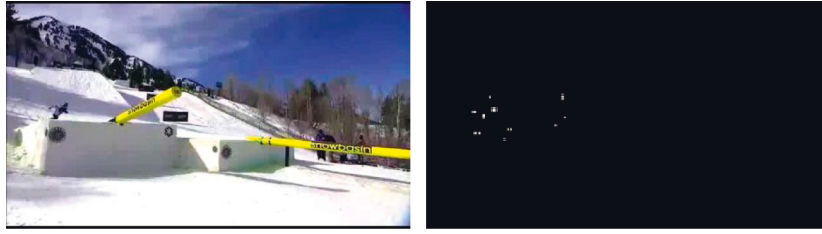


FIGURE 2: Eye tracking data example in DAVSOD dataset.

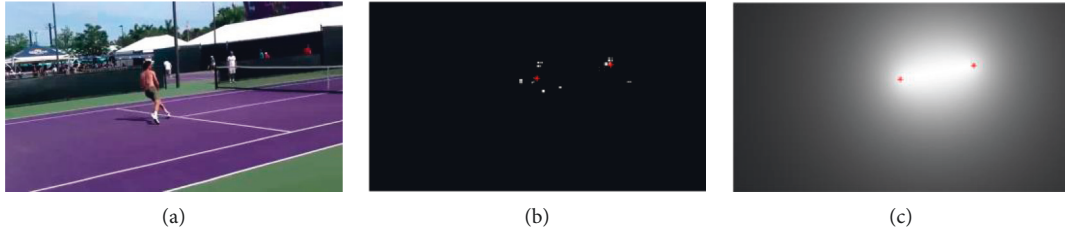


FIGURE 3: Visual attention energy map.

$$e(x_i, y_i) = \frac{\sqrt{W^2 + H^2}}{r_1 + r_2} = \frac{\sqrt{W^2 + H^2}}{\sqrt{(x_i - x_1)^2 + (y_i - y_1)^2} + \sqrt{(x_i - x_2)^2 + (y_i - y_2)^2}}, \quad (1)$$

where W and H are separately the width and height of the video frame. The generated energy map is shown in Figure 3. Figure 3(c) shows the visual attention energy map, which is generated according to the cluster results of eye tracking data as shown in Figure 3(b).

2.2. Motion Estimation. In a video, the background and foreground are usually moving. In addition, the moving direction and speed of background are different from those of the foreground. The human visual system pays greater

attention to the direction where the object is going. For example, in the tennis video, the direction where the players run to will attract more attention. In the racing video, area in front of the car is paid more attention.

Between adjacent video frames, the motion distance and direction of the background and foreground can be calculated to predict the motion trajectory of the salient object. Both current position and the upcoming position of the foreground object are taken as important areas at the same time, which can protect the visual attention areas to reduce the deformation of these important areas in the process of

retargeting and improve the visual effect of retargeting results.

2.2.1. Feature Detection. In the background of a video frame, the mean values of displacement of the feature points are used as the base of moving speed. The same is for the foreground of a video. The position to be reached by the foreground significant object is estimated according to the moving speed. Then, both the current position of the foreground and the position to be reached after motion estimation are regarded as important areas.

SIFT (scale-invariant feature transform) [31] is a computer vision algorithm proposed by Lowe to detect regional features in images. The core idea of SIFT algorithm is to find extreme points in multiple spatial scales and calculate position, rotation, light, and scale invariants to describe the features in images. The SIFT algorithm has good robustness, recognition, expansibility, and efficiency.

In this paper, SIFT algorithm feature detection is used to detect the background and foreground motion information between adjacent frames. Also, 20 feature points with the highest reliability are selected as the basis for motion speed calculation. An example of feature points is shown in Figure 4.

2.2.2. Foreground Separation. In a video frame, salient object is generally the foreground area. By salience detection, the foreground area can be separated from the background. Compared with other algorithms, SSAV [29] can obtain clearer and more accurate result. SSAV [29] is mainly composed of a pyramid deconvolution module and salience transfer perception module. The former is used to robustly learn static salience features. The latter combines the traditional long-term memory convolution network with salience transfer perception attention mechanism. This paper uses the SSAV [29] method to separate the salient foreground object from video frames.

2.2.3. Motion Detection and Estimation. From SIFT feature points, we select n ($n = 20$) point with high reliability as the basis for motion detection and estimation. Concretely, SIFT feature points contained in the background are recorded as $P_{bg}(x_{bg}, y_{bg})$, and the number of those points is n_{bg} . Similarly, SIFT feature points contained in the foreground are recorded as $P_{fg}(x_{fg}, y_{fg})$, and the number is n_{fg} . From frame i to frame $i + 1$, the average moving speed of feature points in the background is recorded as $V_{bg}(dx_{bg}, dy_{bg})$.

$$\begin{cases} dx_{bg} = \frac{\sum_{j=1}^{n_{bg}}(x_{bg}^{i+1} - x_{bg}^i)}{n_{bg}}, \\ dy_{bg} = \frac{\sum_{j=1}^{n_{bg}}(y_{bg}^{i+1} - y_{bg}^i)}{n_{bg}}. \end{cases} \quad (2)$$

Similarly, from frame i to frame $i + 1$, the average value of the moving speed of the feature points in the foreground is denoted as $V_{fg}(dx_{fg}, dy_{fg})$.

$$\begin{cases} dx_{fg} = \frac{\sum_{j=1}^{n_{fg}}(x_{fg}^{i+1} - x_{fg}^i)}{n_{fg}}, \\ dy_{fg} = \frac{\sum_{j=1}^{n_{fg}}(y_{fg}^{i+1} - y_{fg}^i)}{n_{fg}}. \end{cases} \quad (3)$$

For a video, the estimated actual motion speed $V_{act}(dx_{act}, dy_{act})$ of the foreground is defined as the difference between the motion speed of the foreground and background.

$$V_{act} = V_{fg} - V_{bg}. \quad (4)$$

Bring equations (2) and (3) into equation (4).

$$\begin{cases} dx_{act} = \frac{\sum_{j=1}^{n_{fg}}(x_{fg}^{i+1} - x_{fg}^i)}{n_{fg}} - \frac{\sum_{j=1}^{n_{bg}}(x_{bg}^{i+1} - x_{bg}^i)}{n_{bg}}, \\ dy_{act} = \frac{\sum_{j=1}^{n_{fg}}(y_{fg}^{i+1} - y_{fg}^i)}{n_{fg}} - \frac{\sum_{j=1}^{n_{bg}}(y_{bg}^{i+1} - y_{bg}^i)}{n_{bg}}. \end{cases} \quad (5)$$

As shown in Figure 5, after obtaining the salience map of the current frame, we calculate the edge of the salient region by the Canny [32] method. Then, the edge is overlaid with the actual motion speed $V_{act}(dx_{act}, dy_{act})$ as the predicted position of the salient object. The polygon surrounding method [33] is used to obtain the external polygon of both current and predicted object contour. Finally, the area surrounded by the polygon is just the important region after motion estimation. The motion estimation energy map is the binary map of important area after motion estimation, which is shown in Figure 5(d).

When the salient object is too small or the features are not obvious, the first n ($n = 20$) feature points detected by the SIFT algorithm are wholly in the background area. In this situation, the centroid displacement of the salience object detected by SSAV is directly used as the moving speed of the foreground object to predict the position where the foreground will go.

The points in salient object area are denoted as $P_{fg}^*(xc_{fg}, yc_{fg})$. The number of those points is m_{fg} . From frame i to frame $i + 1$, the motion speed of the foreground's centroid is denoted as $V_{fg}^*(dx_{c_{fg}}, dy_{c_{fg}})$, where

$$\begin{cases} dxc_{fg} = \frac{\sum_{l=1}^{m_{fg}^{i+1}}(xc_{fg}^{i+1})}{m_{fg}^{i+1}} - \frac{\sum_{q=1}^{m_{fg}^i}(xc_{fg}^i)}{m_{fg}^i}, \\ dyc_{fg} = \frac{\sum_{l=1}^{m_{fg}^{i+1}}(yc_{fg}^{i+1})}{m_{fg}^{i+1}} - \frac{\sum_{q=1}^{m_{fg}^i}(yc_{fg}^i)}{m_{fg}^i}. \end{cases} \quad (6)$$

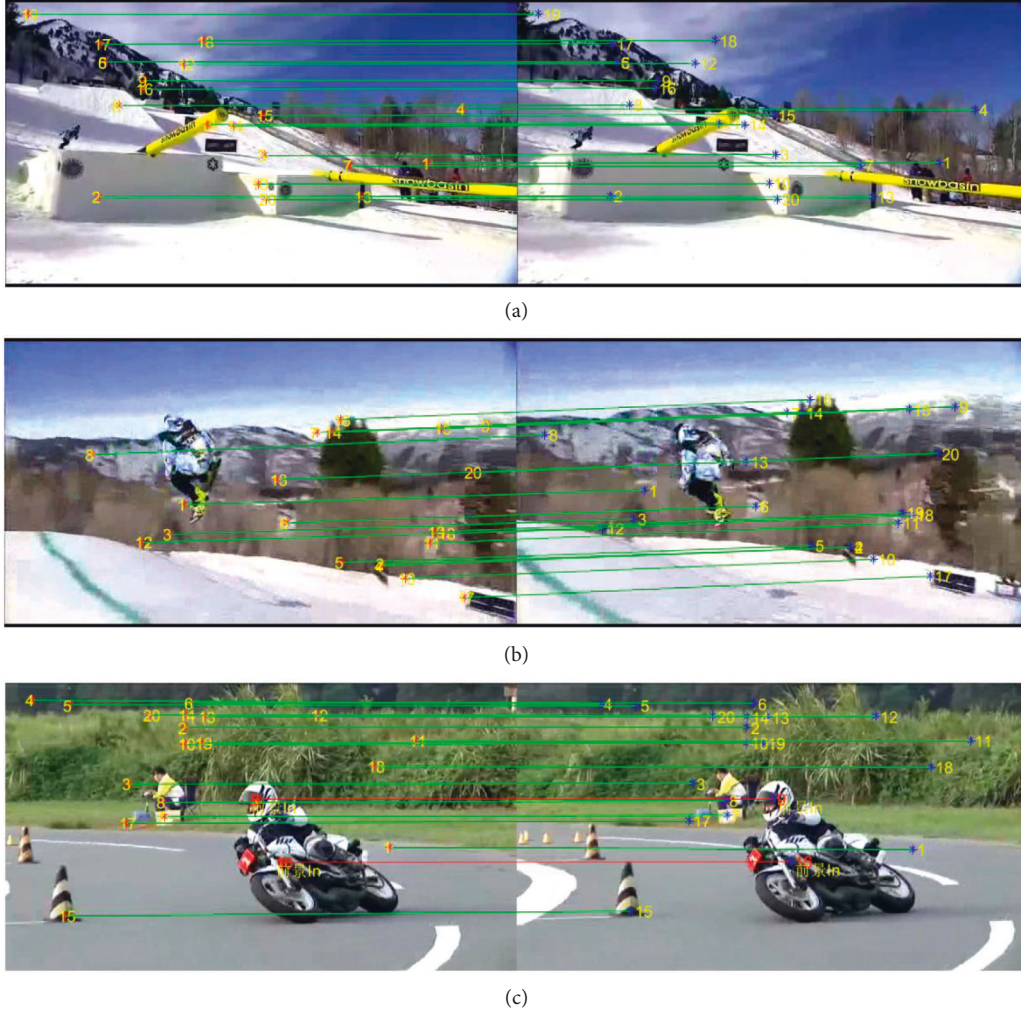


FIGURE 4: SIFT detection between adjacent frames. In (a)–(c), the green lines indicate the connection of feature points in the background, and the red lines indicate the connection of feature points in the foreground. All feature points in (a) and (b) belong to the background area. In (c), only one feature point is in the foreground area and the other feature points are in the background area.

The actual motion speed $V_{act}(dx_{act}, dy_{act})$ of the foreground is the difference between the motion speed of the foreground and the motion speed of the background.

$$V_{act} = V_{fg}^* - V_{bg}. \quad (7)$$

Bring (2) and (6) into (7).

$$\begin{cases} dx_{act} = \left(\frac{\sum_{l=1}^{m_{fg}^{i+1}}(xc_{fg}^{i+1})}{m_{fg}^{i+1}} - \frac{\sum_{q=1}^{m_{fg}^i}(xc_{fg}^i)}{m_{fg}^i} \right) - \frac{\sum_{j=1}^{n_{bg}}(x_{bg}^{i+1} - x_{bg}^i)}{n_{bg}}, \\ dy_{act} = \left(\frac{\sum_{l=1}^{m_{fg}^{i+1}}(yc_{fg}^{i+1})}{m_{fg}^{i+1}} - \frac{\sum_{q=1}^{m_{fg}^i}(yc_{fg}^i)}{m_{fg}^i} \right) - \frac{\sum_{j=1}^{n_{bg}}(y_{bg}^{i+1} - y_{bg}^i)}{n_{bg}}. \end{cases} \quad (8)$$

2.3. Importance Map Fusion. The importance map is the direct basis for image retargeting. The visual attention energy map and motion estimation map obtained in the above steps need to be fused into the importance map.

We denote I_{eye} as the normalized visual attention energy map, I_{grad} as the normalized gradient energy map, I_{motion} as the normalized motion estimation energy map, and I_{imp} as the importance map. The coefficient

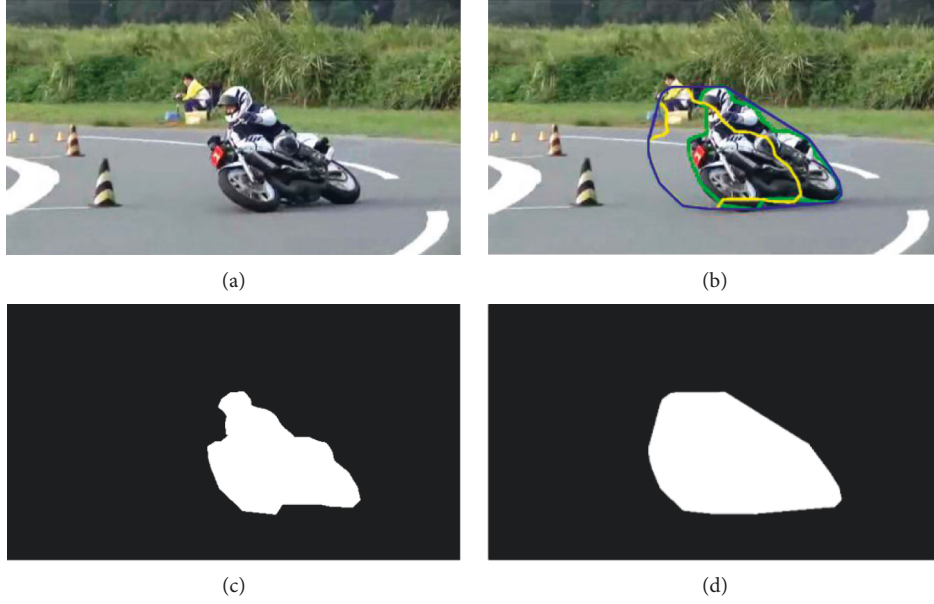


FIGURE 5: Motion estimation. (a) The original frame. In (b), the green contour is the current position of the salient object, the yellow contour is the predicted place of salient object, and the blue contour is the encirclement of both the current and predict areas of salient object after motion estimation. (c) The saliency map of (a). (d) The binary map of important area after motion estimation.

w ($0 \leq w \leq 1$) is the weight of the visual attention energy map in the importance map, over the gradient energy map. Then, the importance map I_{imp} is defined as follows.

$$I_{\text{imp}} = \max\left(\left(w \times I_{\text{eye}} + (1 - w) \times I_{\text{grad}}\right), I_{\text{motion}}\right). \quad (9)$$

The parameter w ($0 \leq w \leq 1$) determines the visual effect of visual attention energy in importance map. The smaller w is, the smaller the proportion of visual attention energy is. Thus, the impact of visual attention on the results in the retargeting process is smaller, and vice versa. The larger w is, the greater the proportion of visual focus energy is. Therefore, the impact of visual attention on the results in the retargeting process is greater. When $w = 0$, the retargeting results only reflect the gradient information and motion estimation information, not the visual attention information. Also, when $w = 1$, the retargeting results only reflect the visual attention information and motion estimation information, not the visual attention information.

2.4. Mesh Deformation. This paper uses Wang's method [18] for mesh deformation to realize video retargeting. The input frame is divided into quadrilateral mesh (V, E, F) . V , E , and F represent the set of vertex, edge, and quadrilateral separately. Each quadrilateral is with a scaling factor s_f . The average importance energy of each quad is w_f . The quad deformation energy is defined as D_u .

$$D_u = \sum_{f \in F} w_f \left(\sum_{(i,j) \in E(f)} \left\| (v'_i - v'_j) - s_f(v_i - v_j) \right\|^2 \right). \quad (10)$$

The grid line bending energy is described as D_l ,

$$D_l = \sum_{(i,j) \in E} \left\| (v'_i - v'_j) - \left(\frac{\|v'_i - v'_j\|}{\|v_i - v_j\|} \right) (v_i - v_j) \right\|^2. \quad (11)$$

The total energy D is the sum of D_u and D_l .

$$D = D_u + D_l. \quad (12)$$

Wang's method [18] uses iterative solver to solve for mesh deformation. In each iteration, the scaling factor s_f of each grid is calculated by local optimization, and then the mesh vertexes are updated by global optimization under the constraint of target image boundary conditions. The iterator will be terminated when the energy is no longer increased or the displacement of mesh vertexes is less than 0.5. The smooth scaling factors s'_f are generated by minimizing the following energy.

$$\sum_{f \in F} w_f \sum_{q \in N(f)} \frac{1}{2} (w_f - w_q) (s'_f - s'_q) + \sum_{f \in F} w_f (s'_f - s_f)^2. \quad (13)$$

2.5. The Algorithm of the Proposed Method. The implementation steps of the proposed methods are shown in Algorithm 1.

3. Results and Analysis

3.1. Experimental Environment and Parameter Settings. To validate the performance of the proposed method, we conduct experiments on a computer with an Intel i7-5500U@2.4GHz CPU and 16GB RAM. The proposed method was implemented in MATLAB R2016a on Windows.

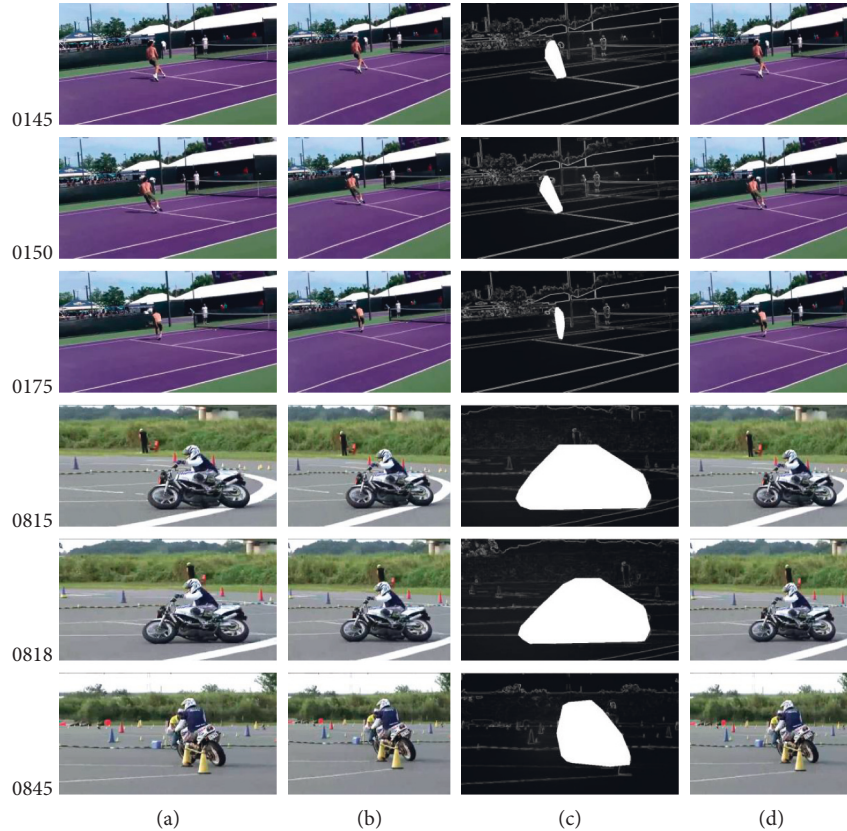


FIGURE 6: Horizontal retargeting to 75%. (a) Input frame. (b) Result of SNS [18]. (c) Our importance map with $w = 0.1$. (d) Our result with $w = 0.1$.

The number of visual attention data cluster k is set as 2. In the important map fusion process, the weight w of visual attention is set as 0.1, 0.5, and 0.9 separately.

In order to illustrate the universality of proposed method, the public dataset DAVSOD [29] is selected as experimental input. DAVSOD is a large-scale video salient object dataset, which mainly serves the evaluation of video salient object detection and video retargeting. DAVSOD contains 226 video sequences and 24000 frames, covering a variety of scenes, object categories, and motion modes. It is marked strictly according to human eye tracking data.

For each dataset, 3 methods were applied for comparison experiments: forward seam carving (FSC) [15], SNS [18], and the proposed VAMEVR.

3.2. Experimental Result and Analysis. We randomly select “select_0115” and “select_0194” videos of DAVSOD as input data of experiment. The data of “select_0115” include a tennis video clip with 105 frames and 640×360 pixels per frame. The data of “select_0194” include a motorcycle race video clip, with 133 frames in total, and the size of each frame is 640×360 . In both of the above video data, the camera is moving during video shooting, that is, the background is moving.

The experimental results are shown in Figures 6 and 7.

From Figures 6 and 7, we can find that the deformation of the salient area is small, especially the area in the direction

the object moves to, which is well protected. As shown in Figure 7(d) concretely, the region, where the tennis ball in “0145” video frame is moving to, is with smaller deformation, and so is the area in front of the motorcycle in “0818” video frame.

The main reason of above results is that the important area is of high energy by visual attention and motion estimation. In “0145” and “0150” of Figure 7(c), it can be seen that people paid more attention to the direction of the ball the player was going to move. Similarly, in “0815” and “0818” of Figure 7(c), people pay more attention to the forward direction of the motorcycle and less attention to the rear direction of the motorcycle.

Specifically, as shown in Figures 6(c), 7(a), and 7(c), the smaller w is, the weaker the effect of the visual attention is. The larger w is, the more obvious the effect of visual attention is.

3.3. Time Analysis. The size of video frames and average processing time of each frame in this paper are shown in Table 1.

It can be seen from Table 1 that the time of FSC is longest, with 6.03 s per frame. The average time per frame of VAMEVR in this paper is 0.53 s. It is 0.24 seconds longer than SNS. The increased time is mainly used to calculate visual attention energy and motion estimation detection.

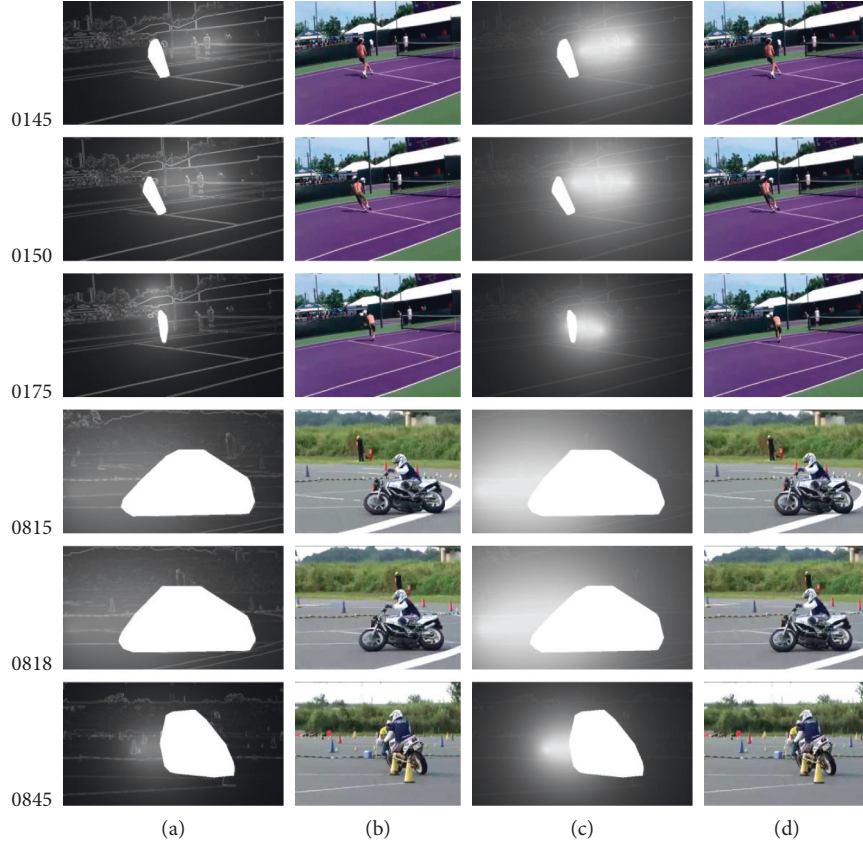


FIGURE 7: Horizontal retargeting to 75%. (a) Our importance map with $w = 0.5$. (b) Our result with $w = 0.5$. (c) Our importance map with $w = 0.9$. (d) Our result with $w = 0.9$.

Input: original video V_{input} , the number of frames K , important map fusion coefficient parameter w

Output: retargeting result video V_{result}

For $i = 1$ to $K - 1$ **do**

 Calculate the two cluster centers of eye tracking data of $Frame_i$ by K-means method

 Use (1) to produce the visual attention energy map I_{eye} of $Frame_i$

 Significant object separation of $Frame_i$ and $Frame_{i+1}$ by SSAV [29] model

 Calculate the position of corresponding features of $Frame_i$ and $Frame_{i+1}$ by SIFT method

 Get the number of trusted feature points in foreground and denote it as n_{fg}

If $n_{fg} \geq 1$

 Calculate the background speed V_{bg} between $Frame_i$ and $Frame_{i+1}$ by (2)

 Calculate the foreground speed V_{fg} between $Frame_i$ and $Frame_{i+1}$ by (3)

 Calculate actual moving speed V_{act} of the salient object by (4) and (5)

Else

 Calculate the background speed V_{bg} between $Frame_i$ and $Frame_{i+1}$ by (2)

 Calculate the foreground speed V_{fg}^* between $Frame_i$ and $Frame_{i+1}$ by (6)

 Calculate actual moving speed V_{act} of the salient object by (7) and (8)

End If

 Estimate the position of foreground $(x_{\text{esti}}, y_{\text{esti}}) = (x_{\text{cur}}, y_{\text{cur}}) + V_{\text{act}}$

 Calculate the circumscribed polygon R_{fg} of both the estimated position and current position of the foreground

 Generate the foreground motion estimation map I_{motion} according to the salient areas S_r in polygon R_{fg}

 Compose importance map I_{imp} from visual attention energy map I_{eye} , foreground motion estimation map I_{motion} , and gradient map I_{grad} by (9)

 Use the mesh deformation method described in Section 2.4 to produce retargeting result of $Frame_i$

End for

Output result V_{result}

ALGORITHM 1: Video retargeting based on visual attention and motion estimation.

TABLE 1: Execution time of different methods.

| Frame size | SNS (s/frame) | FSC (s/frame) | Proposed VAMEVR (s/frame) |
|------------|---------------|---------------|---------------------------|
| 640 × 360 | 0.29 | 6.03 | 0.53 |

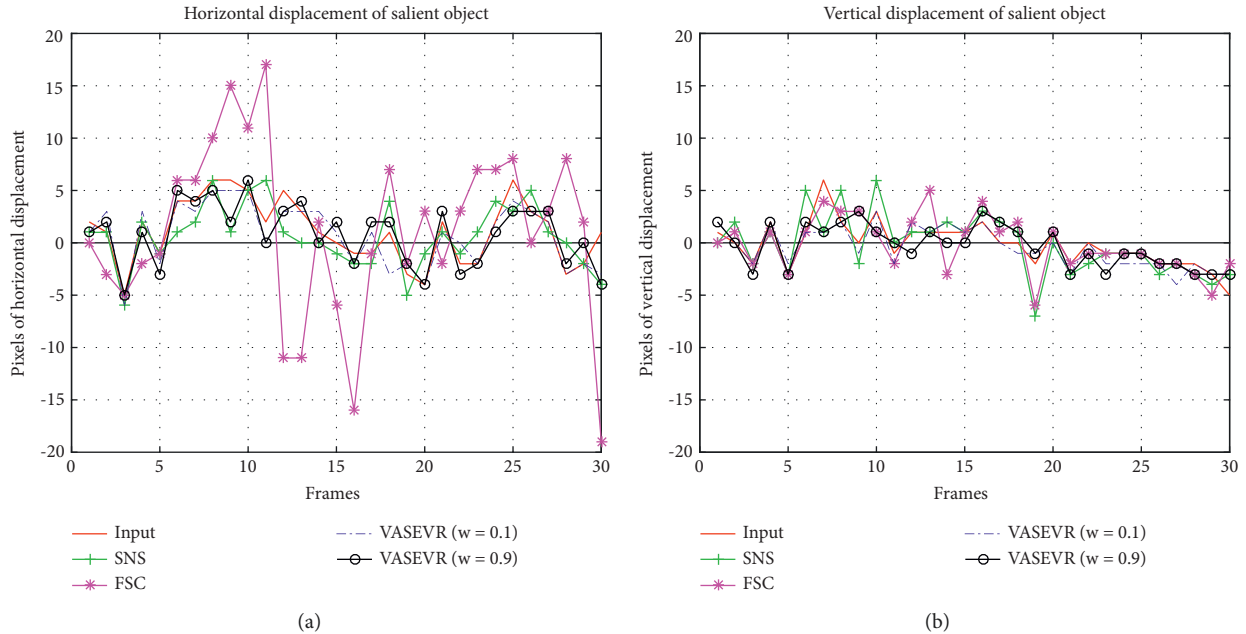


FIGURE 8: Comparison of displacement of salient objects before and after retargeting. (a) Horizontal. (b) Vertical.

TABLE 2: Correlation of salient object displacement before and after retargeting.

| Direction methods | SNS | FSC | Proposed VAMEVR | | |
|-------------------|---------|---------|-----------------|-----------|-----------|
| | | | $w = 0.1$ | $w = 0.5$ | $w = 0.9$ |
| Horizontal | 0.69185 | 0.26246 | 0.89489 | 0.87823 | 0.86542 |
| Vertical | 0.72326 | 0.73599 | 0.84416 | 0.82361 | 0.78051 |

3.4. Discussion. The human visual system is more sensitive to salient objects. The more consistent the displacement of salient objects in adjacent frames before and after retargeting, the lower the content jitter. In this paper, 30 frames of motorcycle racing videos are randomly selected for retargeting.

For the proposed VAMEVR, the centroid displacement of salient object in the retargeting result is basically the same as that of original video. When the weight coefficient of visual attention energy map is 0.1 and 0.9, the comparative analysis of horizontal and vertical displacement is shown in Figure 8.

The displacement correlation of the salient objects can indicate the visual consistency between the original video and the retargeting result. The displacement of the centroid of the significant object in input video and retargeting result is denoted as X and Y separately. The covariance is defined as $cov(X, Y)$, and the standard deviation of X and Y is (σ_X, σ_Y) . The Pearson correlation coefficient $\rho_{X,Y}$ is defined as follows.

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}. \tag{14}$$

As shown in Table 2, for VAMEVR, the displacements of the salient objects before and after retargeting are more positively correlated than SNS and FSC. The visual effects of our results are more consistent with the original video than SNS and FSC.

4. Conclusion

This paper proposes a visual attention and motion estimation-based video retargeting method for medical data security. Firstly, clustering is carried out according to the eye tracking data to generate the visual attention energy map. Then, the motion estimation map is obtained according to the corresponding feature points of the foreground and background between adjacent frames. Thirdly, importance map is generated by composing visual attention energy map, motion estimation map, and gradient map. Finally, video

retargeting is performed by mesh deformation. Experiments show that the proposed method can protect important area concerned by the human visual system. The displacement of a salient object in retargeting results is more close to input video. Therefore, the visual effect is more in line with human visual need. Our future work is to study the multi-object separation method and then study the video retargeting method based on multi-object motion estimation for medical data security.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Hubei Natural Science Foundation under grant no. 2021CFB156 and the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under grant no. JP21K17737.

References

- [1] F. Ding, G. Zhu, Y. Li, X. Zhang, P. Atrey, and S. Lyu, "Anti-forensics for face swapping videos via adversarial training," *IEEE Transactions on Multimedia*, vol. 2021, Article ID 3098422, 2021.
- [2] A. R. Javed, Z. Jalil, W. Zehra, T. Gadekallu, D. Y. Suh, and M. J. Piran, "A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions," *Engineering Applications of Artificial Intelligence*, vol. 106, Article ID 104456, 2021.
- [3] H. Li, Q. Zheng, J. Zhang, Z. Du, Z. Li, and B. Kang, "Pix2Pix-Based grayscale image coloring method," *Journal of Computer-Aided Design & Computer Graphics*, vol. 33, no. 6, pp. 929–938, 2021.
- [4] W. Wang, Q. Chen, Z. Yin et al., "Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8883–8891, 2022.
- [5] W. Wang, M. H. Fida, Z. Lian et al., "Secure-enhanced federated learning for ai-empowered electric vehicle energy prediction," *IEEE Consumer Electronics Magazine*, vol. 2021, Article ID 3116917, 2021.
- [6] H. Li, Q. Zheng, W. Yan, R. Tao, X. Qi, and Z. Wen, "Image super-resolution reconstruction for secure data transmission in Internet of Things environment," *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 6652–6671, 2021.
- [7] L. Tan, K. Yu, L. Lin et al., "Speech emotion recognition enhanced traffic efficiency solution for autonomous vehicles in a 5G-enabled space-air-ground integrated intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2830–2842, 2022.
- [8] A. Balakrishnan, R. Kadiyala, G. Dhiman et al., *A Personalized Eccentric Cyber-Physical System Architecture for Smart Healthcare Security and Communication Networks*, vol. 2021, Article ID 1747077, 2021.
- [9] W. L. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, "Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: a big-data analysis," *Applied Energy*, vol. 285, Article ID 116429, 2020.
- [10] L. Tan, K. Yu, F. Ming, X. Cheng, and G. Srivastava, "Secure and resilient artificial intelligence of Things: a HoneyNet approach for threat detection and situational awareness," *IEEE Consumer Electronics Magazine*, vol. 11, no. 3, pp. 69–78, Article ID 3081874, 2022.
- [11] C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv, and S. Mumtaz, "Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13784–13795, 2020.
- [12] K. Yu, L. Tan, S. Mumtaz et al., "Securing critical infrastructures: deep-Learning-Based threat detection in IIoT," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 76–82, 2021.
- [13] A. Garg, A. Negi, and P. Jindal, "Structure preservation of image using an efficient content-aware image retargeting technique," *Signal, Image and Video Processing*, vol. 15, no. 1, pp. 185–193, 2021.
- [14] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM SIGGRAPH 2007 papers on - SIGGRAPH '07*, vol. 26, no. 3, pp. 10–es, 2007.
- [15] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 1–9, 2008.
- [16] Y. Kim, S. Jung, C. Jung, and C. Kim, "A structure-aware axis-aligned grid deformation approach for robust image retargeting," *Multimedia Tools and Applications*, vol. 77, no. 6, pp. 7717–7739, 2018.
- [17] L. Wolf, M. Guttman, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proceedings of the IEEE 11th International Conference on Computer Vision*, pp. 1–6, IEEE, Rio de Janeiro, Brazil, December 2007.
- [18] Y. S. Wang, C. L. Tai, O. Sorkine, and T. Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM SIGGRAPH Asia 2008 papers on - SIGGRAPH Asia '08*, vol. 27, no. 5, p. 118, 2008.
- [19] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *Computer Graphics Forum*, vol. 28, no. 5, pp. 1257–1268, 2009.
- [20] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 1–11, 2009.
- [21] A. Garg and A. Negi, "Structure preservation in content-aware image retargeting using multi-operator," *IET Image Processing*, vol. 14, no. 13, pp. 2965–2975, 2020.
- [22] Y. Zhou, Z. Chen, and W. Li, "Weakly supervised reinforced multi-operator image retargeting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 1, pp. 126–139, 2021.
- [23] M. Abhayadev and T. Santha, "Multi-operator content aware image retargeting on natural images," *Journal of entific and industrial research*, vol. 78, no. 1, pp. 193–198, 2019.
- [24] H. Nam, D. Park, and K. Jeon, "Jitter-Robust video retargeting with kalman filter and attention saliency fusion network," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 858–862, IEEE, Abu Dhabi, UAE, September 2020.
- [25] S. Wang, Z. Tang, W. Dong, and J. Yao, *Multi-Operator Video Retargeting Method Based on Improved Seam Carving*, IEEE,

- in *Proceedings of the IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 1609–1614, July 2020.
- [26] S. I. Cho and S. J. Kang, “Extrapolation-based video retargeting with backward warping using an image-to-warping vector generation network,” *IEEE Signal Processing Letters*, vol. 27, no. 1, pp. 446–450, 2020.
- [27] H. Kaur, S. Kour, and D. Sen, “Video retargeting through spatio-temporal seam carving using Kalman filter,” *IET Image Processing*, vol. 13, no. 11, pp. 1862–1871, 2019.
- [28] A. Borji, D. N. Sihite, and L. Itti, “What stands out in a scene? A study of human explicit saliency judgment,” *Vision Research*, vol. 91, no. 15, pp. 62–77, 2013.
- [29] D. P. Fan, W. Wang, M. Cheng, and J. Shen, “Shifting more attention to video salient object detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8554–8564, IEEE, Long Beach, CA, USA, January 2019.
- [30] J. A. Hartigan and M. A. Wong, “Algorithm as 136: a K-means clustering algorithm,” *Applied statistics*, vol. 28, no. 1, pp. 100–108, 1979.
- [31] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [32] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [33] K. Hormann and A. Agathos, “The point in polygon problem for arbitrary polygons,” *Computational Geometry*, vol. 20, no. 3, pp. 131–144, 2001.

Research Article

k Nearest Neighbor Similarity Join Algorithm on High-Dimensional Data Using Novel Partitioning Strategy

Youzhong Ma ^{1,2}, Qiaozhi Hua ³, Zheng Wen ⁴, Ruiling Zhang ¹, Yongxin Zhang ¹,
and Haipeng Li ⁵

¹School of Information and Technology, Luoyang Normal University, Luoyang 471934, China

²Henan Key Laboratory for Big Data Processing & Analytics of Electronic Commerce, Luoyang 471934, China

³Computer School, Hubei University of Arts and Science, Xiangyang 441000, China

⁴School of Fundamental Science and Engineering, Waseda University, Tokyo 169-8050, Japan

⁵Capinfo Company Ltd., Beijing 100010, China

Correspondence should be addressed to Qiaozhi Hua; 11722@hbuas.edu.cn

Received 17 January 2022; Accepted 22 March 2022; Published 26 April 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Youzhong Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

k nearest neighbor similarity join on high-dimensional data has broad applications in many fields; several key challenges still exist for this task such as “curse of dimensionality” and large scale of the dataset. A new dimensionality reduction scheme is proposed by using random projection technique, then we design two novel partition strategies, including equal width partition strategy and distance split tree-based partition strategy, and finally, we propose k nearest neighbor join algorithm on high-dimensional data based on the above partition strategies. We conduct comprehensive experiments to test the performance of the proposed approaches, and the experimental results show that the proposed methods have good effectiveness and performance.

1. Introduction

With the rapid development of emerging technologies such as big data [1, 2], Internet of things [3, 4], Deep Learning [5, 6], Adversarial Training [7, 8], Federated Learning [9], and 5G [10–12], the smart healthcare systems are becoming more and more pervasive and necessary in modern hospitals, and massive and diverse medical data have been accumulated by using a great volume of wearable sensors, the Internet of Medical Things [13, 14], or the Internet of Health Things [15, 16]. Medical data analysis, security, and privacy protection [17] are very important for using massive medical data. Many research works have been done, which can be used as the references to analyze the medical data and secure the Internet of Medical Things, such as the blockchain-based security approaches [18, 19], the security technologies in Internet of Internet of Vehicles [20, 21], Industrial Internet of Things [22–25], Artificial Intelligence of Things [26–28], Energy Internet [29], and Intelligent transportation [30, 31]. Similarity join operation plays an important role in medical

data analysis [32, 33]. Threshold-based similarity join query on high-dimensional data can obtain all the data pairs whose distance is less than or equal to the given distance threshold, and it needs to know the distance threshold in advance; however, in many cases, it is hard or even impossible to get the distance threshold in advance, while k nearest neighbor similarity join (k NNJ) does not need to obtain the distance threshold in advance. k NNJ is always used as the pre-processing stage for classification or clustering task and has broad real applications in many data mining tasks, such as multimedia analysis, spatial data mining, time series, data streams, and social network. Taking similar medical image pairs detection as an example, in some cases, it is hard to make a definite diagnosis according to the medical images only; however, it is possible for us to obtain some similar medical images of the existed confirmed cases (k nearest neighbors) to help the doctors to make a final judgement on the unconfirmed diseases.

In many applications, the target objects can be represented as vector forms through feature extraction in order to

facilitate data processing, such as time series, videos, and trajectories, especially for the image processing tasks [34, 35]. With the continuous improvement of the accuracy of data acquisition equipment, the dimensionality of vectors representing the target objects will be very high, maybe hundreds of dimensions or even tens of thousands of dimensions. The calculation of the similarity or distance between the object pairs is a time costly operation. Most of the existed approaches conduct k NN join operation directly on the original dimensional space, so their performance is not ideal. It is an effective way to reduce the time cost of k NN join through reducing the dimensionality of the original data points. There are three contributions in the paper:

- (i) We proposed an effective dimensionality reduction approach that can make sure that the data points in the projected space preserve the location relationship to some extent as in the original space.
- (ii) We proposed two partition strategies, including equal width partition strategy and distance split tree-based partition strategy, and a novel k nearest neighbor join algorithm was proposed by using the above two partition strategies.
- (iii) Comprehensive experiments are conducted, and the final results prove that our approaches have better effectiveness and performance.

The other parts of the paper are arranged as follows. The detailed related research works are described in Section 2. Section 3 displays the notations, problem definitions, and some theorems. The lower bound probability is figured out in Section 4. The k nearest neighbor join algorithm using random projection and partition strategies is described in Section 5. The detailed experimental results of the proposed approaches are displayed in Section 6. Section 7 makes a conclusion of the paper and points out the future research directions.

2. Related Works

Many researchers have conducted in-depth research on similarity join problems because of their broad applications and important role in data mining or machine learning context; several survey articles have conducted a comprehensive and detailed analysis of the similarity join problem literature [36, 37].

k nearest neighbor join: an approximated k NN similarity join approach in metric spaces was proposed by Ferrada et al. [38], its time complexity is $\Theta(n^{3/2})$, and the empirical precision can reach up to 46%. Lu et al. [39] designed a novel approach called PCBJ by using Voronoi diagram, which can deal with exact k NN similarity join problems; however, its performance is not very good as the increase of data dimensionality. Dai et al. [40] proposed two novel k NN join algorithms based on the MapReduce framework, which are DSGMP-J using Distributed Sketched Grid and VDMP-J using Voronoi diagram; DSGMP-J [40] approach is easy to implement, but it ignores the real distribution of the dataset. Zhao et al. [41] designed an effective

data partitioning scheme called k NN-DP, which can solve the load imbalance issues caused by the data skewness problem; two novel schemes called LSH+ and z -value+ were developed based on k NN-DP to deal with k NN join operations under MapReduce framework. Song et al. [42] conducted a detailed analysis of the common workflows of the several existing k NN algorithms and further analyzed their load balancing, accuracy rate, and overall complexity; finally, a choice guideline was given which can help select the suitable methods for a specific case. Song et al. [43] also conducted a detailed comparison among the existing k NN join approaches both theoretically and experimentally on the MapReduce platform. RankReduce approach [44] was proposed by using locality-sensitive hashing with MapReduce for processing k nearest neighbor query. Hu et al. [45] proposed an adaptive νk NN join algorithm by using the Voronoi diagram, which can eliminate many unnecessary computations. There still exist many other research works about k NN join problems in several applications, such as Trajectory Data [46], machine learning [47], and Hilbert R-tree-based k NN join algorithms [48].

Top- k similarity join: Kim et al. [49] proposed two serial algorithms called the “divide-and-conquer algorithm” and “branch-and-bound algorithm” using the MapReduce framework, which can deal with Top- k similarity join problems efficiently. Chen et al. [50] developed a new distance function based on LSH and proposed an RDD-based Top- k similarity join algorithm using the Spark framework, and the test results proved that the RDD-based algorithm has better scalability and effectiveness than that of Hadoop. Ma et al. [51] developed a Top- k threshold estimation approach through sampling and designed an effective filtering solution by using the Symbolic Aggregate Approximation technique and then proposed a SAX-Top- k algorithm, which can deal with Top- k similarity join problem. Lei et al. [52] explored the similarity join problems for massive probabilistic dataset. The main idea of Lei et al. [52] is mapping the probabilistic data from the original space to the reduced dimensional space (one dimension), and then the range query on the one-dimensional space can be instead of the threshold-based similarity join query on the original space. Based on the above schemes, the authors proposed the Top- k Block Nested Loop Join Algorithm and Top- k Data Locality Preserving Join Algorithm, respectively. MELODY-JOIN [53] can improve the efficiency of the Top- k join on the histogram probabilistic dataset by using the standard lower bound space of the EMD distance; however, it cannot deal with the data skew problem efficiently. EMD-MPJ [54] proposed a novel Mapping-based Data Partitioning Framework that can solve the data skew problem. Heads-Join [55] made an extension to MELODY-JOIN [53] so that it can deal with both range similarity join and Top- k similarity efficiently.

Threshold-based similarity join: Cristiani et al. [56] designed a novel randomized set join method whose recall can be up to 100%, and its performance is better than that of the existing approach theoretically and empirically. Gowanlock et al. [57] proposed several novel methods to accelerate the similarity self-join by making full use of the

power and characteristic of GPU. There still exist some other research works which focus on the similarity join problem using GPU [58]. Sandes et al. [59] developed a novel filtering scheme that can speed up the exact set similarity join more efficiently. Ding et al. [60] exploited the privacy preserving problems in similarity join using MapReduce context. Wu et al. [61] proposed a novel parallel framework called SMS-Join which can support similarity join operations in metric space using the MapReduce paradigm. Ma et al. studied the similarity join problems for high-dimensional dataset through developing a novel dimension reduction approach based on the Piecewise Aggregate Approximation technique and proposed two algorithms called SAX-Based HDSJ [62] and Mp-V-SJQ [63].

3. Preliminaries

3.1. Notations. The notations used in this paper are listed in Table 1.

3.2. Problem Definition. The definitions of KNN and KNN join will be described in this subsection. Given two datasets $R \in \mathfrak{R}^d$ and $S \in \mathfrak{R}^d$, $\mathbf{R} = \{r_1, r_2, \dots, r_{n_1}\}$, $\mathbf{S} = \{s_1, s_2, \dots, s_{n_2}\}$, $|\mathbf{R}| = n_1$, and $|\mathbf{S}| = n_2$. r_i is the i_{th} data point from \mathbf{R} , $r_i = \langle r_{i1}, r_{i2}, \dots, r_{id} \rangle$, s_j is the j_{th} data point from \mathbf{S} , $s_j = \langle s_{j1}, s_{j2}, \dots, s_{jd} \rangle$, and the distance measurement used in the paper is the Euclidean distance denoted as $dist(r_i, s_j)$,

$$dist(r_i, s_j) = \sqrt{\sum_{l=1}^d (r_{il} - s_{jl})^2}, \quad (1)$$

where $dist(r_i, s_j) \geq 0$ and $dist(r_i, s_j)$ equals 0 when $r = s$.

Definition 1. K Nearest Neighbor Join (KNN). For a d -dimensional dataset $S \in \mathfrak{R}^d$ and a query data point r , the KNN operation of r on S can be recorded as $knn(r, S, k)$ aiming to obtain the k nearest neighbors of r in S :

$knn(r, S, k) = \{s_1, s_2, \dots, s_k | s_i \in S, 1 \leq i \leq k\}$; for each $\forall s_j \in S - \{s_1, s_2, \dots, s_k | s_i \in S, 1 \leq i \leq k\}$, the distance meets the following requirements:

$$dist(r, s_1) \leq dist(r, s_2) \leq \dots \leq dist(r, s_k) \leq dist(r, s_j).$$

Definition 2. K Nearest Neighbor Similarity Join (KNN Join). Given two datasets $R \in \mathfrak{R}^d$ and $S \in \mathfrak{R}^d$, the KNN similarity join operation on R and S can return the k nearest neighbors

for each data point $r \in \mathbf{R}$ from S , which can be denoted as $knnJ(\mathbf{RS}) = \{(r, knn(r, S, k)) | \text{for each } r \in \mathbf{R}\}$.

3.3. Theorems

Theorem 1. Given two d -dimensional data points \mathbf{s}_1 and \mathbf{s}_2 , then $g(\mathbf{s}_1) - g(\mathbf{s}_2)/dist(\mathbf{s}_1, \mathbf{s}_2) \sim \mathbf{N}(0, 1)$.

Theorem 2. Given two d -dimensional data points \mathbf{s}_1 and \mathbf{s}_2 , then $\Delta_m^2(\mathbf{s}_1, \mathbf{s}_2)/dist^2(\mathbf{s}_1, \mathbf{s}_2) \sim \chi^2(m)$.

Theorem 3. If $dist(\mathbf{s}_1, \mathbf{s}_2) \leq \epsilon$, then the probability that $\Delta_m(\mathbf{s}_1, \mathbf{s}_2)$ is less than or equal to ke will be bigger than or equal to $1 - P(\chi^2 > k^2)$, which can be denoted as follows: $P(\Delta_m(\mathbf{s}_1, \mathbf{s}_2) \leq |ke \text{ dist}(\mathbf{s}_1, \mathbf{s}_2) \leq \epsilon) \geq 1 - P(\chi^2 > k^2)$.

Theorem 1, Theorem 2, and Theorem 3 have been proved by Ma et al. [64]. Theorem 3 indicates that if the Euclidean distance of the original space is less than or equal to ϵ , the probability that the distance of the projected space will be less than or equal to ke has the lower bound; that is, $1 - P(\chi^2 > k^2)$. So we can project d -dimensional data point v into m -dimensional space ($m < d$) through $\pi_m(\mathbf{s}) = \langle g_1(\mathbf{s}), g_2(\mathbf{s}), \dots, g_m(\mathbf{s}) \rangle$.

Theorem 4. Given three d -dimensional data points \mathbf{r}, \mathbf{s}_1 , and \mathbf{s}_2 , $\mathbf{r}, \mathbf{s}_1, \mathbf{s}_2 \in \mathfrak{R}^d$, $\mathbf{U} = \Delta_m^2(\mathbf{r}, \mathbf{s}_1)/dist^2(\mathbf{r}, \mathbf{s}_1) \sim \chi^2(m)$, $\mathbf{V} = \Delta_m^2(\mathbf{r}, \mathbf{s}_2)/dist^2(\mathbf{r}, \mathbf{s}_2) \sim \chi^2(m)$, and then $\mathbf{F} = \mathbf{U}/\mathbf{V}$ obeys the F distribution with degrees of freedom (m, m) ; that is, $\mathbf{F} = \mathbf{U}/\mathbf{V} \sim F(m, m)$.

Proof. According to Theorem 2, $\mathbf{U} = \Delta_m^2(\mathbf{r}, \mathbf{s}_1)/dist^2(\mathbf{r}, \mathbf{s}_1) \sim \chi^2(m)$, and $\mathbf{V} = \Delta_m^2(\mathbf{r}, \mathbf{s}_2)/dist^2(\mathbf{r}, \mathbf{s}_2) \sim \chi^2(m)$; that is to say, \mathbf{U} and \mathbf{V} both obey the χ^2 distribution with freedom m .

Based on the definition of F distribution, $\mathbf{F} = \mathbf{U}/m/\mathbf{V}/m$ obeys the F distribution with degrees of freedom (m, m) ; that is, $\mathbf{F} = \mathbf{U}/\mathbf{V} \sim F(m, m)$. \square

Theorem 5. If $dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)$, then the probability that $\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq k\Delta_m(\mathbf{r}, \mathbf{s}_2)$ is bigger than $1 - P(F > k^2)$; that is, $P(\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq k\Delta_m(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)) > 1 - P(F > k^2)$. F is the distribution with degrees of freedom (m, m) ; that is, $F \sim F(m, m)$.

Proof.

TABLE 1: Notations.

| Notation | Meaning of the notation |
|--|--|
| n_1, n_2 | The data points' number in the dataset. |
| ϵ | The width of each partition under equal width partition strategy. |
| d | The data point's dimensionality. |
| $dist(\mathbf{s}_1, \mathbf{s}_2)$ | The Euclidean distance of data point \mathbf{s}_1 and data point \mathbf{s}_2 . |
| $g(\mathbf{s})$ | $g(\mathbf{s}) = \mathbf{a} \cdot \mathbf{s}$, \mathbf{a} is a d -dimensional vector, and each element is a random variable that obeys \mathbf{p} -stable distribution. |
| $\pi_m(\mathbf{s})$ | $\pi_m(\mathbf{s}) = \langle g_1(\mathbf{s}), g_2(\mathbf{s}), \dots, g_m(\mathbf{s}) \rangle$. |
| $\Delta_m(\mathbf{s}_1, \mathbf{s}_2)$ | $\Delta_m(\mathbf{s}_1, \mathbf{s}_2) = dist(\pi_m(\mathbf{s}_1), \pi_m(\mathbf{s}_2))$. |
| $\chi^2(m)$ | Chi-square distribution with degree of freedom m . |
| $\pi_1(\mathbf{s})_{\min}$ | The minimum projected value of data point \mathbf{s} . |
| $\pi_1(\mathbf{s})_{\max}$ | The maximum projected value of data point \mathbf{s} . |
| len | The width of all the projected values in one-dimensional space; that is, $\pi_1(\mathbf{s})_{\min} - \pi_1(\mathbf{s})_{\max}$. |
| PN | The number of the partitions in one-dimensional space. |
| P_i | The i_{th} partition. |

$$\because \Delta_m(\mathbf{r}, \mathbf{s}_1) \geq 0 \text{ and } \Delta_m(\mathbf{r}, \mathbf{s}_2) \geq 0,$$

$$\begin{aligned} & \therefore P(\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq k\Delta_m(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)) \\ &= P(\Delta_m^2(\mathbf{r}, \mathbf{s}_1) \leq k^2\Delta_m^2(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)) \\ &= P\left(\frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)}{dist^2(\mathbf{r}, \mathbf{s}_1)} \leq \frac{k^2\Delta_m^2(\mathbf{r}, \mathbf{s}_2)}{dist^2(\mathbf{r}, \mathbf{s}_1)} | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)\right) \\ &= \frac{P(\Delta_m^2(\mathbf{r}, \mathbf{s}_1)/dist^2(\mathbf{r}, \mathbf{s}_1) \leq k^2\Delta_m^2(\mathbf{r}, \mathbf{s}_2)/dist^2(\mathbf{r}, \mathbf{s}_1) \text{ and } dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2))}{dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)}, \end{aligned}$$

$$\therefore P(dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)) = 1,$$

$$\therefore P(\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq k\Delta_m(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2))$$

$$= P\left(\frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)}{dist^2(\mathbf{r}, \mathbf{s}_1)} \leq \frac{k^2\Delta_m^2(\mathbf{r}, \mathbf{s}_2)}{dist^2(\mathbf{r}, \mathbf{s}_1)}\right)$$

$$= 1 - P\left(\frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)}{dist^2(\mathbf{r}, \mathbf{s}_1)} > \frac{k^2\Delta_m^2(\mathbf{r}, \mathbf{s}_2)}{dist^2(\mathbf{r}, \mathbf{s}_1)}\right), \quad (2)$$

$$\therefore \frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)}{dist^2(\mathbf{r}, \mathbf{s}_1)} \sim \chi^2(m) \text{ and } dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2),$$

$$\therefore P(\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq k\Delta_m(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2))$$

$$> 1 - P\left(\frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)}{dist^2(\mathbf{r}, \mathbf{s}_1)} > \frac{k^2\Delta_m^2(\mathbf{r}, \mathbf{s}_2)}{dist^2(\mathbf{r}, \mathbf{s}_2)}\right)$$

$$= 1 - P\left(\frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)/dist^2(\mathbf{r}, \mathbf{s}_1)}{\Delta_m^2(\mathbf{r}, \mathbf{s}_2)/dist^2(\mathbf{r}, \mathbf{s}_2)} > k^2\right),$$

$$\text{according to theorem 4, } \frac{\Delta_m^2(\mathbf{r}, \mathbf{s}_1)/dist^2(\mathbf{r}, \mathbf{s}_1)}{\Delta_m^2(\mathbf{r}, \mathbf{s}_2)/dist^2(\mathbf{r}, \mathbf{s}_2)} \sim F(m, m),$$

$$\therefore P(\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq k\Delta_m(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)) > 1 - P(F > k^2). \quad \square$$

According to Theorem 5, when $k = 1$, if $dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)$, then $P(\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq \Delta_m(\mathbf{r}, \mathbf{s}_2) | dist(\mathbf{r}, \mathbf{s}_1) \leq dist(\mathbf{r}, \mathbf{s}_2)) > 1 - P(F(m, m) > 1)$; it indicates that the probability that $\Delta_m(\mathbf{r}, \mathbf{s}_1) \leq \Delta_m(\mathbf{r}, \mathbf{s}_2)$ has the lower bound: $1 - P(F(m, m) > 1)$.

We can conclude that if \mathbf{s}_1 is closer to \mathbf{r} than \mathbf{s}_2 in the original d -dimensional space, $\pi_m(\mathbf{s}_1)$ is still likely to be closer to $\pi_m(\mathbf{r})$ than $\pi_m(\mathbf{s}_2)$ in the projected m -dimensional space with lower bound probability $1 - P(F(m, m) > 1)$.

4. Probability Computation

When $m = 1, k = 1$, the probability $P(F(m, m) > 1)$ can be figured out based on the probability density of F distribution which is described as follows:

$$\psi(y) = \begin{cases} \frac{\Gamma[(n_1 + n_2)/2] (n_1/n_2)^{n_1/2} y^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1 + (n_1 y/n_2)]^{(n_1+n_2)/2}}, & y > 0, \\ 0, & \text{other.} \end{cases} \quad (3)$$

Given the freedom (1,1) and the upper quantile with 1, the probability $P(F(m, m) > 1)$ can be calculated as the follows:

$$\begin{aligned} P(F > 1) &= \int_1^{\infty} \psi(y) dy, \\ &= \int_1^{\infty} \frac{\Gamma[(1+1)/2] (1/1)^{1/2} y^{(1/2)-1}}{\Gamma(1/2)\Gamma(1/2)[1+y]^{(1+1)/2}} dy \\ &= \int_1^{\infty} \frac{\Gamma(1) y^{-1/2}}{(\Gamma(1/2))^2 [1+y]} dy \\ &\because \Gamma(1) = 1, \Gamma(1/2) = \sqrt{\pi} \\ \therefore P(F > 1) &= \int_1^{\infty} \frac{y^{-1/2}}{(\sqrt{\pi})^2 [1+y]} dy \quad (4) \\ &= \frac{1}{\pi} \int_1^{\infty} \frac{1}{\sqrt{y}(1+y)} dy \\ &= \frac{2}{\pi} \int_1^{\infty} \frac{1}{(1+(\sqrt{y})^2)} d\sqrt{y} \\ &= \frac{2}{\pi} \arctan \sqrt{y} \Big|_1^{\infty} \\ &= \frac{2}{\pi} \left(\frac{\pi}{2} - \frac{\pi}{4} \right) = 0.5. \end{aligned}$$

The result of the above computation shows that, given three d -dimensional data points \mathbf{r}, \mathbf{s}_1 , and \mathbf{s}_2 , $\mathbf{r}, \mathbf{s}_1, \mathbf{s}_2 \in \mathcal{R}^d$, they can be reduced to 1-dimensional space through dot product with a d -dimensional vector \mathbf{a} . If \mathbf{s}_1 is closer to \mathbf{r} than \mathbf{s}_2 in the original d -dimensional space, the lower bound probability that $\pi_1(\mathbf{s}_1)$ is still likely to be closer to $\pi_1(\mathbf{r})$ than $\pi_1(\mathbf{s}_2)$ in the projected 1-dimensional space is 0.5.

5. k Nearest Neighbor Join Using Novel Partitioning Strategy

5.1. Algorithm for k Nearest Neighbor Join Using Novel Partitioning Strategy. Theorem 5 shows that if \mathbf{s}_1 is closer to \mathbf{r} than \mathbf{s}_2 in the original d -dimensional space, $\pi_m(\mathbf{s}_1)$ is still

likely to be closer to $\pi_m(\mathbf{r})$ than $\pi_m(\mathbf{s}_2)$ in the projected m -dimensional space with the lower bound probability $1 - P(F(m, m) > 1)$. The conclusion implies that data points in projected space maintain relative location relationship as in original dimensional space. So we proposed k nearest neighbor join algorithm using random projection (RP k NN); it includes two main stages: the first stage is responsible for dimension reduction and space partition, and the second stage is used to conduct k NN join in reduced dimensional space. Figure 1 shows the general framework of k nearest neighbor similarity join algorithm using novel partitioning strategy. The concrete process of the proposed algorithm is described in Algorithm 1. The getPartition routine projects all the data points into one-dimensional space and divides the data points into several partitions according to the specific partition strategy (line 1). For each partition P_i , its corresponding partition \bar{P}_i , which needs to be compared with P_i , can be obtained through lines 3–9. If P_i is the leftmost partition, $\bar{P}_i \leftarrow \cup P_i \cup P_{i+1}$ (line 5). If P_i is the rightmost partition, $\bar{P}_i \leftarrow \cup P_{i-1} \cup P_i$ (line 7); in other cases, $\bar{P}_i \leftarrow P_{i-1} \cup P_i \cup P_{i+1}$ (line 9). Finally, for each data point $v \in P_i$, k NN join routine is used to find its k nearest neighbors from \bar{P}_i (lines 10–12).

5.2. Partition Strategy

5.2.1. Equal Width Partition Strategy. All the data points are divided into several partitions with equal width. The total partition number can be set to $PN = \lfloor \sqrt{n} \rfloor$. Suppose that $\pi_1(\mathbf{s})_{\min}$ is the minimum projected value of s in one-dimensional space and $\pi_1(\mathbf{s})_{\max}$ is the maximum projected value of s in one-dimensional space; that is $\pi_1(\mathbf{s})_{\min} = \min\{\pi_1(s_j), s_j \in R\}$ and $\pi_1(\mathbf{s})_{\max} = \max\{\pi_1(s_j), s_j \in R\}$; len is the width of all the projected values in one-dimensional space, $len = \pi_1(\mathbf{s})_{\min} - \pi_1(\mathbf{s})_{\max}$; ϵ is the width of each partition, $\epsilon = \lfloor len/PN \rfloor$; given a data point s , its corresponding partition number is $P_i = \lfloor \pi_1(s)/\epsilon \rfloor$. The detailed procedure can be shown in Figure 2 and Algorithm 2.

5.2.2. Distance Split Tree-Based Partition Strategy. The previous equal width partition strategy is easy to implement; however, it cannot deal with skewed dataset efficiently. According to our proposed approach, the d -dimensional data point s will be projected into one-dimensional space through $\pi_1(s) = g_1(s) = v \cdot a = \sum_{i=1}^d s_i * a_i$; each element of a obeys standard normal distribution; that is: $a_i \sim N(0, 1), i \in [1, d]$. The projected value $\pi_1(s)$ is subject to normal distribution, so it is skewed. Figure 3 shows the distribution of the projected values.

Theorem 6. Given two d -dimensional data points s and a , $a_i \sim N(0, 1), i \in [1, d]$, $\pi_1(s) = g_1(s) = s \cdot a = \sum_{i=1}^d s_i * a_i$, and then $\pi_1(s) \sim N(0, \sum_{i=1}^d s_i^2)$.

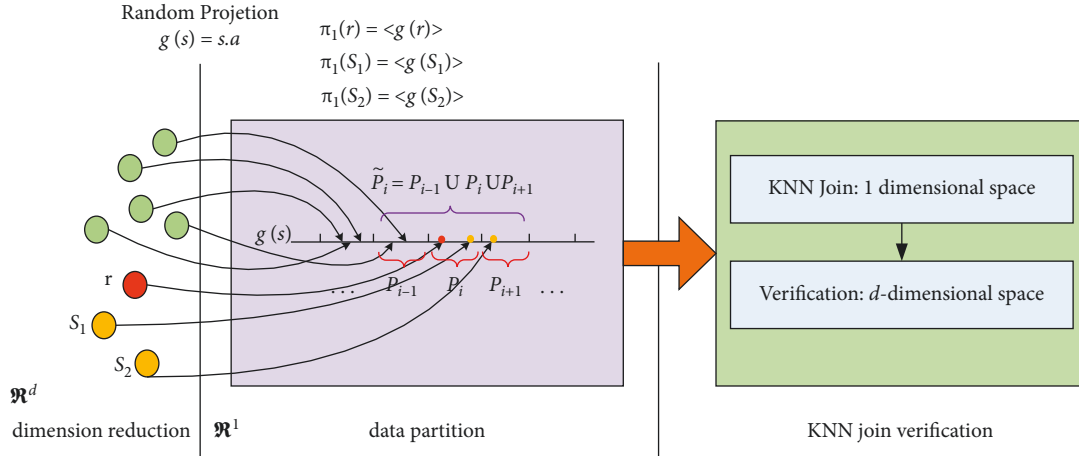


FIGURE 1: Framework of k nearest neighbor join algorithm using novel partitioning strategy.

Input: R, k //dataset, the number of nearest neighbors to find
 Output: res //a set of pairs of data points

- (1) partitions \leftarrow get Partition (R, n);
- (2) $res \leftarrow \emptyset$;
- (3) for $i = 1; i \leq |\text{partitions}|; i++$ do
- (4) if $i = 1$ then
- (5) $\tilde{P}_i \leftarrow \cup P_i \cup P_{i+1}$
- (6) else if $i = |\text{partitions}|$ then
- (7) $\tilde{P}_i \leftarrow \cup P_{i-1} \cup P_i$
- (8) else
- (9) $\tilde{P}_i \leftarrow P_{i-1} \cup P_i \cup P_{i+1}$
- (10) for data point $v \in P_i$ do
- (11) temp $\leftarrow kNN$ Join (k, v, \tilde{P}_i)
- (12) $res \leftarrow res \cup temp$
- (13) return res .

ALGORITHM 1: Algorithm for k nearest neighbor join using novel partitioning strategy.

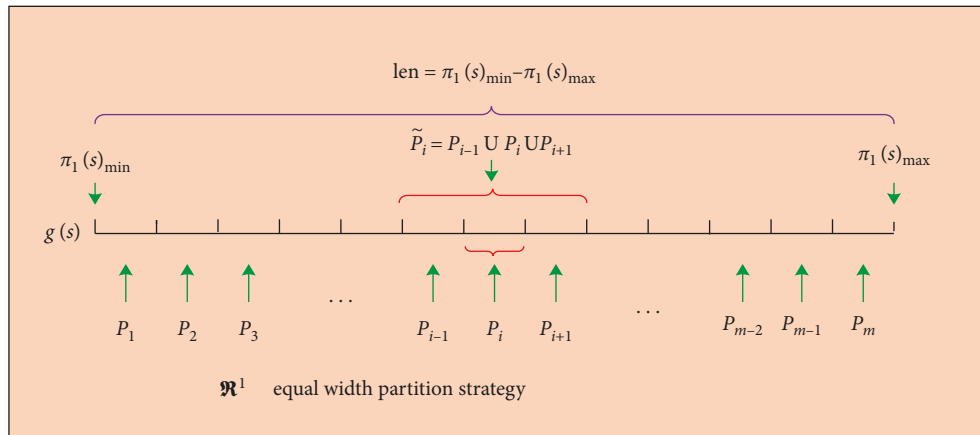
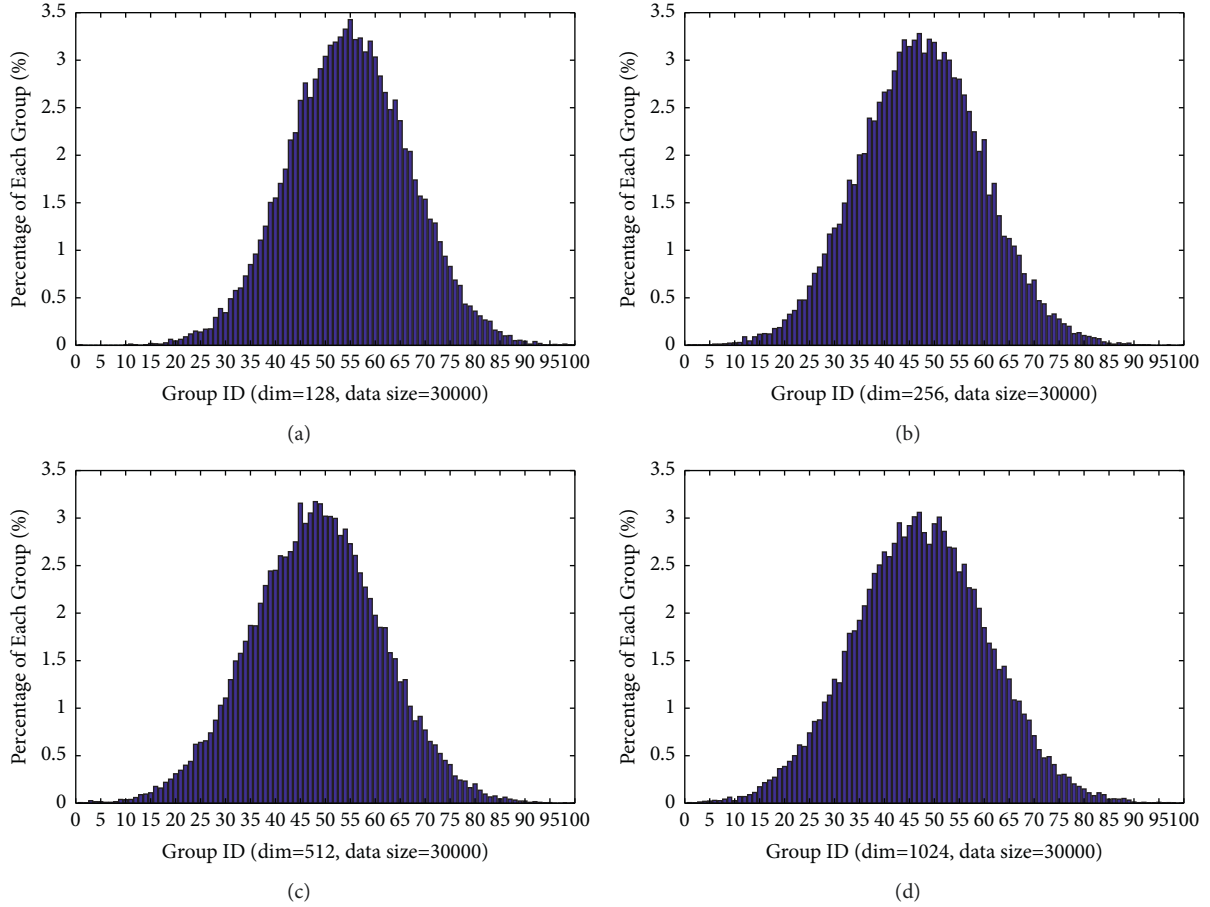


FIGURE 2: Equal width partition strategy.

Input: R, n //dataset, the cardinality of the dataset
Output: partitions //list of the sets that partition the dataset

- (1) $PN = \lfloor \sqrt{n} \rfloor$
- (2) $\text{partitions}[i] \leftarrow \emptyset, \forall i, 0 \leq i \leq PN$
- (3) $\pi_1(s)_{\min} = \min\{\pi_1(s_j), s_j \in R\}$
- (4) $\pi_1(s)_{\max} = \max\{\pi_1(s_j), s_j \in R\}$
- (5) $len = \pi_1(s)_{\min} - \pi_1(s)_{\max}$
- (6) $\epsilon = \lfloor len/PN \rfloor$
- (7) choose one vector randomly, recorded as $a, \forall e \in a \sim N(0, 1)$
- (8) //divide \bar{R} into PN partitions according to P_i , the data points which have the same P_i .
- (9) //belong to the same partition, recorded as $\text{partitions}[1], \text{partitions}[2] \dots, \text{partitions}[PN]$.
- (10) for each data point $s \in R$ do
- (11) $\pi_1(s) \leftarrow \langle g_1(s) \rangle$
- (12) $P_i = \lfloor \pi_1(s)/\epsilon \rfloor$
- (13) $\text{partitions}[P_i] \leftarrow \langle P_i, \pi_1(s), s \rangle$

ALGORITHM 2: Equal width partition strategy.

FIGURE 3: The distribution of the projected values ($\pi_1(s)$): (a) dim = 128 and data science = 30000; (b) dim = 256 and data science = 30000; (c) dim = 512 and data science = 30000; (d) dim = 1024 and data science = 30000.

Proof.

$$\begin{aligned}
&\because a_i \sim N(0, 1), \\
&\Rightarrow s_i a_i \sim N(0, s_i^2), \\
&\because \pi_1(s) = \sum_{i=1}^d s_i * a_i, \\
&\therefore \pi_1(s) \sim N\left(0, \sum_{i=1}^d s_i^2\right), \\
&\Rightarrow \pi_1(s) \text{ is subject to normal distribution and its variance is } \sum_{i=1}^d s_i^2.
\end{aligned} \tag{5}$$

Aiming to deal with the data skew problem, we proposed a novel partition strategy called distance split tree- (DST-) based partition strategy. Figure 4 displays the structure of DST. The main idea of DST is that after the original d -dimensional data points are mapped into one-dimensional space, in the beginning, all the data points are divided into equal width partitions with threshold $\epsilon = \pi_1(s)_{\max} - \pi_1(s)_{\min}/2^c$, c is an adjustable parameter. $\max \text{Num}$ is the upper bound of data point count contained in each partition. Once the data point count in a specific partition exceeds $\max \text{Num}$, the partition will be divided into two new partitions with equal width again and so on, and finally, a distance split tree is formed. For each leaf node, the level of the node's hierarchy, the node number in the specific level, the count of the data, and the corresponding dataset are recorded. Based on the above information, the distance range corresponding to each leaf node can be calculated. The corresponding distance width of each node in the current level can be calculated: $1/2^{\text{level}-1}\epsilon$. The corresponding distance range of the node can be calculated by *or de rNo* in the level $[\text{order No} - 1/2^{\text{level}-1}\epsilon, \text{order No}/2^{\text{level}-1}\epsilon]$. Thus, the distance range corresponding to the N_2 node can be calculated: $[3 - 1/2^{3-1}\epsilon, 3/2^{3-1}\epsilon]$; that is, $[2/4\epsilon, 3/4\epsilon]$.

The construction of distance split tree: the construction process of the distance split tree is as follows: firstly, build a root node N_{root} , for each data point $s_i \in R$, and figure out its projected value in one-dimensional space $\pi_1(s_i)$. Then all the data points are divided into equal width partitions with threshold ϵ , and the corresponding partition number of each vector s_i in the mapping space is obtained $pi\ d \leftarrow \lfloor \pi_1(s_i)/\epsilon \rfloor$. If the node with the number $pi\ d$ does not exist, a new child node with the number $pi\ d$ will be generated. If it already exists, s_i is inserted into the node $pi\ d$, and its count value is increased by 1. Once the amount of data point in a node exceeds a given threshold, such as $\max \text{Num}$, the node will be further divided into two subnodes according to the distance range. Repeat this procedure, and finally, a distance split tree is generated.

Data partitions generated: after the distance split tree is constructed, the partitions set can be obtained through preorder traversal for distance split tree; only the leaf nodes

are left as the member of the final partitions set. Then the obtained partitions set can be used in Algorithm 1. \square

6. Time Complexity Analysis

In this section, we mainly analyze the time complexity of our proposed method. Given the d -dimensional dataset R and $|R| = n$, the total partition number is $PN = \lfloor \sqrt{n} \rfloor$, P_i represents the i_{th} partition, and Cost represents the total computations of the proposed method. The time complexity is as the follows:

$$\text{Cost} = \sum_{i=1}^{PN} |P_i| * |P_i \cup P_{i-1} \cup P_{i+1}| * d. \tag{6}$$

In the best cases, all the data points in R are evenly distributed in each partition; that is, $|P_i| = \lfloor \sqrt{n} \rfloor$, so

$$\begin{aligned}
\text{Cost} &= \sum_{i=1}^{PN} \lfloor \sqrt{n} \rfloor * 3 * \lfloor \sqrt{n} \rfloor * d \\
&= \lfloor \sqrt{n} \rfloor * \lfloor \sqrt{n} \rfloor * 3 * \lfloor \sqrt{n} \rfloor * d \\
&= 3 * n^{3/2} * d.
\end{aligned} \tag{7}$$

The time complexity in the best case can be recorded as $\mathcal{O}(n^{3/2}d)$.

In the worst case, supposing that all the data points are included in one partition, then the time complexity should be $\mathcal{O}(n^2d)$. Because the projected values of the proposed method obey normal distribution, it is between the best case and the worst case, so the time complexity of the proposed method lies in $(\mathcal{O}(n^{3/2}d), \mathcal{O}(n^2d))$.

7. Experimental Analysis

We conducted experiments to test the effectiveness and performance of the proposed methods, k nearest neighbor join algorithm using random projection with equal width partition strategy (RP k NNEW) and k nearest neighbor join algorithm using random projection with distance split tree-based partition strategy (RP k NNDST), and made comparisons between our proposed methods and the existing

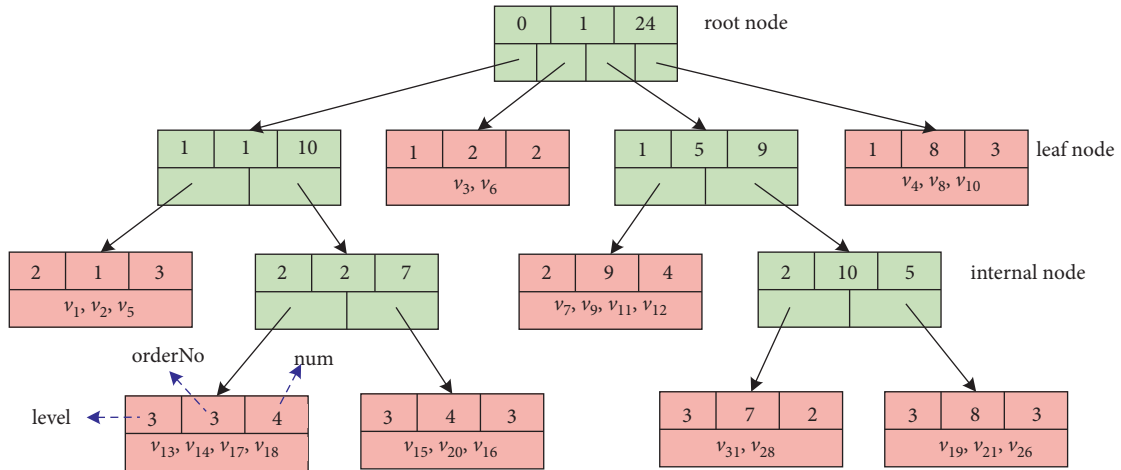


FIGURE 4: Distance split tree-based partition strategy.

methods including a KNN [38] and the brute force method with Block Nested Loop Join Strategy (BNLJ). a KNN [38] is a relatively new research work on k NN similarity join problem, and it also adopted the algorithms based on the partitioning strategy.

Experimental settings: our tests are performed on HP workstation, and the configurations are as follows: CPU: Intel Xeon Gold 6136 @ 3.00 GHz, memory: 128 GB, disk: 2 TB, OS: 64-bit Windows 10, and 12 cores. Table 2 describes the parameters and their values. The bold fonts represent the default values.

Datasets: the datasets adopted in our experiments are synthetic data, the elements of the vector are uniformly distributed in the range $[0, 1]$, and the dimensionality of the datasets includes 128, 256, 512, and 1024. The datasets are listed in Table 3.

7.1. Precision versus Data Size. It can be concluded that RP k NNEW has the best precision among the above three approaches, including a KNN, RP k NNEW, and RP k NNDST; in some cases, the precision of RP k NNEW is more than 50%. However, Figure 5 shows that the precision of a KNN, RP k NNEW, and RP k NNDST is not very stable for different data size, and the precision of RP k NNDST is between that of a KNN and RP k NNEW.

7.2. Precision versus Dimension. Figure 6 displays the precision of a KNN, RP k NNEW, and RP k NNDST under different dimension. The proposed method RP k NNEW has the best precision; although the precision of RP k NNEW varies under different dimension, it is always higher than 40% under all conditions. The precision of RP k NNDST is relatively stable under different dimension, and it is lower than that of RP k NNEW. Because RP k NNDST adopts the distance tree-based partition strategy and all the data points will be distributed into different partitions more evenly, every partition will not contain much more data points, so its precision will decrease to some extent compared with RP k NNEW. The precision of our proposed methods, including

RP k NNEW and RP k NNDST, is better than that of a KNN under all different dimension.

7.3. Precision versus k . Figure 7 displays the precision of the above approaches (including a KNN, RP k NNEW, and RP k NNDST) under different k . The results prove that the precision of RP k NNEW and RP k NNDST is higher than that of the existing method a KNN. The precision of RP k NNEW is higher than that of RP k NNDST; the reason is that RP k NNEW adopted the equal width partition strategy; while the projected values are skewed, several partitions will contain more data points; the precision will be higher accordingly. The precision of a KNN and RP k NNDST is basically stable under different k value, while the precision of RP k NNEW is a little more sensitive to the different k value.

7.4. Precision Distribution. Figure 8 displays the precision distribution of a KNN, RP k NNEW, and RP k NNDST. It can be found that the precision of some data points is very low (less than 5%), and the precision of some data points is very high (more than 80%) by using RP k NNEW. The average precision of RP k NNEW is higher than that of a KNN and RP k NNDST; the percentage of the data points whose precision is more than 80% is 20.6%, 0.05%, and 0.57%, respectively, for RP k NNEW, a KNN, and RP k NNDST. The main reason is that RP k NNEW adopts the equal width partition strategy, which cannot deal with the skewed projected values effectively; however, RP k NNDST adopts distance split tree-based partition strategy, which can distribute all the data points into different partitions more evenly.

7.5. Performance versus Data Size. Figure 9 displays the performance of BNLJ, a KNN, RP k NNEW, and RP k NNDST on the datasets with different size. The run time of a KNN, RP k NNEW, and RP k NNDST is much less than that of the BNLJ method; the run time of BNLJ increases exponentially with the size of the datasets; however, the run time of a KNN, RP k NNEW, and RP k NNDST increases

TABLE 2: Experimental settings.

| Experimental parameters | Values of the parameters |
|-------------------------|---------------------------------------|
| Returned number: k | 10, 20, 30, 40, and 50 |
| Dimensionality: d | 128, 256, 512, and 1024 |
| Data size: N | 10000, 20000, 30000, 40000, and 50000 |

TABLE 3: Datasets descriptions.

| Dataset | Number | Dim. | Data size (M) |
|-------------|--------|------|---------------|
| Data-128-1 | 10,000 | 128 | 11.3 |
| Data-128-2 | 20,000 | 128 | 22.5 |
| Data-128-3 | 30,000 | 128 | 33.8 |
| Data-128-4 | 40,000 | 128 | 45 |
| Data-128-5 | 50,000 | 128 | 56.3 |
| Data-256-3 | 30,000 | 256 | 67.6 |
| Data-512-1 | 10,000 | 512 | 45 |
| Data-512-2 | 20,000 | 512 | 90 |
| Data-512-3 | 30,000 | 512 | 135.1 |
| Data-512-4 | 40,000 | 512 | 180.1 |
| Data-512-5 | 50,000 | 512 | 225.1 |
| Data-1024-3 | 30,000 | 960 | 270 |

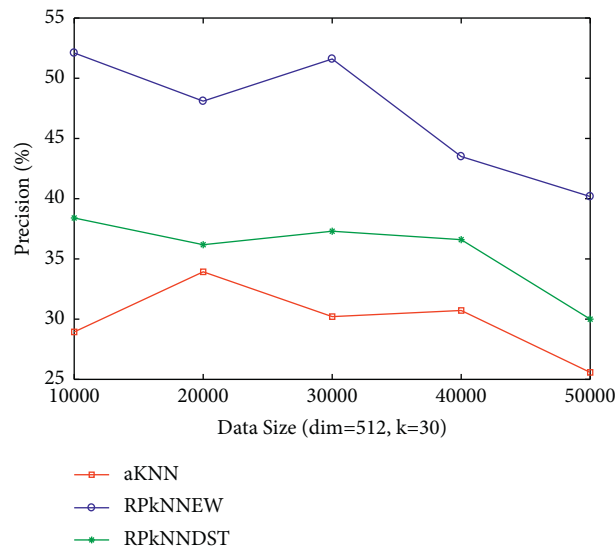


FIGURE 5: Precision versus data size.

linearly with the size of the datasets. The run time of RP k NNEW is a little bit more than that of a KNN and RP k NNDST. While we can find that RP k NNEW has the best precision among all the methods based on the above precision analysis, we can choose the RP k NNEW method when the performance requirements are not very strict; otherwise, RP k NNDST will be the most appropriate choice, because the precision of RP k NNDST is higher than that of a KNN, while its run time is less than that of RP k NNEW.

7.6. Performance versus Dimension. Figure 10 displays the performance of our proposed methods and the existing methods for different dimensions, which are 128, 256, 512,

and 1024, respectively. The time of all algorithms grows with the increase of the dimension, and the reason is the bigger the dimension, the higher the time complexity. The performance of RP k NNEW is the best when the dimension is less than 512, while the run time of RP k NNEW will be slightly higher than that of a KNN when the dimension exceeds 512. The run time of RP k NNEW is higher than that of a KNN and RP k NNEW, while Figure 6 shows that RP k NNEW has the best precision in all cases.

7.7. Performance versus k . The performance of BNLJ, a KNN, RP k NNEW, and RP k NNDST with different k (data size = 30000; dim = 512) is displayed in Figure 11. The run

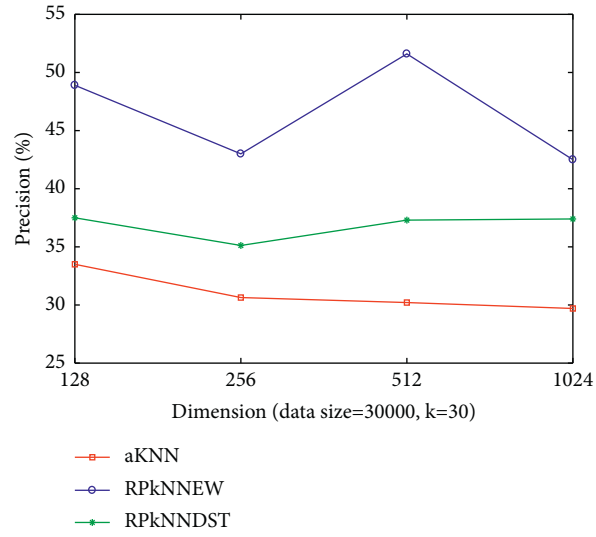


FIGURE 6: Precision versus dimension.

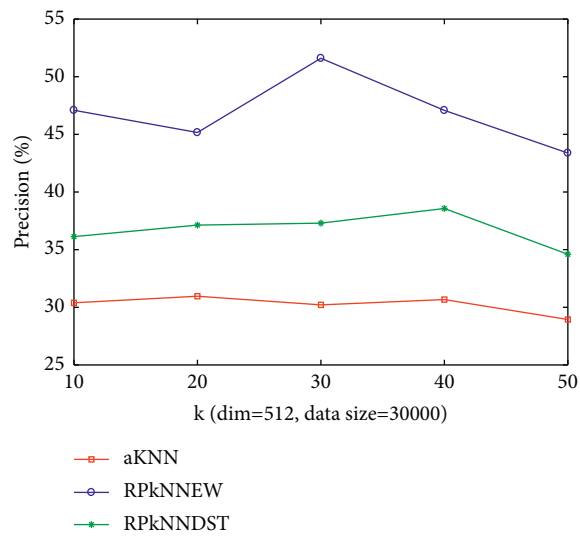
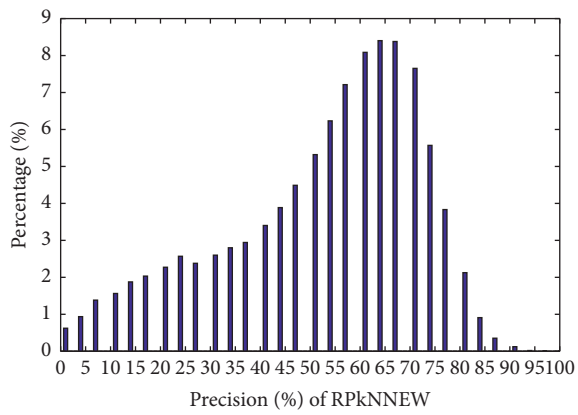
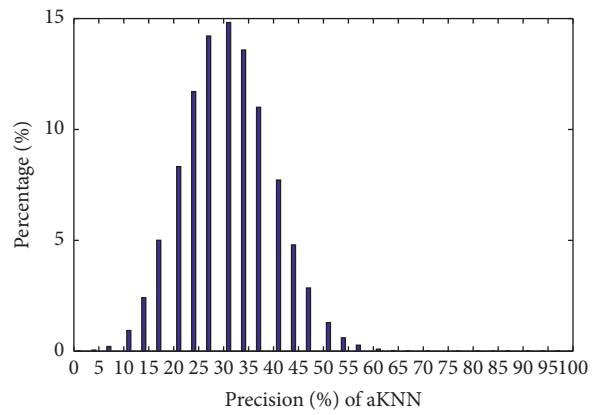


FIGURE 7: Precision versus k .



(a)



(b)

FIGURE 8: Continued.

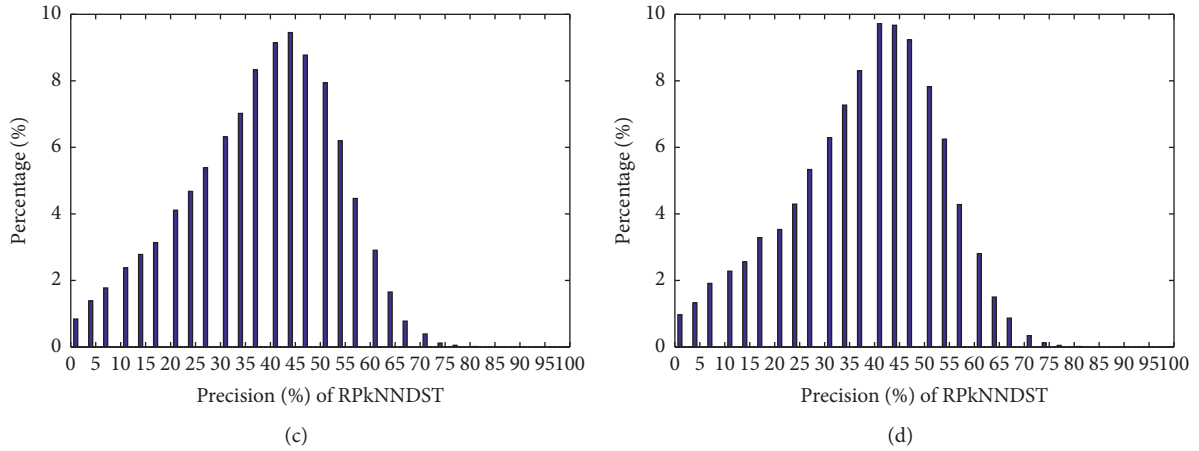


FIGURE 8: The distribution of the precision: (a) dim = 512, data size = 30000, and k = 30; (b) dim = 512, data size = 30000, and k = 30; (c) dim = 512, data size = 30000, and k = 30; (d) dim = 1024, data size = 30000, and k = 30.

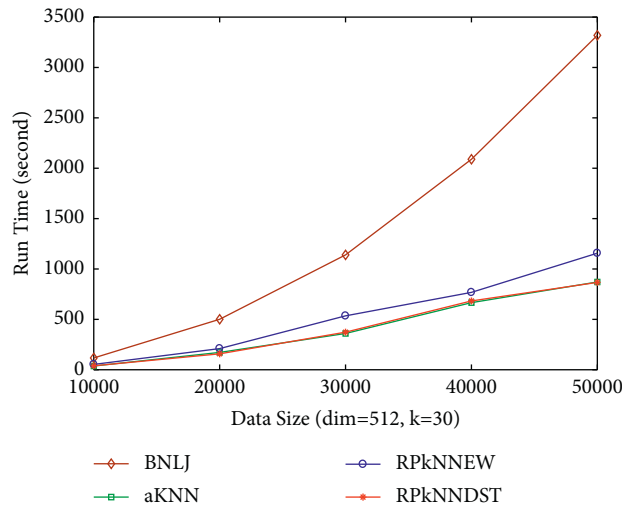


FIGURE 9: Performance versus data size.

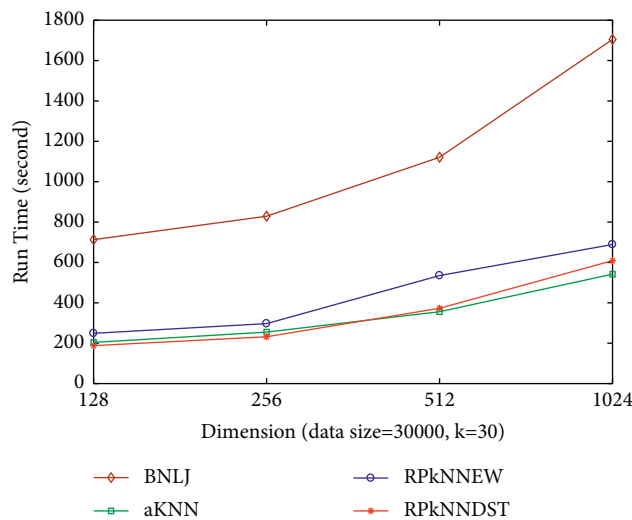


FIGURE 10: Performance versus dimension.

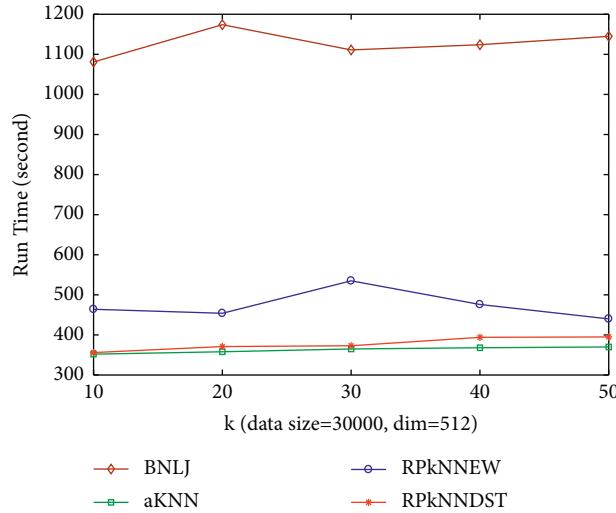


FIGURE 11: Performance versus k .

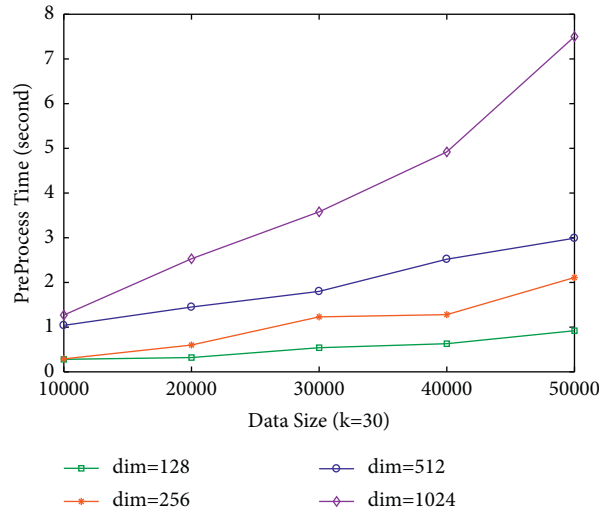


FIGURE 12: Preprocessing time.

time of all the above algorithms changes little with the different value of k . The run time of RP k NNDST is very close to that of a KNN, and the reason has been explained in Section 6; however, the precision of RP k NNDST is better than that of a KNN according to Figure 7. The run time of RP k NNDST and a KNN is less than that of RP k NNEW.

7.8. Preprocessing Time. Figure 12 displays the preprocessing time required by the construction of the distance split tree (DST) in RP k NNDST approach for different dataset. The preprocessing time increases almost linearly with the size of the dataset. Given a dataset with a fixed size, the preprocessing time increases exponentially with the growth of the dimension. Overall, the proportion of preprocessing time in the total time is low and relative stable. The benefit of the distance split tree (DST) can make up for the additional overhead caused by the construction of DST and can make the total run time of RP k NNDST less than that of RP k NNEW.

8. Conclusions

In the above sections, we mainly studied the k nearest neighbor similarity join problem on high-dimensional data. We proposed k nearest neighbor join algorithm using random projection with equal width partition strategy (RP k NNEW) and k nearest neighbor join algorithm using random projection with distance split tree-based partition strategy (RP k NNDST), which can filter out many unnecessary comparisons and ensure the required precision. We also conducted several experiments to test the effectiveness and performance of our proposed approaches, and the test results show that the proposed approaches in this paper have better effectiveness and performance. However, the proposed approaches in this paper have some limitations, and they can only work with the Euclidean distance. In future research works, we are planning to further study the k NN similarity join approaches, which can deal with other similarity measures, other more effective dimension

reduction techniques, and the distributed k nearest neighbor similarity join algorithms.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was partially supported by the grants from the Science and Technology Research Plan Project of Henan Province (202102210357); the Innovative Research Team (in Science and Technology) at University of Henan Province (22IRTSTHN016); the National Natural Science Foundation of China (61602231); the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) (JP21K17737); and the Hubei Natural Science Foundation (2021CFB156).

References

- [1] W. L. Shang, J. Y. Chen, H. B. Bi, Y. Sui, Y. Chen, and H. Yu, "Impacts of covid-19 pandemic on user behaviors and environmental benefits of bike sharing: a big-data analysis," *Applied Energy*, vol. 285, 2021.
- [2] G. T. Reddy, M. P. K. Reddy, and K. Lakshmana, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, Article ID 54776, 2020.
- [3] L. Zhen, Y. K. Zhang, K. P. Yu, N. Kumar, and A. Barnawi, "Early collision detection for massive random access in satellite-based internet of things," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5184–5189, 2021.
- [4] Z. W. Guo, K. P. Yu, A. Jolfaei, F. Ding, and N. Zhang, "Fuz-spam: label smoothing-based fuzzy detection of spammers in internet of things," *IEEE Transactions on Fuzzy Systems*, 2021.
- [5] K. P. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, "Deep-learning-empowered breast cancer auxiliary diagnosis for 5gb remote e-health," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54–61, 2021.
- [6] F. Ding, G. P. Zhu, M. Alazab, X. J. Li, and K. P. Yu, "Deep-learning-empowered digital forensics for edge consumer electronics in 5g hetnets," *IEEE Consumer Electronics Magazine*, vol. 11, 2020.
- [7] F. Ding, G. P. Zhu, Y. C. Li, X. Zhang, P. K. Atrey, and S. Lyu, "Anti-forensics for face swapping videos via adversarial training," *IEEE Transactions on Multimedia*, 2021.
- [8] F. Ding, K. P. Yu, Z. H. Gu, X. J. Li, and Y. Q. Shi, "Perceptual enhancement for autonomous vehicles: restoring visually degraded images for context prediction via adversarial training," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [9] W. Z. Wang, M. H. Fida, Z. T. Lian et al., "Secure-enhanced federated learning for ai-empowered electric vehicle energy prediction," *IEEE Consumer Electronics Magazine*, 2021.
- [10] L. Tan, K. P. Yu, L. Lin, X. Cheng, G. Srivastava, and W. Wei, "Speech emotion recognition enhanced traffic efficiency solution for autonomous vehicles in a 5g-enabled space-air-ground integrated intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, 2021.
- [11] C. S. Feng, B. Liu, K. P. Yu, S. K. Goudos, and S. H. Wan, "Blockchain-empowered decentralized horizontal federated learning for 5g-enabled uavs," *IEEE Transactions on Industrial Informatics*, vol. 18, 2021.
- [12] C. S. Feng, B. Liu, Z. Guo, K. Yu, and Z. Qin, "Blockchain-based cross-domain authentication for intelligent 5g-enabled internet of drones," *IEEE Internet of Things Journal*, vol. 9, no. 8, 2021.
- [13] L. Yang, K. P. Yu, S. X. Y. Yang, C. Chakraborty, and Y. Lu, "An intelligent trust cloud management method for secure clustering in 5g enabled internet of medical things," *IEEE Transactions on Industrial Informatics*, 2021.
- [14] D. W. Wang, Y. X. He, K. P. Yu, L. Nie, and R. Zhang, "Delay sensitive secure noma transmission for hierarchical hap-lap medical-care iot networks," *IEEE Transactions on Industrial Informatics*, 2021.
- [15] H. Li, K. P. Yu, B. Liu, C. Feng, and Z. Qin, "An efficient ciphertext-policy weighted attribute-based encryption for the internet of health things," *IEEE journal of biomedical and health informatics*, 2021.
- [16] Y. Sun, J. Liu, K. P. Yu, M. Alazab, and K. X. Lin, "Pmrss: privacy-preserving medical record searching scheme for intelligent diagnosis in iot healthcare," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1981–1990, 2022.
- [17] L. Tan, K. P. Yu, N. Shi, C. Yang, W. Wei, and H. Lu, "Towards secure and privacy-preserving data sharing for covid-19 medical records: a blockchain-empowered approach," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 271–281, 2022.
- [18] H. Xiong, C. J. Jin, M. Alazab, H. Wang, W. Wang, and C. Su, "On the design of blockchain-based ecdsa with fault-tolerant batch verification protocol for blockchain-enabled iomt," *IEEE journal of biomedical and health informatics*, 2021.
- [19] W. Z. Wang, C. Qiu, Z. M. Yin, G. Srivastava, and C. Su, "Blockchain and puf-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, 2021.
- [20] C. S. Feng, K. P. Yu, M. Aloqaily, Z. Lv, and S. Mumtaz, "Attribute-based encryption with parallel outsourced decryption for edge intelligent iot," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, Article ID 13784, 2020.
- [21] Q. Zhang, K. P. Yu, Z. W. Guo, S. Garg, J. Rodrigues, and M. Guizani, "Graph neural networks-driven traffic forecasting for connected internet of vehicles," *IEEE Transactions on Network Science and Engineering*, 2021.
- [22] K. P. Yu, L. Tan, C. X. Yang, A. K. Bashir, and T. Sato, "A blockchain-based shamir's threshold cryptography scheme for data protection in industrial internet of things settings," *IEEE Internet of Things Journal*, 2021.
- [23] D. Y. Xu, K. P. Yu, and J. A. Ritcey, "Cross-layer device authentication with quantum encryption for 5g enabled iiot in industry 4.0," *IEEE Transactions on Industrial Informatics*, 2021.
- [24] K. P. Yu, L. Tan, S. Mumtaz, S. A. Rubaye, A. A. Dulaimi, and A. K. Bashir, "Securing critical infrastructures: deep-learning-based threat detection in iiot," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 76–82, 2021.
- [25] Y. Gong, L. Zhang, R. P. Liu, K. P. Yu, and G. Srivastava, "Nonlinear mimo for industrial internet of things in cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5533–5541, 2021.

- [26] L. Tan, K. P. Yu, F. P. Ming, X. F. Chen, and G. Srivastava, "Secure and resilient artificial intelligence of things: a honynet approach for threat detection and situational awareness," *IEEE Consumer Electronics Magazine*, 2021.
- [27] K. P. Yu, Z. W. Guo, Y. Shen, W. Wang, and T. Sato, "Secure artificial intelligence of things for implicit group recommendations," *IEEE Internet of Things Journal*, vol. 9, 2021.
- [28] T. Guo, K. P. Yu, M. Aloqaily, and S. H. Wan, "Constructing a prior-dependent graph for data clustering and dimension reduction in the edge of aiot," *Future Generation Computer Systems*, vol. 128, pp. 381–394, 2022.
- [29] Y. H. Peng, A. Jolfaei, and K. P. Yu, "A novel real-time deterministic scheduling mechanism in industrial cyber-physical systems for energy internet," *IEEE Transactions on Industrial Informatics*, 2021.
- [30] L. Zhao, H. Chai, Y. Han, K. Yu, and S. Mumtaz, "A collaborative v2x data correction method for road safety," *IEEE Transactions on Reliability*, 2022.
- [31] Z. Zhou, X. Dong, Z. Li, K. Yu, C. Ding, and Y. Yang, "Spatio-temporal feature encoding for traffic accident detection in vanet environment," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [32] L. Cai, J. Y. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Annals of Translational Medicine*, vol. 8, no. 11, 2020.
- [33] J. X. Zhuang, J. B. Cai, R. X. Wang, J. G. Zhang, and W. S. Zheng, "Deep knn for medical image classification," in *Proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 127–136, Springer International Publishing, Berlin, Germany, 2020.
- [34] H. A. Li, M. Zhang, Z. H. Yu, Z. L. Li, and N. Li, "An improved pix2pix model based on gabor filter for robust color image rendering," *Mathematical Biosciences and Engineering*, vol. 19, no. 1, pp. 86–101, 2021.
- [35] H. A. Li, Q. Y. Zheng, W. J. Yan, and R. Tao, "Image super-resolution reconstruction for secure data transmission in internet of things environment," *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 6652–6671, 2021.
- [36] J. Pang, Y. Gu, J. Xu, and G. Yu, "Research advance on similarity join queries," *Journal of Frontiers of Computer Science and Technology*, vol. 7, no. 1, pp. 1–13, 2013.
- [37] Y. Z. Ma, Z. H. Zhang, and C. J. Lin, "Research progress in similarity join query of big data," *Journal of Computer Applications*, vol. 38, no. 4, pp. 978–986, 2018.
- [38] S. Ferrada, B. Bustos, and N. Reyes, "An efficient algorithm for approximated self-similarity joins in metric spaces," *Information Systems*, vol. 91, 2020.
- [39] W. Lu, Y. Y. Shen, S. Chen, and C. B. Ooi, "Efficient processing of k nearest neighbor joins using mapreduce," *Proceedings of the VLDB Endowment*, vol. 5, no. 10, pp. 1016–1027, 2012.
- [40] J. Dai and Z. M. Ding, "Mapreduce based fast knn join," *Chinese Journal of Computers*, vol. 38, no. 1, pp. 99–108, 2015.
- [41] X. J. Zhao, J. F. Zhang, and X. Qin, "knn-dp: handling data skewness in knn joins using mapreduce," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 3, pp. 600–613, 2018.
- [42] G. Song, J. Rochas, and F. Huet, "Solutions for processing k nearest neighbor joins for massive data on mapreduce," in *Proceedings of the 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*, pp. 279–287, Turku, Finland, March 2015.
- [43] G. Song, J. Rochas, E. L. Beze, and F. Huet, "K nearest neighbour joins for big data on mapreduce: a theoretical and experimental analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2376–2392, 2016.
- [44] A. Stupar, S. Michel, and R. Schenkel, *Rankreduce-processing K-Nearest Neighbor Queries on Top of Mapreduce*, LSDS-IR@SIGIR, 2010.
- [45] Y. Hu, G. Peng, Z. H. Wang, Y. R. Cui, and H. Qin, "Partition selection for large-scale data management using knn join processing," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–14, Article ID 7898230, 2020.
- [46] Y. X. Fang, R. Cheng, W. B. Tang, S. Maniu, and S. X. Yang, "Scalable algorithms for nearest-neighbor joins on big trajectory data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 785–800, 2016.
- [47] G. Chatzigeorgakidis, S. Karagiorgou, S. Athanasiou, and S. Skiadopoulos, "Fml-knn: scalable machine learning on big data using k-nearest neighbor joins," *Journal of Big Data*, vol. 5, no. 1, 2018.
- [48] Q. S. Du and X. F. Li, "A novel knn join algorithms based on hilbert r-tree in mapreduce," in *Proceedings of the 3rd International Conference on Computer Science and Network Technology*, pp. 417–420, Dalian, China, October 2013.
- [49] Y. Kim and K. Shim, "Parallel top-k similarity join algorithms using mapreduce," in *Proceedings of the IEEE 28th International Conference on Data Engineering*, pp. 510–521, Arlington, VA, USA, 2012.
- [50] D. H. Chen, C. G. Shen, J. Y. Feng, and J. J. Le, "An efficient parallel top-k similarity join for massive multidimensional data using spark," *International journal of database theory and application*, vol. 8, no. 3, pp. 57–68, 2015.
- [51] Y. Z. Ma, X. Ci, and X. F. Meng, "Parallel top-k join on massive high-dimensional vectors," *Chinese Journal of Computers*, vol. 38, no. 1, pp. 86–98, 2015.
- [52] B. Lei, J. Xu, Y. Gu, and G. Yu, "Parallel top-k similarity join algorithm on probabilistic data based on earth mover's distance," *Journal of Software*, vol. 24, no. s2, pp. 188–199, 2013.
- [53] J. Huang, R. Zhang, R. Buyya, and J. Chen, "Melody-join: efficient earth mover's distance similarity joins using mapreduce," in *Proceedings of the IEEE 30th International Conference on Data Engineering*, pp. 808–819, Chicago, IL, USA, March 2014.
- [54] X. Jia, B. Lei, Y. Gu, and Z. Zhang, "Efficient similarity join based on earth mover's distance using mapreduce," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2148–2162, 2015.
- [55] J. Huang, R. Zhang, R. Buyya, J. Chen, and Y. W. Wu, "Heads-join: efficient earth mover's distance similarity joins on hadoop," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 6, pp. 1660–1673, 2016.
- [56] T. Christiani, R. Pagh, and J. Sivertsen, "Scalable and robust set similarity join," in *Proceedings of the IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 1240–1243, Paris, France, April 2018.
- [57] M. Gowanlock and B. Karsin, "Accelerating the similarity self-join using the gpu," *Journal of Parallel and Distributed Computing*, vol. 133, no. 9, pp. 107–123, 2019.
- [58] L. N. Yu, T. Z. Nie, D. R. Shen, and Y. Kou, "An approach for progressive set similarity join with gpu accelerating," in *Proceedings of the Web Information Systems and Applications. WISA*, pp. 155–167, 2020.
- [59] F. Shao, G. Chen, L. H. Yu, Y. J. Bei, and J. X. Dong, "Bitmap filtering: an efficient speedup method for xml structural matching," in *Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence,*

- Networking, and Parallel/Distributed Computing (SNPD 2007)*, vol. 3, pp. 756–761, Qingdao, China, July 2007.
- [60] X. F. Ding, W. L. Yang, R. K. K. Choo, X. L. Wang, and H. Jin, “Privacy preserving similarity joins using mapreduce,” *Information Sciences*, vol. 493, pp. 20–33, 2019.
 - [61] J. C. Wu, Y. Zhang, J. Wang, C. Lin, Y. Fn, and C. Xing, “Scalable metric similarity join using mapreduce,” in *Proceedings of the IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 1662–1665, Macao, China, April 2019.
 - [62] Y. Z. Ma, X. F. Meng, and S. Y. Wang, “Parallel similarity joins on massive high dimensional data using mapreduce,” *Concurrency and Computation: Practice and Experience*, vol. 28, no. 1, pp. 166–183, 2016.
 - [63] Y. Z. Ma, S. J. Jia, and Y. X. Zhang, “A novel approach for high dimensional vector similarity join query,” *Concurrency and Computation: Practice and Experience*, vol. 29, no. 5, 2017.
 - [64] Y. Z. Ma, S. J. Jia, and Y. X. Zhang, “Chi-square distribution based similarity join query algorithm on high-dimensional data,” *Journal of Computer Applications*, vol. 36, no. 7, pp. 1993–1997, 2016.

Research Article

Platform Firm's IT Capabilities, External Informal Knowledge Governance, and Green Knowledge Integration in Low-Carbon Economy

Guanghua Fu  and Bencheng Li

Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China

Correspondence should be addressed to Guanghua Fu; ghfu@shmtu.edu.cn

Received 15 January 2022; Revised 15 February 2022; Accepted 17 February 2022; Published 13 April 2022

Academic Editor: G. Thippa Reddy

Copyright © 2022 Guanghua Fu and Bencheng Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Platform ecosystem provides internal firms with an abundant source of green technology knowledge for sustainable development. In a low-carbon economy, green technology could be accumulated via platform, thus utilized to achieve energy conservation and emission reduction. From firm's social capital perspective, this study explores the effects of platform firm's IT capabilities and external informal knowledge governance on green knowledge integration. A theoretical model is constructed about their direct and interactive effects. Based on 372 samples of platform firms, the empirical test results show that IT capabilities have a significant positive impact on collaborative and systematic green knowledge integration; external informal knowledge governance has a significant positive impact on socialized and collaborative green knowledge integration; and their interactive effects pose significant positive impact on socialized, collaborative, and systematic green knowledge integration.

1. Introduction

Low-carbon economy is advocated by an increasing number of countries as climate changes and environmental pollution looms. It has been one development mainstream of the world economy. EU, China, and many other economies are requiring firms to participate into the carbon reduction activities with concrete reduction goals or routines (e.g., carbon neutrality). Besides, from ecological perspective, industry 4.0 emphasizes the efficiency of resource and energy, which is driven by the harmony between economy and ecological environment via green technology innovation (e.g., 5G, AI) [1, 2]. In these contexts, firms, main carbon emitters, should assume key responsibilities of reducing carbon emission. In the last decades, various platforms are prevailing as firms are directly and indirectly linked by online transaction based on digital technology, especially in the era of industry 4.0 and globalization. They exist in various forms (e.g., e-commerce platform, smartphone platform). And they are usually initiated or

created by key stone firms (e.g., Amazon, Alibaba Inc.). Platform, as a popular organization cluster, provides a key micro context for collaboration, cooperation and coevolution among those firms surrounding the key stone firms or dominator firms [3]. With aims of peak carbon dioxide by 2030 and carbon neutrality by 2060 committed by Chinese government, Chinese platform firms have developed rapidly in recent years. These platform firms are associated and coordinated by surrounding customer value, introducing platform ecological effects that could complement each other. The symbiosis environment in platform ecosystem provides good external conditions and abundant resources for innovation and sustainable development by searching, sharing, integrating, and creating technical knowledge inside and outside the internal firms [4, 5]. Green technology knowledge integration helps firms integrate green technology into the link of opportunity identification and innovation, thereby promotes energy conservation, emission reduction, and low-carbon development [6].

In green knowledge integration (GKI) of platform firm, many factors (e.g., knowledge characteristics, market mechanism failure) will affect technological innovation [7]. For example, the deeper the tacitness and embedding of knowledge, the more difficult its observation, which will affect the knowledge sharing, dissemination, and absorption among different platform firms. And information asymmetry, sharing dilemma, and evaluation difficulties also hinder knowledge exchange and collaboration. These problems will lead to knowledge hiding, “free riding”, and other risks, so GKI should be conducted from the perspective of governance [8, 9]. In platform ecosystem, firms are linked via symbiosis. Due to the lack of formal mechanism (e.g., bureaucracy), platform firm’s external informal knowledge governance (EIKG) is needed to guide green knowledge activities [10]. Different from internal informal governance tools or methods (e.g., corporate culture, personal relationship), firm’s external knowledge governance primarily relies on organizational relationship, external network structure, etc. [11, 12]; however, existing studies on knowledge governance mainly focus on firm’s internal teams or departments, and external firm’s EKG still needs further exploration.

In platform ecosystem, low-carbon economy requires efficient coordination among platform firms. This process is accompanied by information circulation (e.g., data, knowledge), which is inseparable from information technology. Therefore, platform firm’s IT capabilities (ITC) are key elements to improve the level of green knowledge management [13]. In low-carbon economy, environmental capacity is limited and environmental awareness is constantly increasing. Thus, platform firms would utilize information technologies (e.g., cloud computing, AI) to acquire more green knowledge and improve the efficiency of green knowledge creation and application. In firm’s knowledge activities, ITC could promote knowledge sharing by providing data processing system and knowledge sharing platform [14]. However, the relationship between ITC and GKI needs further exploration especially in the context of platform ecosystem, such as effects of ITC on GKI and EIKG on GKI in the platform.

To explore these effects in platform firms, this study introduces social capital into firm’s EIKG, constructs a conceptual model about these effects, collects 372 sample data of platform firms from China, and verifies their relationships through statistical analysis. This paper makes the following main contributions:

- (1) To explore the effects of platform firm’s ITC on the subdimensions of GKI and verify these effects
- (2) To explore the effects of platform firm’s EIKG on the subdimensions of GKI from firm’s social capital perspective and discover the different roles of EIKG in the subdimensions of GKI
- (3) To explore the interactive effects of platform firm’s ITC and EIKG on the subdimensions of GKI, verify these effects, and analyze the moderating effects in the interaction

The rest of the paper is organized as follows. Section 2 briefly reviews the relevant literature and introduces the conceptual model. Section 3 describes the methodology used for data collection and measurement. Section 4 presents model test and analysis. Section 5 summarizes the conclusion and presents possible future work.

2. Literature Review and Concept Model Construction

2.1. Platform Firm’s IT Capabilities and Green Knowledge Integration. Platform ecosystem is a super-organization composed of interrelated groups of organizations that share common vision and achieve coevolution and sustainable development through cooperative innovation [15]. It emphasizes symbiosis and openness, and provides environmental support for open innovation and low-carbon development [16]. In low-carbon economy, relying on various symbiotic relationships in platform ecosystem (e.g., mutualism, commensalism, and parasitism symbiosis), firms could achieve low-carbon-related technology and knowledge sharing, complementation, optimization, and collaboration with other members, then green technology knowledge integration and application to achieve energy conservation and emission reduction. Therefore, the efficiency of green technology knowledge activities would not only affect platform firm’s green technology innovation and competitive advantages but also affect the energy conservation and emission reduction of the entire platform ecosystem.

Green knowledge is judged from its long-term positive impact on environment, and generally includes green technologies and shared vision that could protect the environment and promote firm’s sustainable development [17]. Usually, platform firm’s external green technology knowledge is mainly from active sharing, mutual beneficial exchange, or transaction among internal firms of the platform ecosystem, which is produced in the process of firm allying and common business transaction. In addition, core firm could enrich green knowledge of the platform ecosystem by attracting firms with complementary knowledge to join and actively promote their knowledge sharing and exchange. Noncore firm (e.g., dominator and niche firms) will actively participate in the knowledge activities to obtain environment-friendly knowledge [18]. Because of the strong complementarity of platform firms’ knowledge and the small green knowledge distance, it is easy for platform firms to share and integrate their green knowledge under the symbiosis vision and the leadership of core firms in platform ecosystem.

Green knowledge integration (GKI) refers to the dynamic identification, collection, reconstruction, and optimization of green knowledge by firms, so that green knowledge could be systematically linked and integrated, which is conducive to knowledge creation and innovation [19]. It involves changes about knowledge forms, organization, and transferring [20]. GKI could not only promote firms’ internal green technology innovation but also help them perceive the market demand for low-carbon economy

and energy conservation. GKI urges firms to cross organizational boundaries to innovate, so firms' external knowledge reconfiguration is inevitable. From the perspective of integration method, firm's green knowledge integration could be divided into socialized green knowledge integration (SOGKI), collaborative green knowledge integration (COGKI), and systematic green knowledge integration (SYGKI). Inside platform firm, SOGKI is embodied by the establishment of shared visions and values to promote knowledge integration. It has characteristics of high integration efficiency, narrow scope, and low flexibility. COGKI is embodied by firms' participation in common business activities to promote knowledge exchange and integration between organizations. It has low integration efficiency, wide scope, and high flexibility. SYGKI is reflected in firm's existing knowledge coding, reorganizing, flowing, and other standardized knowledge processes. It has high integration efficiency, narrow scope, and low flexibility [21]. In platform ecosystem, firm's GKI has many problems (e.g., knowledge vagueness, embeddedness, tacitness, context dependence, specificity, and pricing difficulty, sparse distribution, rapid changing). Some problems lead to the failure of firm's formal governance mechanism. While the low-carbon economy poses high requirements for their green knowledge integration and innovation.

IT not only affects the knowledge network structure of platform ecosystem, but also affects the acquisition, creation, and application of knowledge by influencing information collection, storage, transmission, and processing [14]. Firm's ITC is a complex collection of internal and external IT-related resources, skills, coordination activities, and business practice knowledge of using IT assets to achieve expected goals [22, 23], involving the comprehensive transferring and deploying IT-based and relevant resources [24]. Firm's ITC could significantly promote knowledge sharing and integration with suppliers and customers [25]. Therefore, platform firm could facilitate the sharing and integration of green knowledge by building ITC.

In knowledge activities, firms mainly apply ITC to connect dispersed information carriers, integrate multilevel information, and optimize knowledge networks to enhance their ability to integrate, match, and innovate business processes including knowledge activities [26]. Regarding IT roles in interfirm business, Rai and Tang [27] divided firm's ITC into IT integration capabilities (ITIC) and IT reset capabilities (ITRC) from the perspective of organizational cooperation. ITIC are the abilities to integrate business data, communication technology, and business collaboration systems among firms via IT. They achieve the integration of business processes including knowledge among firms by resolving differences in the information business level (e.g., differences in data structures and business processes) and coordinating differences in software and hardware (e.g., differences in hardware infrastructure and information system compatibility). ITRC are the abilities to expand and reorganize IT resources according to external business needs. They are based on information system modularization, interface standardization, then realizes the support of information technology for adjustment of firm's knowledge activities.

In platform ecosystem, firm's ITC are embedded in the organizational business processes and could promote value chain development via supporting and optimizing business processes: firms utilize ITC to connect knowledge nodes on the platform ecological chain, optimize knowledge network structure and business processes, and improve the efficiency of knowledge activities to support the business ecological chain [28]. During these processes, core firms will actively use IT to establish a knowledge sharing platform to promote business collaboration and knowledge integration. Noncore firms could also use IT to increase their participation in knowledge activities. With the assistance of ITC, platform firms could link external knowledge nodes with external knowledge networks, meanwhile perform explicit coding and rapid transmission of information to achieve rapid aggregation, sharing, and transfer of knowledge. Platform firm's ITC could also realize the intelligent processing, mining, and efficient retrieval of knowledge to facilitate the systematic integration of knowledge [29]. Therefore, ITIC could positively affect knowledge management including GKI. When the demand for energy conservation and emission reduction related to green innovation changes, ITRC could help firms quickly adjust external knowledge networks and coordinate knowledge activities [30], so as to quickly obtain green knowledge needed to support the low-carbon development. Based on the above analysis, this article proposes the following hypotheses:

Hypothesis 1a (H1a): platform firm's IT capabilities could positively affect the socialized green knowledge integration.

Hypothesis 1b (H1b): platform firm's IT capabilities could positively affect the collaborative green knowledge integration.

Hypothesis 1c (H1c): platform firm's IT capabilities could positively affect the systematic green knowledge integration.

2.2. External Informal Knowledge Governance and Green Knowledge Integration

2.2.1. Platform Firm's External Knowledge Governance.

The complexity and characteristics (e.g., vagueness, embeddedness, tacitness, and context dependence) of knowledge activities hinder knowledge sharing and transferring. However, in the platform ecosystem, there is a lack of restraints and incentives from contracts, hierarchical organization, culture, economy, etc. Therefore, platform firms need to use relationship and common cognition, trust, agreements, etc. to promote external knowledge transfer [31]. This determines that knowledge organizing and managing need coordination and optimization from governance perspective. Knowledge governance is the process that an organization coordinates knowledge exchanging and sharing between internal and external knowledge nodes to optimize knowledge acquisition, creation, and distribution [32]. Foss and Michailova [33] divide knowledge governance into formal knowledge governance (governance through the

application of formal bureaucratic mechanisms such as power and institutions) and informal knowledge governance (governance through the application of informal mechanisms such as culture and relationship). Firms should choose a more effective combination of governance mechanisms based on specific context to positively affect knowledge sharing and integration.

In platform ecosystem, firms lack formal mechanisms (e.g., relevant rules and regulation, contracts) for external knowledge activities, so their restraint and incentive effects are insufficient. And some problems (e.g., knowledge hiding and pricing difficulty) also hinder the knowledge exchange through market mechanisms. Therefore, informal mechanisms (e.g., relationship governance, network governance) should be applied to influence external knowledge activities [34, 35]. Firm's EIKG is mainly achieved by tapping its social capital (e.g., mutual trust, network relationship, consensus among platform firms) [12, 19, 25]. Firm's social capital is the sum of relationship, cognition, and structure that could affect external information activities and could be used to coordinate knowledge acquisition and sharing among firms [25, 31, 36, 37]. It could be divided into structural capital, cognitive capital, and relational capital [38]. Thus, external informal knowledge governance from social capital perspective for platform firms could be conducted from three dimensions: structure, cognition, and relationship. (1) On the structural dimension, informal governance mainly establishes the niche of the firm through the positioning mechanism on the business ecological chain and affects the platform ecosystem. (2) On the cognitive dimension, informal governance mainly promotes knowledge sharing and open innovation among internal members through the establishment of symbiosis mechanism. (3) On the relational dimension, informal governance mainly relies on trust mechanism establishment among members to maintain and promote the formation and development of the business ecological chain, thereby promoting cooperation between knowledge nodes. Therefore, EIKG of platform firm could be mainly launched from three aspects: positioning mechanism (PM), symbiosis mechanism (SM), and trust mechanism (TM).

2.2.2. External Informal Knowledge Governance and Green Knowledge Integration. As for PM, platform firms mainly have keystone, dominator, and niche positioning [39]. Keystone positioning means platform firm is the initiator and leader of the platform who builds and optimizes the business ecological chain through core technologies and business models, and it creates business niches for other positioning. Dominator positioning firms build key niches in platform ecosystem through business collaboration via key technologies and business links. Niche firms are small ones that are in the gap or on the edge of platform ecosystem who diversifies platform development.

To occupy or re-establish the business niche through technological and business innovation, PM would utilize

internal and external resources, business models, etc., to establish business relationship with other system members [40, 41]. This process is accompanied by the formation, optimization, and reconstruction of firm's external knowledge network. Firm's positioning with greater influence needs more green knowledge, while such positioning could help them consolidate the favorable position in the knowledge network. This positioning would conversely affect their external green knowledge acquisition, absorption, and integration [42].

As for SM, firm's symbiosis is based on labor division and coevolution of the value chain in platform ecosystem [43]. Platform firms establish corresponding symbiotic infrastructure (e.g., industry-university-research platform, supply chain platform). They would collaborate with each other to achieve the low-carbon development of the entire platform. Therefore, they should establish common knowledge activity standards (e.g., knowledge exchanging and sharing process) to facilitate green knowledge sharing and collaboration. In addition, the symbiosis vision could promote platform firms to reach a consensus on cooperation in the organizing, transferring, and sharing of green knowledge. Firm's symbiotic relationship with higher mutual benefit tends to increase green knowledge sharing and open innovation to obtain more green knowledge.

As for TM, trust is the subjective belief that the two parties do not expect opportunistic behaviour from each other, which could prompt firms to advance potential transactions according to expectations without worrying about being misused by other parties [44]. Trust could effectively reduce the risk and uncertainty in green knowledge activities [45]. More trust could also increase the relationship strength between organizations, thereby increasing their green knowledge cooperation, sharing or exchange [46]. Additionally, trust could reduce knowledge transaction costs and increase benefits of knowledge activities [47]. To sum up, EIKG from the perspective of social capital could positively affect green knowledge transfer among platform firms, thus affecting their GKI. Therefore, this article proposes the following hypotheses:

Hypothesis 2a (H2a): external informal knowledge governance could positively affect socialized green knowledge integration of platform firms.

Hypothesis 2b (H2b): external informal knowledge governance could positively affect collaborative green knowledge integration of platform firm.

Hypothesis 2c (H2c): external informal knowledge governance could positively affect systematic green knowledge integration of platform firms.

2.3. Interactive Effects of IT Capabilities and External Informal Knowledge Governance. The impact of IT capabilities (ITC) on external informal knowledge governance (EIKG). In the

process of applying EIKG, firstly, ITIC could help platform firms establish a knowledge transmission network which could improve the efficiency of green knowledge transfer and reduce costs. Secondly, ITIC could help platform firms systematically integrate data, knowledge in different formats and levels through systematic coding and standardized processing. Furthermore, ITIC could help platform firms adjust their knowledge network to optimize business processes. Therefore, the roles of ITC in the construction of firms' green knowledge network and systematic knowledge processing will directly affect their ecological positioning and knowledge acquisition; the roles of ITC in business collaboration and the establishment of symbiotic relationship could promote the communication in platform ecosystem chain, and strengthen the symbiotic relationship and trust relationship among platform firm. The impact of EIKG on ITC. PM in EIKG would guide platform firm's ITC. The IT strategic thinking under SM could prompt platform firms to use IT in green knowledge transfer and collaboration to attract external firms, thereby expanding green knowledge network in platform ecosystem and green knowledge source. The sustainable development under the SM could guide the construction and application of ITRC (e.g., IT upgrade, business process reengineering). Finally, the application of TM could improve the effects of ITC by reducing the risk of opportunism. Therefore, EIKG could guide and promote the establishment, application, and upgrade of ITC in the process of GKI.

In summary, this article proposes the following hypotheses:

Hypothesis 3a (H3a): the interactive effect of IT capabilities and external informal knowledge governance has a positive impact on platform firm's socialized green knowledge integration.

Hypothesis 3b (H3b): the interactive effect of IT capabilities and external informal knowledge governance has a positive impact on platform firm's collaborative green knowledge integration.

Hypothesis 3c (H3c): the interactive effect of IT capabilities and external informal knowledge governance has a positive impact on platform firm's systematic green knowledge integration.

Based on the hypotheses above, this paper constructs a conceptual model (Figure 1) about the impacts of platform firm's ITC, EIKG, and their interaction on the sub-dimensions of green knowledge integration.

3. Methodology

To verify the conceptual model and its hypotheses proposed above, empirical research is conducted whose process is shown in Figure 2. In Figure 2, initial scales from literature review are adopted and revised, final scales are formed via preliminary investigation, and sample data are collected via sampling questionnaire survey. These data are subjected to reliability and validity check, then analyzed via multiple

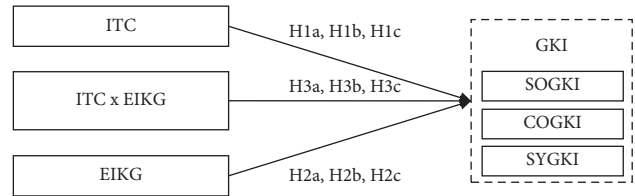


FIGURE 1: Conceptual model.

linear regression method, and further moderating effects are tested for analyzing the results of hypothesis verification.

3.1. Data Collection. This study uses empirical analysis to verify the proposed conceptual model. Chinese platform ecosystems have developed rapidly in recent years, and they are accelerating their low-carbon transformation, and actively conducting green knowledge integration. Therefore, they provide sufficient samples for this study. The survey subjects are selected from three e-commerce platform ecosystems (Taobao, JD, and Pinduoduo). These three platform ecosystems are relatively mature, and the number of their internal firms is very large. Four types of subjects in these platform ecosystems (core firms, flagship stores, ordinary merchants, and logistics firms) cover the roles of keystone, dominator, and niche firms. There are two ways to issue questionnaires: (1) issuing questionnaires to managers and first-line workers from different departments in core firms of these platforms, and a total of 100 questionnaires are collected; (2) issuing questionnaires to those people from clothing, home appliances, food, book flagship stores, logistics firms via platform instant messenger, and 600 questionnaires are collected. Questionnaire survey was conducted from May to December 2020. Except invalid questionnaires, a total of 372 valid questionnaires are collected.

3.2. Measurement. The questionnaire contains two parts: background information and variable scales. Initial items of variable scales mainly come from existing mature scales, and the indicators use Likert's 7-point scale (1 means "completely disagree/do not conform," 7 means "completely agree/conform"). We firstly used the Delphi method to improve the initial scale by inviting 5 experts to modify the scales for three rounds; then conducted a preliminary investigation on one e-commerce platform (collecting 61 samples of data). Via reliability and validity analysis of the scales, we eliminated two items with underloading factors to get the final scales.

- (1) *Green Knowledge Integration (GKI)*. The measurement items refer to the research of Long [12], Prieto-Pastor [25], and Mehta [19], including 4 socialized green knowledge integration items (SOGKI1—SOGKI4), 4 collaborative green knowledge integration items (COGKI1—COGKI4), and 4 systematic green knowledge integration items (SYGKI1—SYGKI4).

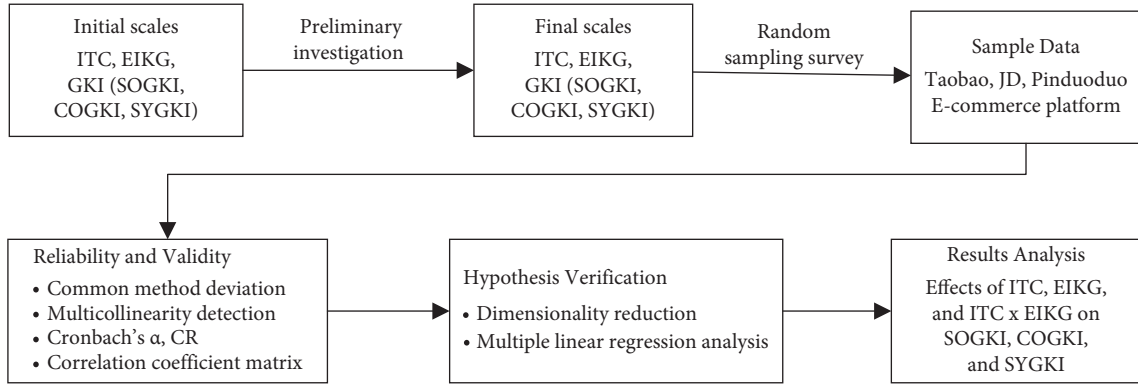


FIGURE 2: Methodology of model verification.

- (2) *IT Capabilities (ITC)*. The measurement items come from the scale of Rai et al. [27], including IT integration capabilities (ITIC) and IT reset capabilities (ITRC). ITIC includes 3 measurement items (ITIC1-ITIC3). ITRC includes 3 measurement items (ITRC1-ITRC3).
- (3) *External Informal Knowledge Governance (EIKG)*. The measurement items refer to the research of Iansiti [39], Prieto-Pastor [25], etc., including 3 positioning mechanism items (PM1-PM3), 5 symbiosis mechanism items (SM1-SM5), and 4 trust mechanism items (TM1-TM4).
- (4) *Controls*. The age and scale of platform firms are controlled for their possible effects on knowledge integration according to the existing literature [48].

4. Model Test and Analysis

4.1. Validity Test. In multiple linear regression analysis, common method deviation and multicollinearity are two common threats to the validity of research model. And their prevention and check are as follows:

- (1) *Common Method Deviation*. After adopting common method deviation prevention technology (e.g., anonymity of questionnaires, terminology annotation, item order changing, 3 reverse questions, and item expression improvement), this study uses Harman's single factor test to check the common method deviation: the KMO value without rotation is 0.917, the chi-square value is 146.172, and the significance level is 0.000. The factor analysis of all items shows that the contribution value of the first principal component factor is 29.743%, so the impact of common method deviation in the study is not significant.
- (2) *Multicollinearity Detection*. This study uses variance inflation factor (VIF) as the indicator to check the multicollinearity among variables. The VIFs of ITIC, ITRC, PM, SM, and TM are 4.247, 4.324, 1.751, 2.988, 2.056, respectively. They are all less than 5, so there is no multicollinearity problem among related variables.

4.2. Reliability and Validity Analysis. In this study, exploratory factor analysis and confirmatory factor analysis were applied to check the reliability and validity of the variable scales. The results are shown in Table 1. The Cronbach's α of subdimensions of ITC, EIKG, and GKI in Table 1 are all greater than 0.7, so the scales have good internal consistency. In addition, the minimum CR (construct reliability) of each variable subdimension is 0.806 (greater than 0.7), so the scales have good construction reliability. In summary, the reliability of the scales is acceptable.

In terms of validity, the designing process of the questionnaire ensures the content validity of the scales. In the confirmatory factor analysis, the minimum factor loading of each subdimension is 0.718 (greater than 0.6), the maximum cross-factor loading is 0.336 (less than 0.4), and the minimum value of AVE of all subdimensions in Table 1 is 0.574 (greater than 0.5). In addition, the correlation coefficient values between the variables in Table 2 are all less than 0.7. In summary, the scales have good convergence validity and discriminative validity.

4.3. Hypothesis Verification

- (1) *Dimensionality Reduction of Independent Variables*. Since independent variables (ITC, EIKG) are all high-dimensional variables, we firstly use the projection pursuit method [49] to reduce their dimensionality. The projection pursuit method could project non-normal distributed and non-linear high-dimensional data into a one-dimensional space that could reflect their principal information or characteristics and avoid possible mutual influence. Its operation are as follows [50]: (i) Normalize the item data of independent variables with minimum-maximum method: $x^*(i, j) = (x(i, j) - x_{\min}(j)) / (x_{\max}(j) - x_{\min}(j))$ ($x(i, j)$ is the value of j th question of sample i , and $x_{\max}(j)$, $x_{\min}(j)$ are the maximum and minimum of the j th question in all samples, respectively). (ii) Construct the projection index function that is the projection direction. (iii) Optimize the projection index function via genetic algorithm. (iv) Calculate the one-dimensional independent variable value

TABLE 1: Reliability and validity analysis of the scales.

| Variable | Subdimension | Cronbach's α | CR | AVE |
|----------|--------------|---------------------|-------|-------|
| ITC | ITIC | 0.789 | 0.849 | 0.651 |
| | ITRC | 0.747 | 0.854 | 0.661 |
| EIKG | PM | 0.782 | 0.806 | 0.581 |
| | SM | 0.791 | 0.870 | 0.574 |
| | TM | 0.802 | 0.859 | 0.605 |
| GKI | SOGKI | 0.834 | 0.877 | 0.640 |
| | COGKI | 0.797 | 0.895 | 0.680 |
| | SYGKI | 0.827 | 0.869 | 0.625 |

TABLE 2: Correlation coefficient matrix and square root of average variance extraction value ($N = 372$).

| Variables | ITIC | ITRC | PM | SM | TM | SOGKI | COGKI | SYGKI |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| ITIC | 0.807 | | | | | | | |
| ITRC | 0.167* | 0.813 | | | | | | |
| PM | 0.201 | 0.213 | 0.762 | | | | | |
| SM | 0.102* | 0.094* | 0.431 | 0.758 | | | | |
| TM | 0.262 | 0.309 | 0.101 | 0.205* | 0.778 | | | |
| SOGKI | 0.104 | 0.239 | 0.512** | 0.357* | 0.413** | 0.801 | | |
| COGKI | 0.351** | 0.287** | 0.245 | 0.401** | 0.351** | 0.155* | 0.824 | |
| SYGKI | 0.574** | 0.197** | 0.303 | 0.198** | 0.339 | 0.203 | 0.112 | 0.791 |

Notes: (1) Diagonal elements are the square roots of AVE; (2) * $p > 0.05$; (3) ** $p < 0.05$.

based on the optimal projection direction. Use Python program to calculate the optimal projection directions of ITC and EIKG. They are $a_1 = (0.513, 0.401, 0.351, 0.336, 0.412, 0.502)$ and $a_2 = (0.310, 0.323, 0.312, 0.216, 0.122, 0.149, 0.287, 0.251, 0.314, 0.402, 0.301, 0.350)$, which are all unit length vectors.

- (2) *Multiple Linear Regression Analysis*. To verify the hypotheses proposed above, based on dimensionality reduction of independent variables, this study followed the recommendation of Baron and Kenny [51], step-by-step introduced control variables, independent variables (ITC, EIKG), interactive effect of independent variables, and analyzed the multiple linear regression models with three dependent variables (SOGKI, COGKI, SYGKI). The multiple regression analysis results are shown in Table 3.

As shown in Table 3, the p values corresponding to the F values in all models with independent variables are all less than 0.01, so all models are effective in statistics. After introducing two independent variables and their interactive effects in sequence into the models with three dependent variables, the values of adjusted R^2 are all improved. Therefore, the fit of the model containing interactive effects is better.

In SOGKI submodels, no significant correlation exists between ITC and SOGKI (Model 1), so hypothesis 1a is not supported. While EIKG alone is positively correlated with SOGKI since the coefficient is 0.469 ($p < 0.001$) in Model 2. This supports hypothesis 2a. As for interactive effect of independent variables on EIKG in Model 3, the correlation is highly significant ($p < 0.001$) and positive (the coefficient is 0.191), so hypothesis 3a is supported.

In COGKI submodels, ITC are significantly and positively correlated with COGKI in Model 5 (the coefficient is 0.313, and $p < 0.01$), thus hypothesis 1b is supported. And EIKG has a highly significant positive correlation with COGKI in Model 5 since the coefficient is 0.395, and $p < 0.01$. This provides support for hypothesis 2b. In terms of interactive effect of independent variables in Model 6, the correlation is highly significant ($p < 0.01$) and positive (the coefficient is 0.469), which supports hypothesis 3b.

In SYGKI submodels, ITC and SYGKI are highly significantly positively correlated in Model 8 since the correlation coefficient is 0.465, and $p < 0.001$. Thus, hypothesis 1c is supported. While the correlation between EIKG and SYGKI is not significant, thus hypothesis 2c is not supported. In terms of interactive effect of independent variables in Model 9, the correlation is highly significant ($p < 0.01$) and positive (the coefficient is 0.142). This result yields support for hypothesis 3c.

4.4. Result Analysis

- (1) ITC could promote COGKI and SYGKI, but there is no positive correlation of ITC with SOGKI. IT could improve the efficiency of explicit knowledge collection, standardized storage, and transmission. Therefore, the application of ITC could enhance the collaborative and systematic integration of explicit green knowledge among platform firms. The reason for the failure of the relationship between ITC and SOGKI may be that the latter requires the guidance of shared cognition such as symbiotic vision. Therefore, without the aid of EIKG, pure ITC may not be able to promote the SOGKI of platform firms.

TABLE 3: Multiple linear regression analysis.

| | SOGKI | | | COGKI | | | SYGKI | | |
|----------------------------|---------|-----------|-----------|---------|----------|----------|---------|-----------|----------|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 |
| <i>Controls</i> | | | | | | | | | |
| Age | 0.175 | 0.111 | 0.135 | 0.113 | 0.081 | 0.092 | 0.074 | 0.069 | 0.079 |
| Scale | 0.304* | 0.227* | 0.282* | 0.412 | 0.411 | 0.405 | 0.185 | 0.173* | 0.170* |
| <i>IV</i> | | | | | | | | | |
| ITC | | 0.182 | 0.177 | | 0.313** | 0.308** | | 0.465*** | 0.436** |
| EIKG | | 0.469*** | 0.437*** | | 0.395** | 0.342** | | 0.118 | 0.142 |
| <i>Interactive Effects</i> | | | | | | | | | |
| ITC X | | | 0.191** | | | 0.237** | | | 0.305** |
| <i>EIKG</i> | | | | | | | | | |
| R^2 | 0.054 | 0.501 | 0.623 | 0.035 | 0.634 | 0.782 | 0.022 | 0.401 | 0.537 |
| ΔR^2 | 0.054 | 0.498 | 0.620 | 0.035 | 0.621 | 0.763 | 0.022 | 0.386 | 0.528 |
| F | 2.103 | 49.554*** | 42.968*** | 2.583 | 28.526** | 25.287** | 1.331 | 35.108*** | 29.701** |

Notes: (1) * $p < 0.05$; (2) ** $p < 0.01$; (3) *** $p < 0.001$; (4) Two-tailed test.

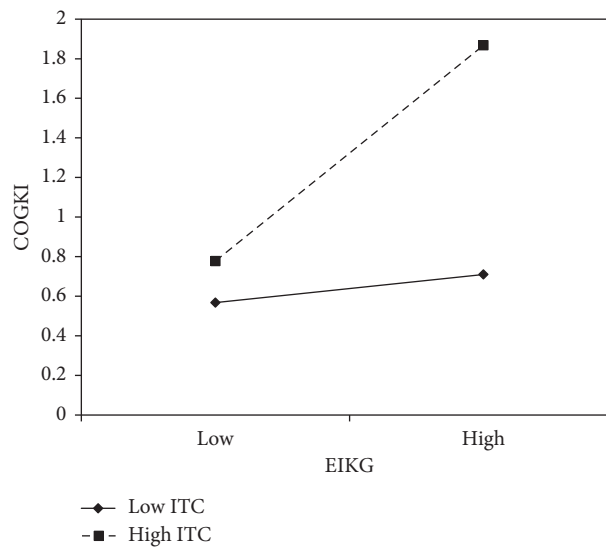


FIGURE 3: Interactive effect of EIKG and ITC on COGKI.

In addition, SOGKI also involves the learning and dissemination of tacit knowledge. These issues may be greatly influenced by factors such as trust, culture, etc., and are not closely related to ITC.

- (2) EIKG could promote SOGKI and COGKI, but it has no positive correlation with SYGKI. In platform ecosystem where formal governance mechanism is insufficient, firms could use social capital to influence GKI in terms of ecological positioning, vision, and trust establishment: ecological positioning helps firm occupy the target position of the platform ecosystem (appropriate knowledge network nodes), establish the structural basis for green knowledge integration; the implementation of the symbiosis vision and the establishment of trust will help firm reach a consensus on green knowledge sharing, transaction, etc., reduce uncertainty and opportunistic behavior, thus laying the foundations of cognition and relationship for green knowledge

integration. The construction of these foundations will help platform firms to form a consensus, so as to realize socialized and collaborative green knowledge integration. The reason for the insignificant correlation between EIKG and SYGKI may be that the latter is more dependent on standardized knowledge activity rules. SYGKI may require the guidance of formal governance mechanism such as clear rules and regulations. In addition, large-scale systematic knowledge integration including coding and processing may also require the assistance of information technology such as computer and database.

- (3) The interactive effect of ITC and EIKG has a significant positive impact on socialized, collaborative, and systematic GKI. We further analyze these interactive effects. In COGKI submodel (Model 6), the moderating effects of (i) independent variables are mutual, and the interactive effect is shown in

Figure 3. Among them, the correlation coefficients of EIKG and COGKI under high- and low-level ITC are significant. As shown in Figure 3, when ITC are high, EIKG is positively related to COGKI; when ITC are low, this positively significant relationship becomes weaker. The effect of EIKG on COGKI is the same. Hence, COGKI needs not only the guidance of EIKG but also ITC to improve the efficiency of GKI. The former mainly promotes the sharing and exchange of tacit and explicit green knowledge through vision sharing, knowledge network, and trust establishment, thereby reducing the uncertainty and transaction costs in collaborative green knowledge integration; the latter could promote the efficiency and frequency of the COGKI via information system, AI, and other information technology. (ii) In SOGKI submodel (Model 3), interactive effect is only reflected in the moderating role of ITC that positively regulates the relationship between EIKG and SOGKI ($\gamma = 0.191$, $p < 0.01$). This might be because the roles of IT in knowledge activities are mainly reflected in the processing, storage, and transmission of green knowledge. And ITC only have the advantage in knowledge collaboration and exchange. (iii) In SYGKI submodel (Model 9), interactive effect is embodied as the moderating effect of EIKG that positively moderates the relationship between ITC and SYGKI ($\gamma = 0.305$, $p < 0.01$). EIKG could guide the IT integration and reset process from the aspects of external knowledge network construction and the purpose of knowledge systematization, thereby affecting the relationship between ITC and EIKG. To sum up, interactive effect of the two independent variables could be complementary (embodied in a one-way moderating effect) or enhanced (embodied as a two-way moderating effect) in platform firm's green knowledge integration.

5. Conclusion

In the context of the platform ecosystem, from the perspective of firm's social capital, this article explores the impact of platform firm's IT capabilities, external informal knowledge governance and their interaction on green knowledge integration by building a conceptual model about their relationships, and uses 372 sample data to test this model. The final results show that: (i) platform firm's IT capabilities could positively affect the collaborative and systematic green knowledge integration; (ii) platform firm's external informal knowledge governance could positively affect the socialized and collaborative green knowledge integration; (iii) the interactive effect of IT capabilities and external informal knowledge governance has a positive impact on platform firm's socialized, collaborative, and systematic green knowledge integration.

However, there are some possible limitations in the research. Further in-depth study could be conducted from the following aspects:

- (1) Distinguish the types of platform firms. Platform firms surveyed in this research are mainly from e-commerce platform ecosystems. Further study could be expanded to those platform firms from other industries.
- (2) Conduct in-depth research about the effects of the subdimensions of external informal knowledge governance and ITC on green knowledge integration. Such research could help platform firms build specific capabilities or informal governance to promote the green knowledge integration.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 62176150).

References

- [1] P. Kumar, R. Maddikunta, Q. V. Pham et al., "Industry 5.0: a survey on enabling technologies and potential applications," *Journal of Industrial Information Integration*, vol. 26, Article ID 100257, 2021.
- [2] H. R. Jhaveri, S. V. Ramani, G. Srivastava, T. R. Gadekallu, and V. Aggarwal, "Fault-resilience for bandwidth management in industrial software-defined networks," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 4, pp. 3129–3139, 2021.
- [3] M. Iansiti and R. Levien, "Keystones and dominators: framing the operational dynamics of business ecosystems," *The Operational Dynamics of Business Ecosystems*, pp. 3–19, The Pennsylvania State University, Pennsylvania, 2002.
- [4] L. Yu, Y. Duan, and K. Li, "A real-world service mashup platform based on data integration, information synthesis, and knowledge fusion," *Connection Science*, vol. 33, no. 2, pp. 463–481, 2021.
- [5] Y. Zhang, J. Li, and T. W. Tong, "Platform governance matters: how platform gatekeeping affects knowledge sharing among complementors," *Strategic Management Journal*, vol. 43, no. 3, pp. 599–626, 2022.
- [6] W. Jiang and X. Chen, "Optimal strategies for low carbon supply chain with strategic customer behavior and green technology investment," *Discrete Dynamics in Nature and Society*, vol. 2016, Article ID 9645087, 13 pages, 2016.
- [7] F. J. Contractor and W. Ra, "How knowledge attributes influence alliance governance choices," *Journal of International Management*, vol. 8, no. 1, pp. 11–27, 2002.
- [8] Y. Cao and Y. Xiang, "The impact of knowledge governance on knowledge sharing," *Management Decision*, vol. 50, no. 3–4, pp. 591–610, 2012.
- [9] S.-C. Fang, C.-W. Yang, and W.-Y. Hsu, "Inter-organizational knowledge transfer: the perspective of knowledge

- governance,” *Journal of Knowledge Management*, vol. 17, no. 6, pp. 943–957, 2013.
- [10] S. A. Zahra, D. O. Neubaum, and J. Hayton, “What do we know about knowledge integration: fusing micro- and macro-organizational perspectives,” *The Academy of Management Annals*, vol. 14, no. 1, pp. 160–194, 2020.
- [11] L. Zhang, J. Cheng, and D. Wang, “The influence of informal governance mechanisms on knowledge integration within cross-functional project teams: a social capital perspective,” *Knowledge Management Research and Practice*, vol. 13, no. 4, pp. 508–516, 2015.
- [12] Y. Long, P. Li, and B. You, “Knowledge transfer, governance mechanisms in alliance and environmental uncertainty,” *Chinese Management Studies*, vol. 8, no. 3, pp. 438–472, 2014.
- [13] M. S. Akram, M. A. Goraya, A. Malik, and A. M. Aljarallah, “Organizational performance and sustainability: exploring the roles of it capabilities and knowledge management capabilities,” *Sustainability*, vol. 10, no. 10, pp. 1–20, 2018.
- [14] M. Huysman and V. Wulf, “It to support knowledge sharing in communities, towards a social capital analysis,” *Journal of Information Technology*, vol. 21, no. 1, pp. 40–51, 2006.
- [15] G. Koenig, “Business ecosystems revisited,” *Management*, vol. 15, no. 2, pp. 209–214, 2012.
- [16] J. F. Moore, “Business ecosystems and the view from the firm,” *Antitrust Bulletin*, vol. 51, no. 1, pp. 31–75, 2006.
- [17] J. R. Wang, Y. J. Xue, X. L. Sun, and J. Yang, “Green learning orientation, green knowledge acquisition and ambidextrous green innovation,” *Journal of Cleaner Production*, vol. 250, pp. 119–128, 20 March 2020.
- [18] M. Iansiti and R. Levien, *The New Operational Dynamics of Business Ecosystems: Implications for Policy, Operations and Technology Strategy*, Division of Research, Boston, MA, USA, pp. 35–67, 2002.
- [19] A. Mehta and N. Mehta, “Knowledge integration and team effectiveness: a team goal orientation approach,” *Decision Sciences*, vol. 49, no. 3, pp. 445–486, 2018.
- [20] R. Guo, L. Cai, and Y. Fei, “Knowledge integration methods, product innovation and high-tech new venture performance in China,” *Technology Analysis & Strategic Management*, vol. 31, no. 3, pp. 306–318, 2019.
- [21] M. D. Boer, F. Bosch, and H. Volberda, “Managing organizational knowledge integration in the emerging multimedia complex,” *Journal of Management Studies*, vol. 36, no. 3, pp. 237–246, 1999.
- [22] M. D. Stoel and W. A. Muhanna, “It capabilities and firm performance: a contingency analysis of the role of industry and it capability type,” *Information & Management*, vol. 46, no. 3, pp. 181–189, 2009.
- [23] G. Yenduri and T. R. Gadekallu, “Firefly-based maintainability prediction for enhancing quality of software,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. Supp02, pp. 211–235, 2021.
- [24] A. S. Bharadwaj, “A resource-based perspective on information technology capability and firm performance: an empirical investigation,” *MIS Quarterly*, vol. 24, no. 1, pp. 169–196, 2000.
- [25] I. Prieto-Pastor, V. Martín-Pérez, and N. Martín-Cruz, “Social capital, knowledge integration and learning in project-based organizations: a ceo-based study,” *Journal of Knowledge Management*, vol. 22, no. 8, pp. 1803–1825, 2018.
- [26] M. J. Lyver and T. J. Lu, “Sustaining innovation performance in smes: exploring the roles of strategic entrepreneurship and it capabilities,” *Sustainability*, vol. 10, no. 2, pp. 1–27, 2018.
- [27] A. Rai and X. Tang, “Leveraging it capabilities and competitive process capabilities for the management of interorganizational relationship portfolios,” *Information Systems Research*, vol. 21, no. 3, pp. 516–542, 2010.
- [28] M. Tarafdar and S. R. Gordon, “Understanding the influence of information systems competencies on process innovation: a resource-based view,” *The Journal of Strategic Information Systems*, vol. 16, no. 4, pp. 353–392, 2007.
- [29] T. Cui, Y. Tong, H.-H. Teo, and J. Li, “Managing knowledge distance: it-enabled inter-firm knowledge capabilities in collaborative innovation,” *Journal of Management Information Systems*, vol. 37, no. 1, pp. 217–250, 2020.
- [30] K. Trantopoulos, G. V. Krogh, G. von Krogh, M. W. Wallin, and M. Woerter, “External knowledge and information technology: implications for process innovation performance,” *MIS Quarterly*, vol. 41, no. 1, pp. 287–300, 2017.
- [31] K. V. Lins, H. Servaes, and A. Tamayo, “Social capital, trust, and firm performance: the value of corporate social responsibility during the financial crisis,” *The Journal of Finance*, vol. 72, no. 4, pp. 1785–1824, 2017.
- [32] N. J. Foss, “The emerging knowledge governance approach: challenges and characteristics,” *Organization*, vol. 14, no. 1, pp. 29–52, 2007.
- [33] N. J. Foss and S. Michailova, “Knowledge governance: processes and perspectives: processes and perspectives,” *Knowledge governance: what have we learned? And where are we heading?*, Oxford University Press, Oxford, UK, pp. 272–288, 2009.
- [34] K. Shahzad, T. Ali, M. Kohtamäki, and J. Takala, “Enabling roles of relationship governance mechanisms in the choice of inter-firm conflict resolution strategies,” *Journal of Business & Industrial Marketing*, vol. 35, no. 6, pp. 957–969, 2020.
- [35] M. Park, M. Kim, and S. Ryu, “The relationship between network governance and unilateral governance in dynamic consumer demand,” *Industrial Marketing Management*, vol. 84, no. 1, pp. 194–201, 2020.
- [36] H. H. Chang and S.-S. Chuang, “Social capital and individual motivations on knowledge sharing: participant involvement as a moderator,” *Information & Management*, vol. 48, no. 1, pp. 9–18, 2011.
- [37] M. M. Wasko and S. Faraj, “Why should I share? Examining social capital and knowledge contribution in electronic networks of practice,” *MIS Quarterly*, vol. 29, no. 1, pp. 35–57, 2005.
- [38] J. Nahapiet and S. Ghoshal, “Social capital, intellectual capital, and the organizational advantage,” *Academy of Management Review*, vol. 23, no. 2, pp. 242–266, 1998.
- [39] M. Iansiti and R. Levien, *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability*, pp. 30–26, Harvard Business Press, Boston, USA, 2004.
- [40] M. Masucci, S. Brusoni, and C. Cennamo, “Removing bottlenecks in business ecosystems: the strategic role of outbound open innovation,” *Research Policy*, vol. 49, no. 1, pp. 1–17, 2020.
- [41] S. A. Zahra and S. Nambisan, “Entrepreneurship and strategic thinking in business ecosystems,” *Business Horizons*, vol. 55, no. 3, pp. 219–229, 2012.
- [42] A. Nerkar and S. Paruchuri, “Evolution of R&D capabilities: the role of knowledge networks within a firm,” *Management Science*, vol. 51, no. 5, pp. 771–785, 2005.
- [43] G. Herczeg, R. Akkerman, and M. Z. Hauschild, “Supply chain collaboration in industrial symbiosis networks,” *Journal of Cleaner Production*, vol. 171, pp. 1058–1067, January 2018.
- [44] P. A. Pavlou, “Institution-based trust in interorganizational exchange relationships: the role of online B2b marketplaces

- on trust formation,” *The Journal of Strategic Information Systems*, vol. 11, no. 3, pp. 215–243, 2002.
- [45] N. Lazaric and L. Edward, “Introduction: the learning dynamics of trust, reputation and confidence,” *Journal des Economistes et des Etudes Humaines*, vol. 8, no. 2, pp. 353–362, 1998.
- [46] G. R. Jones and J. M. George, “The experience and evolution of trust: implications for cooperation and teamwork,” *Academy of Management Review*, vol. 23, no. 3, pp. 531–546, 1998.
- [47] W. Rutten, J. Blaas-Franken, and H. Martin, “The impact of (low) trust on knowledge sharing,” *Journal of Knowledge Management*, vol. 20, no. 2, pp. 199–214, 2016.
- [48] J. Wei and L. Xu, “Knowledge network dual embedding, knowledge integration and cluster enterprise innovation ability,” *Journal of Management Science*, vol. 17, no. 2, pp. 34–47, 2014.
- [49] C. Croux and A. Ruiz-Gazen, “High breakdown estimators for principal components: the projection-pursuit approach revisited,” *Journal of Multivariate Analysis*, vol. 95, no. 1, pp. 206–226, 2005.
- [50] L. O. Jimenez and D. A. Landgrebe, “Hyperspectral data analysis and supervised feature reduction via projection pursuit,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653–2667, 1999.
- [51] R. M. Baron and D. A. Kenny, “The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations,” *Journal of Personality and Social Psychology*, vol. 51, no. 6, pp. 1173–1182, 1986.

Research Article

Optimization of Urban Waste Transportation Route Based on Genetic Algorithm

Yanling Zhang, Xu Luo , Xiaoxuan Han, Yongxing Lu, Jiacheng Wei, and Chunyu Yu

School of Mechanics and Construction Engineering, Jinan University, Guangzhou, China

Correspondence should be addressed to Xu Luo; tluoxu@jnu.edu.cn

Received 15 January 2022; Revised 6 February 2022; Accepted 8 February 2022; Published 11 April 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Yanling Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under the normal conditions of the new economy, people's living standards have improved; the amount of urban domestic waste has increased suddenly; and the differentiation of waste has become more and more complicated. However, the garbage transportation method at this stage has been difficult to meet the needs of reality. Therefore, it is particularly urgent to reform the original and simple treatment methods and adopt the latest and more reasonable garbage transportation. The urban waste transportation route optimization described in this article is quite complicated in the specific problems faced in the optimization of the urban domestic waste transportation route. This paper proposes to combine computer science to study the system optimization model and combine information management and other disciplines to develop and apply research on the dynamic management software of urban waste logistics systems. In the modern service industry, logistics occupies an important position. The optimization of urban waste transportation routes based on genetic algorithms also promotes the development of high-end modern service industries and is an important task to promote urban development and maintained hygiene and improve the healthcare of the citizens. The optimization of urban waste transportation routes is conducive to creating a clean and tidy urban environment.

1. Introduction

Urban domestic waste mainly includes kitchen cabinet waste, waste paper, fabrics, household utensils, glass and ceramic fragments, waste electrical appliances, waste plastic products, coal ash, and waste vehicles. With the acceleration of the pace of life and the improvement of living standards, the output of domestic garbage has increased rapidly, and the proportion of inorganic substances such as coal ash and slag in the garbage has gradually decreased. The proportion of organic matter, combustibles, and recyclables such as products in the garbage is gradually increasing. The classification and treatment of garbage at the source determines whether the process of resource utilization and reduction of garbage can be carried out effectively, and it is the most critical part of garbage classification and transportation. In recent years, the continuous improvement of domestic waste removal has reached 261.04 million tons, and the national urban waste is increasing at an annual rate of 8%~10%. The

domestic garbage that has not been effectively collected and processed not only pollutes the air and water sources but also poses a threat to the lives and health of urban residents. The use of "Internet" technology to establish a coordinated transportation mechanism for the government, private enterprises, social organizations, and the public is an important way to achieve the reduction, harmlessness, and resource disposal of domestic waste. The problem of urban domestic garbage has become an increasingly prominent problem [1, 2]. The amount of garbage generated is greater than the amount of clearing and transportation, and the amount of harmless treatment is smaller. The transport link of municipal solid waste is an important part of the waste treatment system. In the cost of waste treatment, the cost of the collection and transportation accounts for a considerable proportion. For example, the total annual waste treatment cost in the United States is about 20 billion US dollars. Collection and transportation costs have exceeded 10 billion US dollars. Therefore, it is necessary to optimize the

collection and transportation route of garbage trucks to reduce the cost of collection and transportation system and reduce environmental pollution and social impact.

At present, in the transportation of urban domestic waste, due to the lack of information disclosure mechanism and information communication mechanism, it is difficult for different transportation entities to achieve interconnected information communication and information sharing, which has led to the existence of information between various government departments, government departments and enterprises, and the public. The municipal urban management committee and the district urban management bureaus are responsible for the treatment, and the trade committees are responsible for the garbage recycling [3]. The garbage infrastructure construction and related transfer work are responsible for the relevant enterprises. Since the information exchange platform among various departments has not been established. Among different departments, it is difficult for relevant functional departments, enterprises, and the public to understand the progress of their respective work, and it is difficult to achieve collaborative operations, which makes it difficult to improve the effect of waste classification. In addition, due to the lack of information-sharing platforms, residents also lack an expression and feedback mechanism for the actual needs of urban domestic waste transportation, which makes it difficult to stimulate the public's enthusiasm for waste classification. According to the forecast of the growth for waste generated in the past 10 years, it will be 409 million tons by 2030 and 528 million tons by 2050. At present, China has accumulated nearly 7 billion tons of municipal solid waste, covering an area of more than 500 million square meters. About two-thirds of large and medium-sized cities are surrounded by garbage, and about one-fourth have developed to no suitable place for stacking. Municipalities and provincial capitals account for an important proportion of waste generation [4]. Sixty percent of the domestic waste output is concentrated in 52 key cities with a population of more than 500,000. Such a current situation makes the collection and transportation system of domestic wastes more and more important in the management of domestic waste. China's municipal solid waste has the following characteristics. (1) The generation source is scattered, and the generation amount is large. Domestic waste is mainly produced in households, so the sources of production are spread across all residential areas. In the past 20 years, the process of urbanization in our country has been accelerating year by year, and the amount of municipal solid waste produced has gradually increased. (2) The composition is complex, and the nature is unstable. Due to the diversity of residents' lives, there are many types of domestic waste generated, resulting in a complex composition of domestic waste. Especially new materials and new products emerge in an endless stream, making the composition of domestic waste more and more complicated. (3) The amount, composition, and nature of domestic waste are related to many factors, for example,

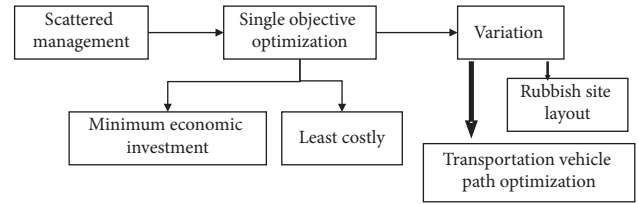


FIGURE 1: Optimized status of urban waste transportation.

residents' living standards, living habits, climate, geographic location, and so on. (4) Municipal solid waste has potential economic value. Many components in urban domestic waste are useful resources that can be recycled and reused, such as waste paper, waste batteries, waste plastics, and so on, and thus exhibit great economic value.

The development of cities and the continuous expansion of the city's scale have made the urban area covered by the waste collection and transportation system larger and larger. The collection and transportation system of municipal solid waste is composed of the three links of collection, transportation, and transfer in the disposal system. Collection and transportation are common to each collection and transportation system. It is determined according to the transportation distance from the source of garbage to the treatment site, the transportation capacity of garbage collection vehicles, and the amount of garbage. The distance between garbage treatment and disposal facilities and the source of municipal solid waste is getting farther and farther. Therefore, the overall optimization of the modern urban domestic waste collection and transportation system is very necessary. How to achieve scientific and effective collection and processing through classification and how to classify are always the focus of domestic waste transportation. Many developed countries in the world have issued corresponding laws, regulations, policy plans, and transportation measures in accordance with their national conditions, for example, Germany's "garbage economy" legislation, two-way recycling system, and other measures; Belgium's garbage classification "family compulsory course"; and so on. In addition, urban household waste has the characteristics of large quantity and wide range, complex composition, rising harmful types, and large regional differences [5], making the classification and treatment of urban domestic waste in our country face tremendous pressure. The main work of this paper is as follows: it introduces the research status of urban waste transportation routes, which needs to optimize. This paper proposes the GA optimization method and develops and applies the dynamic management software of the urban waste logistics system. System software promotes the important tasks of urban development, maintaining hygiene and improving the health of citizens. Based on the standardized transportation of urban domestic waste in some areas in the early stage, it is of great significance to conduct research on the classification and treatment of urban domestic waste from the perspective of a more systematic standardized transportation theory.

2. Related Theoretical Research

2.1. Literature Review. The optimization of transportation routes of municipal solid waste started late. It is in a weak link in the solid waste environmental management system and lacks scientific planning and research. This has caused a lot of labor loss, material resources, and financial resources as well as useful resources. The optimization route of garbage vehicles based on the basic situation of our country is drawing on the results of foreign research [6]. The optimization model from single- to multi-objective is shown in Figure 1, which is from a simple function to an uncertain multi-objective mathematical model.

Initially, due to the small amount of garbage generated, the impact on the environment was not obvious, and the management of urban domestic garbage was in a state of fragmented management, allowing it to dissipate naturally in the environment. The single objective is optimized by only one factor. Multi-objective realization and multi-factor have been optimized by multi-dimensional space realization. With the growth of the population, the development of cities, and the improvement of people's living standards, the output of domestic waste is increasing. People gradually realize that if domestic waste is not managed effectively, it will bring serious environmental and health problems [7]. The waste management model has been researched and tested, and the initial research method is a single-objective optimization method. The single-objective optimization model focuses on economic input and is simple and easy to implement. However, the consideration is one-sided and suitable for the management model of small and medium-sized cities. For large cities, environmental impact and social effects must have been considered. Therefore, for the collection and transportation of domestic waste in large cities, single-objective optimization mode performance is also very limited. Jia [8] took the cold fresh meat logistics distribution path of S enterprise as the research object, constructed a cold chain logistics path optimization decision model with the goal of minimizing the total cost, and used the improved genetic algorithm to optimize the distribution route. The optimized plan is in various distribution costs are significantly reduced. Hadipour [9] established an improved multi-objective, mixed-integer planning model for collecting vehicle routes and solved the problem of collecting vehicle routes and schedules. GIS technology is used to create, store, retrieve, analyze, and display spatial information under complex spatial and geographic relationships and has been widely used in many aspects of the environmental field. For example, use GIS technology to simulate surface water flow and surface water pollution and manage water distribution networks. Isnafitri's research [10] shows that by combining GIS technology, mathematical planning software, and associated database management systems, it is possible to analyze and compare the available garbage collection schemes under the circumstances of changes in the environment and municipal planning. In addition, a large number of scholars continue to experiment and research on this issue.

2.2. Genetic Algorithm. Genetic algorithm is generated by simulating the concept of biological survival in nature. Generally speaking, the genetic algorithm includes three operators, namely selection, crossover, and mutation. The function of the selection operator is to increase the average moderate value of the entire population. In the entire population, individuals with high evaluation values are selected to form the main group of the mating pool: The main function of the crossover operator is to select the good genes in the mating pool to be inherited to the next. In the first generation, individuals in the mating pool are paired, and then some genes are exchanged purposefully to generate individuals with more complex genetic traits. The mutation operator is to invert one or several individual binary characters according to a certain small probability, so as to realize the simulation of gene mutation phenomenon in nature. In this algorithm, every problem that needs to solve has been coded and designed as a "chromosome" as much as possible. Multiple chromosomes can then form a population. In this process, genetic operations such as selection, mutation, crossover, and duplication will occur. When the genetic algorithm is initially set, an initial value, namely a population, is first randomly generated, then the individuals in the population are processed and evaluated according to the function of the algorithm, and the corresponding value of the environmental fitness is generated [11–13]. Then the algorithm will select excellent individuals for next-generation derivation based on these fitness values, and then mutate and cross-process the selected excellent individuals. As shown in Figure 2, genetic algorithms are widely used in the path design of robots; the scope of application is not only in the travel of a single robot but also in the cooperation of multiple robots; and they have achieved good results.

The genetic algorithm is a robust query algorithm applied to the optimization of complex systems. Compared with other optimization algorithms, optimization genetic algorithm has the following characteristics:

- (1) Decisive variables are encoded, and a code is used as the object of algorithm processing
- (2) In the algorithm, the calculated fitness value is used to query other data information
- (3) The query process of the genetic algorithm starts from a population, not from an individual
- (4) The query of a genetic algorithm is a query based on probability, not a query based on a certain value

3. Optimization of Urban Waste Transportation Route

3.1. Existing Problems and Optimization Ideas. The municipal solid waste collection and transportation system is a large-scale logistics system, and its problems are very complex and involve a wide range of areas. However, it is mainly applied to the specific situation of foreign municipal solid waste management. Due to different national conditions, the current situation of municipal solid waste management in China is quite different from that of foreign

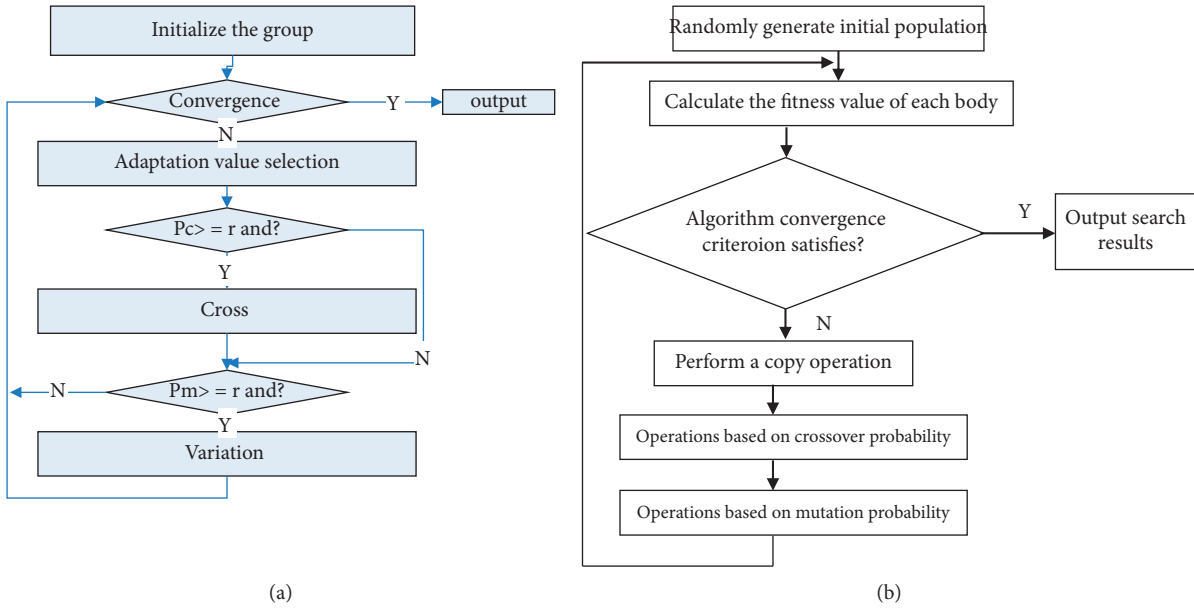


FIGURE 2: Workflow of two methods: (a) standard genetic algorithm and (b) optimization genetic algorithm.

countries. In terms of optimization methods, some optimization methods (such as fuzzy comprehensive evaluation and analytic hierarchy process) have strong subjectivity and cannot express the economics of optimization results [14]. The calculation results of some optimization methods cannot be better. The actual situation and complex calculations such as nonlinear programming, dynamic programming, and fuzzy programming restrict its application in practical problems.

In terms of optimization algorithms, heuristic methods can simultaneously meet the needs of the detailed description of the problem and solution, and more accurate optimization methods are more practical. The disadvantage is that it is difficult to know when a good heuristic solution has been obtained. Although the saving method lists between each point and construct the path from large to small according to the savings, it has the advantage of fast calculation speed, but it has the problem of messy uncombined points and difficulty in combining edge points. If the genetic algorithm wants to obtain a more satisfactory solution set, it is at the cost of prolonging the calculation time.

After referring to a large number of domestic and foreign documents, the author of this paper proposes research routes and methods for the optimization of urban domestic waste collection and transportation routes, which is shown in Figure 3.

This method integrates model research, optimization methods, evaluation, and diagnosis technologies. The theory application of pollution loss and uncertainty factors used to optimize the collection and transportation routes of municipal solid waste based on the economic quantification of environmental impacts, explore the optimal mode and method for collection and transportation of municipal solid waste, and establish a set of measures for municipal solid waste [15]. A comprehensive, multi-level comprehensive evaluation, diagnosis, and management system provides

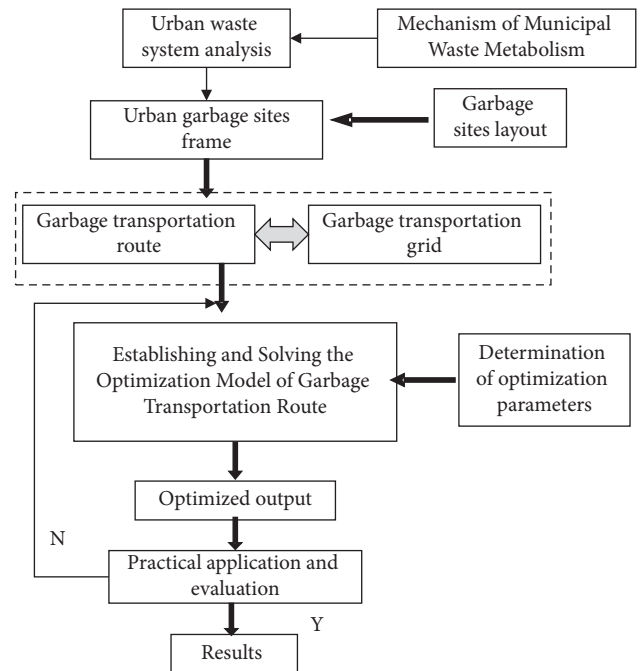


FIGURE 3: The optimization of urban domestic waste collection and transportation routes.

decision support for municipal solid waste management and promotes a benign interaction between solid waste, social, and economic development. In addition, comprehensively consider various factors in the optimization process from both vertical and horizontal aspects. On the one hand, the systematic analysis of municipal solid waste, the setting of garbage points, the selection of collection routes, and the selection of transportation routes need to study in detail to determine the best optimization plan. On the other hand, the formulation of urban domestic waste collection and

transportation routes has closely related to factors such as economy, population, consumption, and society. Any change in one of them will affect the entire optimization model. Various factors are intricately intertwined and mutually restrict each other. The related works are in Table 1.

3.2. Mathematical Model of Urban Garbage Transportation Vehicle Scheduling Problem. The scheduling problem of urban garbage transportation vehicles can be described as follows: from a logistics center with multiple garbage transportation vehicles to multiple customers, each customer's location and demand for goods are certain, and the load capacity of each garbage transportation vehicle is certain. The maximum driving distance of the distribution is fixed; it is required to arrange the vehicle distribution route reasonably.

Suppose that the logistics center has K garbage transport vehicles, a car goes to n garbage sites to pull goods, all the routes are repeated and then back to the starting point, so that the distance traveled is the shortest. The path optimization problem is expressed as a directed graph of M garbage sites $G = (m, b)$.

$$\begin{aligned} M &= \{1, 2, \dots, m\}; \\ B &= \{(i, j) | i, j \in M\}. \end{aligned} \quad (1)$$

The distance between junk sites is as follows:

$$(e_{ij})_{m \times m}. \quad (2)$$

The objective function is as follows:

$$f(uv) = \sum_{l=1}^m e_{li_{l+1}}, \quad (3)$$

where

$$uv = (i_1, i_2, \dots, i_m). \quad (4)$$

It is an arrangement of garbage sites $1, 2, \dots, m$. When $i_{m+1} = i_1$, $f(uv)$ is the garbage compression transfer facility. The construction of garbage compression facilities improves the efficiency of garbage transportation, garbage transportation generally uses large vehicles, and the starting point of garbage transportation cannot be too scattered [16]. Moreover, garbage should have compressed before transportation to reduce the volume of garbage and save transportation costs. All cities should make reasonable planning and construction of garbage compression stations as an indispensable supporting infrastructure for urban construction.

The probability that s garbage trucks are randomly placed at m garbage sites and the k -th garbage truck at site i chooses the next site j is as follows:

$$P^s(i, j) = \begin{cases} \frac{[\tau(i, j)]^\alpha \cdot [\eta(i, j)]^\beta}{\sum_{s \notin \text{tabu}_s} [\tau(i, k)]^\alpha \cdot [\eta(i, k)]^\beta}, & \text{if } j \notin \text{tabu}_s, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

TABLE 1: The related works.

| NO. | Contents |
|-----|--|
| 1 | Existing problems |
| 2 | Construction of the mathematical model |
| 3 | GA path optimization |
| 4 | Case analysis |

where

$\tau(i, j)$ indicates the pheromone concentration on edge (i, j)

$\eta(i, j) = 1/d(i, j)$ is heuristic information, where d is the distance between cities i and j

α and β reflect the relative importance of pheromone and heuristic information

tabu_s indicates the list of cities that ant K has visited

$$\begin{aligned} \tau_{ij}(t+n) &= \rho \cdot \tau_{ij}(t) + \Delta\tau_{ij}, \\ \Delta\tau_{ij} &= \sum_{k=1}^m \Delta\tau_{ij}^k. \end{aligned} \quad (6)$$

ρ is a constant less than 1, indicating the persistence of information.

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L_k}, & ij \in l_k, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where Q is a constant, l_k represents the path taken by the k -th garbage truck in this iteration, and L_k is the path length.

3.3. Path Optimization Design. The standard genetic algorithm includes group initialization, selection, crossover, and mutation operations. The main steps can be described in Figure 4. The transportation of urban garbage vehicles has neighborhood characteristics, and the candidate window is set, and the window size should be a reasonable value [17]. Garbage vehicles always prioritize the cities in the candidate window. After the search is over, the path is optimized according to the candidate window. If the node in the candidate window is switched to the vicinity of the current node and the distance is shorter, then the mutation is performed.

Explanation:

- (1) The judgment end criterion of this algorithm is to fix the number of iterations. When the algorithm reaches the number of iterations, the algorithm ends, and the current optimal solution is output.
- (2) When calculating and selecting according to the adaptation value, the recorded current optimal value is added to the new group after the mutation to ensure that the TSP solution is getting better and better (will not get worse) in the new iterative cycle.

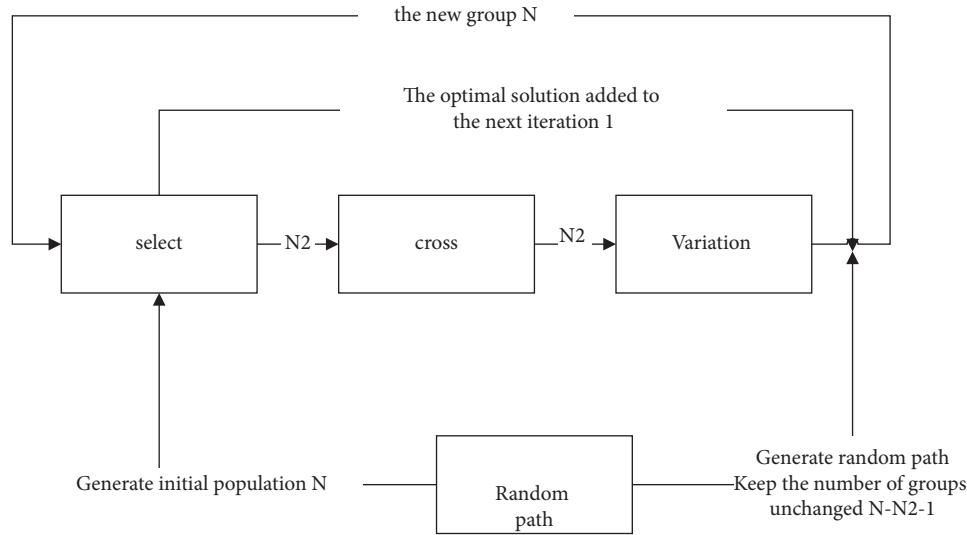


FIGURE 4: Route optimization ideas for garbage transportation vehicles.

- (3) An operation in the selection is to replace the worst K individuals with the best K individuals. In this example, the selection is based on the fit value, and the number of groups is reduced. After each mutation operation, a random path is generated to supplement the population; the number of the population remains unchanged; and the cycle is repeated. To a certain extent, prevent falling into the local optimum due to the selection problem of the initial group.

4. Case Analysis

The transportation of urban domestic garbage requires professional vehicles and so on. Individuals and small collectives do not have the conditions for garbage transportation, mainly relying on the unified completion of the urban management (environmental sanitation) department. Large-scale garbage disposal sites are generally far away from cities. The larger the city, the greater the amount of garbage, the more transportation vehicles, and the more transportation routes there will be. The transportation routes should have been planned reasonably to avoid empty running of vehicles or insufficient capacity, which may cause labor and material resources. Garbage transportation vehicles should maintain a neat appearance, and it is strictly forbidden to expose garbage, abide by traffic rules, and pay attention to safety precautions during night transportation [18]. This paper takes Guangzhou as an example to optimize the route of garbage transportation vehicles by using an ant colony algorithm.

4.1. Establish a Garbage Transportation Network. This paper uses GIS to establish a Guangzhou garbage transportation network. The process is as follows: (1) import the map of Guangzhou New Area into GIS and select the appropriate geographic coordinate system and projected coordinate system. (2) A new Shapefile file will enter and exit the main

line of Guangzhou. The route map is identified in the GIS, including national highways, provincial highways, and expressways, and the same geographic coordinate system and projected coordinate system as above have been selected to establish a garbage transportation network in Guangzhou Port. The shortest GIS path of n sites or searches for a subset of natural numbers $X = \{1, 2, \dots, n\}$ (the elements x represent the number of n cities). The subset of the sites is $\pi(X) = \{V1, V2, \dots, Vn\}$. $len = \sum d(Vi, Vi + 1) + d(V1, Vn)$ is set to take the minimum value, where $d(Vi, Vi + 1)$ means the distance from city Vi to city $Vi + 1$.

This paper uses spatial registration vectorization to operate the Guangzhou garbage transportation network, design the corresponding electronic map for follow-up research, and import it into the personal geographic information database. Establish a topological layer for the map; set topological rules, that is, no hanging points, no overlapping; and check the correctness of the connections between paths according to the rules. It is assumed that the working space of the garbage truck is a two-dimensional structured space, the location and size of the obstacles are known, and the location and size of the obstacles do not change during the movement of the garbage truck. If there are no obstacles in a certain road size range, the road is called a free road. Otherwise, it is called an obstacle road. Both free space and obstacles can be expressed as a collection of roadblocks. Number the divided roads. The working space of the garbage truck after the division is shown in Figure 5. The shaded areas in the figure are obstacles.

4.2. Road Data Collection. After modeling the garbage transportation network in Guangzhou, it is necessary to collect road-related data. The road-related data in this article is taken from a traffic analysis research report on the starting and ending points of motor vehicles on the Guangzhou highway network. The results of the study pointed out that the number of vehicles within 24 hours of a day has a certain

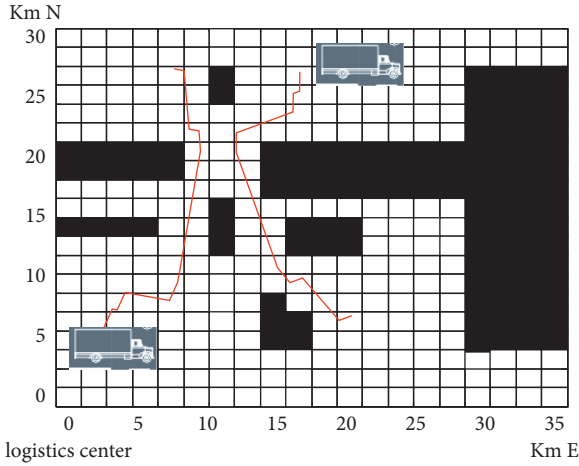


FIGURE 5: Working space of garbage truck after division.

pattern in the distribution of the city and analyzed the traffic adaptability of different road sections. It also listed the number of vehicles in each hour of the 24 hours during the survey process. Converting the traffic volume, through comprehensive analysis, the corresponding time-space matrix can be established based on the data of each road section. This article assumes that the ratio of the number of vehicles in different time periods of each road section is the same, and the ratio is only related to the time period. Based on the data, the vehicle number ratio of each period is sorted out as shown in Table 2. These data are substituted into the ant colony algorithm to optimize the transportation route.

4.3. Path Optimization. In this paper, a map of a certain area in the center of Guangzhou is collected as an example, as shown in Figure 6 for delivery. C language program is used to simulate the driving path of the supply trolley, and the optimized driving path of the supply trolley is obtained. This map contains one shipping point, ten shipping points, and fourteen intersections. These points are regarded as the vertices of the mixed graph $G=(V, E, W)$, where 1 is the shipping point. The ten points from 2 to 11 are delivery points, and the others are intersections. The path that the delivery car travels meets the requirements: each delivery point is required to have a car for delivery, and only one delivery is required. Make each driving route as short as possible that is as optimal as possible. The solutions are based on the above-mentioned ideas to solve the problem of the optimal travel path of the supply car.

4.4. Result and Analysis

4.4.1. Optimization Results. When optimizing garbage transportation routes, there must be clear goals and basic principles. The choice of the garbage transportation route plan can be considered from the following aspects: (1) the garbage transportation has the highest benefit or the garbage transportation cost is the lowest. The benefit is the main goal pursued by an enterprise, and it can be simplified to express it in terms of profit or take profit maximization as the goal.

TABLE 2: The situation of road vehicles in various periods in Guangzhou.

| Period | Vehicle ratio (%) | Period | Vehicle ratio (%) |
|--------|-------------------|-----------|-------------------|
| 0~2 | 1.92 | 12~14 | 6.34 |
| 2~4 | 2.01 | 14~16 | 6.76 |
| 4~6 | 1.84 | 16~18 | 8.23 |
| 6~8 | 7.12 | 18~20 | 7.15 |
| 8~10 | 7.62 | 20~22 | 3.7 |
| 10~12 | 6.42 | 22~24 (0) | 2.64 |

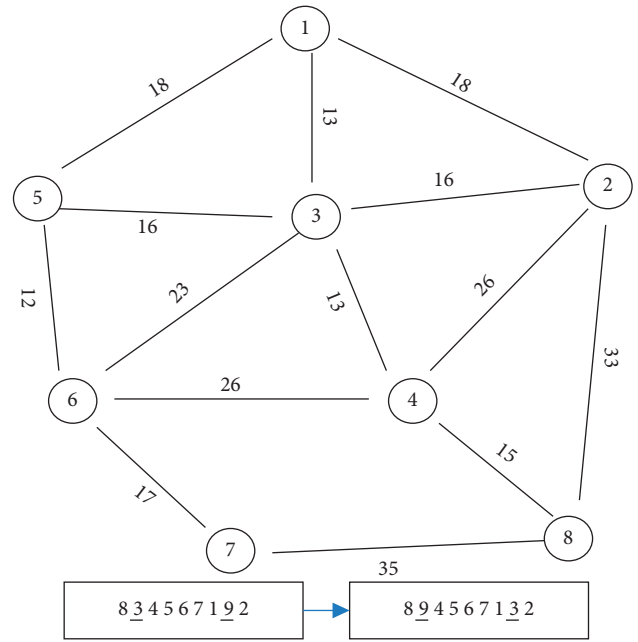


FIGURE 6: Road map of garbage transport vehicles.

The cost has a direct impact on enterprise benefits, and choosing cost minimization as the target value is directly related to the former. (2) The garbage transportation mileage is the shortest. If the garbage transportation cost has a strong correlation with the garbage transportation mileage, but the correlation with other factors is weak, the shortest garbage transportation mileage is essentially the lowest garbage transportation cost. The shortest garbage transportation mileage can be considered as the target value, which can greatly simplify route selection and vehicle scheduling methods. (3) The level of garbage transportation service is the best. If the requirement for punctual garbage transportation becomes the first priority or when the cost is needed to ensure the service level, the service level should be the first choice within the maximum tolerance of the cost. The loss of this cost may have been compensated by other aspects, such as high-quality services that can adopt a higher price strategy. (4) The consumption of garbage transportation is minimal. That is, the goal is to minimize the consumption of materialized labor and living labor. In many cases, such as labor shortages, fuel shortages, and vehicles and equipment are tight, the scope of selection of garbage transportation operations is limited, and the garbage transportation needs can have considered. Labor, vehicles, or other related resources are used as the target value. Path

simulation of garbage transportation vehicles is shown in Figure 7.

For practical problems, due to the increase of restrictions on the garbage transportation path, many related problems can be derived. The main optimization purpose of this article is the (2) mentioned above, that is, the shortest garbage transportation mileage. In this example, a partial matching crossover strategy is adopted for crossover, and the basic steps are as follows:

Step 1: Randomly select two intersections

Step 2: Exchange the gene segments between the two intersections

Step 3: Replace the parts other than the exchanged gene segment that conflict with the elements in the exchanged gene segment with the corresponding position of another parent until there is no conflict

In this example, the path instance as shown in the figure, the intersection points are 2, 7, and after the matching segment is exchanged, there are 7, 6, and 5 conflicts in A. In the matching segment of B, find the value 7-3 at the corresponding position in the matching segment of A. The point 6-0 and the point 5-4 continue to detect conflicts until there are no conflicts. Do the same for B to get the result.

4.4.2. Analysis and Discussion. The author separately compiled the traditional transportation method program for logistics distribution vehicle scheduling problem and the genetic algorithm program for transportation route optimization in C language, which addressed the problem of a distribution center by 2 vehicles to transport garbage to 8 demand points in the literature [9]. Experimental calculations are performed (in addition to the original problem, the constraint condition that the maximum driving distance of the vehicle at one time is 20 km * 30 km is added in the calculation). The following parameters are used in the experiment: the population size is 5; the evolutionary algebra is 50; the crossover probability and mutation probability of the traditional transportation method are 0.92 and 0.1, respectively; and the maximum number of gene transpositions of the genetic algorithm for transport path optimization is 4. The computer programs are run 20 times randomly, and the calculated results are shown in Figure 6.

It can be seen from Figure 8 that the 10 calculation results of the traditional transportation method and the genetic algorithm of transportation route optimization are better than the result of the saving method by 78.3 km, and the genetic algorithm of transportation route optimization obtains better results. In 10 calculations, the genetic algorithm for transport path optimization got the optimal solution of 57.5km for the problem 3 times and got the suboptimal solution of 69 km 3 times. Before optimization, the total distribution cost was 3,301.25 yuan. After multi-objective optimization, the number of garbage trucks will be reduced by 1, and the number of transportation mileage will be reduced by 0.6 km, saving distribution cost 550.37 yuan, which has a good optimization effect. The transmission efficiency has increased by 23%. However, the traditional

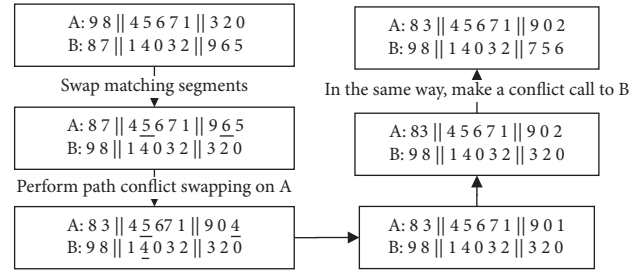


FIGURE 7: Path simulation of garbage transportation vehicles.

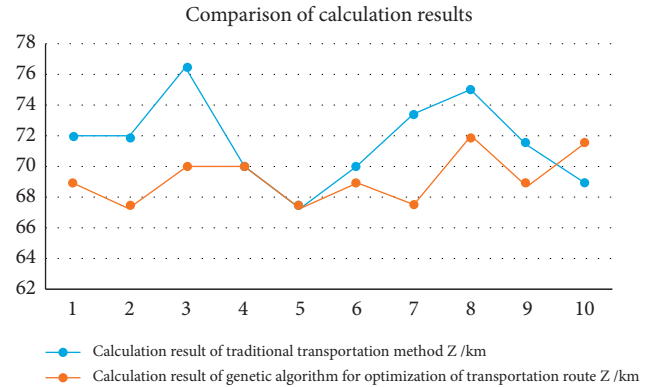


FIGURE 8: Comparison of calculation results of garbage transportation vehicle scheduling problem.

transportation method only got the optimal solution once, and the genetic algorithm for transport route optimization can overcome the “premature convergence” problem of traditional transportation methods, which avoids falling into local optima. This algorithm adds mutation operation so that the whole process can jump out of the local optimum and reach the global optimum.

In addition, the author also performed experimental calculations on an example of a certain distribution center using three vehicles to transport garbage to ten demand points in the literature [18]. The calculations show the same effect. The genetic algorithm for the optimization of transportation routes can be easily used. Find two optimal solutions to the problem. The total length of the distribution path is 80 km, and one of the solutions is the same as the calculation result of the economy method. Due to the relatively strong constraint conditions of this problem, the traditional transportation method is used to solve it, and sometimes, even a feasible solution cannot be obtained.

5. Conclusions

Unreasonable stacking of urban garbage is one of the main environmental problems facing humans today. Due to its serious harm, it has received widespread attention from all over the world. This article describes the current situation of urban garbage transportation, analyzes its impact on the environment and society and optimizes garbage transportation routes through computers so that garbage can be cleared and transported in time, and the city is clean and hygienic. While enjoying the urban civilization, people are

also suffering from the troubles caused by urban waste, of which construction waste accounts for a considerable proportion, accounting for about 30% to 40% of the total waste. Therefore, how to deal with and utilize more and more construction waste has become an important issue faced by government departments at all levels and construction waste treatment units. This paper simulates the accident rate and accident consequences of urban waste transportation; establishes a comprehensive, dynamic, multi-objective urban waste transportation route optimization model based on the transportation cost; and uses the GA model to obtain travel time of different road sections. Then, the coding method in the GA algorithm module and the method of solving the objective function in the custom problem module were designed and improved to make it more suitable in this paper. Finally, an empirical study was carried out with the main line entering and leaving Guangzhou as the research area. By collecting corresponding road data and combining it with the accident threat area obtained by modeling the urban garbage accident scenario, the model and algorithm established in this paper were used to select optimal urban garbage. The transportation route proved the feasibility of the study. This work will apply the optimization models of municipal solid waste collection and transportation routes in other cities. It can also optimize the collection and transportation routes of municipal solid waste in the future. The work of this paper also has certain limitations. It only focuses on the optimization of garbage transportation within individual cities, and cooperative transportation between cities is indeed considered, as well as the consideration of the occurrence of emergencies [19].

Data Availability

All data included in tables are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was supported by the project of the Guangzhou Science and Technology Association (K20210702019).

References




- [1] B. Zhang, X. Zhang, W. Chen, X. U. Shunlian, Y. Deng, and W. Jiang, "Sensor location optimization of large span bridge based on nested-stacking genetic algorithm," *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, vol. 40, no. 4, 2016.
- [2] O. Rızvanođlu, S. Kaya, M. Ulukavak, and M. İ. Yeşilnacar, "Optimization of municipal solid waste collection and transportation routes, through linear programming and geographic information system: a case study from Şanlıurfa, Turkey," *Environmental Monitoring and Assessment*, vol. 192, no. 1, pp. 9–1, 2020.
- [3] O. A. Lebedeva and J. O. Poltavskaya, "Cost optimization of intermodal freight transportation in the transport network," *Journal of Physics: Conference Series*, vol. 1680, no. 1, Article ID 012033, 2020.
- [4] C. Wang, Z. Ye, and W. Wang, "A multi-objective optimization and hybrid heuristic approach for urban bus route network design," *IEEE Access*, vol. 8, no. 99, pp. 12154–12167, 2020.
- [5] S. Singh, S. N. Behera, and K. Dhamodharan, *Development of GIS-based optimization method for selection of transportation routes in municipal solid waste management, Advances in Waste Management* p. 3, Springer, New York, NY, UA, 2019.
- [6] P. Nowakowski and M. Wala, "Challenges and innovations of transportation and collection of waste," *Urban Ecology*, vol. 23, pp. 457–478, 2020.
- [7] Y. Chen, X. Zheng, Z. Fang, Y. Yu, Z. Kuang, and Y. Huang, "Research on optimization of tourism route based on genetic algorithm," *Journal of Physics: Conference Series*, vol. 1575, no. 1, Article ID 012027, 2020.
- [8] P. Jia, S. Fu, Z. Li, and H. He, "Low-carbon optimization of spatial pattern in shenfu new district based on genetic algorithm," *Journal of Physics: Conference Series*, vol. 1419, no. 1, Article ID 012039, 2019.
- [9] M. Hadipour, M. Mirzaaghaee, S. Pourebrahim, M. Mokhtar, and M. Naderi, "Environmental optimization of urban transportation network, using gis and genetic algorithm," *Arabian Journal of Geosciences*, vol. 13, no. 5, 2020.
- [10] M. F. Isnafitri, C. N. Rosyidi, and A. Aisyati, "A truck allocation optimization model in open pit mining to minimize investment and transportation costs," *IOP Conference Series: Materials Science and Engineering*, vol. 1096, no. 1, Article ID 012024, 2021.
- [11] J. Li and L. Li, "Study on optimization of coal logistics network based on hybrid genetic algorithm," *International Journal of Innovative Computing Information and Control*, vol. 15, no. 6, pp. 2321–2339, 2019.
- [12] S. Agrawal, S. Sarkar, M. Alazab, P. K. R. Maddikunta, T. R. Gadekallu, and Q. V. Pham, "Genetic CFL: Hyperparameter Optimization in Clustered Federated Learning," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 7156420, 10 pages, 2021.
- [13] G. T. Reddy, M. P. K. Reddy, K. Lakshmana, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evolutionary Intelligence*, vol. 13, no. 2, pp. 185–196, 2020.
- [14] Z. Avdagic, A. Smajevic, S. Omanovic, and I. Besic, "Path route layout design optimization using genetic algorithm: based on control mechanisms for on-line crossover intersection positions and bit targeted mutation," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 2, pp. 835–847, 2021.
- [15] L. Zhu, H. Li, S. Chen et al., "Optimization analysis of a segmented thermoelectric generator based on genetic algorithm," *Renewable Energy*, vol. 156, pp. 710–718, 2020.
- [16] A. Amrane, F. Debbat, and K. Yahyaoui, "Gpu-based hybrid cellular genetic algorithm for job-shop scheduling problem," *International Journal of Applied Metaheuristic Computing*, vol. 12, no. 2, pp. 1–15, 2021.
- [17] B. Tayibia and Z. Sherin, "Genetic algorithm based optimization in peer to peer cloud networks," *International Journal of Sensors, Wireless Communications & Control*, vol. 7, no. 3, pp. 226–231, 2018.
- [18] P. Ngae, H. Kouichi, P. Kumar, A. A. Feiz, and A. Chpoun, "Optimization of an urban monitoring network for emergency response applications: an approach for characterizing the source of hazardous releases," *Quarterly Journal of the*

Royal Meteorological Society, vol. 145, no. 720, pp. 967–981, 2019.

- [19] P. Agrawal and T. Ganesh, “Solution of stochastic transportation problem involving multi-choice random parameter using Newton’s divided difference interpolation,” *Journal of Information and Optimization Sciences*, vol. 42, no. 1, pp. 77–91, 2021.

Research Article

Risk Analysis of Distal Metastasis in Chondrosarcoma and the Development and Validation of a Novel Clinical Prediction Model: A Clinical Study Based on the SEER Database

Wenle Li,^{1,2} Rong Li,³ Wanying Li,² Chan Xu,² Minmin Ma,² Haiwen Peng,⁴ Bing Wang ,² Qiang Liu ,¹ and Chengliang Yin ⁵

¹Department of Orthopedics, Xianyang Central Hospital, Xianyang 712000, China

²Clinical Medical Research Center, Xianyang Central Hospital, Xianyang 712000, China

³Shaanxi University of Traditional Chinese Medicine, Xianyang 712046, China

⁴Orthopaedic Department, the Fourth Medical Center of PLA General Hospital, Beijing 100853, China

⁵Faculty of Medicine, Macau University of Science and Technology, Macau 999078, China

Correspondence should be addressed to Bing Wang; nsadsice@163.com, Qiang Liu; m13992079668@163.com, and Chengliang Yin; yinchengliang@301hospital.com.cn

Received 11 January 2022; Revised 14 February 2022; Accepted 21 February 2022; Published 7 April 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Wenle Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. Distal metastasis in chondrosarcoma is associated with a poor prognosis. The aim of this study was to develop and validate columnar maps to predict the risk of distal metastasis in patients with chondrosarcoma, thereby contributing to clinical diagnosis and treatment. **Methods.** Data from chondrosarcoma patients obtained from the Surveillance, Epidemiology, and End Results (SEER) database from 2010 to 2016 were then screened by univariate and multifactorial logistic regressions to construct a predictive distal metastasis risk. The model discrimination of nomogram was assessed by calibration plots, while the predictive accuracy and clinical values were measured by decision curve analysis (DCA). In addition, predictive column line plots were validated in an in-house test set. **Results.** A total of 1,290 patients were included and randomized in a 7 to 3 ratio into a training group ($n=906$) and a test group ($n=384$). After logistic regression analysis, the significant variables were gender, tumor pathological grade, laterality, primary tumor stage, regional lymph node metastasis, surgical treatment, and chemotherapy. Calibration curves showed agreement between column line graph predictions and actual observations, while DCA showed the clinical utility of the nomogram. In addition, ROC showed good discrimination and calibration in the training (AUC = 0.937, 95% CI 0.919–0.952) and validation groups (AUC = 0.91, 95% CI 0.877–0.937). **Conclusions.** The nomogram for distal metastasis risk in patients with chondrosarcoma can effectively predict the individualized risk of distal metastasis and provide clinicians with enlightening information to optimize treatment options.

1. Introduction

Chondrosarcomas (CSs) are a group of heterogeneous bone malignancy with diverse histopathological and clinical features characterized by the production of a cartilaginous stroma [1], which are the second most frequent type of bone malignancy after osteosarcoma, accounting for approximately 20% of all types of bone malignancy. They are commonly found in male adults within flat bones; the pelvis and femur are two common sites of involvement, although

any bone may be affected. Most chondrosarcomas exhibit indolent, with approximately 90% below intermediate grade, and stably behave and rarely metastasize [1]. Approximately 8% of the patients with chondrosarcoma developed distant metastasis [2]. Recently, there are no common therapies such as radiotherapy or chemotherapy regimens, and targeted drugs are still under basic research and clinical trials [3]. Fortunately, CS is insensitive to chemotherapy and radiotherapy, after complete surgical resection at present. The prognosis of combined pre- and postsurgical

radiotherapy and chemotherapy is significantly better than that of surgical treatment alone [3]. Some studies have shown that distant metastases occur in 8–38% of patients with chondrosarcoma, which can greatly affect the execution of surgery and make complete resection of the tumor extremely difficult [4]. Pulmonary metastases with local recurrence are the most common reason for death in CS. Prospective research treatment of chondrosarcoma is, therefore, necessary [5], and clinicians treating patients with chondrosarcoma must determine the likelihood of metastasis [6], and it is necessary to identify risk factors for distal metastases, requiring an expansion of treatment options and approaches to improve clinical outcomes. For the current studies, predicting survival in individual patients remains difficult [7].

Since chondrosarcomas with lower incidence are relatively rare, studies assessing prognostic factors are difficult to carry out while requiring a large sample size. The National Cancer Institute's SEER program is a comprehensive source of population-based cancer incidence and survival information from the United States, collecting 18 population-based cancer registries representing approximately 27.8% of the total U.S. population [8]. This study is based on the SEER database, which does not require patient authorization, to investigate risk factors for distal metastasis of chondrosarcoma at the time of initial diagnosis, considering that this database has been providing site-specific data on metastatic tumors since 2010, by collecting information on the demographic and clinical characteristics of chondrosarcoma.

Nomogram is commonly used to generate the likelihood of clinical events through complex computational formulas [4, 9]. With nomograms, clinicians can assess the risk of clinical events, personalize individual treatment alternatives, and optimize treatment strategy. Considering the important role of distal metastasis in the chondrosarcoma prognosis, this research study aimed to assess patients at high risk of distal metastasis from chondrosarcoma through the use of a nomogram.

2. Methods

2.1. Data Sources and Inclusion Criteria. This study was based on the clinical data of patients with chondrosarcoma from the SEER database. Patients with a pathological diagnosis of chondrosarcoma in the SEER database were retrieved through the SEER * Stat software, along with the third edition of the International Classification of Oncology (ICDO-3), morphology code (9220) used to identify chondrosarcoma. The data in this study included patients diagnosed with chondrosarcoma from 2010 to 2016. Exclusion criteria were as follows: (1) patients with no positive pathology; (2) patients with unknown survival; (3) tumors that were not the first occurrence; (4) more than one primary tumor; (5) distal metastasis information unknown; and (6) regional lymphatic fluid information was incomplete for lymph node metastases.

Data were extracted from the SEER database including age, sex, race, primary site, survival duration, laterality,

tumor pathology grade, primary tumor stage, surgical treatment, radiotherapy, chemotherapy, and lymph node metastasis. These data were derived from the variable "CS site-specific factor 6." In addition, less than 20 tumor sites were categorized as "other."

2.2. Nomogram Construction, Validation, and Clinical Application. Patients with chondrosarcoma who met the inclusion criteria were randomly divided into training and validation groups in a ratio of 7 to 3. Subsequently, the following variables were selected for the study: age, race, gender, laterality, survival time, tumor pathological grade, primary tumor metastasis, primary site, surgical treatment, radiotherapy, chemotherapy, and lymph node metastasis. Univariate and multivariate binary logistic regressions were applied to identify independent risk factors with a forward stepwise regression method. Nomogram was constructed based on the results of logistic regression analysis. Calibration plot chart of clinical prediction model (calibration plot) and ROC curve were plotted, and ROC was used to estimate the predictive performance of column line plot. The larger area under the ROC curve (AUC) means the better the discriminatory ability or prognostic accuracy of the variable. In addition, decision curve analysis (DCA) plots the net benefit (NB) under a range of reasonable risk thresholds that were consistent with clinical practice and were used to assess the clinical utility of column line plots in decision-making.

2.3. Statistical Methods and Software. Continuous variables were expressed as mean \pm standard deviation (SD), and categorical variables were expressed as proportions. Continuous and categorical variables were compared by *t*-test and chi-square test of SPSS, respectively. IBM SPSS Statistics version 26.0 (SPSS Inc., Chicago, IL, USA) and R software version 4.0.5 (<http://www.r-project.org>) performed the above statistical methods, and several R packages (including *regplot*, *rms*, *rmda*, and *pROC*) were applied to plot graphs, such as nomogram, calibration plot, DCA plot, and ROC curve. KM curves were plotted by GraphPad Prim 8.0. All *P* values were bivariate, values of $P < 0.05$ were considered statistically significant, and confidence intervals (CIs) were expressed at the 95% confidence level.

3. Results

3.1. Results of Single-Factor and Multifactor Logistic Regression. A total of 1,290 patients were included in the statistics, and by the univariate and multifactorial logistic regressions, the extracted variables were first subjected to univariate logistic regression analysis, which showed that age, survival time, tumor pathology grading classification, gender, laterality, primary tumor stage, surgical treatment, and chemotherapy were prognostic factors affecting distal metastasis ($P < 0.05$). Further multifactorial logistic risk was performed, resulting in seven factors as independent prognostic factors for distal metastasis (Table 1), such as gender (female: dominance ratio (OR) 0.368, 95% CI (0.150–0.901), $P < 0.05$), tumor pathological grade (GIII:

TABLE 1: Univariate and multifactorial logistic regression analysis of risk factors for metastases in patients with chondrosarcoma.

| Variables | Univariate OR (95% CI) | P value | Multivariate OR (95% CI) | P value |
|--|------------------------|---------|--------------------------|---------|
| Age (years) | 1.018 (1.002–1.035) | <0.05 | 1.005 (0.983–1.027) | 0.656 |
| Survival time (month) | 0.959 (0.944–0.975) | <0.001 | 0.985 (0.965–1.005) | 0.146 |
| Race | | | | |
| White | Ref | Ref | Ref | Ref |
| Black | 1.337 (0.512–3.491) | 0.553 | — | — |
| Other | 0.583 (0.138–2.467) | 0.464 | — | — |
| Sex | | | | |
| Male | Ref | Ref | Ref | Ref |
| Female | 0.275 (1.36–0.554) | <0.001 | 0.368 (0.150–0.901) | <0.05 |
| Primary site | | | | |
| Limb bones | Ref | Ref | Ref | Ref |
| Axis of a bone | 1.569 (0.858–2.872) | 0.144 | — | — |
| Other | 1.794 (0.580–5.554) | 0.311 | — | — |
| Grade | | | | |
| Well differentiated; grade I | Ref | Ref | Ref | Ref |
| Moderately differentiated; grade II | 4.897 (1.413–16.972) | <0.05 | 3.047 (0.689–13.474) | 0.142 |
| Poorly differentiated; grade III | 15.267 (4.253–54.803) | <0.001 | 6.921 (1.410–33.983) | <0.05 |
| Undifferentiated; anaplastic; grade IV | 21.957 (4.935–97.697) | <0.001 | 3.006 (0.439–20.588) | 0.262 |
| Unknown | 14.559 (4.162–50.923) | <0.001 | 6.216 (1.385–27.891) | <0.05 |
| Laterality | | | | |
| Left | Ref | Ref | Ref | Ref |
| Right | 3.186 (1.533–6.622) | <0.01 | 2.674 (1.029–6.945) | <0.05 |
| Other | 2.384 (1.026–5.542) | <0.05 | 1.890 (0.584–6.115) | 0.288 |
| T | | | | |
| T1 | Ref | Ref | Ref | Ref |
| T2 | 8.456 (3.652–19.578) | <0.001 | 4.698 (1.732–12.738) | <0.01 |
| T3 | 42.515 (8.463–213.564) | <0.001 | 59.117 (8.152–428.709) | <0.001 |
| TX | 9.101 (3.588–23.082) | <0.001 | 4.074 (1.221–13.594) | <0.05 |
| N | | | | |
| N0 | Ref | Ref | Ref | Ref |
| N1 | 25.901 (6.699–100.184) | <0.001 | 9.168 (1.572–53.669) | <0.05 |
| NX | 8.290 (3.441) | <0.001 | 6.743 (1.874–24.269) | <0.01 |
| Surgery | | | | |
| No | Ref | Ref | Ref | Ref |
| Yes | 0.102 (0.057–0.183) | <0.001 | 0.237 (0.101–0.555) | <0.01 |
| Radiation | | | | |
| No | Ref | Ref | Ref | Ref |
| Yes | 0.841 (0.327–2.165) | 0.720 | — | — |
| Chemotherapy | | | | |
| No | Ref | Ref | Ref | Ref |
| Yes | 23.505 (11.957–46.205) | <0.001 | 19.188 (7.554–48.740) | <0.001 |

OR = 6.921, 1.410–33.983, $P < 0.05$; unclear tumor grading: OR = 6.216, 1.385–27.891, $P < 0.05$), laterality (right: OR = 2.674, 1.029–6.945, $P < 0.05$), primary tumor stage (T2: OR = 4.698, 1.732–12.738, $P < 0.01$; T3: OR = 59.117, 8.152–428.709, $P < 0.001$; TX: OR = 4.074, 1.221–13.594, $P < 0.05$), regional lymph node metastasis (positive: OR = 9.168, 1.572–53.669, $P < 0.05$; unknown: 6.743, 1.874–24.269, $P < 0.01$), surgical treatment (OR = 0.237, 0.101–0.555, $P < 0.01$), and chemotherapy (OR = 19.188, 7.554–48.740, $P < 0.001$). The specific results are shown in Table 1 as single- and multifactor logistic regression analysis.

3.2. Baseline Characteristics of Patients with Nonmetastatic and Metastatic Chondrosarcoma. The 1,290 patients were grouped by M0 and M1 and are then summarized in Table 2. Age, tumor pathological grade, laterality, primary tumor stage, regional lymph node metastasis, surgical treatment,

chemotherapy, and survival time were shown to be prognostic factors affecting distal metastasis ($P < 0.05$).

3.3. Population Baseline Characteristics. The 1,290 patients were randomized into a training group ($n = 906$, 7 to 3 ratio) and a validation group ($n = 384$), which are then summarized in Table 3. There were no statistically significant differences between the training and validation groups ($P > 0.05$).

3.4. Construction and Validation of Nomogram for Distal Metastasis of Chondrosarcoma. The results of univariate and multivariate logistic regressions were used to construct the nomogram of distal metastasis (Figure 1(a)). The primary tumor stage was the best predictor, followed by chemotherapy, tumor pathological grade, lymph node metastasis,

TABLE 2: Baseline of chondrosarcoma patients without and with metastases.

| Variable | Level | Overall (N=1290) | M0 (N=1215) | M1 (N=75) | P |
|-----------------------------|--|------------------|---------------|---------------|--------|
| Race (%) | Black | 96 (7.4) | 90 (7.4) | 6 (8.0) | 0.752 |
| | Other | 77 (6.0) | 74 (6.1) | 3 (4.0) | |
| | White | 1117 (86.6) | 1051 (86.5) | 66 (88.0) | |
| Age (mean (SD)) | NA | 53.44 (18.12) | 52.96 (18.06) | 61.16 (17.43) | <0.001 |
| Sex (%) | Female | 571 (44.3) | 550 (45.3) | 21 (28.0) | 0.005 |
| | Male | 719 (55.7) | 665 (54.7) | 54 (72.0) | |
| Primary site (%) | Axis bone | 677 (52.5) | 629 (51.8) | 48 (64.0) | 0.068 |
| | Bone of limb | 544 (42.2) | 522 (43.0) | 22 (29.3) | |
| | Other | 69 (5.3) | 64 (5.3) | 5 (6.7) | |
| Laterality (%) | Left | 496 (38.4) | 482 (39.7) | 14 (18.7) | 0.001 |
| | Not a paired site | 293 (22.7) | 268 (22.1) | 25 (33.3) | |
| | Right | 501 (38.8) | 465 (38.3) | 36 (48.0) | |
| Grade (%) | Moderately differentiated; grade II | 518 (40.2) | 492 (40.5) | 26 (34.7) | <0.001 |
| | Poorly differentiated; grade III | 129 (10.0) | 115 (9.5) | 14 (18.7) | |
| | Undifferentiated; anaplastic; grade IV | 39 (3.0) | 32 (2.6) | 7 (9.3) | |
| | Unknown | 169 (13.1) | 146 (12.0) | 23 (30.7) | |
| | Well differentiated; grade I | 435 (33.7) | 430 (35.4) | 5 (6.7) | |
| T (%) | T1 | 716 (55.5) | 704 (57.9) | 12 (16.0) | <0.001 |
| | T2 | 389 (30.2) | 351 (28.9) | 38 (50.7) | |
| | T3 | 13 (1.0) | 9 (0.7) | 4 (5.3) | |
| | TX | 172 (13.3) | 151 (12.4) | 21 (28.0) | |
| | N0 | 1237 (95.9) | 1178 (97.0) | 59 (78.7) | |
| N (%) | N1 | 11 (0.9) | 5 (0.4) | 6 (8.0) | <0.001 |
| | NX | 42 (3.3) | 32 (2.6) | 10 (13.3) | |
| | No | 177 (13.7) | 133 (10.9) | 44 (58.7) | |
| Surgery (%) | Yes | 1113 (86.3) | 1082 (89.1) | 31 (41.3) | <0.001 |
| | No | 1213 (94.0) | 1142 (94.0) | 71 (94.7) | |
| Lymph node dissection (%) | Yes | 77 (6.0) | 73 (6.0) | 4 (5.3) | 1 |
| | No | 1149 (89.1) | 1081 (89.0) | 68 (90.7) | |
| Radiation (%) | Yes | 141 (10.9) | 134 (11.0) | 7 (9.3) | 0.79 |
| | No/unknown | 1231 (95.4) | 1181 (97.2) | 50 (66.7) | |
| Chemotherapy (%) | Yes | 59 (4.6) | 34 (2.8) | 25 (33.3) | <0.001 |
| | NA | 34.19 (24.16) | 35.38 (24.00) | 14.99 (18.03) | |
| Survival months (mean (SD)) | NA | 34.19 (24.16) | 35.38 (24.00) | 14.99 (18.03) | <0.001 |

laterality, surgical treatment, and gender. The calibration plots of the nomogram (Figures 1(b) and 1(c)) showed that the apparent curves, in both training and validation groups, showed good consistency. The AUC values of nomogram were 0.937 (95% CI 0.919–0.952) and 0.91 (95% CI 0.877–0.937) in the training and validation groups, respectively (Figures 2(a) and 2(b)). According to the ROC curves in the training set, the values of nomogram were more important than other variables, including tumor pathological grade (AUC = 0.733, 95% CI 0.703 to 0.762), laterality (0.591, 0.558 to 0.623), primary tumor stage (AUC = 0.608, 95% CI 0.575 to 0.640), gender (AUC = 0.635, 95% CI 0.603 to 0.666), and surgical treatment (0.718, 95% CI 0.688 to 0.748). Similarly, nomogram values were higher in the test set than in the univariate, as shown in Table 4.

3.5. Clinical Usefulness of the Distal Metastasis Nomogram. Kaplan-Meier survival curves for overall survival (OS) were plotted for the 1,290 patients enrolled in the study (Figure 3(a)), and there was a significant difference between the Kaplan-Meier survival curves for the two groups ($P < 0.001$), suggesting that patients with chondrosarcoma who are expected to develop distal metastases would have a significant survival disadvantage. Subsequently, the DCA

curve shown (Figure 3(b)) with a threshold of approximately 0–0.7 was the maximum gain for distal metastasis.

4. Discussion

This study is a study to analyze the risk of distal metastasis in chondrosarcoma based on the SEER database. According to the study, approximately 8% of patients with chondrosarcoma develop distal metastases [10]. Since patients with chondrosarcoma combined with distal metastases have a poor prognosis, it is necessary to identify relevant factors to identify patients with chondrosarcoma at high risk for distal metastases [11]. The results of this study showed that patients with tumor pathological grade 3, i.e., hypofractionated tumor and tumor with an unclear grade, laterality to the right, higher stage of primary tumor, lymph node metastasis, male, nonsurgical treatment, and chemotherapy alone were at higher risk of distal metastasis.

This study found that tumors with grade 3 pathology versus unclear grade had a higher risk of distal metastasis. In both univariate and multifactorial logistic regression analyses, tumor pathology grade 3 and unclear grade were associated with the risk of distal metastasis, with OR values of 6.921 and 6.216, respectively, representing a 6.921- and 6.216-fold higher risks of distal metastasis in

TABLE 3: Baseline data table of the training group and the validation group.

| Variable | Level | Overall (N=1290) | Training group (N=906) | Validation group (N=384) | P |
|--------------------------------|--|-------------------------|-------------------------|--------------------------|--------|
| Race (%) | Black | 96 (7.44) | 65 (7.17) | 31 (8.07) | 0.6624 |
| | Other | 77 (5.97) | 57 (6.29) | 20 (5.21) | |
| | White | 1117 (86.59) | 784 (86.53) | 333 (86.72) | |
| Age (median (IQR)) | NA | 54.000 [41.000, 67.000] | 54.000 [41.000, 67.000] | 55.000 [41.000, 67.000] | 0.6463 |
| Sex (%) | Female | 571 (44.26) | 401 (44.26) | 170 (44.27) | 1 |
| | Male | 719 (55.74) | 505 (55.74) | 214 (55.73) | |
| Primary site (%) | Axis bone | 677 (52.48) | 471 (51.99) | 206 (53.65) | 0.6012 |
| | Bone of limb | 544 (42.17) | 383 (42.27) | 161 (41.93) | |
| | Other | 69 (5.35) | 52 (5.74) | 17 (4.43) | |
| Laterality (%) | Left | 496 (38.45) | 353 (38.96) | 143 (37.24) | 0.6818 |
| | Not a paired site | 293 (22.71) | 200 (22.08) | 93 (24.22) | |
| | Right | 501 (38.84) | 353 (38.96) | 148 (38.54) | |
| Grade (%) | Moderately differentiated; grade II | 518 (40.16) | 346 (38.19) | 172 (44.79) | 0.1098 |
| | Poorly differentiated; grade III | 129 (10.00) | 99 (10.93) | 30 (7.81) | |
| | Undifferentiated; anaplastic; grade IV | 39 (3.02) | 28 (3.09) | 11 (2.86) | |
| | Unknown | 169 (13.10) | 127 (14.02) | 42 (10.94) | |
| | Well differentiated; grade I | 435 (33.72) | 306 (33.77) | 129 (33.59) | |
| T (%) | T1 | 716 (55.50) | 503 (55.52) | 213 (55.47) | 0.8912 |
| | T2 | 389 (30.16) | 272 (30.02) | 117 (30.47) | |
| | T3 | 13 (1.01) | 8 (0.88) | 5 (1.30) | |
| | TX | 172 (13.33) | 123 (13.58) | 49 (12.76) | |
| N (%) | N0 | 1237 (95.89) | 869 (95.92) | 368 (95.83) | 0.6182 |
| | N1 | 11 (0.85) | 9 (0.99) | 2 (0.52) | |
| | NX | 42 (3.26) | 28 (3.09) | 14 (3.65) | |
| M (%) | M0 | 1215 (94.19) | 853 (94.15) | 362 (94.27) | 1 |
| | M1 | 75 (5.81) | 53 (5.85) | 22 (5.73) | |
| Surgery (%) | No | 177 (13.72) | 124 (13.69) | 53 (13.80) | 1 |
| | Yes | 1113 (86.28) | 782 (86.31) | 331 (86.20) | |
| Lymph node dissection (%) | No | 1213 (94.03) | 854 (94.26) | 359 (93.49) | 0.6848 |
| | Yes | 77 (5.97) | 52 (5.74) | 25 (6.51) | |
| Radiation (%) | No | 1149 (89.07) | 807 (89.07) | 342 (89.06) | 1 |
| | Yes | 141 (10.93) | 99 (10.93) | 42 (10.94) | |
| Chemotherapy (%) | No/unknown | 1231 (95.43) | 859 (94.81) | 372 (96.88) | 0.14 |
| | Yes | 59 (4.57) | 47 (5.19) | 12 (3.12) | |
| Survival months (median (IQR)) | NA | 31.000 [13.000, 53.000] | 31.000 [13.000, 54.000] | 30.500 [13.000, 51.000] | 0.2844 |

hypofractionated tumors and tumors with unclear grade than in highly differentiated tumors. It has also been shown that the main prognostic factor for patients with chondrosarcoma is the grade of the tumor, with increased pathological grade suggesting a poor prognosis [12]. Approximately 85% of these tumors are low grade, and overall survival is favorable [13], due to the fact that low-grade chondrosarcomas have abundant cartilage stroma, low cell density, are easily locally confined, and have a good prognosis after surgical resection, whereas high-grade tumors have little cartilage stroma, high cell density, and are prone to metastasis, leading to a poor prognosis [1]. Therefore, high-grade tumors have been considered as an independent risk factor for metastasis and death [14], which is also consistent with the findings of this study. Regarding laterality at diagnosis, this study concluded that the more right-sided tendency at diagnosis, the greater the risk of developing metastasis. The results of logistics analysis in

Table 1 show that with right-sided tendency, the risk of developing metastasis is 2.674 times greater than with left-sided tendency. Chondrosarcoma is often a lateral growth, and there are fewer related studies. This study found that right-sided laterality has a higher risk of developing distal metastasis, and the reasons for this need further study. It was found that the higher the primary tumor stage, the higher the risk of distal metastasis. In the univariate and multifactorial logistic regression analyses, T2, T3, and TX were associated with the risk of distal metastasis, with OR values of 4.698, 59.117, and 4.074, respectively, representing that the risk of developing distal metastasis will be higher when the primary tumor is limited to the bone cortex, exceeds the bone cortex, or cannot be determined 4.698, 59.117, and 4.074 times, with the highest risk of distal metastasis at T3 stage. Possible reasons for this phenomenon are that as the tumor volume increases, the depth of infiltration and the extent of collateral tissue involvement increase, and the tumor becomes more

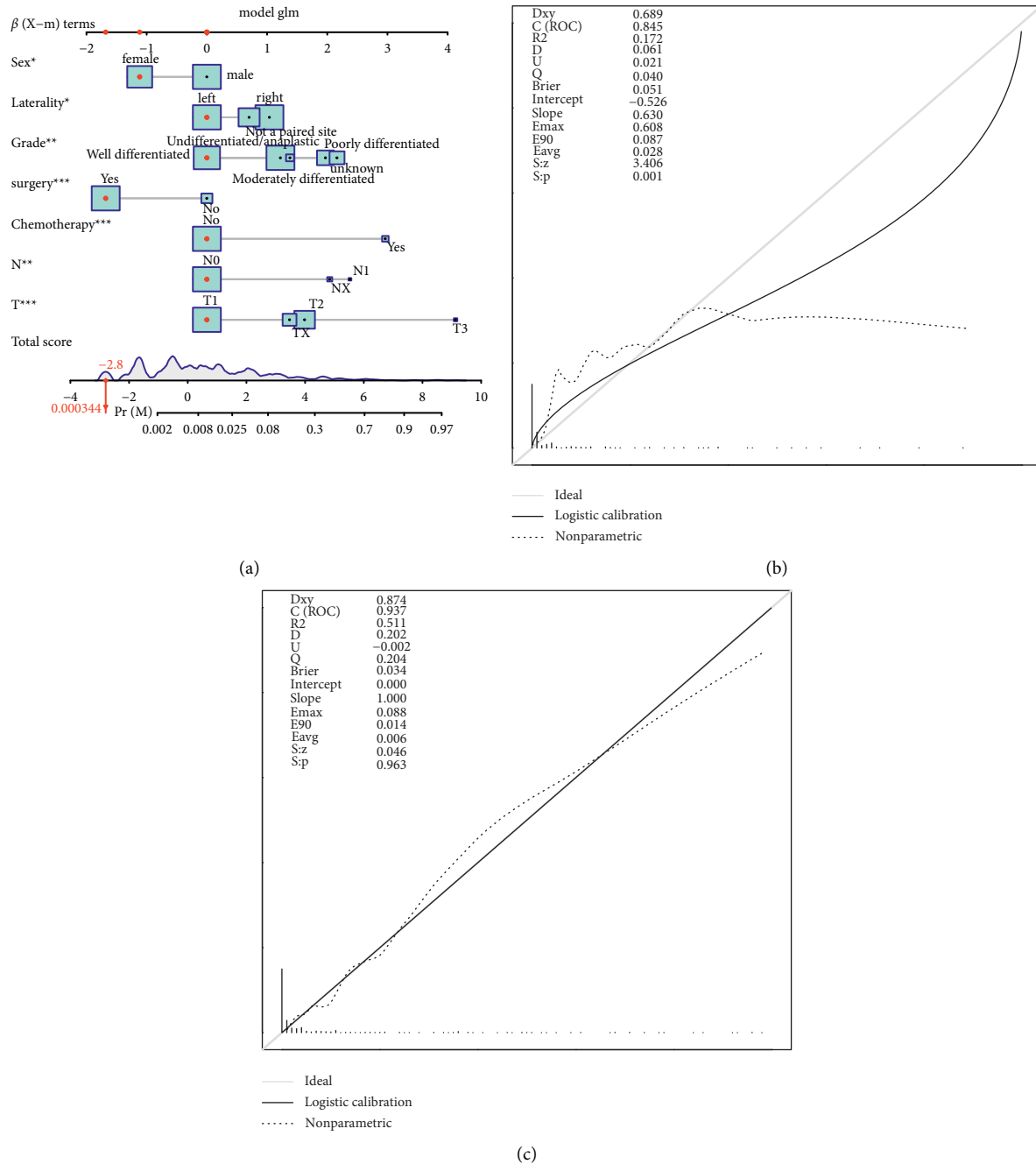


FIGURE 1: Line plots and calibration curves to predict the risk of distal metastasis in patients with chondrosarcoma. Seven features are included in the nomogram (a) and illustrated by mapping their values to covariate scales for patients. The calibration plot that predicts the training group (b) and the test group (c) is shown on the right.

aggressive and also predicts a higher degree of malignancy, thus increasing the likelihood of metastasis [15]. It has been suggested that larger chondrosarcomas may be a predictor of poor survival expectations [16], which is also consistent with the findings of this study.

It is also noteworthy that this study found studies showing that inability to undergo surgical resection and chemotherapy alone has a higher risk of metastasis, with an OR of 19.188 in univariate and multifactorial logistic regression analyses, representing an inability to resect and a

19.188-fold higher risk of distal metastasis after chemotherapy alone. A related study found chemotherapy to be an important risk factor, and the results indicated that patients receiving chemotherapy were more likely to have higher tumor grade, larger tumor size, and greater tumor extent [17], consistent with the findings of this study. This may be related to the ineffective delivery of chemotherapeutic agents, with some studies showing that tumors poorly respond to chemotherapy, and no significant improvement in disease-free survival or overall survival was seen compared

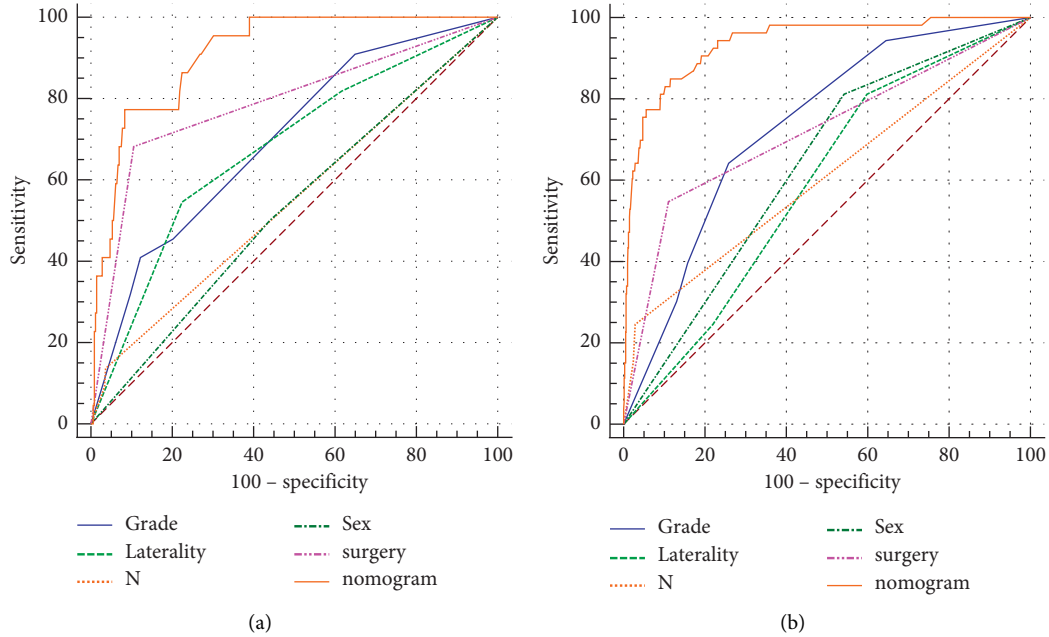


FIGURE 2: ROC curve analysis of nomogram for indicating the discriminative ability of nomogram. In the nomogram of training (a) and testing (b) groups, the AUC was 0.937 (95% CI 0.919–0.952) and 0.91 (95% CI 0.877–0.937), respectively, which proved that the nomogram had a good predictive ability.

TABLE 4: AUC of the training group and validation group.

| Variable | Training group | | | Validation group | | |
|------------|----------------|--------|----------------|------------------|--------|----------------|
| | AUC | SE | 95% CI | AUC | SE | 95% CI |
| Grade | 0.733 | 0.03 | 0.703 to 0.762 | 0.699 | 0.0534 | 0.650 to 0.745 |
| Laterality | 0.591 | 0.0329 | 0.558 to 0.623 | 0.678 | 0.0589 | 0.628 to 0.724 |
| N | 0.608 | 0.0298 | 0.575 to 0.640 | 0.55 | 0.0374 | 0.498 to 0.600 |
| Sex | 0.635 | 0.0284 | 0.603 to 0.666 | 0.53 | 0.0561 | 0.479 to 0.581 |
| Surgery | 0.718 | 0.0349 | 0.688 to 0.748 | 0.788 | 0.0515 | 0.744 to 0.828 |
| Nomogram | 0.937 | 0.0173 | 0.919 to 0.952 | 0.91 | 0.025 | 0.877 to 0.937 |

to cohorts that did not receive chemotherapy [18]. The role of chemotherapy in the treatment of patients with localized and advanced chondrosarcoma is unclear. Although its use in conventional chondrosarcoma has been largely ineffective, recent data suggest that it may play a role in certain chondrosarcoma subtypes, particularly dedifferentiated and mesenchymal variants, and reports suggest that chemotherapy may provide benefit for this particular subtype [19]. Therefore, further research is needed on the role of chemotherapy in the treatment of chondrosarcoma. In this study, we found that the OR of the population treated with surgery compared to the population not treated with surgery showed that the risk of distal metastases was only 0.2 times higher in patients who underwent surgery. Surgical resection is the primary treatment for both primary and metastatic chondrosarcoma [20]. The goal of resection is to remove the primary tumor with clear margins to limit recurrence and metastasis. Currently, 90% to 95% of patients with osteosarcoma of the extremities successfully avoid amputation through limb-preserving surgery [21]. It has also been demonstrated in relevant studies that high-grade chondrosarcoma (grade 2 and above) is best treated with

extensive resection [22]. Approximately half of the patients have a good or very good prognosis [23], which is consistent with the findings of this study. This study showed that male patients with chondrosarcoma had a higher and statistically different risk of developing distal metastases compared to female patients. OR values showed that female patients had only 0.4 times the risk of developing LM compared to male patients. A related study found that men were an independent risk factor for survival in patients with chondrosarcoma [24]. Considering that men have a higher risk of distal metastasis, it may, therefore, affect survival expectations.

According to the results of the regression analysis in Table 1 logistics, patients with lymphatic metastases had an approximately 9-fold higher risk of distal metastasis than patients without lymphatic metastases (OR = 9.168), and patients with unknown lymphatic metastases had an approximately 7-fold higher risk of distal metastasis than patients without lymphatic metastases (OR = 6.743). The prevalence of regional lymph node involvement in patients with chondrosarcoma is 1.3% due to the lack of lymphatic vessels in the bone, which rarely metastasize through lymph

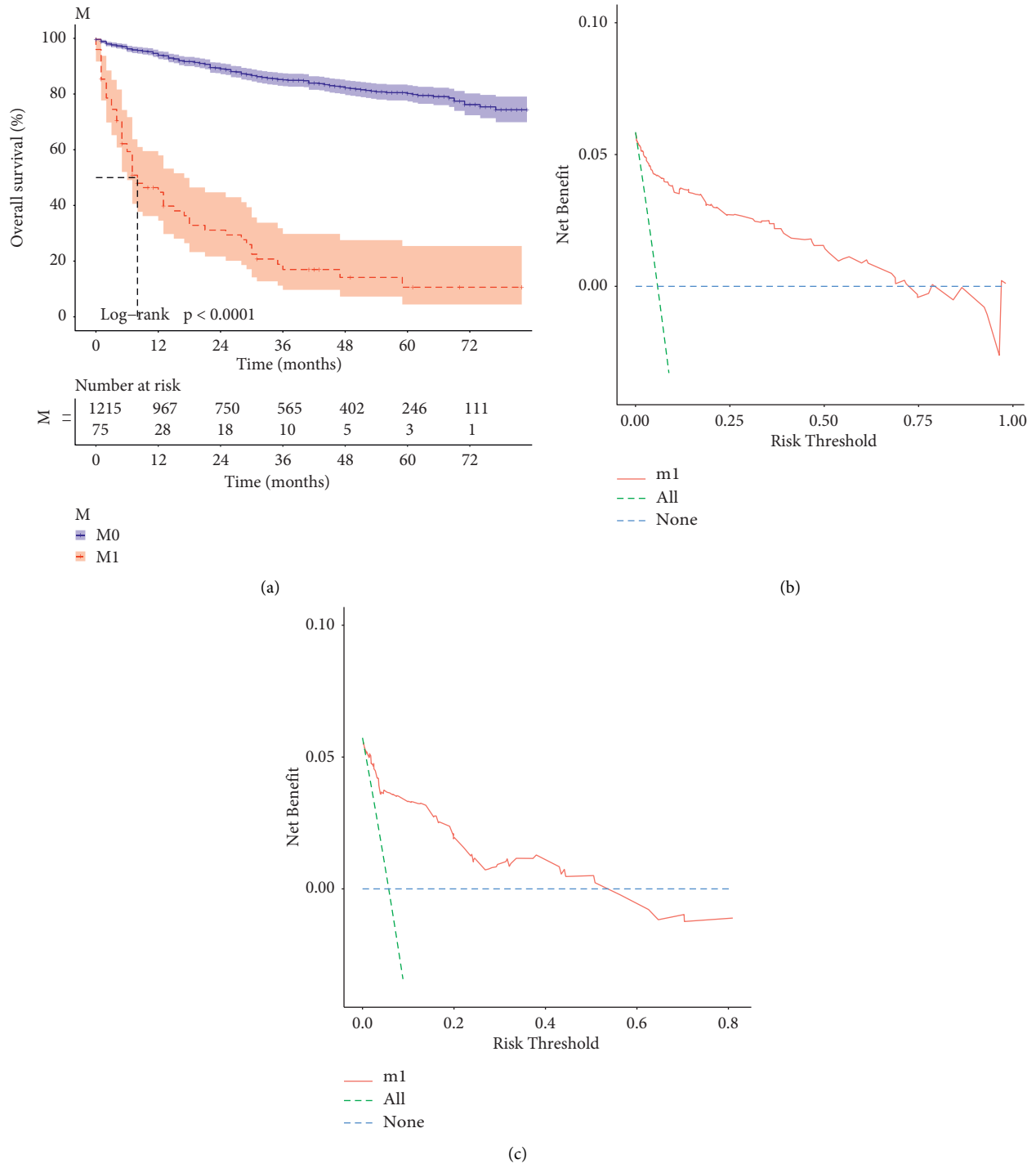


FIGURE 3: (a) Kaplan-Meier survival curve, (b) decision curve analysis (DCA) internal, and (c) decision curve analysis (DCA) external.

nodes. Therefore, lymph node spread in chondrosarcoma is extremely rare, and lymph node metastasis, although extremely rare in primary osteochondrosarcoma, has been shown to have an overall 5-year survival rate of 28% and 77% in patients with and without regional lymph node metastasis, respectively. Patients with chondrosarcoma with lymph node metastases had a worse prognosis than those who did not report regional lymph node metastases [25, 26], consistent with the findings of this study. This study also

found that patients with regional lymph node metastases had different tumor characteristics compared to those without regional lymph node metastases. Most importantly, the finding of regional lymph node involvement independently indicates a lower survival rate in patients with chondrosarcoma; this may be important when planning treatment or advising patients on their condition, so clinicians should more carefully examine patients with chondrosarcoma who have lymph node metastases [27].

Therefore, factors related to lymphatic metastasis and distal metastasis still need to continue to be studied in depth.

Nomogram is a quantitative mathematic tool to assess risk and benefit and has been widely used in the medical field for clinical decision-making in a variety of diseases [28, 29]. In previous studies, several nomograms have been developed and validated to predict specific survival and overall survival in chondrosarcoma [30]. However, a nomogram for predicting distal metastasis has not been reported. In this study, 1,290 chondrosarcoma cases were obtained from the SEER database, and patient prognostic factors (i.e., gender, tumor pathology grading classification, laterality, primary tumor stage, regional lymph node metastasis, surgical treatment, and chemotherapy) based on seven of the logistics regression analyses were used to establish a nomogram for predicting the distal nomogram for distal metastasis that showed better diagnostic efficiency compared to other individual variables as evidenced by calibration plots and ROC curves (Figures 1 and 2, Table 4). This all proves the value of the use of nomograms in this study, which can be further applied and improved in clinical work, and clinicians can choose better medical tests and optimize treatment plans with the help of nomograms.

This study still has limitations. First, this study is a retrospective analysis, and the data are biased and indeed lack systematic and prospective data. Also, as a single-center study, even though it was divided into training and validation groups, it still lacks external validation from other institutions, which may lead to overfitting of the nomogram for predicting distal metastasis.

5. Conclusions

A large population-based cohort from the SEER dataset was screened for this study and statistically analyzed to conclude that age, gender, survival time, tumor pathological grade, laterality, primary tumor stage, and whether or not surgical treatment and chemotherapy were prognostic factors affecting distal metastasis; higher or unclear tumor pathological grade, laterality to the right, higher primary tumor stage, male, lymph node metastasis, chemotherapy, and not the higher tumor pathology grade or rightward, higher primary tumor stage, male, lymph node metastasis, chemotherapy, and no surgical treatment were independent risk factors for distal metastasis. A nomogram was further constructed based on the results of statistical analysis to predict distal metastasis in patients with chondrosarcoma. Based on the results of internal validation, DCA curves and clinical impact maps, the nomogram in this study can effectively predict the individualized risk of distal metastasis.

Data Availability

SEER database within the article is a public dataset.

Ethical Approval

The SEER database is a comprehensive data source developed based on population data and annually updated since

its launch in 1973. It is public and identifiably accessible that data analysis is treated as nonhuman subjects by the Office for Human Research Protections. As such, no institutional review board approval and informed consent were required.

Conflicts of Interest

All authors declare that they have no conflicts of interest in this paper.

Authors' Contributions

Wenle Li, Rong Li, and Wanying Li have equally contributed to this work. CLY, RL, QL, and HWP jointly carried out the entire research design. WLL, RL, WYL, CX, and MMM participated in the research and collected and analyzed data. WLL, RL, WYL, and BW drafted the manuscript. CLY and QL provided expert consultation and advice. All authors conceived this research, participated in its design and coordination, and helped polish the language. All authors reviewed the final version of the manuscript.

References

- [1] J. C. Chen, Y. C. Fong, and C. H. Tang, "Novel strategies for the treatment of chondrosarcomas: targeting integrins," *BioMed Research International*, vol. 2013, Article ID 396839, 11 pages, 2013.
- [2] K. Song, X. Shi, and H. Wang, "Can a nomogram help to predict the overall and cancer-specific survival of patients with chondrosarcoma?," *Clinical Orthopaedics and Related Research*, vol. 476, no. 5, pp. 987–996, 2018.
- [3] D. Landini, F. Montanari, and F. Rolla, "Visual changes following cyberknife radiosurgery for skull base chordomas and chondrosarcomas," *Journal of Neurological Surgery Part B: Skull Base*, vol. 79, no. 1, pp. S1–S188, 2018.
- [4] W. Li, S. Dong, H. Wang, R. Wu, and C. Yin, "Risk analysis of pulmonary metastasis of chondrosarcoma by establishing and validating a new clinical prediction model: a clinical study based on SEER database," *BMC Musculoskeletal Disorders*, vol. 22, no. 1, p. 529, 2021.
- [5] J. S. Whelan and L. E. Davis, "Osteosarcoma, chondrosarcoma, and chordoma," *Journal of Clinical Oncology*, vol. 36, no. 2, pp. 188–193, 2018.
- [6] L. W. O'Neal and L. V. Ackerman, "Chondrosarcoma of bone," *Cancer*, vol. 5, no. 3, pp. 551–577, 2015.
- [7] X.-B. Wang, G.-H. Lv, M.-X. Zou, and L. Jing, "Prognostic factors in spinal chordoma: a systematic review," *Clinical Neurology and Neurosurgery*, 2015.
- [8] K. M. Doll, A. Rademaker, and J. A. Sosa, "Practical guide to surgical data sets: surveillance, epidemiology, and end results (SEER) database," *Jama Surgery*, vol. 153, no. 6, pp. 588–589, 2018.
- [9] L. I. Wenle, W. Hao-Sheng, N. Li-Jun, G. Sen, and H. U. Zhaohui, "Risk analysis for pulmonary metastasis of chondrosarcoma and establishment and validation of novel clinical prediction models: a clinical study based on the SEER database," *BMC Musculoskeletal Disorders*, vol. 22, no. 1, p. 529, 2020.
- [10] A. Y. Giuffrida, J. E. Burgueno, L. G. Koniaris, J. C. Gutierrez, R. Duncan, and S. P. Scully, "Chondrosarcoma in the United States (1973 to 2003): an analysis of 2890 cases from the SEER

- database,” *Journal of Bone & Joint Surgery American*, vol. 91, 2009.
- [11] E. Q. Wu, D. Hu, P. Y. Deng, Z. Tang, and H. Ren, “Non-parametric bayesian prior inducing deep network for automatic detection of cognitive status,” *IEEE Transactions on Cybernetics*, vol. 55, pp. 5483–5496, 2020b.
- [12] O. Toshifumi, N. Lindner, A. Hillmann, and S. Blasius, “Influence of intralesional surgery on treatment outcome of chondrosarcoma,” *Cancer*, vol. 77, no. 7, pp. 1292–1297, 1996.
- [13] P. Lanzkowsky, *Manual of Pediatric Hematology and Oncology*, Academic Press, Cambridge, MA, USA, 5 edition, 2010.
- [14] Z. Tang, R. Zhu, P. Lin et al., “A hardware friendly unsupervised memristive neural network with weight sharing mechanism,” *Neurocomputing*, vol. 332, pp. 193–202, 2019.
- [15] S. Nota, Y. Braun, J. Bramer, and J. Schwab, *The Identification of Reliable Prognostic Factors of Chondrosarcoma: A Systematic Review*, 2015.
- [16] Z. Wang, G. Chen, X. Chen et al., “Predictors of the survival of patients with chondrosarcoma of bone and metastatic disease at diagnosis,” *Journal of Cancer*, vol. 10, no. 11, pp. 2457–2463, 2019.
- [17] E. Q. Wu, P. Y. Deng, X. Y. Qu, Z. Tang, and R. Sheng, “Detecting fatigue status of pilots based on deep learning network using EEG signals,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, pp. 575–578, 2020a.
- [18] I. D. Dickey, P. S. Rose, B. Fuchs et al., “Dedifferentiated chondrosarcoma: the role of chemotherapy,” *J Bone Joint Surg Am*, vol. 86, pp. 2412–2418, 2008.
- [19] R. F. Riedel, N. Larrier, L. Dodd, D. Kirsch, S. Martinez, and B. E. Brigman, “The clinical management of chondrosarcoma,” *Current Treatment Options in Oncology*, vol. 10, no. 1-2, pp. 94–106, 2009.
- [20] J. L. P. Ferguson and S. P. Turner, “Bone cancer: diagnosis and treatment principles,” *American Family Physician*, vol. 98, pp. 205–213, 2018.
- [21] J. R. Lieberman, “AAOS Comprehensive Orthopaedic Review 2”, *American Academy of Orthopaedic Surgeons*, Rosemont, IL, USA, 2009.
- [22] F. Fiorenza, A. Abudu, R. J. Grimer, S. R. Carter, and A. M. Davies, “Risk factors for survival and local control in chondrosarcoma of bone,” *Journal of Bone & Joint Surgery-british Volume*, vol. 84, no. 1, pp. 93–99, 2002.
- [23] J. Bruns, M. Elbracht, and O. Niggemeyer, “Chondrosarcoma of bone: an oncological and functional follow-up study,” *Annals of Oncology*, vol. 12, no. 6, pp. 859–864, 2001.
- [24] Z. Tang, R. Zhu, R. Hu, Y. Chen, and S. Chang, “A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, pp. 645–656, 2020.
- [25] M. R. Claxton, G. Reynolds, D. E. Wenger, P. S. Rose, and M. T. Houdek, “Extraskeletal myxoid chondrosarcoma: a high incidence of metastatic disease to lymph nodes,” *Journal of Surgical Oncology*, vol. 122, no. 8, pp. 1662–1667, 2020.
- [26] V. Kurisunkal, A. Gulia, A. Puri, and B. Rekhi, “Lymph node metastasis in extremity chondrosarcomas: a series of four cases,” *South Asian Journal of Cancer*, vol. 9, no. 1, p. 1, 2020.
- [27] L. M. Nystrom, “CORR insights: regional lymph node involvement is associated with poorer survivorship in patients with chondrosarcoma,” *Clinical Orthopaedics and Related Research*, vol. 477, no. 11, p. 1, 2019.
- [28] G. Li, M. L. Tian, Y. T. Bing, H. Y. Wang, and D. R. Xiu, “Nomograms predict survival outcomes for distant metastatic pancreatic neuroendocrine tumor: a population based STROBE compliant study,” *Medicine (Baltimore)*, vol. 99, no. 13, Article ID e19593, 2020.
- [29] K. Song, J. Song, and H. Wang, “Development and validation of Nomograms predicting overall and cancer-specific survival of spinal chondrosarcoma patients,” 2018.
- [30] X. Zhou, W. Liang, W. Li, K. Yan, and I. K. Wang, “Hierarchical adversarial attacks against graph neural network based IoT network intrusion detection system,” *IEEE Internet of Things Journal*, vol. 1, no. 99, p. 1, 2021.

Research Article

Agent-Based Data Extraction in Bioinformatics

Shakir Ullah Shah ^{1,2}, Abdul Hameed,¹ Abdulwahab Ali Almazroi,³
and Mohammed A. Alqarni⁴

¹Department of Computer Science, Iqra University, Islamabad, Pakistan

²National University of Computer and Emerging Sciences, Peshawar, Pakistan

³College of Computing and Information Technology, College of Computer Science and Engineering at Khulais,
Department of Information Technology, University of Jeddah, Jeddah, Saudi Arabia

⁴College of Computer Science and Engineering at Khulais, Department of Information Technology, University of Jeddah,
Jeddah, Saudi Arabia

Correspondence should be addressed to Shakir Ullah Shah; shahshakir@yahoo.com

Received 5 January 2022; Revised 24 January 2022; Accepted 31 January 2022; Published 26 March 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Shakir Ullah Shah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bioinformatics is an active and important research discipline in which molecular data is exponentially growing in complex nature. Because of the substantial research in this field, researchers are faced with critical issues such as bandwidth, storage, and complexity in order to retrieve molecular data. It becomes very difficult to conduct research using low computational devices such as Internet of things and sensors. We are employing migration of the agent technique to decrease network traffic and to mitigate the client's limited resource problem by utilizing server-side resources to perform large-scale computation. Our proposed solution does not necessitate additional storage or processing power on the client's side which makes it cost effective. In the proposed solution, (i) an agent visits service provider containing biological data, say sequences requested by the client, (ii) agent fetches the required data, and on the server side it will manipulate the data, and (iii) returns along with the required results to its source platform. Thus, it solves the bandwidth, storage, and computational issues without involving the low resources of the client. For the proof of concept, Java Agent Development (JADE) framework is used as an implementation tool and the results are compared with Java Remote Method Invocation (RMI). It is important to note that our findings reveal that our strategy saves the user up to 16.25% of average time with respect to bandwidth. On the other hand, our approach takes 46.82% less time than the other with respect to data that the agent carries. In addition to the previous contributions, our approach acts as a mashup, to collect data in different format from several service providers, and converts it in any required format. Thus, it solves the problem of complexity hidden in the nature of the data to increase the researchers' productivity.

1. Introduction

The volume and complexity of data over multiple service providers are generated exponentially. Now the issue is extraction, retrieval, and processing of relevant information which has made obvious the need for a system to facilitate users. Data coming in from multiple domains needs to be integrated together to provide a cohesive view. One technology that massively helps facilitate this goal is the concept of mashups [1, 2]. Mashups help users get an integrated user-oriented view of data and code from multiple heterogeneous

sources. Trendsmap, housingmaps, Wheel of Lunch, and InstantWatcher are some examples of mashup. A mashup is a web application that stitches together the contents, presentations, and application functionalities from multiple sources and gives them a new and useful look. In other words, it combines multiple services into a single one [3, 4]. In this paper, we apply the concept of mashup through multiagent paradigm to the domain of bioinformatics.

The bioinformatics [5] and [6] is about understanding biological data and is a growing field of research. With the advance in technology, the amount of biological data is

growing at a tremendous pace. This makes the field of bioinformatics important to the society. They have also made strides in understanding how they interact with other proteins, which is known as protein-protein interactions. For example, data reports [8] and [9] also include both structural data [8] and other data [10], [11] about proteins. A wide range of computational techniques has proposed, primarily for image manipulation and pattern detection. With NGS, analysts are analyzing precious data and doing a lot of intensive work to find patterns. More importantly, it is considerably difficult for them to create complex maps from heterogeneous sources.

There are various experimental approaches [18] which are very expensive and time consuming because they require a lot of resources and time to measure the physical interaction among proteins. They have a high possibility of error because experiments are purely carried out in the lab and are not standardized. The adoption of agent technologies and multiagent systems constitutes an emerging area in bioinformatics where data is quite big (that is, in gigabytes) and complex in nature. Researchers face the problem of data retrieval due to low (1) bandwidth, (2) storage, and (3) computation on their machines. For the processing, analyzing and transportation of multilevel complexity of molecular data require high bandwidth and storage. Thus, it becomes very difficult for researchers to conduct research with low power researches. This study proposed agent migration-based approach [19–21].

The main advantage of an agent-based approach is that the agent will get the request from the requester and visit the service provider, which saves the time for the data provider to make the data available, and agents will return back to the requester. Another advantage of using agent-based approach is that we will be able to transfer data to the machine which has high computational power. The computation is done at the service side. And the results are available in the same format as they were at service side. This will help to reduce the data size and transfer the data to the machine which has high computational power. This will give better efficiency, reduce network congestion, and transfer the data to machine with high computational power. The main theme of this paper is to use agent-based methodology which tackles processing and network issues of multiagent systems. It processes all the necessary steps on the client side with low computational resources.

The rest of the paper is organized into different sections. Literature review about bioinformatics, multiagent systems, and mashup is provided in Section 2. Section 3 provides a list of complete steps of our agent-based solution along with details. In Section 4, as a proof of concept, a reference implementation is listed. Section 5 provides the analysis and discussion of the proposed solution. Section 6 concludes this study along with future directions.

2. Literature Review

In this section, literature about bioinformatics mashup and multiagent systems is presented. In each section, the

importance of each domain is provided. We first turn to the target domain, namely, biofoundation.

2.1. Bioinformatics. Bioinformatics is an interdisciplinary field that mostly uses computer as a computational tool for solving issues related to the biological data. Such computing devices are used for the analysis of the internal structure and biological functions of living organisms. The main purpose of computing devices is giving an efficient structure to the data so that it could be interpreted accurately. Mainly it deals with genome and protein. One of the important characteristics of bioinformatics is personalized medicines. It is the application of computer processing techniques to the field of genetics and biochemistry. This is a branch of computer science that deals with the storage, retrieval, and analysis of biological data [11]. It is a classification of data in a standard manner. The data are analyzed in order to determine their structure and content [12]. The bioinformatics includes the research in the field of genetics and genomics. This branch of science is implemented in various fields such as the study of evolution and phylogeny of various species [13]. The data generated by the bioinformatics are stored at various data centers. These data can be used for the diagnosis of the diseases and for the treatment and prevention of the diseases [14].

The protein-protein interactions [22] are of extreme importance because they play a vital role in many biological processes, such as signal transduction and transcription regulation. They also act as protein-based modules that are extensively used by nature to build complex systems. Therefore, they are a primary objective of many bioinformatics algorithms. However, in the field of protein-protein interactions, the problem of interactions between proteins in the context of the entire proteome has received less attention. In this context, a recent study has been carried out by proposing a new protein-protein interaction network. The network is based on the comparison of the entire proteome between two different organisms.

Protein-protein interactions [8, 9, 23] are a key element in the study of molecular biology, as molecular interactions are at the foundation of all biological systems. In this regard, the interactions between proteins have been extensively studied [10, 24]. Protein Interactions by Structural Matching (PRISM) [25–27] is an online web tool for predicting protein-protein interactions with high confidence. It is based on the structural and functional domain similarity of proteins. The first step in the PRISM-2 algorithm is to generate a structural alignment of two proteins. The proteins are compared by hand-determined structural similarity, and this similarity is used to generate a structural alignment.

There is a gigantic development in the organic succession where a huge amount of data is being made and transferred on the sites/servers. Presently to get the information we would have to communicate with the connection point utilizing electronic inquiries [28]. This means that the user has to click on a single link to access all the linked websites. This is very time consuming and tedious to stay online each query. The idea of a mashup is to integrate multiple datasets

into a single system. The paradigm of a mashup combines the data from multiple heterogeneous data sources. It is a great way to combine diverse data sources into a single view. Mashups are a great way to implement the existing data into a new structure. It is used in situations where the data from multiple heterogeneous data sources are required to be combined into a single view. The main idea behind the mashup is to integrate the data from multiple data sources into one system.

2.2. Mashup. Information retrieval is becoming a challenging task due to rapid proliferation of data. It becomes more complex when the required information is scattered on multiple service providers. This complexity demands an efficient system to retrieve the desired results in an appropriate manner. There are different approaches to retrieve the information, combine it, and give a desired look; mashup is one of such approaches [29]. Mashup gives entirely a new and different look or some added value to the existing data for end users. Service providers provide APIs which act as an interface for data and services. Some APIs are free, and some are proprietary in nature that need authentication and authorization. Asynchronous JavaScript and XML (AJAX), Representational State Transfer (REST), and Services-Oriented Access Protocol (SOAP) are some state-of-the-art technologies that have influenced the mashup architecture [30]. REST, screen scraping, and RSS feed/widget are used to retrieve the contents from other websites. It is widely developed for web applications such as social networking, e-government, enterprise resource management, real state, and more [31, 32].

A number of tools exist to create mashup such as Yahoo Pipes [33] or IBM Mashup Center [34]. Traditionally, a mashup runs inside a web browser, but there are also some other environments for it. Two important styles of mashup are server-side mashup and client-side mashup [35]. The difference between server-side and client-side mashups is the way the data is processed. In a server-side mashup, also called a proxy-style mashup, a web server serves the mashup to retrieve all the data from multiple web hosts, and stitching takes place on server side and is rendered on client's web browser. In a client-side mashup, opposite to server-side mashup, stitching of the services and contents takes place on the client, namely, within the web browser. These are also called Rich Internet Applications (RIAs) and have the added advantage of prompt response over server-side mashup. A mashup can be either a consumer or enterprise [36]. A consumer mashup also known as service or client mashup integrates data from multiple public sources inside the browser, for example, iGuide; server-side mashup is the target of this study. Both styles of mashup have their own obvious benefits, as both provide new insight into existing resources. But using such mashup tools, users must trust them. So user data is not secure, since it has to be released to the third parties. We address this issue using the multiagent paradigm.

2.3. Multiagent Systems (MAS). Multiagent system is the collection of multiple software agents [37]. A software agent is a piece of code that works autonomously and communicates with other agent-oriented and non-agent-oriented software [38, 39]. The basic building blocks of an agent consist of code, data, and state. The data part represents the data structure to preserve important information about the expression before and after evaluation. The configuration of the agent is stored in its data and state parts. It contains information about platforms which changes dynamically when it travels from one node to another node within a network. The code part of an agent is the collection of ordered statements that remains nearly constant during the execution though it can change when required. It represents the actual logic of the agent. The state part of an agent represents the current status of the data part. Basically, state is the collection of information of all data structures.

2.4. Significance of Multiagent Systems. Agent-oriented software paradigm has become a promising technology which is widely used in distributed environments such as e-commerce [40], network management [41], data mining [42], robotics [43, 44], and information extraction [45]. Some interesting applications of agent systems can be found in healthcare system [46, 47] for patient scheduling, storing medical records of patients, and sharing them with concerns. Agent-based system, also called multiagent system [48], is the system in which multiple agents interact, cooperate, and coordinate with each other. Such system, loosely coupled, enhances the capabilities of monolithic system to perform different tasks which are beyond the scope of individual agent. It is widely used to share or get resources over the network among agents. The resources might be computational, logic to solve the problem, software or expertise distributed temporally and spatially. Normally, systems are categorized into two categories: client (to make a request) and server (to server) but multiagent systems combine the benefits of both in a social, proactive, and reactive manner.

2.5. Design Issues in Multiagent Systems. The most important design issue for multiagent systems is how they will communicate among each other and with other entities. The starting point is to select any tool or middleware to facilitate developers to get the core benefits of this technology rather than to resolve the basic issues of communication. So some standards are needed prior to deploying such system. The Foundation for Intelligent Physical Agents (FIPA) [49], AGENTLINK [50], and OMG Agent Platform Special Interest Group (PSIG) [51] are the leading standardization bodies to promote agent technology. This study focuses on FIPA for agent reference and development model. There are various tools for agent-based modeling like NOMADS [52, 53], AgentScape [54], Agentcities [55], Aglets [56], Voyager [57], Janus [58], TACOMA [59, 60], Grasshopper [61], JADE [62, 63], JaCaMo [64], Adresse Jason [65], and ABLE [66].

JADE [62] is used to launch an agent platform. The most important reason is that it is an open source middleware under the Library Public License (LGPL). It is entirely implemented in the Java language, which makes it more portable and smarter. It is one of the most popular middleware types within the research community. It alleviates the implementation of multiagent systems (MAS). It provides a set of graphical tools which make it very easy to deploy agent platform on a standalone system as well as over a distributed network. This study highly recommends JADE as agent middleware. Its infrastructure is very flexible and agent community is adding different add-ons to enhance its features. It is compliant with the FIPA-IEEE computer society specifications. The core concept of FIPA is to resolve the issue of interoperability and it has extended FIPA's model in multiple areas. According to the JADE specification, the mobility of an agent can be categorized into two types: inter- and intraplatform. In intraplatform mobility, an agent migrates itself between containers of the same platform but cannot move to containers of the different platform. In intraplatform mobility [67], an agent moves among different platforms. In interplatform mobility, the agent leaves its own main container and joins another main container of another platform. The main focus of this study is interplatform mobility; see step 1 for details of each and how an agent can migrate from one platform to another.

2.6. Bioinformatics and Multiagent Systems. This study also explores the area of bioinformatics as a real application of multiagent systems and explores how a mobile agent can operate in a highly dynamic environment for data dissemination. A mobile agent visits different itineraries to collect the required information and stitch it together to provide a new shape. We propose agent migration to mitigate the aforementioned issues by moving the agent to the server side to perform computations [21, 68]. In a nutshell, this study proposes agent migration characteristic to make it a mashup. Hence, it can be used for data dissemination.

3. Solution

We propose mobility characteristic of an agent to find a solution. The solution provides accurate and fine-grained result even though the bandwidth and storage of the client might be low. It also deals with complexity present in the nature of the data as well as in the dynamic environment. In the agent migration approach, an agent is executed in a client machine; the agent migrates to another machine when the original machine is overloaded. To migrate an agent from one machine to another, the agent must be able to traverse the network. The abstract details are in Figure 1.

Java Remote Method Invocation (RMI) [69] is a way to extract data from the server as it allows remote access to Java objects on a remote host. It is light weight communication protocol. The payload of Java RMI is the Java object that contains references to the remote methods. A Java object on the remote host is a Java object that is created on the remote host. A client stub object, which is a proxy object that

contains references to the remote object, is used to access remote object. The complete steps which were used in this study are listed in Figure 2.

Mobile agents use the resources of the system to complete the tasks and then get back to the system where they started their execution [10, 11]. Agent-based systems are a powerful and effective way to develop intelligent systems because of their simplicity and extensibility. They are more effective in handling with the problems related to distributed, parallel, and autonomous systems.

It is also effective in handling with the problem of complex systems. The reason behind it is that they are not heavy in computation and they can be used on multiple systems at the same time. This makes it easy to design and implement these systems. The steps which are carried out in this study are mentioned in Figure 3.

4. Reference Implementation

To deploy agents, various agent frameworks are available [70, 71]. Java Agent Development (JADE) framework is FIPA Agent Markup Language (FAM), a language designed to be used to model agent systems. It is compliant with the FIPA Agent Communication Language (ACL) specification and with the FIPA Agent Communication Framework (ACF). We provide the source of our own reference implementation at <https://github.com/BioAgent>.

For the testbed configuration, two personal computers were used: one as a server and the other one as a client. Both systems were connected through the 4G Huawei E5573s-320 which is a pocket WiFi router. Table 1 shows both hardware and software details of both personal computers.

5. Results and Discussion

This section presents the study carried out on the performance of mobile agents and Java RMI. A detailed discussion of the results is carried out in this section. Java RMI and agent migration approached are compared. Due to the fact that mobile agents are not dependent on the host application and can be independently transferred to another host, an effective approach to large-scale agent migration has been proposed.

Table 2 shows the amount of network load made by client using Java RMI and our agent-based approach. The agent approach is more efficient than Java RMI approach because it decreases the number of network calls made by the client. The agent approach is more efficient because the agent is migrated to the server only when there is a need for it. As a result, the client makes fewer network calls, and the overhead of the network calls is reduced. Therefore, it is clear that the Java RMI approach has more network load than the agent approach. In the agent approach, the number of network calls is reduced by migrating the agent. It is important to note that, in the agent approach, the agent is only migrated if there is an urgent need for it.

The Java RMI approach is a lot more mature compared to the agent-based approach. Table 2 provides a summary of the results of Java RMI and agent approaches based on

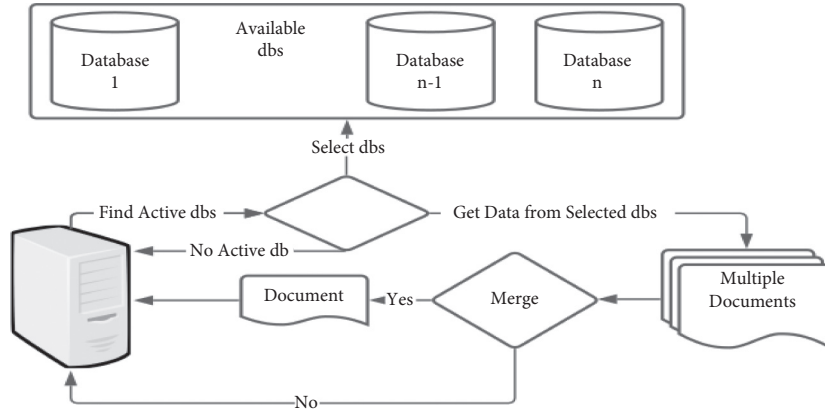


FIGURE 1: System architecture.

Algorithm 1 P2P Interactions using JAVA RMI

Require: $len(Avail_{dbs}) \geq 1$
Ensure: $Download_{dbs} = Selected_{dbs}$
 $Download_{dbs} \leftarrow []$
 $Avail_{dbs} \leftarrow fetch_{dbs}(URL, Proten_{id})$
while $Avail_{dbs} \neq 0$ **do**
 $Selected_{dbs} \leftarrow select_{dbs}(Avail_{dbs})$
 $i \leftarrow 1$
 $current_{dbs} \leftarrow Selected_{dbs}[i]$
 while $i \neq len(Selected_{dbs})$ **do**
 $inter_s \leftarrow interactions(current_{dbs})$
 while $inter_s \neq 0$ **do**
 $down_{db} \leftarrow interactions(inter_s, current_{dbs})$
 $Download_{dbs}.append \leftarrow down_{db}$
 $inter_s \leftarrow inter_s - 1$
 end while
 $i \leftarrow i + 1$
 $current_{dbs} \leftarrow Selected_{dbs}[i]$
 end while
end while

FIGURE 2: Protein-protein interactions using Java RMI.

network load. The agent migration approach does not have any network load as it needs only interconnectivity.

In Figure 4, the x -axis represents the size of the extracted data from multiple databases, while the y -axis shows the network load consumed by both approaches. According to Figure 4, we conclude that Java RMI is showing an increasing trend in result size. But, at the same time, the curve shows an increasing trend in network load. Thus, while achieving high results, network load also increases with time. That is why the curve has been shown in the graph. It shows a direct relationship between result size and network load. The result size is dependent on network load. On the other hand, an agent-based system shows high results with a constant value of network load. That is why the graph of an agent-based system is a straight line.

In Figure 5, we can see that the agent-based approach gives better results as compared to Java RMI when the result size increases from 2 kB. Similarly, the response time is only 10 seconds at result size of 5 kB. Furthermore, 10 kB result size is achieved at a response time of only 15 seconds.

From Figure 5, which is based on Table 3, we can conclude that, in the Java RMI system, result size is directly proportional to response time. As the size of the result increases, the response time also increases. It will take more response time to achieve a high volume of results. On the other hand, an agent-based system shows a high return size with 46.82% less response time than the other with respect to data that the agent carries. The average of Java RMI approach is 20.21785714 while the average time of our approach is 10.75047619. The difference of both approaches is 9.467380952.

In Figure 6, the blue line shows the agent graph and the red line shows the graph of Java RMI. We can clearly see that if we decrease the bandwidth, our agent is computing faster as compared to Java RMI.

According to Table 4, the average responses of both approaches are 20.21785714 and 10.75047619. It is important to note that our findings reveal that our strategy saves the user up to 16.25% of the average time with respect to bandwidth.

Algorithm 2 P2P Interactions using Agent Migration

Require: $Avail_{dbs} \neq 0, AgentPlatform, MobilityService_{(client+dbs)} \leftarrow True$

Ensure: $Download_{dbs} = Selected_{dbs}$

```

bioAgent  $\leftarrow Agent$ 
bioAgent.address  $\leftarrow client$ 
Downloaddbs  $\leftarrow [ ]$ 
Availdbs  $\leftarrow fetch_{dbs}(URL, Proten_{id})$ 
while Availdbs  $\neq 0$  do
  Selecteddbs  $\leftarrow select_{dbs}(Avail_{dbs})$ 
  i  $\leftarrow 1$ 
  currentdb  $\leftarrow Selected_{dbs}[i]$ 
  while i  $\neq len(Selected_{dbs})$  do
    AMSr  $\leftarrow AID("ams@current_{db}/JADE", AID.ISGUID)$ 
    destination  $\leftarrow AMS_r.addAddresses("http://current_{db}:7778/acc")$ 
    bioAgent.doMove(destinationdb)
    downdb  $\leftarrow interactions(inters, current_{db})$ 
    inters  $\leftarrow interactions(current_{db})$ 
    while inters  $\neq 0$  do
      downdb  $\leftarrow interactions(inter_s, current_{db})$ 
      Downloaddbs.append  $\leftarrow down_{db}$ 
      inters  $\leftarrow inter_s - 1$ 
    end while
    i  $\leftarrow i + 1$ 
    currentdbs  $\leftarrow Selected_{dbs}[i]$ 
  end while
  bioAgent.doMove(client)
end while

```

FIGURE 3: Protein-protein interactions using agent migration.

TABLE 1: Testbed configuration.

| Feature | Server | Client |
|------------------|---|---|
| Operating system | Windows 10 (64 bits) | Windows 10 (64 bits) |
| Model | Toshiba satellite L50-B1380 core i5 6 GB RAM 1 TB HDD | Dell inspiron 15 5570 core i5 4 GB RAM 1 TB HDD |
| JADE version | 4.5.0 | 4.5.0 |

TABLE 2: Comparison of network load between Java RMI and agent-based approach.

| Result size (byte) | Java RMI Network load (byte) | Agent-based approach Network load (byte) |
|--------------------|---------------------------------|---|
| 0 | 0 | 0 |
| 50 | 0.0625 | 0 |
| 100 | 0.125 | 0 |
| 150 | 0.25 | 0 |
| 200 | 0.5 | 0 |
| 250 | 0.75 | 0 |
| 300 | 1.00 | 0 |
| 350 | 1.0625 | 0 |
| 400 | 1.25 | 0 |
| 450 | 1.75 | 0 |
| 500 | 2.00 | 0 |
| 550 | 2.25 | 0 |
| 600 | 2.65 | 0 |
| 650 | 3.01 | 0 |
| 700 | 3.50 | 0 |
| 750 | 3.96 | 0 |
| 800 | 4.25 | 0 |
| 850 | 4.98 | 0 |
| 900 | 5.25 | 0 |
| 950 | 5.97 | 0 |
| 1000 | 6.125 | 0 |

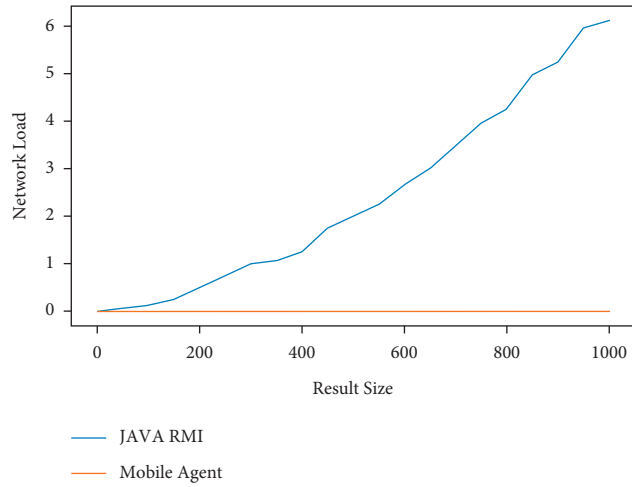


FIGURE 4: Client-side network load and result size.

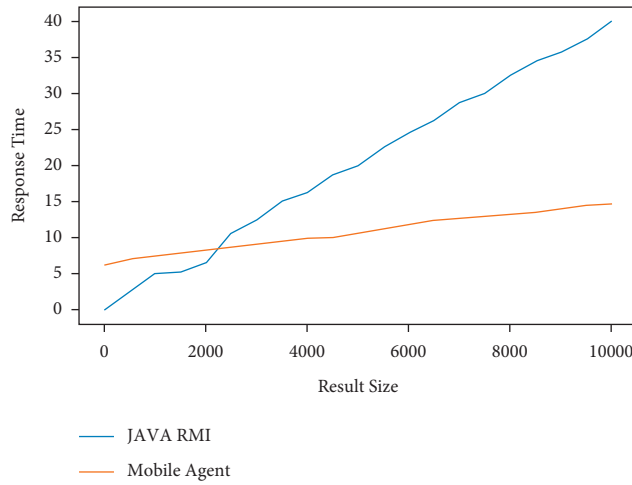


FIGURE 5: Response time according to agent size.

TABLE 3: Java RMI vs. agent, based on response time.

| Result size (byte) | Java RMI Response time (seconds) | Agent-based approach Response time (seconds) |
|--------------------|-------------------------------------|---|
| 0 | 0 | 6.25 |
| 500 | 2.5 | 7.00 |
| 1000 | 5.0 | 7.42 |
| 1500 | 5.25 | 7.84 |
| 2000 | 6.5 | 8.26 |
| 2500 | 10.625 | 8.68 |
| 3000 | 12.5 | 9.10 |
| 3500 | 15.0 | 9.52 |
| 4000 | 16.25 | 9.94 |
| 4500 | 18.75 | 10.00 |
| 5000 | 20.0 | 10.60 |
| 5500 | 22.5 | 11.20 |
| 6000 | 24.5 | 11.80 |
| 6500 | 26.25 | 12.40 |
| 7000 | 28.75 | 12.75 |
| 7500 | 30.0 | 13.00 |
| 8000 | 32.5 | 13.25 |

TABLE 3: Continued.

| Result size (byte) | Java RMI Response time (seconds) | Agent-based approach Response time (seconds) |
|--------------------|-------------------------------------|---|
| 8500 | 34.5 | 13.50 |
| 9000 | 35.7 | 14.00 |
| 9500 | 37.5 | 14.50 |
| 10000 | 40.0 | 14.75 |

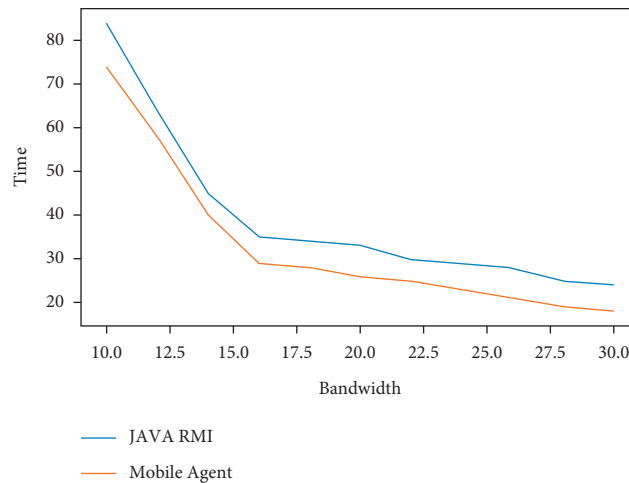


FIGURE 6: Response time according to agent size.

TABLE 4: JAVA RMI vs. agent based on bandwidth.

| Bandwidth (kbs) | Java RMI Time (seconds) | Agent-based approach Time (seconds) |
|-----------------|----------------------------|--|
| 10 | 84 | 74 |
| 12 | 64 | 58 |
| 14 | 45 | 40 |
| 16 | 35 | 29 |
| 18 | 34 | 28 |
| 20 | 33 | 26 |
| 22 | 30 | 25 |
| 24 | 29 | 23 |
| 26 | 28 | 21 |
| 28 | 25 | 19 |
| 30 | 24 | 18 |

6. Conclusions and Future Work

In this study, we have designed an agent migration approach for transferring the information between the clients. The agents migrate from client to client to collect the data and transfer it to a central server. The client uses the agent's services. Feedback service of the agent is used to ask the client for any information required by the agent. The client can provide information to the agent to ask the server for any service the client requires. The agent can migrate between the client and the server. The client can also request the agent to migrate to any other client. This approach has many advantages. The agents are intelligent, and they can even work well in low network areas. They can be used for many

generic purposes. The agents can be used to find out the interactions between proteins. This approach can be used for many bioinformatics problems like finding out the similarity of sequences, or even finding the missing sequence in known sequences. The findings also show that mobile agent technology leverages network load and storage on the client side and heterogeneous data can be converted into homogeneous format. The main limitation of this study is the deployment of agent environment on client and service side. This approach does not demand the availability of the user online for a full time period. Our research can be modified to make it work on different bioinformatics problem, like viewing the interaction of sequences. It can also be used to find out the similarity of sequences. By modifying the approach, one can

also find out the similarity of proteins, or even find the missing sequence in known sequences. It is also possible to find out the similarities between different organisms.

Data Availability

All relevant code samples can be found at GitHub-shahshakir/BioAgent (<https://github.com/shahshakir/BioAgent/>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References



- [1] A. Koschmider, V. Torres, and V. Pelechano, "Elucidating the mashup hype: definition, challenges, methodical guide and tools for mashups," in *Proceedings of the 2nd Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web at WWW2009*, pp. 1–9, Madrid, Spain, January 2009.
- [2] A. Jhingran, "Enterprise information mashups: integrating information, simply," in *Proceedings of the The 32nd International Conference on Very Large Data Bases*, pp. 3–4, Seoul, Korea, September 2006.
- [3] B. Hartmann, L. Wu, K. Collins, and S. R. Klemmer, "Programming by a sample: rapidly creating web applications with d. mix," in *Proceedings of the ACM Proceedings of the 20th annual ACM symposium on User interface software and technology*, pp. 241–250, Newport Rhode Island, USA, October 2007.
- [4] N. Zang, M. B. Rosson, and V. Nasser, "Mashups: who? What? Why," in *Proceedings of the ACM CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pp. 3171–3176, Florence Italy, April 2008.
- [5] A. D. Baxevanis, G. D. Bader, and D. S. Wishart, *Bioinformatics*, John Wiley & Sons, New Jersey, NY, USA, 2020.
- [6] R. Stevens, C. A. Goble, and S. Bechhofer, "Ontology-based knowledge representation for bioinformatics," *Briefings in Bioinformatics*, vol. 1, no. 4, pp. 398–414, 2000.
- [7] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, "Essential dynamics of proteins," *Proteins: Structure, Function, and Genetics*, vol. 17, no. 4, pp. 412–425, 1993.
- [8] D. F. Waugh, "Protein-protein interactions," *Advances in Protein Chemistry*, vol. 9, pp. 325–437, 1954.
- [9] T. Berggård, S. Linse, and P. James, "Methods for the detection and analysis of protein–protein interactions," *Proteomics*, vol. 7, no. 16, pp. 2833–2842, 2007.
- [10] Q. C. Zhang, D. Petrey, L. Deng et al., "Structure-based prediction of protein-protein interactions on a genome-wide scale," *Nature*, vol. 490, no. 7421, pp. 556–560, 2012.
- [11] S. Das and S. Chakrabarti, "Classification and prediction of protein–protein interaction interface using machine learning algorithm," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.
- [12] K. A. Theofilatos, C. M. Dimitrakopoulos, A. D. Likothanassis, T. Papadimitriou, and P. Mavroudi, "Computational approaches for the prediction of protein-protein interactions: a survey," *Current Bioinformatics*, vol. 6, no. 4, pp. 398–414, 2011.
- [13] R. B. Russell, F. Alber, P. Aloy et al., "A structural perspective on protein-protein interactions," *Current Opinion in Structural Biology*, vol. 14, no. 3, pp. 313–324, 2004.
- [14] B. Suter, X. Zhang, C. G. Pesce, A. R. Mendelsohn, S. P. Dinesh-Kumar, and J. H. Mao, "Next-generation sequencing for binary protein-protein interactions," *Frontiers in Genetics*, vol. 6, Article ID 346, 2015.
- [15] R. Carter, A. Luchini, L. Liotta, and A. Haymond, "Next-generation techniques for determination of protein-protein interactions: beyond the crystal structure," *Current pathobiology reports*, vol. 7, no. 3, pp. 61–71, 2019.
- [16] D. S. Horner, G. Pavesi, T. Castrignano et al., "Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing," *Briefings in Bioinformatics*, vol. 11, no. 2, pp. 181–197, 2010.
- [17] J. K. Kulski, "Next-generation sequencing—an overview of the history, tools, and "omic" applications," *Next generation sequencing-advances, applications and challenges*, vol. 10, pp. 3–60, 2016.
- [18] N. Safari-Alighiarloo, M. Taghizadeh, M. Rezaei-Tavirani, B. Goliaei, and A. A. Peyvandi, "Protein-protein interaction networks (ppi) and complex diseases," *Gastroenterology and Hepatology from bed to bench*, vol. 7, no. 1, 2014.
- [19] M. Wooldridge and N. R. Jennings, "Intelligent agents: theory and practice," *The Knowledge Engineering Review*, vol. 10, no. 2, pp. 115–152, 1995.
- [20] A. Omicini, A. Ricci, and M. Viroli, "Artifacts in the A&A meta-model for multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 17, no. 3, pp. 432–456, 2008.
- [21] L. Gao, H. Dai, T. L. Zhang, and K. C. Chou, "Remote data retrieval for bioinformatics applications: an agent migration approach," *PLoS One*, vol. 6, no. 6, Article ID e20949, 2011.
- [22] T. Simonson, "Electrostatics and dynamics of proteins," *Reports on Progress in Physics*, vol. 66, no. 5, pp. 737–787, 2003.
- [23] M. Shatsky, R. Nussinov, and H. J. Wolfson, "A method for simultaneous alignment of multiple protein structures," *Proteins: Structure, Function, and Bioinformatics*, vol. 56, no. 1, pp. 143–156, 2004.
- [24] J. Zahiri, J. Bozorgmehr, and A. Masoudi-Nejad, "Computational prediction of protein-protein interaction networks: algorithms and resources," *Current Genomics*, vol. 14, no. 6, pp. 397–414, 2013.
- [25] U. Ogmen, O. Keskin, A. S. Aytuna, R. Nussinov, and A. Gursoy, "Prism: protein interactions by structural matching," *Nucleic Acids Research*, vol. 33, pp. W331–W336, 2005.
- [26] N. Tuncbag, A. Gursoy, R. Nussinov, and O. Keskin, "Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using prism," *Nature Protocols*, vol. 6, no. 9, pp. 1341–1354, 2011.
- [27] Y. Ding and L. Gao, "Macrodynamics analysis of migration behaviors in large-scale mobile agent systems for the future internet," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 5, pp. 1032–1036, 2011.
- [28] K.-C. Chou, "Some remarks on protein attribute prediction and pseudo amino acid composition," *Journal of Theoretical Biology*, vol. 273, no. 1, pp. 236–247, 2011.
- [29] A. Ranganathan, A. Riabov, and O. Udrea, "Mashup-based information retrieval for domain experts," in *Proceedings of the ACM Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 711–720, Hong Kong, China, November 2009.
- [30] D. Merrill, *Mashups: The New Breed of Web App*, pp. 1–13, IBM Web Architecture Technical Library, 2006, http://scholar.google.com/scholar_lookup?hl=en&publication_year=2006&pages=1-13&journal=IBM+Web+Architecture+Technical+Library&author=Duane+Merrill&title=Mashups%3A+The+new+breed+of+Web+app.

- [31] G. Nachouki and M. Quafafou, "Mashup web data sources and services based on semantic queries," *Information Systems*, vol. 36, no. 2, pp. 151–173, 2011.
- [32] P. de Vrieze, L. Xu, A. Bouguettaya, J. Yang, and J. Chen, "Building enterprise mashups," *Future Generation Computer Systems*, vol. 27, no. 5, pp. 637–642, 2011.
- [33] Y Incorporation, Yahoo Pipes, <https://pipes.yahoo.com/>.
- [34] IBM. Corporation, IBM Mashup Center, <https://www.ibm.com/software/info/mashup-center/>.
- [35] J. Yu, B. Benatallah, F. Casati, and F. Daniel, "Understanding mashup development," *IEEE Internet Computing*, vol. 12, no. 5, pp. 44–52, 2008.
- [36] V. Hoyer and M. Fischer, "Market overview of enterprise mashup tools," *Service-Oriented Computing-ICSOC 2007*, vol. 5364, pp. 708–721, 2008.
- [37] M. Luck, P. McBurney, and C. Preist, *Agent Technology: Enabling Next Generation Computing (A Roadmap for Agent Based Computing)*, AgentLink, Louisville, USA, 2003.
- [38] N. R. Jennings, K Sycara, and M. Wooldridge, "A roadmap of agent research and development," *Autonomous Agents and Multi-Agent Systems*, vol. 1, no. 1, 1998.
- [39] M. El Fissaoui, A. Beni-hssane, S. Ouhmad, and K. El Makkaoui, "A survey on mobile agent itinerary planning for information fusion in wireless sensor networks," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 1323–1334, 2021.
- [40] M. B. Hasan and P. W. C. Prasad, "A review of security implications and possible solutions for mobile agents in e-commerce," in *Proceedings of the Innovative Technologies in Intelligent Systems and Industrial Applications, CITISIA 2009*, Kuala Lumpur, Malaysia, July 2009.
- [41] A. Kolioussis and J. Sventek, "A trustworthy mobile agent infrastructure for network management," in *Proceedings of the 2007 10th IFIP/IEEE International Symposium on Integrated Network Management*, pp. 383–390, USA, May 2007.
- [42] M. Yubao and D. Renyuan, "Mobile agent technology and its application in distributed data mining," in *Proceedings of the 2009 First International Workshop on Database Technology and Applications*, pp. 151–155, Wuhan, China, April 2009.
- [43] S. C. Banik, K. Watanabe, M. K. Habib, and K. Izumi, "An emotion-based task sharing approach for a cooperative multiagent robotic system," in *Proceedings of the Mechatronics and Automation*, pp. 77–82, Takamatsu, Japan, August 2008.
- [44] A. S. Gazafroudi, T. Pinto, F. Prieto-Castrillo et al., "Energy flexibility assessment of a multi agent-based smart home energy system," in *Proceedings of the Ubiquitous Wireless Broadband (ICUWB)*, Salamanca, Spain, September 2017.
- [45] G. S. Narula, "An approach for information extraction using jade: a case study," *Journal of Global Research in Computer Science*, vol. 4, no. 4, pp. 186–191, 2013.
- [46] F. Bergenti, A. Poggi, and M. Tomaiuolo, *Handbook of Research on ICTs for Human-Centered Healthcare and Social Care Services*, IGI Global, USA, pp. 549–567, 2013.
- [47] N. Benhajji, D. Roy, and D. Ancaux, "Patient-centered multi agent system for health care," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 710–714, 2015.
- [48] V. Julian and V. Botti, "Multi-agent systems," *Applied Sciences*, vol. 1402, 2019.
- [49] O. James and N. Marian, "The foundation for intelligent physical agents," 2003, <https://www.fipa.org/docs/input/f-in-00085/f-in-00085.pdf>.
- [50] M. Luck, P. McBurney, O. Shehory, and S. Willmott, *Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)*, AgentLink, Louisville, USA, 2005.
- [51] APSIG, Agent platform special interest group, <https://agent.omg.org/>.
- [52] N. Suri, J. M. Bradshaw, M. R. Breedy et al., "Nomads: toward a strong and safe mobile agent system," in *Proceedings of the Fourth International Conference on Autonomous Agents*, pp. 163–164, Barcelona Spain, June 2000.
- [53] N. Suri, J. M. Bradshaw, M. R. Breedy et al., "An overview of the nomads mobile agent system," in *Proceedings of the Workshop On Mobile Object Systems in association with the 14th European Conference on Object-Oriented Programming (ECOOP 2000)*, Cannes, France, June 2000.
- [54] F. M. T. Brazier, D. G. A. Mobach, B. J. Overeinder, S. van Splunter, M. van Steen, and N. Wijngaards, "Agent-escape: middleware, resource management, and services," in *Proceedings of the 3rd international SANE Conference*, pp. 1–3, Maastricht, The Netherlands, May 2002.
- [55] S. Willmott, J. Dale, B. Burg, P. Charlton, and P. O'Brien, *Agentcities: A Worldwide Open Agent Network*, Agentlink News, Louisville, USA, 2001.
- [56] D. B. Lange and O. Mitsuru, *Programming and Deploying Java Mobile Agents Aglets*, Addison-Wesley Longman Publishing Co., Inc., MA, USA, 1998, <https://en.wikipedia.org/wiki/Boston>.
- [57] G. Glass, "ObjectSpace voyager-the agent ORB for Java," in *Proceedings of the International Conference on Worldwide Computing and Its Applications*, pp. 38–55, Tsukuba, Japan, March 1998.
- [58] S. Galland, N. Gaud, S. Rodriguez, and V. Hilaire, *Janus: Another yet General-Purpose Multiagent Platform* Seventh AOSE Technical Forum, Paris, France, 2010.
- [59] N. P. Sudmann, *Tacoma-fundamental Abstractions Supporting Agent Computing in a Distributed Environment*, pp. 33–59, Department of Computer Science, University of Tromso, Norway, 1996.
- [60] D. Johansen, F. B. Schneider, and R. V. Renesse, "What tacoma taught us," *Mobility, Mobile Agents and Process Migration—An Edited Collection*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1999.
- [61] C. Bäumer and T. Magedanz, *Grasshopper? a mobile Agent Platform for Active Telecommunication Networks*, pp. 19–32, Springer International Workshop on Intelligent Agents for Telecommunication Applications, Paris, France, 1999.
- [62] F. L. Bellifemine, G. Caire, and D. Greenwood, *Developing Multi-Agent Systems with JADE*, Wiley, Hoboken, NY, USA, 2007.
- [63] Y. El-Gamal, K. El-Gazzar, and M. Saeb, "A comparative performance evaluation model of mobile agent versus remote method invocation for information retrieval," *Proceedings of World Academy of Science*, pp. 1–6, 2007.
- [64] M. Cossentino, S. Lopes, A. Nuzzo, G. Renda, and L. Sabatucci, "A comparison of the basic principles and behavioural aspects of akka, jacamo and jade development frameworks," in *Proceedings of the 19th Workshop from Objects to Agents (WOA)*, pp. 133–141, New Jersey, USA, April 2018.
- [65] R. H. Bordini and J. F. Hübner, "Bdi agent programming in agentspeak using jason," in *Proceedings of the International Workshop on Computational Logic in Multi-Agent Systems*, pp. 143–164, London, UK, June 2005.
- [66] J. P. Bigus, D. A. Schlosnagle, J. R. Pilgrim, W. N. Mills III, and Y. Diao, "Able: a toolkit for building multiagent autonomic systems," *IBM Systems Journal*, vol. 41, no. 3, pp. 350–371, 2002.

- [67] F. B. G. Caire, A. Poggi, and G. Rimassa, "Jade. a white paper," *Telecom Italia Lab*, vol. 3, pp. 1–14, 2003.
- [68] K. Miller, G. Mansingh, and G. Mansingh, "Comparing the use of mobile intelligent agents vs. client server approach in a distributed mobile health application," *Journal of Computers*, vol. 10, no. 6, pp. 365–373, 2015.
- [69] G. A. Aderounmu, B. O. Oyatokun, and M. O. Adigun, "Remote method invocation and mobil agent: a comparative analysis," *Issues in Informing Science and Information Technology*, vol. 3, 2006.
- [70] R. H. Bordini, L. Braubach, M. Dastani, A. E. F. Seghrouchni, and J. J. Gomez-Sanz, "A survey of programming languages and platforms for multi-agent systems," *Informatica*, vol. 30, no. 1, pp. 33–44, 2006.
- [71] K. Kravari and N. Bassiliades, "A survey of agent platforms," *The Journal of Artificial Societies and Social Simulation*, vol. 18, no. 1, Article ID 11, 2015.

Research Article

Data Mining Method under Model-Driven Architecture (MDA)

Jiangning Xie ^{1,2}, Feng Xu,² Zhen Li,³ and Xueqing Li ⁴

¹Graduate School, Shandong University, Jinan, China

²School of Management, Shandong University, Jinan, China

³Informatization Office of ShanDong University, Jinan, China

⁴School of Software, Shandong University, Jinan, China

Correspondence should be addressed to Jiangning Xie; xjn@sdu.edu.cn

Received 14 January 2022; Revised 9 February 2022; Accepted 10 February 2022; Published 22 March 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Jiangning Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of university information technology, how to mine and visually analyze the data of the existing separated information system will become an important research topic. The current university information system is a combination of some proprietary business systems characterized by poor data separation and storage and data analysis power. In addition, the data mining methods based on cloud computing will make customers gradually lose the ability to control the data. Because of the above problems, this paper proposes a university data mining method based on the MDA idea by constructing a data analysis and visualization framework, including multidimensional data modeling, data extraction, and data display based on visualization technology. The framework makes full use of the design idea of MDA and models multidimensional data, data extraction, and data display, respectively. The multidimensional data model module, data extraction module, and data visualization module provide efficient solutions for data analysis and visualization in universities.

1. Introduction

With the development of computer technology and network technology, various university business information management systems based on the network, including enrollment management, academic affairs management, graduation management, and financial management, have been widely used in universities. The scale and function are constantly expanding. With the continuous increase of the number of colleges and universities, information management more and more reflects its unique advantages. Many universities have gradually launched the student management information system. Functions cover student enrollment, training management, course performance management, graduation employment management, etc. Realize the standardization, informatization, and network of student management. At the same time, it has also accumulated a lot of information and data in its daily work. Also, it has the storage, backup, query, and simple statistical functions of massive data, but there are still the following problems:

First, a large amount of data is accumulated in each management stage of the system, but most data are separated into different system databases. Second, the management system realizes the management function of the data. Still, the data analysis function is relatively weak. The lack of multiangle analysis and data statistics is not enough to excavate the valuable information hidden in the massive data not to provide enough decision support for the school business managers. Third, the statistical analysis function in the management system is relatively simple, and the analysis results are primarily displayed in the form of reports and data tables, not intuitive enough. The present analysis is limited to the simple number of people, grades, courses, and so on, less to give the problems reflected by the data. It is difficult to recall the internal relationship between the data. Therefore, using statistical analysis and data mining methods to analyze the student management data from multiple angles and how to use the analysis results to provide accurate, intuitive, and good decision support for various management departments has become the focus and key of the current research.

The literatures [1, 2] proposed a data analysis method combining data mining for information management in colleges and universities, but just for a particular business problem to put forward a basic solution of ideas, lack of technical solutions for overall business data analysis. At the same time, the lack of data visual analysis function cannot provide intuitive visual analysis function. A solution that organically combines cloud computing technology and data mining is proposed in the literature [3, 4]. This solution can effectively solve the problem of the massive data and the limited computing power of the traditional data mining systems caused by the exponential growth of the data. The crow search algorithm has been successfully applied for the optimization of the data mining process, based on the characteristic of less parameter settings, easy implementation, and strong optimization capacity [5]. The performance of the system is tested through the mining and analysis of the electronic literature access log data set of college teachers and students in the library. The real-time performance and reliability of the system are verified.

For the problems of data analysis and data visualization, this paper comprehensively studies data mining [6] and data visualization technology [7, 8]. It proposes the analysis and visualization framework for college information data based on MDA [9] ideas. The framework makes full use of the design idea of MDA and models multidimensional data, data extraction, and data display, respectively. The multidimensional data model module, data extraction module, and data visualization module provide efficient solutions for data analysis and visualization in universities. Businesses and developers can quickly complete functional development for a thorough data analysis and visualization business through this framework.

2. The MDA-Based Analysis and Visualization Framework

For college students' data analysis and visualization problems, this paper designs and implements the data acquisition, analysis, and visualization framework based on MDA [10], which mainly includes three submodules of analysis data visualization modeling, data acquisition, and visualization display. The overall design structure diagram of the framework is shown in Figure 1.

In analyzing data and visualization, it is first necessary to visually model the relevant business data to build the data model. Then, data extraction and cleaning are done according to the established data model to obtain all the data sets to be displayed. Finally, the data is visually visualized and analyzed through the interactive defined display model.

2.1. Data Analysis Modeling. The data model is a typical performance of data structure. According to the needs of user-oriented, data models gradually establish different degrees of detail and refinement, which is an understanding of various degrees of abstraction in the real world. The data model in the traditional business processing system is a relational data model, which cannot effectively reflect the

structure and semantic information between the data. The primary purpose of a data analysis system is to analyze operations on a particular topic, which are called facts or measures. In contrast, the various angles of the analysis are called dimensions. Therefore, a multidimensional data model is used to model the data analysis and visualization systems in this paper. The system can complete a trend analysis, a continuous time subset of data sections, and quickly create a new view representing this section.

The establishment of the multidimensional data analysis model can integrate various kinds of data details and summarize the comprehensive information to meet the needs of the decision support system. However, the establishment of the model can be effectively organized in various parts, form complete and summary data for decision analysis, and provide strong decision analysis support for the leadership decision-making layer. The logical structure design of multidimensional data model is mainly the structure design of dimension table and fact table. For the logical definition of relational patterns, the patterns should be divided according to the current implemented topics to form multiple specific dimension tables and fact tables and determine the relationship patterns of each table.

Multidimensional data models organized through the dimensional table-fact table structure form of multidimensional phenotypes can be expressed in star mode, snowflake mode, or fact constellation pattern form. Multidimensional data models often regard the data as the form of a data cube, and the data cube is defined by dimension and facts. Dimension is about the perspective or entity that an organization wants to record and collects the same class of data. This paper adopts the star mode to ensure the performance of the data query and the easy understanding of the model. Its multidimensional data model will be established as follows:

Step 1: to determine the analysis topic, assuming that the analysis topic is to analyze the teacher's performance in the past year, which is also a fact or measure in the multidimensional data model.

Step 2: to determine the analysis dimension, including scientific research dimension, teaching dimension, guidance student dimension, post-level assessment dimension, and academic reputation dimension, in which each size can be divided according to the actual situation, into smaller dimensions, to slice the data cube and other operations.

Step 3: according to the analysis of the first two steps, a multidimensional data model of the star pattern representation can be obtained.

2.2. Data Procurement. For a specific business analysis topic, after completing the creation of the multidimensional data model, the data extraction next needs to obtain the necessary data to analyze and visualize the data from different data sources. To provide clean, complete, accurate, uninformative data for the data analysis process, improve the efficiency of the analysis process, and ensure the rapid generation of the

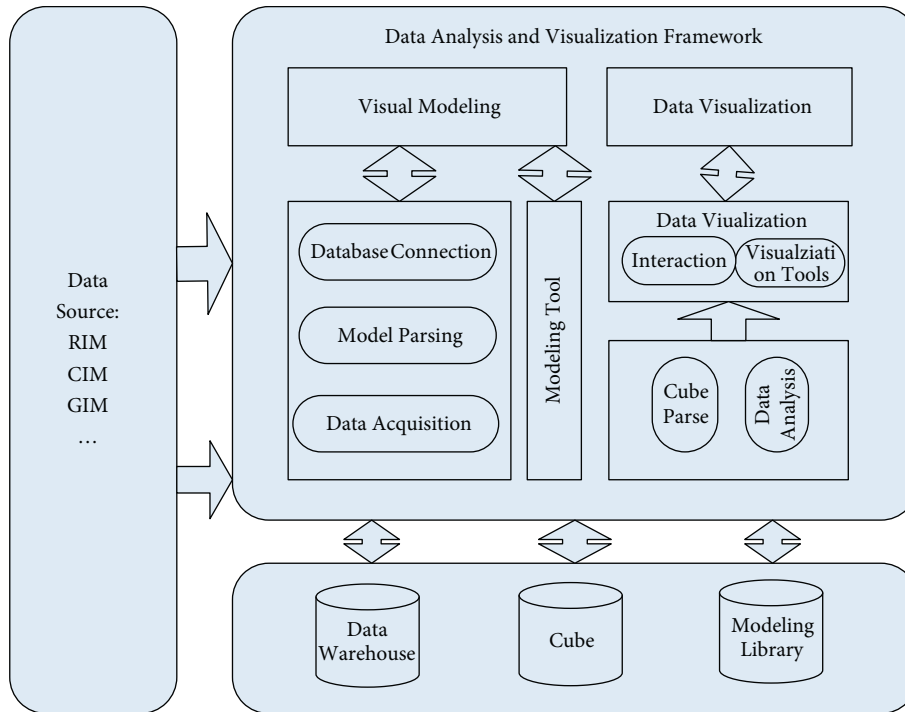


FIGURE 1: Data analysis and visualization framework.

presented results, this paper designs a method of data extraction, data conversion, and data loading from a data source to a cube (namely, the ETL process). The framework for the data acquisition is shown in Figure 2. In this framework, the data obtained from the source is not directly written to the cube. Instead, the data is preprocessed first, then data conversion and cleaning according to the correspondence between source and target data. Convert good data as intermediate data. The intermediate data is then stored into the cube after profound transformation.

As shown in Figure 2, the model divides the entire application system into three layers: data source layer, intermediate data layer, and multidimensional data layer. The data source layer can be divided into structured and unstructured data according to the data characteristics. This paper adopts different data acquisition methods for the above two different types of data. The paper uses traditional data extraction, data transformation, and data loading processes for structured data. This paper uses the interface-based design for unstructured data, using four typical interface methods: Web Service, intermediate library, file, TCP/UDP message transmission. Transparency to heterogeneous data acquisition is achieved by interface mode. Then, the data acquisition is completed in the transformation and loading mode of the heterogeneous data obtained in different ways.

2.3. Data Presentation. Data visualization is a theory, method, and technique for using computer graphics and image processing techniques to convert data into images or images displayed on the screen and perform interactive processing. Data visualization changes the traditional way of

showing data relationships through the relationship tables, allowing people to make more intuitive and efficient observations of the relationship between data. After completing the multidimensional data modeling and data acquisition process, this paper presents the data in a visual way. The general way of data display is to convert the obtained data into a vector chart or bitmap and display the bar chart pie chart on the vector chart or bitmap. For example, SVG generates a vector map for data amount visualization. The Scalable Vector Graphics (SVG) describes a vector drawing standard for two-dimensional vector graphs in the XML language, including rectangles, circles, and polygons. It has the advantages of high graphics quality, small files, and rich performance effect, but SVG plugin must be installed at the user's browser end, which inconveniences customer browsing.

A Flex technology-based data presentation model is chosen to improve the applicability of data presentation to different architecture designs, including B/S and C/S architectures. The data display can meet the display requirements of both B/S and C/S architecture and can meet users' needs. Through the data display mode of Flex technology, the overall data display process includes data transmission, the display mode of interactive processing, and data binding. The result data is obtained from the callback mechanism of front-end Flex communication with Java. During the data conversion between Flex and java, Flex implements the data transformation and binding via a binary AMF protocol. This paper provides a visual analysis and display of data to support high-dimensional data, including pie charts, bar charts, line charts, scatter plots [11], and parallel axes [10, 12].

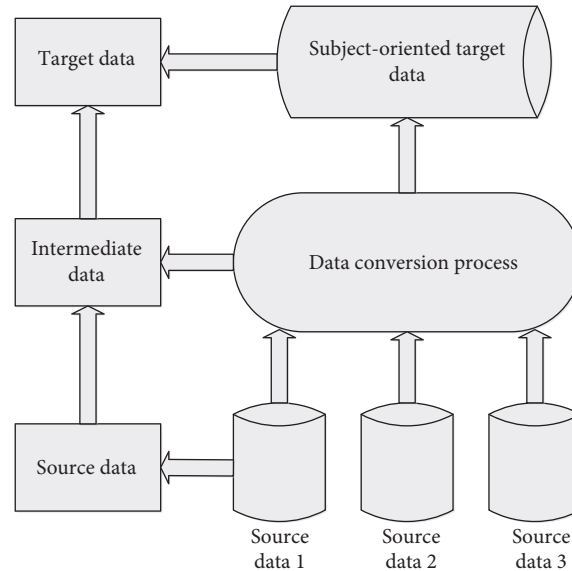


FIGURE 2: Data acquisition framework.

3. Example Analysis

Step 1. Determine the theme of the analysis. The theme is to analyze the school performance of the doctoral students in Shandong University to provide the school doctoral student training policy and establish an effective decision-making system. Mainly according to the daily routine of post-graduate students in school, the doctoral data is analyzed to predict the proportion of general, reasonable, and excellent academic performance. All administrative departments of the school can adopt appropriate policies to improve the research performance of doctoral students from the categories of general and good performance.

Step 2. First, obtain the behavioral data of the doctoral student performance, and then, according to the doctoral student data, determine the analysis dimension, including the scientific research dimension, the degree dimension, and the performance dimension, in which each dimension can be divided into smaller dimensions according to the actual situation. In the downward subdivided dimension, the results of various subjects to calculate the achievement dimension are used. In this example, the three main courses of doctoral students are used for analysis. The research dimensions consider published papers and influencing factors. Because of a degree dimension, it mainly evaluates the tutor evaluation results of the doctoral thesis, whether to delay defense or repeat defense.

Step 3. The model is designed from the analysis of the first two steps. In this example, the K-Means algorithm and the PCA algorithm are used to build a model framework for the cluster analysis of doctoral data. This model first clusters the doctoral student data through K-Means, after which each doctoral student corresponds to a class. Later, the doctoral student data is reduced through the PCA algorithm. The

results obtained from clustering are combined with the source data and analyzed using visualization techniques.

In this example, *Python* is a programming language compatible with many platforms that support both process-oriented and object-oriented programming and include various standard libraries. The data clustering analysis is mainly used for the sklearn library and visualized to the matplotlib library.

3.1. Data Procurement. In this example, we establish excellent data standards for school data and definition of metadata, data items, data classes, and datasets. In the data standards, 15 data categories, including departmental units, student management, teaching management, staff management, and scientific research management, are established, and each data class contains its respective subclass. The data interaction is related to huge amount of data of undergraduate schools, graduate schools, and colleges through the data standards. We extracted the daily behavior performance data with doctoral students.

Through the school system, we collect data on doctoral students' daily behavior and scientific research performance and use the traditional data extraction, data transformation, and data loading process for processing. Scientific research information, degree information, and results are all structured data.

The collected data are 3579 pieces of doctoral data. The doctoral degree types include general master and post-graduate degrees, general professional doctors, general doctors, and general direct degrees. The disciplines come from medicine, science, law, engineering, and other fields. In the college distribution, these doctors come from many colleges, including the School of Pharmacy, the Law School, the Business School, and the Institute of Economics. This example mainly considers information from 11 dimensions to analyze it.

The dimensions that assess doctoral performance are 11. These mainly include the following:

- (1) Degree information is divided into MidCheck, answer, comment1, comment2, comment3, commentNum. It represents the degree information in the doctoral thesis review, and the larger the number of comments made by the reviewing supervisor, the worse the doctoral student's performance.
- (2) Achievement dimension information is divided into Course1, course2, course3. They represent the results of three doctoral subjects. The higher the performance, the better the doctoral student performs.
- (3) The scientific research dimension information is as follows: DisserNum, impactFactor. It represents the sum of the number of published papers and the influence factors of the papers, respectively. The larger the number, the better the doctoral student's performance.

3.2. K-Means Model. We send the extracted doctoral data into the cluster model for clustering to obtain the label values for their categories when analyzing the doctoral data. We use the K-Means algorithm [13] for clustering statistics. The K-Means showed fast convergence, excellent clustering effect, and strong interpretable algorithm, so we used K-Means as a model for clustering. In the model, we randomly divided the data into K groups and selected K objects as the initial clustering center. The distance between each data point and each cluster center is calculated after being assigned to the nearest cluster center. When each data point is assigned to the corresponding data center, the clustered cluster center is recomputed based on the existing cluster data point distribution. This step is repeated continuously until all points are clustered. The clustering result has the smallest sum of error.

In the K-Means, the distance measure used is the square of the Euclidean distance:

$$\begin{aligned} d(x, y)^2 &= \sum_{i=1}^n (x_i - y_i)^2 \\ &= \|x - y\|_2^2, \end{aligned} \quad (1)$$

where x, y represent two different samples, and n represents the dimension of the sample. The problem with the Euclidean algorithm is that the sum of squares of error (SSE) in the cluster is minimized with the following formula:

$$SSE = \sum_{i=1}^n \sum_{j=1}^m w^{(i,j)} = \|x^{(i)} - \mu^{(j)}\|_2^2, \quad (2)$$

where $\mu^{(j)}$ represents the central point of the cluster j .

We determine the K cluster using the elbow method. As the number of clusters, K, increases, the sample division will be gradually detailed, the aggregation degree of each class of clusters will gradually increase, and the resulting sum of errors (SSE) will gradually decrease. When K is smaller than the real number of clusters, the SSE will drop greatly because

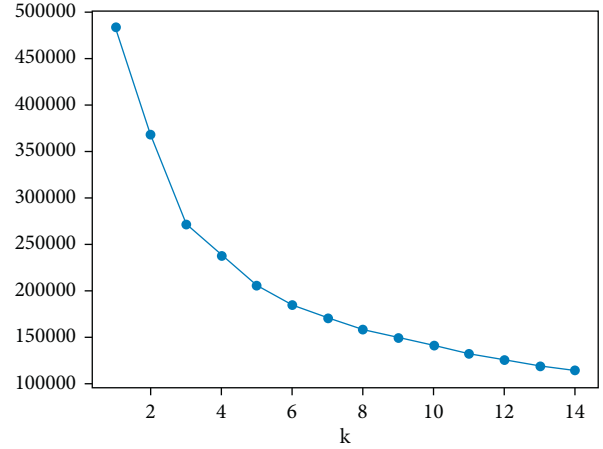


FIGURE 3: Elbow method image of doctoral student data.

the K greatly increases the aggregation of each cluster. When the K reaches the real cluster number, the aggregation obtained by the K will decrease rapidly, so the decline of SSE decreases sharply. Then, the curve will gradually flatten as the K value increases. The curve is similar to the elbow shape, with the corresponding K value as the true cluster number of the data.

Classification images of doctoral students are drawn by elbow method, as shown in Figure 3.

The horizontal axis of the image is K, and the vertical axis is SSE. The image shows that the inflection point is $k=3$ when the clustering class is 3.

After determining the K value, we set the K to 3 and send it to the K-Means model for clustering operations, in order to obtain the best clustering data.

The algorithm steps for K-Means are in Algorithms 1 and 2.

α_j First, the initialized k samples are selected as the initial cluster center $\alpha = \alpha_1, \alpha_2, \dots, \alpha_k$.

- (1) For each sample x_i in the dataset, its distance to the k cluster center is calculated and divided into the class corresponding to the cluster center with the smallest distance.
- (2) For each class α_j , recalculate its cluster center $\alpha_j = 1/|c_j| \sum_{x \in c_j} x$ (the center of mass of all samples belonging to the class) for each class;
- (3) Repeat the above 1 and 2 steps until some abort condition (number of iterations, minimum error change, etc.) is reached.

After completing the multidimensional data modeling and data acquisition process, we obtained 3579 doctoral data sample labels and clustering results of the data, followed by dimensionality reduction and presentation of the data in a visual manner.

3.3. PCA Dimension Reduction. The main idea of PCA [14] is to map n -dimensional features to k -dimensions. This k -dimensional feature is an entirely new orthogonal feature, called the principal component, and is a k -dimensional

```

n × mArrn×m Input: dimension array, iteration t
Output: The label value Labeln
Begin
  Automatic_Random_Generate() pointk
  while(t)
    for(int i = 0; i < n; i++)
      for(int j = 0; j < k; j++)
        Calculate_Distance() Arripointj
    for(int i = 0; i < k; i++)
      Find_All_Data_Points_belong_Cluster() Arrn×mpointk
      Modify_Coordinate_to_Center_Coo_Points() Arrn×mpointk
  End

```

ALGORITHM 1: K-Means clustering.

feature reconstructed based on the original n -dimensional feature. PCA's work is to find a set of mutually orthogonal axes from the original space sequentially, and the selection of the new axes is closely related to the data itself. The first new axis selection is the direction of the largest variance in the original data. The second new axis selection is the plane orthogonal to the first axis and the largest in the plane orthogonal to the 1 and 2 axes. By this, in turn, n such axes can be obtained. With the new axes obtained this way, we find that most of the variances are contained in the preceding k axes, and the latter axis contains a variance of almost 0. Therefore, we can ignore the remaining axes and retain only the first k axes containing most of the variance. We retain only the dimensionality features containing most of the variance in the doctoral data, while ignoring the feature dimension containing the variance of almost 0, achieving the dimensionality reduction of the data features.

PCA algorithm based on eigenvalue decomposition covariance matrix is in Algorithm 2.

We transform the doctoral data into m n -dimensional vectors $X_{m \times n}$ for and performed zero-mean value (the average of this column) for it. The covariance matrix C is found, which further obtains its eigenvalue and eigenvector A . The eigenvectors λ form the matrix with the corresponding eigenvalues from large to small and then take the first k to form the matrix P . $Y=PM$ is a matrix of a new k -dimensional size.

By calculating the covariance matrix of the data matrix, the eigenvector of the eigenvalues of the covariance matrix is obtained, and the matrix composed of the eigenvectors corresponding to the k features with the largest eigenvalue is selected. Transfer data matrix into a new space to achieve dimensionality reduction of data features. The covariance is

$$\begin{aligned} \text{COV}(X, Y) &= E[(X - E(Y))(Y - E(Y))] \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \end{aligned} \quad (3)$$

For the n -dimensional matrix, the covariance matrix is

$$C = (c_{ij})_{n \times m} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nm} \end{bmatrix}, \quad (4)$$

$c_{ij} = \text{Cov}(X_i, X_j)$, where $i, j = 1, 2, 3, \dots, n$.

The divergence matrix is defined as

$$S = \sum_{k=1}^n (x_k - m)(x_k - m)^T, \quad (5)$$

where m is the average vector: $m = 1/n \sum_{k=1}^n x_k$.

Vector v is the eigenvector of matrix A and can be expressed in the following form:

$$Av = \lambda v. \quad (6)$$

Among them, λ is the eigenvalue corresponding to the eigenvector v , and a set of eigenvectors of the matrix is a set of orthogonal vectors.

For matrix A , there is a set of eigenvectors v , which are orthogonalized to obtain a set of orthogonal unit vectors. Eigenvalue decomposition decomposes the matrix A into the following formula:

$$A = Q \sum Q^{-1}. \quad (7)$$

Among them, Q is a matrix composed of the eigenvectors of the matrix A , and \sum is a diagonal array, and the elements on the diagonal are the eigenvalues.

In this example, we use the PCA algorithm and divide it into two dimension reduction methods:

- (1) We reduce the degree information midCheck, answer, comment1, comment2, comment3, commentNum and research dimension disserNum, impactFactor to 1 dimension information, representing the degree level and research level of doctoral students. The achievement dimension information course1, course2, and course3 are reduced to 1-dimension information, which indicates the performance level. After the dimensionality reduction

```

Input:  $m$   $n$ -dimensional vectors  $X_{m \times n} = \{x_1, x_2, x_3, \dots, x_m\}$ 
Output:  $k$   $n$ -dimensional vectors  $Y_{k \times n} = \{y_1, y_2, y_3, \dots, y_k\}$ 
begin
  for( $i = 0; i < n; i++$ )
     $M_i = \text{Minus\_Average } x_i$ 
   $C = 1/mMM^T$ 
   $\lambda, A = \text{Solve\_covariance}(C)$ 
   $P = \text{Array\_Dwindle}(A[1:k])$ 
   $Y = PM$ 
end

```

ALGORITHM 2: PCA algorithm.

through the PCA algorithm model, we reduce the 11-dimensional intake to 2.

- (2) We reduce the degree information midCheck, answer, comment1, comment2, comment3, and commentNum to 1-dimension information representing the degree level of a doctoral student. The achievement dimension information course1, course2, and course3 are reduced to 1-dimension information, which indicates the performance level. Reduce the scientific research dimension disserNum and impactFactor to 1-dimensional details, representing the scientific research level. After the dimensionality reduction through the PCA algorithm model, we reduce the 11-dimensional intake to 3.

After the doctoral data underwent PCA dimension reduction, we combined the source data for analysis and presented them with visualization techniques.

3.4. Data Display. After the model operation, we can get the picture effect of the doctoral student data clustering. The 2D clustering effect is shown below, and the total number of doctoral students is 3579, including 2052 in class 0 (blue), 1179 in class 1 (green), and 348 in class 2 (red). The drawing invokes the matplotlib graphics library for the python programming language. matplotlib is a library dedicated to developing 2D charts and 3D charts, realizing data visualization gradually and interactively, strong control over image elements, and output multiple formats including PNG, PDF, SVG, and EPS, as shown in Figure 4.

From the results of data clustering image classification, combined with the analysis of Tables 1 and 2, compared with the mean value, class 0 (blue) doctoral students perform the best, and their scientific research, degree, and performance are relatively excellent. The number of papers (disserNum) and paper impact factors (impactFactor) exceeded the other two doctoral students. The impact factors are nearly five times more than the two different categories. Among the achievement items, class 0 students had the highest grades in course 1 and course 3, with course 2 at the middle level.

There are some problems in class 2 (red) doctoral students—their scientific research ability performance in general. The number of papers and paper influence factors is relatively small, but their performance is good, the best in

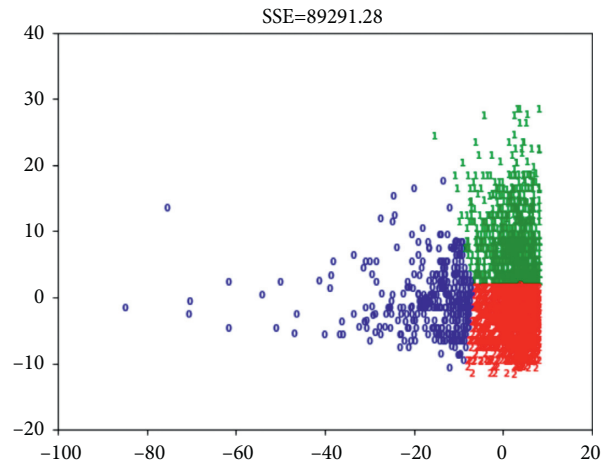


FIGURE 4: Doctoral cluster image of doctoral students, horizontal axis represents research degree information and vertical axis represents achievement information.

course2 and course3. Their degree performance is in the middle level. Class 1 (green) doctoral students have the least number of scientific research papers. Still, the impact of the papers is relatively high, with the worst performance, and their degree performance is in the middle level.

Based on the above data, we continue to explore the classification results of $k=4$ with $k=5$. After data analysis, we found that the subdivided types of doctoral students are more detailed in terms of scientific research, degree, and performance.

In the $k=4$ classification, 9 doctoral students with extra categories of 2 (red part) in Figure 5(a) have better degree information, good grades, the best number of published papers, and the highest impact factors, because fused Ph.D. data from class 0 in $k=3$. In the classification of $k=5$, according to the source data analysis, the influence factors of the doctoral data with class 4 in Figure 5(b) papers are significant, but their academic performance is low.

After collecting the above data analysis results, we further explored that we standardize the doctoral data and processed the standardized data through K-Means and PCA. Then, two-dimensional images are then obtained, as shown in Figure 6. From the data distribution perspective, the data distribution after the standardization is somewhat more uniform.

TABLE 1: PhD student data $K=3$ cluster research and achievement mean.

| Classify | Disser Num | Impact factor | Course 1 | Course 2 | Course 3 |
|----------|------------|---------------|----------|----------|----------|
| 0 class | 4.6041 | 24.2042 | 89.7470 | 88.5654 | 87.2619 |
| 1 class | 2.0813 | 5.4582 | 88.1742 | 80.8182 | 86.1568 |
| 2 class | 2.2335 | 5.0030 | 88.3930 | 91.9798 | 88.1305 |

TABLE 2: PhD student data $K=3$ cluster degree information mean.

| Classify | MidCheck | Answer | Comment 1 | Comment 2 | Comment 3 | Comment Num |
|----------|----------|--------|-----------|-----------|-----------|-------------|
| 0 class | 1.2321 | 2.7351 | 1.9375 | 1.7202 | 1.3184 | 7.9315 |
| 1 class | 1.0871 | 2.4655 | 1.9145 | 1.6506 | 1.2663 | 7.7668 |
| 2 class | 1.1133 | 2.4170 | 1.9156 | 1.6496 | 1.2654 | 7.7894 |

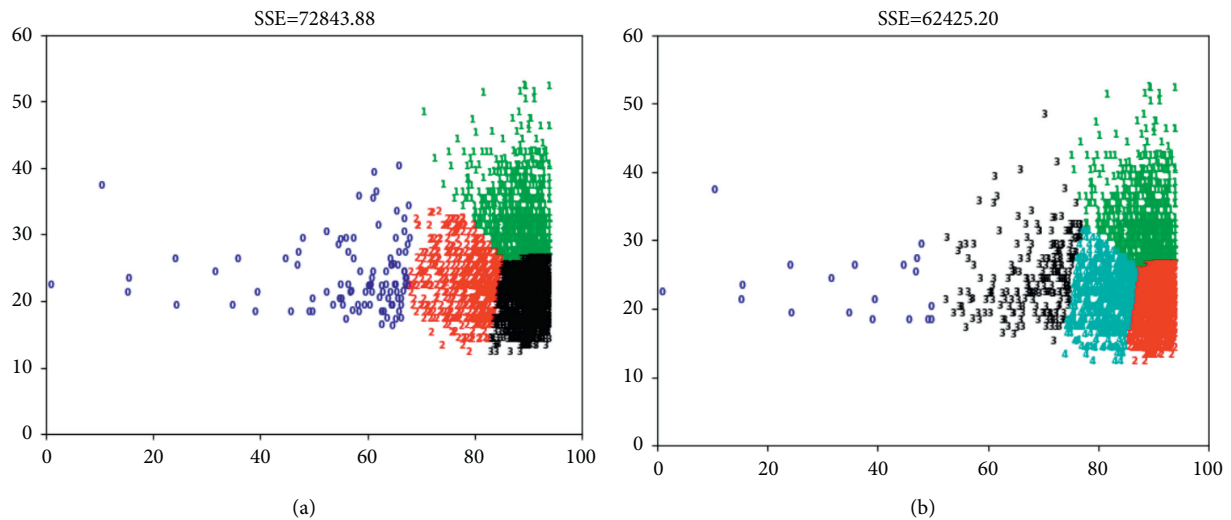


FIGURE 5: (a) PhD student cluster images, $k=4$. The horizontal axis represents scientific research degree information, and the vertical axis represents achievement information. (b) PhD student cluster images, $k=5$. The horizontal axis represents the scientific research degree information, and the vertical axis represents the achievement information.

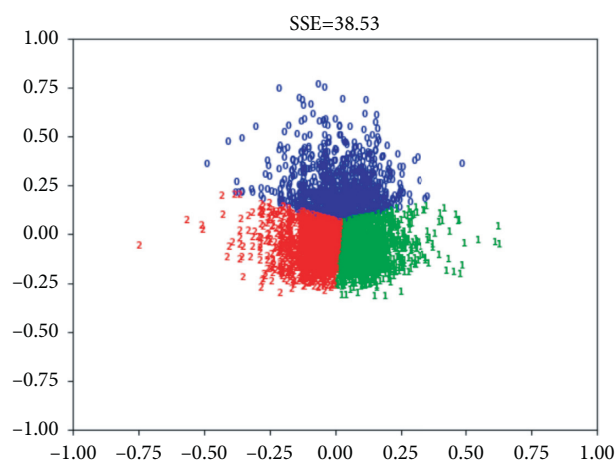


FIGURE 6: Cluster images of doctoral students after standardization: horizontal axis represents scientific research degree information, and vertical axis represents achievement information.

We analyze the data and averaged the clustering results according to the data clustering results. It can be found from Table 3 that class 0 (blue) doctoral students are at the middle

level. Their grades are at the downstream level. Class 1 (green) doctoral students have the best scientific research, degree, and performance. These kinds of students are

TABLE 3: Data mean values after normalization.

| Average value | Scientific research degree | Mark |
|---------------|----------------------------|-------------|
| 0 class | 0.350780663 | 0.654025096 |
| 1 class | 0.407664049 | 0.784348448 |
| 2 class | 0.302289483 | 0.773064877 |

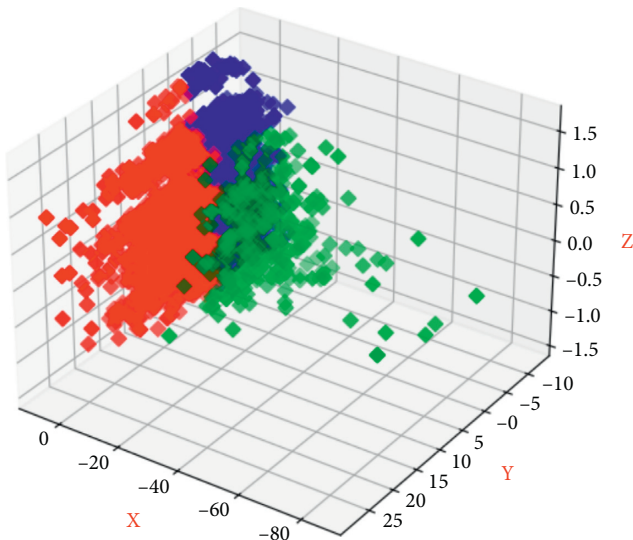


FIGURE 7: 3 *D* doctoral student cluster image: *x*-axis represents scientific research information, *y*-axis represents achievement information, and *z*-axis represents degree information.

relatively excellent. Class 2 (red) has a low scientific research and degree life, but their grades are in the middle level.

We further expand based on the analysis of the above 2 *D* data. When we use PCA to reduce 11 *D* doctoral information to 3 *D*, we can build 3 *D* graphics, as shown in Figure 7.

From the analysis of data clustering results combined with doctoral performance data, we see that green doctoral students have the best performance. Their scientific research, degree, and performance are relatively excellent. Red doctoral students have general scientific research ability. Their performance is the worst, and their degree is middle. Blue doctoral students have general scientific research ability, good performance, and degree performance.

From the above analysis, it is learned that appropriate data analysis applied to doctoral student performance can effectively extract effective information from a large amount of data. It can be used for the school management decision-making process. Statistical analysis of doctoral data helps provide adaptive learning guidance. Preserving the students' performance and behavior in advance will help the relevant departments of the school to take appropriate adjustment measures to cultivate talents better and build a talent system.

4. Conclusion

Data analysis and visualization technology is an emerging and promising research field in university information management. With the development of university information technology, how to mine and visually analyze the

data of the existing separated information system will become an important research topic.

This paper presents a complete set of solutions based on MDA ideas for analyzing and visualizing information data in universities. The proposed framework includes multidimensional data modeling and analysis modules, data extraction and cleaning modules, and data display modules based on data visualization techniques.

Through this framework, the business analysis and developers can quickly conduct the data visualization business's modeling analysis and programming implementation. In the existing framework, the interface is provided for data mining analysis. Moreover, the data mining algorithms and multidimensional data visualization techniques will be deeply studied and applied to the existing data analysis and visualization frameworks in future work.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- [1] Y. Xu and D. Xiong, "Research on the information construction of college student management in the era of big data," *Journal of Lanzhou Education College*, vol. 34, no. 1, p. 3, 2018.
- [2] Q. Zhang and F. Rui, "Application of data mining in university information management," *Data Mining*, vol. 9, no. 1, p. 7, 2019.
- [3] R. Zhong and H. Wang, "Specific data query technology in the university cloud computing management system based on data mining," *Modern Electronic technology*, vol. 41, no. 2, p. 3, 2018.
- [4] G. Shen, "Application of big data analysis in smart education in universities," *Modern Electronics Technology*, vol. 42, no. 4, pp. 105–108, 2019.
- [5] C. Qu and Y. Fu, "Crow search algorithm based on neighborhood search of non-inferior solution set," *IEEE Access*, vol. 7, 2019.
- [6] A. R. Raut and S. P. Khandait, "Review on data mining techniques in wireless sensor networks," in *Proceedings of the IEEE Sponsored 2nd International Conference on Electronics and Communication System (ICECS 2015)*, February 2015.
- [7] D. M. Schug, P. H. Taylor, S. Iudicello, and J. H. Swasey, "Using online data visualization and analysis to facilitate public involvement in management of catch share programs," *Marine Policy*, vol. 122, 2020.

- [8] G. Ralitz, B. Eugenia, S. Patricia et al., "Data visualization tools of tobacco product use patterns, transitions and sex differences in the PATH youth data," *Nicotine & Tobacco Research*, vol. 22, no. 10, pp. 1901–1908, 2020.
- [9] B.-K Park and W.-S. Jang, "MDA(Model driven architecture)," *Journal of Platform Technology*, vol. 7, 2019.
- [10] L. Lu, W. Wang, and Z. Tan, "Double-arc parallel coordinates and its axes re-ordering methods," *Mobile Networks and Applications*, vol. 25, no. 4, pp. 1376–1391, 2020.
- [11] S. Cao, Y. Zeng, S. Yang, and S. Cao, "Research on Python data visualization technology," *Journal of Physics: Conference Series*, vol. 1757, no. 1, pp. 012122–012128, 2021.
- [12] G. Richer, J. Sansen, F. Lalanne, D. Auber, and R. Bourqui, "HiePaCo: scalable hierarchical exploration in abstract parallel coordinates under budget constraints," *Big Data Research*, vol. 17, no. 8, 2019.
- [13] J. Li, S. Xu, Wan Can, Y. Lu, and S. Wang, "Analysis of power load characteristics based on the adaptive k-means + algorithm," *China Southern Power Grid Technology*, vol. 13, no. 2, p. 7, 2019.
- [14] J. Xie, X. Li, L. Wang, and Y. Niu, "A MDA-based campus data analysis and visualization framework," in *Proceedings of the ETCS '11: Proceedings of the 2011 Third International Workshop on Education Technology and Computer Science - Volume International Journal of Education and Management Engineering*, vol. 2, no. 10, Washington, D C USA, March 2011.

Review Article

Research Contribution and Comprehensive Review towards the Semantic Segmentation of Aerial Images Using Deep Learning Techniques

P. Anilkumar  and P. Venugopal 

School of Electronics Engineering, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India

Correspondence should be addressed to P. Venugopal; venugopal.p@vit.ac.in

Received 22 December 2021; Revised 19 January 2022; Accepted 31 January 2022; Published 20 March 2022

Academic Editor: Mamoun Alazab

Copyright © 2022 P. Anilkumar and P. Venugopal. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Semantic segmentation is a significant research topic for decades and has been employed in several applications. In recent years, semantic segmentation has been focused on different deep learning approaches in the area of computer vision, which has aimed for getting superior efficiency while analyzing the aerial and remote-sensing images. The main aim of this review is to provide a clear algorithmic categorization and analysis of the diverse contribution of semantic segmentation of aerial images and expects to give the comprehensive details associated with the recent developments. In addition, the emerged deep learning methods demonstrated much improved performance measures on several public datasets and incredible efforts have been dedicated to advancing pixel-level accuracy. Hence, the analysis on diverse datasets of each contribution is studied, and also, the best performance measures achieved by the existing semantic segmentation models are evaluated. Thus, this survey can facilitate researchers in understanding the development of semantic segmentation in a shorter time, simplify understanding of its latest advancements, research gaps, and challenges to be used as a reference for developing the new semantic image segmentation models in the future.

1. Introduction

Semantic segmentation is an image analysis task, which assigns a label for each pixel in input images for describing the class of its encircled region [1]. Semantic segmentation of aerial images represents the assignment of one land cover category to each pixel, which is a complex task owing to the huge variations in the appearances of ground objects. Several works have been presented in recent years [2]. The state-of-the-art approaches in semantic segmentation are focused on the hand-crafted features, which fail to get the satisfactory performances and are restricted through the depiction ability of features [3]. When compared with object detection and image classification, semantic segmentation is used as the highest level of the image analysis process, which permits complete scene information of the complete input image [4]. In several remote-sensing tasks, semantic segmentation is

considered as pixel-wise classification [5]. Semantic segmentation of aerial imagery has been employed in diverse applications such as hazard identification and avoidance, traffic management and evaluation, and urban area planning and monitoring [6]. However, the growth of semantic segmentation techniques was stopped years ago due to the lower accuracy rate of existing image analysis methods focused on the extraction of hand-crafted features [7].

Aerial and satellite imagery have been utilized in different applications such as regional planning, cartography, landscaping, and agriculture [8]. In 2020, Maddikunta et al. [9] have focused on applications, requirements, and challenges of UAV images which were captured from UAV vehicles for smart agriculture system. Multirotor UAVs are usually used for airborne surveillance, photography, and other similar tasks. These are the simplest to produce and the least expensive of all types of UAVs. These images have

different visible colors and other spectra. There is also elevation imagery, which is generally prepared through light detection and ranging (LiDAR) and radar images [10]. Moreover, along with the emergence of satellite and aerial images, remote sensing is also implemented. Remotesensing images are gathered from the remote object through a device, which cannot be physically contacted the object [11]. In recent years, the data analysis and interpretation are still performed by human experts. Although, semantic segmentation offers superior abilities in object detection, it suffers from implementing it into the real use cases [12]. In 2020, Ch et al. [13] have suggested the security and privacy of UAV data using blockchain technology. The value of virtual circuit (VC)-based devices—UAVs, drones, and similar other IoT-based devices—has grown tremendously in recent years. These gadgets are mostly utilized for aerial surveying in sensitive and isolated locations. The object detection in aerial images is complex due to the bird’s-eye view of aerial images, which have huge variations in orientation, high nonuniform object densities, large aspect ratios, and scale variations of objects. Moreover, several challenges are presented in the detection of objects using aerial images, which are low GPU memory capacity, downsampling a large image, and lack of inference on large images [14]. In aerial images, several sensor and resolution are considered as the factors for producing the dataset biases [15]. The standard dataset is prepared by collecting the images from different platforms and sensors through several resolutions including aerial images, satellite images, Gaofen-2 (GF-2) Satellite, and Google Earth [16].

Currently, many DL applications are being used all over the world. Healthcare, social network analysis, audio and speech processing (such as recognition and enhancement), visual data processing methods (such as multimedia data analysis and computer vision), and NLP (translation and sentence classification) are examples of these applications. These applications are divided into five groups: classification, localization, detection, segmentation, and registration. Although each of these jobs has its own aim, as seen in Figure 1, there is significant overlap in the pipeline implementation of these applications.

The semantic segmentation is adopted by deep learning approaches in recent years, which has attained high efficiency in diverse conventional computer vision applications and consists of detection and classification of objects and semantic segmentation [15]. These approaches have automatically derived features, which are customized for classification tasks that create these approaches to offer suitable options for managing complex cases [17]. The huge achievement in other fields makes the extension and adoption of deep learning approaches for solving the challenges in remote-sensing fields. Although, deep learning offers noteworthy performance, it suffers from allocating significant labels to the components of remote-sensing image [18]. Due to the large number and enormous quantity of modalities of the remote-sensing data, the deep neural network has been facilitated for feature extraction [19]. It has also offered great benefits to practitioners and researchers, which require less programming intensive tools for high-

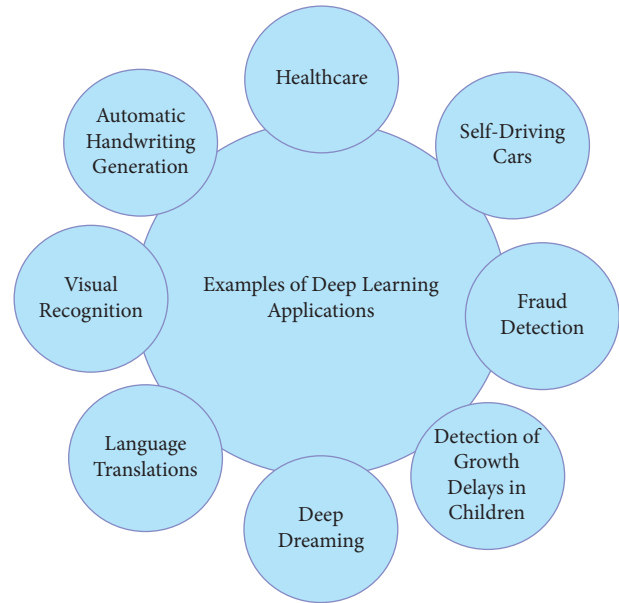


FIGURE 1: Examples of DL applications.

level data analysis and are understandable in geosciences [20]. In 2021, Kumar et al. [21] have given a secured privacy preserving framework for smart agriculture unmanned aerial vehicles for both blockchain and nonblockchain frameworks. Balamurugan et al. [22] have given a direction-of-arrival (DOA) tracking for seamless connectivity in beamformed IoT-based drones, and their communication and beamformed performances were increased.

The primary and significant deep learning approach consists of restricted Boltzmann machines, autoencoders, and convolutional neural networks (CNNs) which have focused on understanding the satellite imagery or aerial imagery [23]. Hence, this study has reviewed several semantic segmentation models with diverse deep learning algorithms for future works.

The major contribution of this survey is (i) to design a detailed survey on existing semantic segmentation models on diverse imaging modalities in recent years by gathering the noteworthy information from each and every semantic segmentation model along with diverse algorithms on machine learning and deep learning, (ii) to present a comprehensive study about datasets, simulation platforms, chronological review, performance metrics, features, and challenges of the conventional semantic segmentation models and their algorithms focused, and (iii) to give the appropriate research gap with the limitations present in existing semantic segmentation systems for motivating the researchers to design a new semantic segmentation model.

The remaining sections of this survey are depicted here. Section 2 discusses the literature review on state-of-the-art semantic segmentation models. Section 3 presents the algorithmic categorization and features and challenges of existing semantic segmentation models. Section 4 describes the simulation platforms and dataset description for conventional semantic segmentation models. Section 5 demonstrates the performance measures and best accuracy rate

attained by the conventional semantic segmentation models. Section 6 gives the research gaps and challenges. Section 7 concludes this survey.

2. Literature Review on State-of-the-Art Semantic Segmentation Models

2.1. Literature Survey. In 2015, Saito et al. [24] have utilized CNN for training the pixel labeling to get the extracted building areas for determining the semantic segmentation of aerial images. Then, they have used Dijkstra's algorithm for discovering the optimal seam line to get shortest path on the map. In 2016, Marmanis et al. [25] have described the semantic segmentation model using high-resolution aerial images and using ENSEMBLE OF CNNs named FCN and modified CNNs to show the superior efficiency on standard dataset. In 2017, Holliday et al. [26] have addressed the semantic segmentation model by applying the model compression techniques for getting the superior segmentation accuracy, which has also used ConvNet to determine the significance of segmentation.

In 2018, Chen et al. [27] have suggested shuffling CNNs for realizing the aerial images for semantic segmentation in a periodic way, which has also proposed a field-of-view improvement for improving the predictions. This model has attained effective and promising results for two datasets. In 2018, Yu et al. [28] have designed an end-to-end scheme for semantically segmenting the high-resolution aerial images by considering the CNN structure with pyramid pooling phase for extracting the feature maps at diverse scales. In 2018, Chen et al. [29] have presented the digital surface models (DSMs). They have presented the deeply supervised shuffling convolutional neural network (DSCNN) for efficient upsampling of feature maps, and furthermore, the multiscale features were attained. In 2018, Volpia and Tuia [30] have suggested a semantic segmentation model using aerial images for learning the shallow-to-deep visual features, semantic boundaries across classes, and semantic class likelihoods through a multitask CNN. Here, the top-down and bottom-up information were combined and encoded with a conditional random field model. In 2018, Sun et al. [31] have implemented a new semantic segmentation model from LIDAR data and high-resolution aerial images through a multifilter CNN for offering multiresolution segmentation. It has also delineated the object boundaries to reduce the salt and pepper artifacts. In 2018, Kemker et al. [32] have designed a semantic segmentation method using DCNNs from multispectral remote-sensing images for getting the efficient performance on RIT-18 dataset. In 2018, Marmanis et al. [33] have designed a semantic segmentation model from high-resolution aerial images by applying DCNN for representing and extracting the boundaries among the regions of diverse semantic classes. In 2018, Vo and Woong [34] have designed a semantic segmentation method through investigating the effects of deep network and cascaded framework of dilated convolutions, which has improved the localization efficiency. This model has trained efficiently.

In 2019, Peng et al. [35] have presented a new architecture by combining the "dense connection and fully convolutional networks (FCN)" for providing the fine-grained semantic segmented maps for remote-sensing images. The suggested model has achieved the traditional efficiency on two datasets without any postprocessing and pretraining. In 2019, Luo et al. [36] have proposed a new deep FCN with channel attention mechanism (CAM-DFCN) for semantic segmentation using high-resolution aerial images, which has included encoder-decoder architecture. The integration of multilevel feature maps has also facilitated. It has also offered accurate segmentation for offering spatial location information and weight semantic information. In 2019, Li et al. [37] have designed a road segmentation system with the combination of "adversarial networks with multiscale context aggregation." This study has focused on extracting the road by utilising the UAV remote-sensing images. This model has used morphological techniques for getting the results with the elimination of small independent patches. In 2019, Azimi et al. [38] have designed a symmetric FCN improved with wavelet transform for doing the segmentation of lane marking from aerial imagery. This model has used a customized loss function for improving the accuracy of pixel-wise localization. In 2019, Wang et al. [39] have designed a semantic segmentation from UAV-taken images for generating the defect detection outcomes through applying matrix operations with segment connection technique for connecting the segment features of objects. It has also used an artificial contour segment feature generator with a background filter which was used for line accessory detection that has enhanced the detection efficiency. In 2019, Cao et al. [40] have suggested a digital surface fusion models (DSMF) for improving the semantic segmentation results along with four end-to-end networks named DSMFNets to get the overall accuracy on segmenting the high-resolution aerial images. In 2019, Nguyen et al. [41] have suggested a MAVNet for semantic segmentation with the use of deep neural network on microaerial vehicles (MAVs). It has demonstrated the superior efficiency on standard datasets. In 2019, Guo et al. [42] have integrated the super-resolution approaches for improving the segmentation efficiency using "efficient subpixel convolutional neural network (ESPCN) and UNet" using remote-sensing imagery. It has significantly attained more precise and high accurate segmentation results. In 2019, Igonina and Tiumentseva [43] have focused on identifying the known neuroarchitectures to solve the problems persists in remote sensing of Earth's surface, which has also focused on semantic segmentation of UAV images. In 2019, Wu et al. [44] have studied attention dilation-linknet (AD-linknet) neural network by adopting the encoder-decoder framework along with pretrained encoder, channel-wise attention scheme, and serial-parallel integrated dilated convolution for semantic segmentation of high-resolution satellite images. In 2019, Masouleh and Shah-Hosseini [45] have presented a Gaussian-Bernoulli restricted Boltzmann machine (GB-RBM) for the semantic segmentation of UAV-based thermal infrared images, which has evaluated the efficiency on average processing time and average precision concerning with

the extraction of ground vehicles in road. In 2019, Audebert et al. [46] have introduced a regression-based semantic segmentation regularization model through a distance transform, in which the FCN was trained for both continuous and discrete spaces through learning the distance regression and joint classification. In 2019, Mohammadi et al. [47] have implemented a semantic segmentation model from polarimetric synthetic aperture radar images using FCN architecture, which has extracted the discriminative polarimetric features for finding the wetland on complex land cover ecosystem. In 2019, Hua et al. [48] have presented a CNN for processing the extracted features for enhancing the efficiency of semantic segmentation of aerial images, which has used two modules such as patch attention module and attention embedding module for getting the significant information of low level features. In 2019, Panboonyuen et al. [49] have designed a global convolutional network (GCN) for semantic segmentation of remotely sensed images for extracting the multiscale features from diverse phases of the network.

In 2020, Liu et al. [50] have proposed a semantic segmentation model for high-resolution remote-sensing images using a multichannel segmentation network termed DAPN that has completely extracted the multiscale features of the images and retained the spatial features of the object. In 2020, Mou et al. [51] have considered two efficient networks called channel and spatial relation module for learning and reasoning about the global correlations among the feature maps or positions. The suggested model was termed as relation module-equipped FCN. In 2020, Wang et al. [52] have designed a “context and semantic enhanced high-resolution network (CSE-HRNet)” with two comprehensive processes for tackling the intraclass heterogeneity problem and for enhancing the representational ability of multiscale contexts. In 2020, Martinez-Soltero et al. [53] have utilized CNN for terrain detection using aerial images, which has aimed for solving the navigation tasks and robot mapping along with the pixel-level segmentation for generating a high detailed map. In 2020, Jiawe et al. [54] have proposed a real-time semantic segmentation model by designing a new “asymmetric depth-wise separable convolution network (ADSCNet)” for offering the better prediction efficiency. In 2020, Deng et al. [55] have developed a semantic segmentation network from UAV images for real-time weed mapping for reducing the time gap among the herbicide treatment and image collection. This model has focused on implementing a hardware system with combined processes. In 2020, Niu et al. [56] have designed a new “hybrid multiple attention network (HMANET)” for adaptive capturing of global relationships, which has computed the category-based relationship and recalibrated the class level details. This study has introduced an efficient region shuffle attention (RSA) module for enhancing the effectiveness of semantic segmentation. In 2020, Chai et al. [57] have proposed the semantic segmentation model from high-resolution aerial images that has addressed the problem of learning spatial context through Deep CNNs (DCNNs). This model has predicted the distance map rather than the score map for every class that has enhanced the segmentation efficiency. In

2020, Song et al. [58] have offered the sunflower lodging detection method from remote-sensing images by considering the deep semantic segmentation and image fusion from UAV, which has attained by improved SegNet. In 2020, Diakogiannis et al. [59] have suggested a reliable framework with “ResUNet-a” for semantic segmentation of high-resolution aerial images along with dice loss function through UNet encoder-decoder network. In 2020, Ye et al. [60] have introduced Uavid dataset for semantic segmentation of urban scenes through ensemble learning including multispectral dilation with feature space optimization (FSO). In 2020, Bianco et al. [61] have suggested a semantic segmentation model for detecting the road participants and road lane through a multitask instance segmentation neural network. This model has developed an ad-hoc training process for composing the final annotations utilized to train the suggested model by applying the CNN. In 2020, Mi and Chen [62] have introduced “superpixel-enhanced deep neural forest (SDNF)” for improving the classification capability from remote-sensing images along with the semantic segmentation, which has also designed a “superpixel-enhanced region module (SRM)” for reducing the noises and improves the edges of ground objects. In 2020, Zhang et al. [63] have proposed a new fused network with the model-agnostic metalearning (MAML) and FCNN for semantic segmentation of remote sensing based on RGB images along with the optimization algorithm, particle swarm optimization (PSO) algorithm. In 2020, Boonpook et al. [64] have proposed a multifeature semantic segmentation from images of UAV photogrammetry using the deep learning method, in which the accuracy of building extraction has improved with help of SegNet. In 2020, Yang et al. [65] have focused on understanding the pixel-level information from high-spatial resolution remote-sensing images using end-to-end network called residual network (ResNet), which has also considered several additional losses for enhancing the suggested model with optimization of multilevel features. In 2020, Mehra et al. [66] have suggested a semantic segmentation method for classifying the land cover through “six deep learning architectures such as pyramid scene parsing, UNet, and deeplabv3, path aggregation network, encoder-decoder network, and feature pyramid network,” which has attained superior results. In 2020, Tasar et al. [67] had proposed a semantic segmentation method by using color mapping GAN named ColorMAPGAN, which has also used element-wise matrix manipulation to learn the transformation of colors in the training data to the colors of the test data. In 2020, Venugopal [68] has suggested “a feature learning method named deep lab dilated CNN (DL-DCNN)” for automatic semantic segmentation for determining the correlation among two images, which has shown the superior efficiency over existing methods.

In 2021, Girisha et al. [69] have an improved encoder-decoder-based CNN architecture termed Uvid-Net for semantic segmentation from UAV video frames. This architecture was used to incorporate the temporal smoothness, which has captured the correlation among the sequence of frames using multibranch CNNs. In 2021, Huang et al. [70] have suggested an attention-guided label refinement

network (ALRNet) to enhance the semantic labeling of very high-resolution remote-sensing images with the encoder-decoder framework. Here, attention-guided feature fusion (AGFF) module was significantly developed for declining the semantic gap among diverse levels of features. In 2021, Abdollahi et al. [71] have suggested a GAN for segmenting the roads from high-resolution aerial imagery. This model has also used a modified UNet model (MUNet) for attaining the suitable results. In 2021, Alam et al. [72] have suggested an integrated framework using CNN with enhanced UNet and “encoder-decoder CNN structure SegNet with index pooling” for semantic segmentation of remote-sensing images, which has attained appropriate segmentation results on multitargets. In 2021, Anagnostis et al. [73] have suggested a semantic segmentation approach for obtaining the orchard trees from aerial images, which has used UNet for improving the efficient performance in terms of accuracy. This designed model has focused on automatic localization and detection of the canopy of orchard trees on different constraints. In 2021, Li et al. [74] have proposed a semantic segmentation model for analyzing the properties of photovoltaic, which has also enhanced the recommendations of segmenting the PV. It has revealed the high nonconcentrated and class imbalance distribution of photovoltaic panel image data through hard sampling and soft sampling. In 2021, Wang et al. [75] have designed a real-time semantic segmentation of high-resolution aerial images named an aerial bilateral segmentation network (Aerial-BiseNet) for offering superior accuracy. This suggested model has used two modules termed “feature attention module (FAM) and channel attention-based feature fusion module (CAFFM)” for analyzing the features. In 2021, Vasquez-Espinoza et al. [76] have suggested a semantic segmentation scheme using indoor imagery through the exploitation of details offered with the metadata utilized in the training stage of UNet. In 2021, Chen et al. [77] have considered different existing approaches such as “deeplabv3, generative adversarial network Pix2Pix, and UNet” for semantic segmentation of partially occluded apple trees, which has provided more details on branch paths, where the recovery of finer details from occlusions was offered. In 2021, Tasar et al. [78] have suggested a coined DAUGNet for the semantic segmentation of satellite images, including a data augmentor and classifier, which have performed on life-long, multitarget, multisource, single-source, and single-target problems. In 2021, Li et al. [79] have recommended a “dual attention deep fusion semantic segmentation network of large-scale satellite remote-sensing images (DASSN_RSI)” for getting the significant results which have also analyzed the challenges of conventional semantic segmentation approaches using remote-sensing images. In 2021, Jiang [80] has suggested a semantic segmentation model using high-resolution remote-sensing images through CNN and mask generation, in which the NN architecture was intended for obtaining a precise mask. In 2021, Liu et al. [81] have designed a new semantic segmentation model using remote-sensing images using Inceptionv-4 network for getting the enhanced classified information. This model has introduced the fusion of features for solving the classification of edge of objects. In 2021,

Zheng et al. [82] have implemented an “end-to-end CNN network named GAMNet” for balancing the controversies among the local and global information, which has also realized the boundary recovery and multiscale feature extraction. In 2021, Ouyang and Li [83] have offered a new DSSN called attention residual U-shaped network (AttResUNet) for encoding the feature maps and refining of features through attention module, which has also used GCN for classification.

2.2. Chronological Review. The chronological review on semantic segmentation models through deep learning approaches in the past years is given in Figure 2. The semantic segmentation is emerged as a major research area after 2015, and thus, this survey is prepared by gathering a set of research works from the year of 2015 to 2021. In the years of 2015, 2016, and 2017, the total number contributions is taken as 1.67% for each. Similarly, at 2018, 13.3% of the research works are gathered for analysis. In the year of 2020, 31.6% of the contributions are considered for evaluation. Likewise, while considering the 2019 and 2021, the number of research works is taken as 25%, respectively.

2.3. Security and Privacy Issues in Deep Learning. Many applications of deep learning in everyday life are self-driving cars, biometric security, health prediction, speech processing, financial technology, and retail [84]. Depending on the nature of the data and the user’s intent, each application has its own set of requirements. Many models were offered by the researchers to fit the application needs, users, and features of each sort of application, including LeNet, VGG, GoogleNet, Inception, and ResNet. Despite the fact that many studies on both attacking and safeguarding users’ privacy and security measures have been published, they remain fragmented. Tramèr evaluated different attack strategies based on FGSM and GAN before proposing the R-FGSM algorithm [85]. Xiaoyong Yuan also discusses security vulnerabilities in the deep learning approach. [86]. The preceding research has solely focused on the security of the deep learning model and does not provide an overview of preserving privacy in the deep learning model [87, 88].

In this work, we cover current studies on model security and data privacy that have led to the development of a secure and private artificial intelligence (SPAI). To address the demand for strong artificial intelligence (AI) systems, we compiled fragmented results and methodologies with the goal of delivering insights important to future study.

To conclude, we examine current research on privacy and security problems related to DL in the areas listed below.

- (1) DL model attacks: the two primary forms of DL attacks are evasion and poisoning attacks, with evasion attacks involving the inference phase and poisoning attacks involving the training phase
- (2) Defense of DL models: the different defense mechanisms presented may be divided into two broad categories based on the kind of attack, evasion and poisoning; tactics applied against evasion assaults

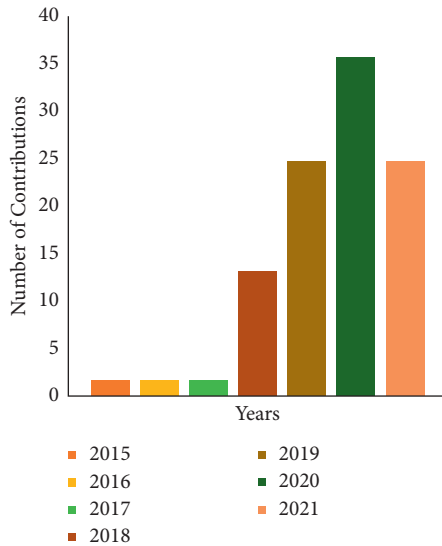


FIGURE 2: Chronological review of the semantic segmentation model using deep learning.

can be further divided into empirical (e.g., gradient masking, robustness, and detection) and certified approaches

- (3) Privacy attacks on AI systems: the potential privacy threats to DL-based systems arising from service providers, information silos and users
- (4) Defense against a privacy breach: the most modern cryptographic protection approaches, such as homomorphic encryption, safe multiparty computing, and differential privacy

According to training and testing stages in deep learning model security, attack techniques are categorized. This research emphasises on threats at the testing. Furthermore, the categorization is based on the attacker's expertise as well as the attacker's pattern of assaulting black boxes and white boxes. Attack strategies are classed in order to safeguard user privacy based on the system design and the attacker's knowledge. Attack strategies are divided into two categories in system architecture: centralized and distributed. According to the information, the attacker is also split into white box and black box attacks. Based on the stages of the deep learning model, defensive techniques are classified.

The assumptions for implementing certain threats in deep learning security are based on situations. The threat models are classified depending on the adversary's knowledge, the goal of the attacker, and the frequency of attacks.

2.3.1. The Adversary's Knowledge. A black box attack occurs when the attacker lacks knowledge of the system, in case of which the attacker submits input and receives output without understanding the system parameters. In contrast, in the event of a white box attack, the attacker has access to all system information, including the model's structure and parameter values.

2.3.2. Attacker's Target. Targeted attacks detect certain data or object types that misclassify this data collection. These types of attacks are common when categorization systems are used. In face recognition or authentication systems, for example, an attacker selects a certain face, one of which is misclassified among hostile samples. Nontargeted attacks, on the contrary, choose arbitrary data and are simpler to execute than targeted attacks.

2.3.3. Frequency of Attacks. One-time attacks require only one hostile example to be created. Otherwise, repeated attacks build adversarial instances through multiple updates. Iterative attacks outperform one-time attacks every time, but they need more queries to the deep learning system and take longer.

Deep learning security threats are classified into two types: adversarial and poisoning. We will concentrate on adversarial assaults in this research. During a system query, an adversarial attack introduces noise to the usual data. When the attacker receives the reported results, he or she utilizes this information to generate adversarial instances. This type of assault may be found in image processing, audio processing, and virus detection. It can trick deep learning machines, but not humans, particularly in the field of image processing. The gap between the source data and the adversarial example is represented by the noise value.

2.4. Limitations and Alternate Solutions of Deep Learning. Several challenges are frequently taken into account when adopting DL. Those that are more difficult are mentioned next, with various viable solutions supplied.

2.4.1. Training Data. Because it also requires representation learning, DL is tremendously data-hungry. To produce a well-behaved performance model, DL necessitates a massive quantity of data, i.e., as the data accumulates, an even more well-behaved performance model may be achieved. Most of the time, the supplied data are adequate to generate a solid performance model. However, there are situations when there is insufficient data to use DL directly. There are three proposed techniques for dealing with this issue. The first entails using the transfer-learning idea after collecting data from similar activities. While the transmitted data will not directly enhance the real data, it will aid in improving both the original input data representation and its mapping function. The model's performance is improved as a result. Another method is to use a well-trained model from a comparable assignment and fine-tune the end of two layers, or even one layer, depending on the limited original data. The second option involves data augmentation. Because picture translation, mirroring, and rotation frequently do not modify the image label, this activity is extremely useful for supplementing image data. In contrast, it is critical to exercise caution while using this approach in some circumstances, such as with bioinformatics data. When mirroring an enzyme sequence, for example, the resulting data may not represent the real enzyme sequence. In the third

way, simulated data may be used to increase the size of the training set. If the problem is sufficiently understood, it is sometimes possible to construct simulators based on the physical process. As a result, the end product will comprise the simulation of as much data as is required.

2.4.2. Transfer Learning. Deep CNNs, which provide ground-breaking help for solving numerous classification issues, have been widely used in recent research. Deep CNN models, in general, need a large amount of data in order to function well. The most prevalent problem with employing such models is a lack of training data. Gathering a big number of data is a demanding task, and no viable solution is currently available. As a result, the undersized dataset problem is now being addressed utilising the TL approach, which is very efficient in handling the lack of training data issue. The TL technique entails training the CNN model with vast amounts of data. The model is then fine-tuned for training on a small request dataset.

The student-teacher interaction is an effective method for explaining TL. The first step is to learn everything there is to know about the subject. The teacher then gives a “course” by imparting the material over time through a “lecture series.” Simply put that the instructor transmits information to the pupil. More specifically, the expert (teacher) imparts knowledge (information) to the learner (student). Similarly, the DL network is trained using a large amount of data and learns the bias and weights during training. These weights are then transmitted to several networks in order to retrain or test a comparable unique model. As a result, the innovative approach can pretrain weights rather than requiring training from beginning.

2.4.3. Data Augmentation Techniques. Data augmentation techniques are one viable answer if the aim is to expand the quantity of accessible data while avoiding overfitting. These strategies are data-space solutions to any problem with little data. Data augmentation refers to a set of approaches for improving the properties and quantity of training datasets. As a result, when these strategies are used, DL networks perform better. Following that, we will go through some other data augmentation solutions.

- (i) Flipping: vertical axis flipping is a less prevalent procedure than horizontal axis flipping. On datasets such as ImageNet and CIFAR10, flipping has been shown to be beneficial. Furthermore, it is really simple to implement. Furthermore, it is not a label conserving transformation on datasets involving text recognition (such as SVHN and MNIST).
- (ii) Color space: as a dimension tensor, encoding digital picture data is often utilized (height \times width \times colour channels). Performing enhancements in the colour space of the channels is an alternate method that is particularly practical for implementation. Color augmentation is as simple as isolating a channel of a certain colour, such as red, green, or blue. By dividing that matrix and introducing extra double

zeros from the remaining two colour channels, you may quickly transform a picture utilising a single-color channel. Furthermore, the picture brightness may be increased or decreased by utilising simple matrix operations to modify the RGB values. Additional better colour augmentations can be acquired by generating a colour histogram that represents the image. Lighting changes can also be done by altering the intensity values in histograms similar to those used in photo-editing software.

- (iii) Cropping: cropping a prominent region of every single image is a technique used as a specialised processing step for image data with combined dimensions of height and width. Furthermore, random cropping can be used to achieve the same effect as translations. The distinction between translations and random cropping is that translations preserve the image’s spatial dimensions, but random cropping decreases the input size. The label-preserving transformation may not be addressed because to the cropping reduction threshold that was chosen.
- (iv) Rotation: rotation augmentations are created by rotating a picture left or right from 0 to 360° around the axis. The rotation degree parameter has a significant impact on the applicability of rotation augmentations. Small rotations (from 0 to 20°) are quite useful in digit identification tasks. When the rotation degree rises, however, the data label cannot be kept post-transformation.
- (v) Translation: shifting the picture up, down, left, or right is a highly important transformation for avoiding positional bias in image data. For example, it is typical for all of the photos in a dataset to be centred; also, the tested dataset should be fully composed of centred images in order to test the model. It is worth noting that, after translating the starting pictures in a certain direction, the remaining space should be filled with Gaussian or random noise, or a constant value such as 255 s or 0 s. Using this padding, the spatial dimensions of the picture after augmentation are kept.
- (vi) Noise injection: this method entails introducing a matrix of arbitrary values. A Gaussian distribution is typically used to generate such a matrix. Injecting noise into photos allows the CNN to learn more robust features.

2.4.4. Interpretability of Data. DL approaches are occasionally studied to serve as a black box. They can, in fact, be interpreted. Many areas, such as bioinformatics, have a requirement for a way of interpreting DL, which is utilized to acquire the valuable motifs and patterns detected by the network. It is necessary not only to understand just the illness diagnosis or prediction findings of a trained DL model but also how to improve the certainty of the prediction outcomes, as the model bases its choices on these verifications. To do this, each section of the specific example

can be assigned a weighted score. Backpropagation-based techniques or perturbation-based approaches are employed in this solution. A fraction of the input is altered in the perturbation-based techniques, and the effect of this modification on the model output is monitored. This notion has a high computational complexity, yet it is easy to grasp. With contrast, in backpropagation-based approaches, the signal from the output propagates back to the input layer to verify the score of the relevance of distinct input sections.

2.4.5. Overfitting. Because of the large number of parameters involved, which are complexly interrelated, DL models have an extremely high risk of resulting in data overfitting during the training stage. Such circumstances limit the model's capacity to perform well on the tested data. This issue is not just restricted to a single field, but also encompasses a variety of duties. As a result, while proposing DL approaches, this issue should be thoroughly examined and handled correctly. According to current research, the inherent bias of the training process helps the model to overcome critical overfitting concerns in DL. Nonetheless, strategies for dealing with the overfitting problem must be developed. An examination of the various DL algorithms for easing the overfitting problem may be divided into three categories. The first class contains the most well-known methods, such as weight decay, batch normalisation, and dropout, and it operates on both the model architecture and model parameters. Weight decay is the default approach in DL, and it is used widely as a universal regularizer in practically all ML algorithms. The second class is concerned with model inputs such as data corruption and data augmentation. One cause of overfitting is a paucity of training data, which causes the learnt distribution to differ from the true distribution. Data augmentation increases the size of the training data. In contrast, marginalised data corruption improves the solution solely through data augmentation. The last class is concerned with the model's output. For regularising the model, a recently developed method penalises overconfident outputs. This approach has been shown to be capable of regularising RNNs and CNNs.

2.4.6. Vanishing Gradient Problem. In general, when utilising backpropagation- and gradient-based learning approaches with ANNs, an issue known as the vanishing gradient problem emerges, particularly, during the training stage. In further detail, during each training iteration, each weight of the neural network is updated depending on the current weight and is proportionately relevant to the partial derivative of the error function. However, owing to a vanishingly tiny gradient, this weight update may not occur in some situations, implying that no more training is feasible and the neural network would cease entirely. In contrast, the sigmoid function, such as other activation functions, compresses a huge input space to a compact input region. As a result of the huge fluctuation at the input resulting in a little variation at the output, the derivative of the sigmoid function will be small. Only a few layers in a shallow network employ these activations, which is not a big deal. While

having additional layers causes the gradient to become very tiny during the training stage, the network operates effectively in this scenario. The gradients of neural networks are determined using the backpropagation approach. Initially, this approach identifies the network derivatives of each layer in reverse order, beginning with the most recent layer and moving back to the first. The next step is to multiply the derivatives of each layer along the network in the same way that the previous step was done. When there are N hidden layers, for example, multiplying N small derivatives together requires an activation function such as the sigmoid function. As a result, the gradient decreases exponentially as it propagates back to the first layer. Because the gradient is modest, the biases and weights of the initial layers cannot be updated efficiently during the training stage. Furthermore, because these early layers are typically vital in detecting the main aspects of the input data, this circumstance reduces total network accuracy. However, by using activation functions, such an issue may be avoided. These functions lack the squishing attribute, which allows them to squish the input space to a tiny space. The ReLU is the most preferred choice for mapping X to \max since it does not provide a modest derivative that is useful in the field. Another option is to use the batch normalisation layer. As previously stated, the difficulty arises when a huge input space is squeezed into a tiny space, resulting in vanishing the derivative. Using batch normalisation mitigates this problem by simply normalising the input, i.e., the expression $|x|$ does not achieve the sigmoid function's outside borders. The normalisation procedure causes the majority of it to fall into the green region, ensuring that the derivative is large enough for future activities. Furthermore, faster hardware, such as that supplied by GPUs, can address the above issue. In comparison to the time necessary to notice the vanishing gradient problem, this enables normal backpropagation over many deeper levels of the network.

2.4.7. Exploding Gradient Problem. The gradient problem is the inverse of the vanishing problem. Specifically, during backpropagation, huge error-gradients accrue. The latter will result in extraordinarily big modifications to the network's weights, causing the system to become shaky. As a result, the model's capacity to learn successfully will deteriorate. Moving backward in the network during backpropagation causes the gradient to expand exponentially by repeatedly compounding gradients. As a result, the weight values may get extremely big and may overflow to produce a not-a-number (NaN) value. Some potential solutions include

- (1) Using different weight regularization techniques
- (2) Redesigning the architecture of the network model

2.4.8. Underspecification. In 2020, a Google team of computer scientists found a new difficulty known as underspecification. When evaluated in real-world applications such as computer vision, medical imaging, natural language processing, and medical genomics, machine learning models, particularly, deep learning models, frequently

exhibit startlingly low performance. Underspecification is to blame for the poor performance. It has been demonstrated that modest changes may push a model to an entirely new solution and result in different predictions in deployment domains. There are several methods for dealing with the issue of underspecification. One of them is to create “stress tests” to see how well a model performs on real-world data and to identify potential problems. Nonetheless, this necessitates a solid grasp of the process, as the model can perform incorrectly. “Designing stress tests that are well-matched to application criteria and that give adequate “covering” of probable failure modes is a huge problem,” the researchers concluded. Underspecification severely limits the trustworthiness of ML predictions and may necessitate some reconsideration of some applications. Because ML is tied to humans through applications such as medical imaging and self-driving automobiles, it will necessitate careful consideration of this issue.

2.5. Computational Approaches and Comparison between Different Aspects Related to Devices. Complex ML and DL algorithms have quickly emerged as the most significant techniques for computationally exhausting applications, and they are widely applied in a variety of domains. The creation and refinement of algorithms, together with the capabilities of well-behaved computational performance and massive datasets, allow for the successful execution of various applications that were previously either impossible or difficult to conceive.

2.5.1. CPU-Based Approach. The CPU nodes’ well-behaved performance frequently aids robust network connectivity, storage capabilities, and huge memory. Although CPU nodes are more general purpose than FPGA or GPU nodes, they lack the ability to compete in raw compute facilities since this demands improved network capability and a bigger memory capacity.

2.5.2. GPU-Based Approach. GPUs are exceptionally effective for various fundamental DL primitives, including highly parallel-computing operations such as activation functions, matrix multiplication, and convolutions. Incorporating HBM-stacked memory onto modern GPU models dramatically improves bandwidth. This enhancement enables a wide range of primitives to make efficient use of all available computational resources on GPUs. In the case of dense linear algebra computations, the boost in GPU performance over CPU performance is typically 10–20:1.

2.5.3. FPGA-Based Approach. FPGA is widely used in a variety of functions, including deep learning. FPGA is widely used to create inference accelerators. The FPGA can be effectively configured to reduce the number of unnecessary or overhead functions in GPU systems. The FPGA, in comparison to the GPU, is limited to both poor-behaved floating-point performance and integer inference. The key FPGA feature is the ability to dynamically modify the array

characteristics (at run-time), as well as to configure the array using effective design with little or no overhead. Table 1 [89] represents the comparison between different aspects related to the devices.

3. Algorithmic Categorization and Features and Challenges of Existing Semantic Segmentation Models

3.1. Algorithmic Classification. This section presents different deep learning approaches utilized for developing a semantic segmentation model as given in Figure 3.

The semantic segmentation models mostly use deep learning algorithms for getting superior accuracy with better quality. The techniques have been categorized into two sections, namely, deep learning and miscellaneous approaches. In deep learning, CNN architectures play a major role for semantic segmentation, which is extended by adopting different convolutional layers or other frameworks.

Supervised learning: in this model, training data consist of both input and desired results. These supervised learning algorithms are often accurate and fast. It has the ability of generalization that gives the precise results while processing new data without knowing a priori about the target.

CNN [24, 48, 53, 61, 80] inspires the researchers because of the superior efficiency in the area of computer vision, which has been adopted in diverse applications such as object detection, image recognition, and other fields. Figure 4 represents the architecture of convolutional neural network This architecture enhances accuracy of prediction or classification due to the large number of training samples along with building neural networks with several layers. CNN is a hierarchical system, which takes the input data as raw data through stacking a set of operations such as mapping of nonlinear activation functions, convolution, and pooling operations. This procedure is named as “feedforward operation.” Due to this effective operation of CNN, it has attained superior results in the data mining and natural processing tasks when compared with the deep neural networks. Owing to the efficiency of CNN architectures, multiple CNN-based approaches are designed by integrating many ideas or integration of FCN architecture. This adoption of several networks into one framework is named as ensemble learning [60, 66, 77], which has attained superior results compared to single architecture because of the utilization of multiple layers. Ensemble of CNNs [25] is adopted by utilising several layers of CNN architecture to reduce the computational cost and avoids aliasing problem. It provides promising performance when compared to the existing models. DP-DCN [35] focuses on extracting the significant features from DSM data and spectral channels for fusing them through an encoder-decoder framework. The extended version of CNN consists of Shuffling CNNs [27], DSMFNets [28], UVid-Net [69], ESPCN [42], neuro-architectures [72], ensemble of CNNs [25], ADSCNet [54], DSCNN [29], DCNN [32, 33, 57], multitask CNN [30], multifilter CNN [31], ConvNet [26], GAMNet [82], DL-DCNN [68], and GCN [49, 83]. This modified or integrated

TABLE 1: A comparison between different aspects related to the devices.

| Feature | Assessment | Leader |
|------------------------|--|----------|
| Development | CPU is the easiest to program, then GPU, and then FPGA | CPU |
| Size | Both FPGA and CPU have smaller volume solutions due to their lower power consumption | FPGA-CPU |
| Customization | Broader flexibility is provided by FPGA | FPGA |
| Ease of change | Easier way to vary application functionality is provided by GPU and CPU | GPU-CPU |
| Backward compatibility | Transferring RTL to novel FPGA requires additional work; furthermore, GPU has a less stable architecture than CPU | CPU |
| Interfaces | Several varieties of interfaces can be implemented using FPGA | FPGA |
| Processing/\$ | FPGA configurability assists utilization in wider acceleration space; due to the considerable processing abilities, GPU wins | FPGA-GPU |
| Processing/watt | Customized designs can be optimized | FPGA |
| Timing latency | Implemented FPGA algorithm offers deterministic timing, which is in turn much faster than GPU | FPGA |
| Large data analysis | FPGA performs well for inline processing, while CPU supports storage capabilities and the largest memory | FPGA-GPU |
| DCNN inference | FPGA has lower latency and can be customized | FPGA |
| DCNN training | Greater float-point capabilities provided by GPU | GPU |

concept of CNN is designed for efficient semantic segmentation.

FCN (see [38, 46, 47, 51]): the basic idea of FCN includes processes such as “multilayer convolution, deconvolution, and fusion,” where the convolutional layers are replaced with the fully connected layers. The image score is computed by using pixel-wise convolution. UNet [42, 73, 76] is a type of FCN that is efficient for small training dataset, which includes convolution and deconvolution layers with filters along with ReLU activation function. The modified versions of FCN are given here as integrated algorithm [71], ResU-Net-a [59], CAM-DFCN [36], relation module-equipped FCN [52], FCN-Alexnet model [55], and AD-LinkNet [44]. Improved SegNet [58] and SegNet [64] follow a FCN structure with encoder and decoder network. SegNet saves the element index in the upsampling process of the decoder network for solving the ambiguous spatial information in the resultant of deeper layers. Figure 5 depicts the architecture of fully connected Network.

FCN introduces many significant ideas: (i) end-to-end learning of the upsampling algorithm via an encoder/decoder structure that first downsamples the size of the activations and then upsamples it again, (ii) using fully convolutional architecture allows the network to take images of arbitrary size as input since there is no fully connected layer at the end that requires a specific size of the activations, and (iii) introducing skip connections as a way of fusing information from different depths in the network for multiscale inference.

GAN (see [37, 71]): generative adversarial network (GAN) model considers a softmax layer, in which the discriminator of the GAN produces label types for efficient classification of unlabeled samples and labeled examples. The architecture of generative adversarial network is shown in Figure 6. The modified version of ColorMapGAN [67] has aimed at minimizing the computational complexity and improving the accuracy.

DNN: deep neural networks (DNN) focus on semantic segmentation of high-resolution images which consist of several parameters that need a large number of labeled

examples for training. A general scheme for constructing a deep network to process a rich dataset is complex. The improved DNN models are modified inceptionV-4 network [50], NDRB [52], ResNet101-v2 [39], ALRNet [70], HMANET [56], ResNet [65], inceptionV-4 network [81], MAVNet [41], and SDNF [61], which are aimed to enhance the superior accuracy on segmentation.

Unsupervised learning: this model is not offered with the precise results during training, which can be employed for clustering the input data in classes through statistical properties.

DAugNet [78]: DAugNet generates the precise maps and has provided life-long adaptation settings for giving the superior semantic segmentation results. GB-RBM [73] is introduced for enhancing the segmentation results and improving the speed and accuracy. Figure 7 gives the training procedure of data augmentation network.

3.2. Features and Challenges. The features and challenges of the conventional semantic segmentation model using deep learning techniques are listed in Table 2. This description provides the researchers for focusing on a new semantic segmentation model on aerial images for solving the existing challenges through adopting deep learning techniques.

4. Simulation Platforms and Dataset Description for Conventional Semantic Segmentation Models

4.1. Simulation Platforms. The simulation environments used for implementing a semantic segmentation model with different imaging modality is presented in Figure 8. Here, some of the tools such as CUDA version 8.0, Edge Detection and Image Segmentation (EDISON) library, MXNet, TensorRT, and two-fold validation tool are used in 1.7% of the contributions, respectively. MATLAB and Tesla use 3.3% of the contributions for implementation and Pascal and Keras utilize 8.3% of the research works with the TensorFlow as a platform, respectively. TensorFlow is used as the simulation

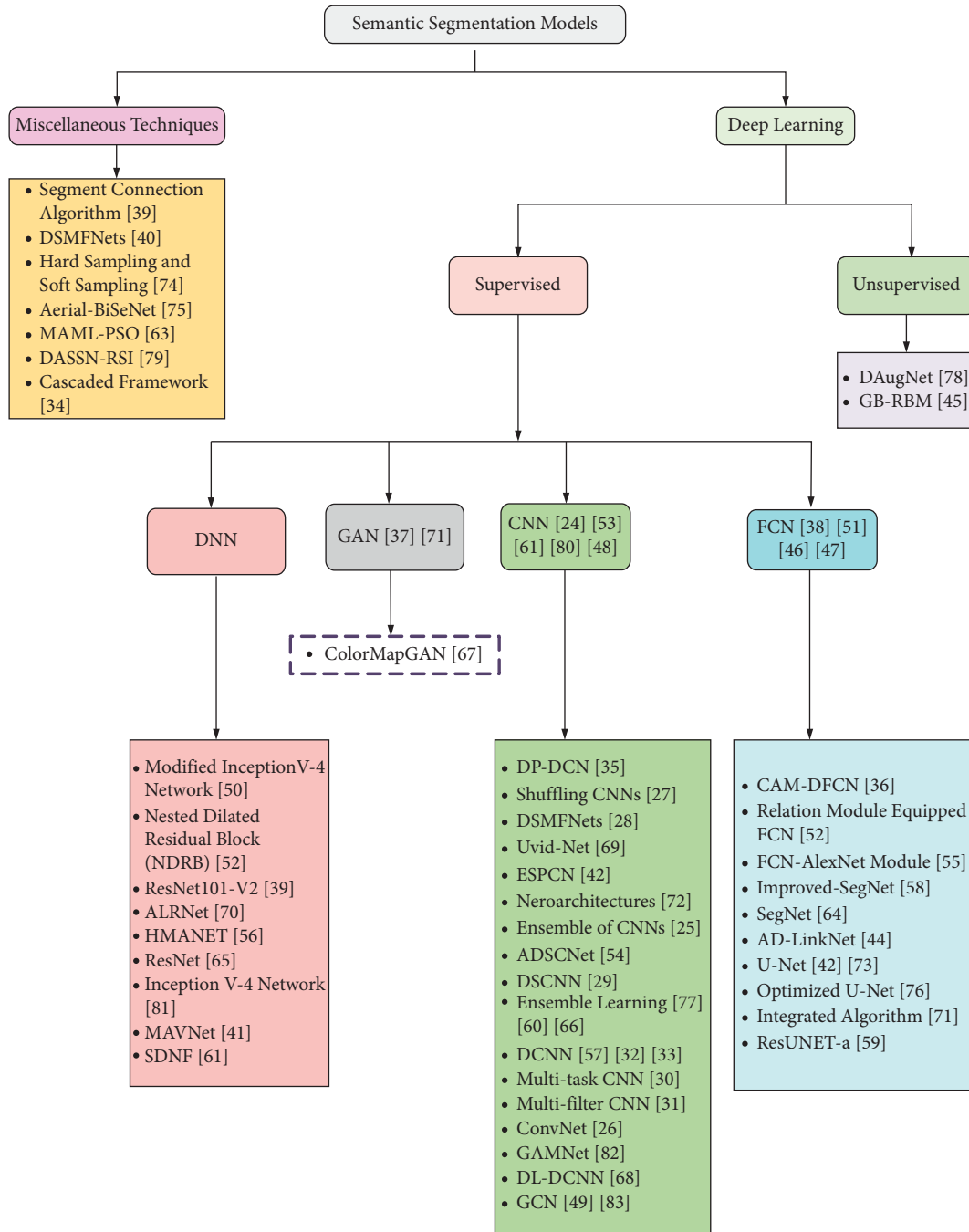


FIGURE 3: Algorithmic categorization of existing semantic segmentation models.

environment for 18.3% of the works and NVIDIA is considered in 5% of the contributions. Finally, the python tool is used in 6.6% of the research works and other platform environments are taken in 20% of the contributions.

4.2. Dataset Description and Imaging Modalities Focused. The dataset used for implementing the semantic segmentation model along with different imaging modalities is given in tabular forms (Tables 3–7). Most of the contributions are

considered aerial images for semantic segmentation, which is used in 23.3% of the work, s, and high-resolution aerial imagery is taken in 16.6% of the contributions. Similarly, the remote-sensing and high-resolution remote-sensing images are taken in 25% of the research works. Unoccupied aerial vehicles’ (UAVs) images are gathered in 11.7% of the contributions.

Multiscale and multispatial resolution images are included in 1.7% of the research papers, respectively, and satellite images are taken in 5% of the contributions. Other high-resolution images are taken in 13.4% of the research works.

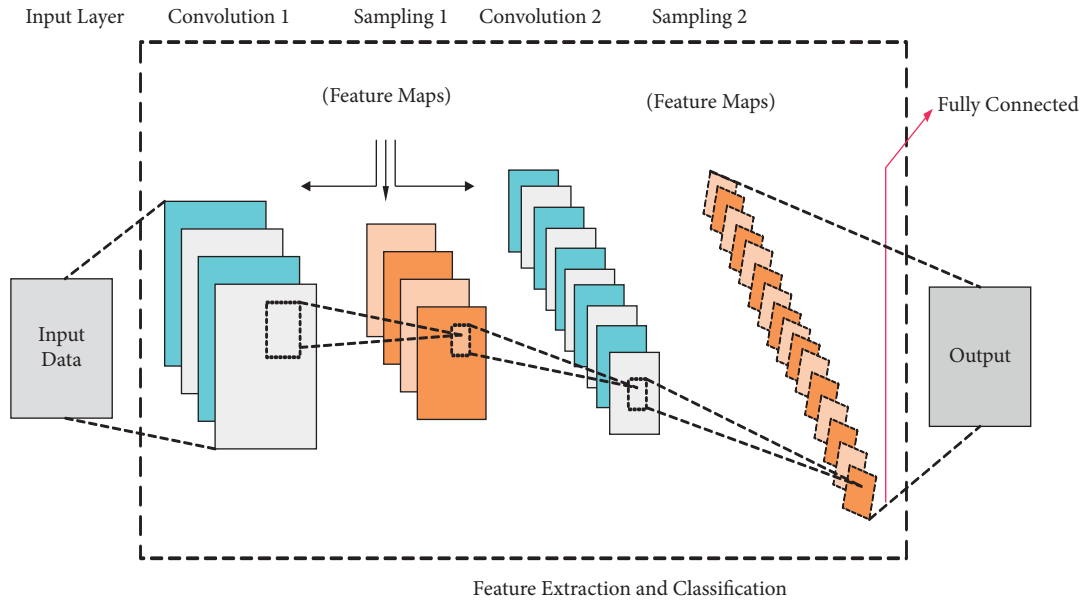


FIGURE 4: CNN architecture.

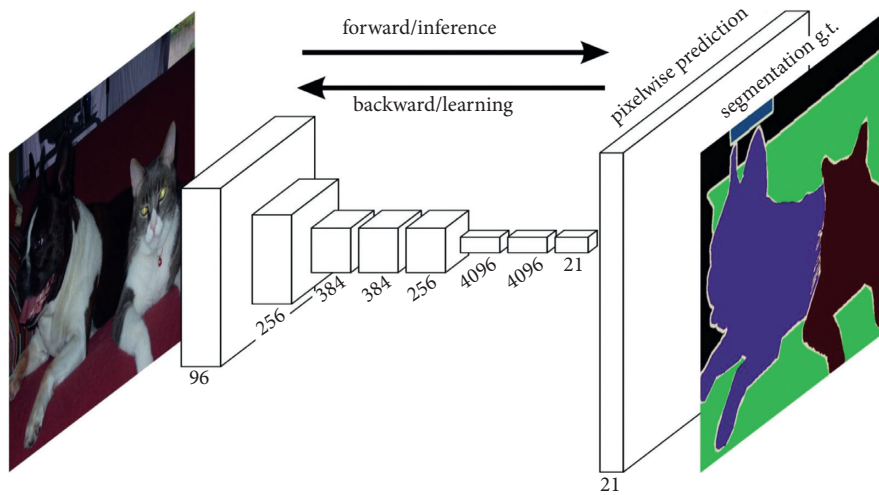


FIGURE 5: FCN architecture [90].

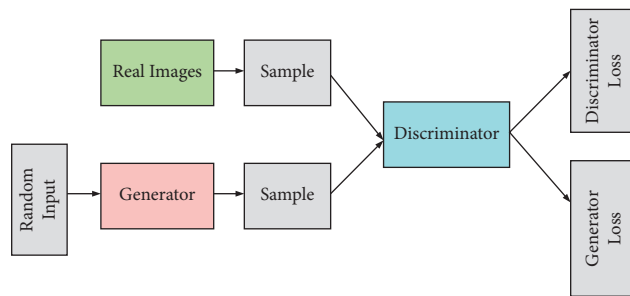


FIGURE 6: GAN architecture.

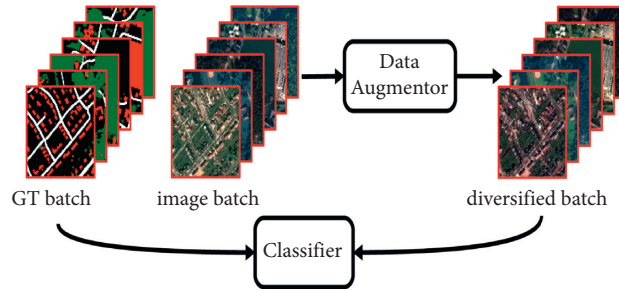


FIGURE 7: The training procedure of DAugNet that comprises a data augmentor and a classifier. In each training iteration, the classifier learns from the diversified batch generated by the data augmentor [78].

TABLE 2: Merits and demerits of existing semantic segmentation model using deep learning approaches.

| Citation number | Methodology | Features | Challenges |
|-----------------|---|--|--|
| [24] | CNN | It accurately extracts the round objects using CNN | It requires more cost for getting pixel intensities on shadow regions |
| [25] | Ensemble of CNNs | It reduces the computational cost and avoids aliasing problems; it provides promising performance when compared to the existing models | Conversely, multicore parallelization over diverse scenes is complex |
| [26] | ConvNet | It gets efficient segmentation performance with better sophistication | This model is not suitable for unlabeled data |
| [27] | Shuffling CNNs | This model is limited to use ensemble approaches | This model is limited to use ensemble approaches |
| [28] | ResNet101-v2 and a pyramid pooling module | It has offered an effective network framework with superior performance | However, the segmentation accuracy is limited while considering the large spectral similarities among imperious surfaces and buildings |
| [29] | DSCNN | It provides enhanced and smoother identifications for different objects | It does not offer superior numerical outcomes |
| [30] | Multitask CNN | It has offered a principled and flexible structure for providing the efficient segmentation results | It does not preserve the geometrical features and complex for segmentation |
| [31] | Multifilter CNN | It has achieved the highest overall accuracy and removed the noise | This paper does not investigate how diverse data sources from other sensors are integrated in deep CNN |
| [32] | DCNN | This model gets superior efficiency with discriminative frameworks; it avoids overfitting problems | However, this model does not fully remove the salt and pepper noise |
| [33] | DCNN | The DCNN achieves superior effectiveness on a standard dataset | However, it is a tedious and small problem that affects the segmentation quality |
| [34] | Cascaded framework | It has improved the prediction with object boundaries and removed the isolated false positives | It has high computational costs |
| [35] | Dual-path densely convolutional networks (DP-DCN) | It avoids the vanishing gradient problem and strengthens the information flow among the layers by a dense connection | However, it requires less test time and training time |
| [36] | CAM-DFCN | This model has attained mainstream performance; it also promotes the segmentation results with efficient feature selection | The performance of the suggested CAM-DFCN was not improved significantly while comparing to the CNN + RF + CRF |
| [37] | GANs with multi-scale context aggregation | This network has improved the accuracy of road extraction and offered superior visual effects | It is a computationally inefficient one, which has to enhance the segmentation precision |

TABLE 2: Continued.

| Citation number | Methodology | Features | Challenges |
|-----------------|---|---|--|
| [38] | FCN | This model has recovered the lost data to get high robustness and accuracy | This model is not applicable for processing the shadow areas |
| [39] | Segment connection algorithm | It enhances the detection efficiency that enhances the applicability of the framework | The precision rate of lost vibration damper identification is less which gives a lower F1 score |
| [40] | DSMFNets | It shows superior fusion results using DSMFNets with efficient performance | However, the effectiveness can be affected due to the restriction on the feature extraction module |
| [41] | MAVNet | It has shown a better tradeoff between performance and inference time | This model is not applicable to apply modestly sized networks |
| [42] | ESPCN and UNet | It enhances the segmentation and improves the robustness | It suffers from insufficient training samples |
| [43] | Neuroarchitectures including (a) MultiNet, (b) SegNet, and (c) UNet | This model has improved the quality of object segmentation | However, the implementation is restricted |
| [44] | AD-LinkNet | The suggested AD-LinkNet boosts the efficiency on segmentation | The designed model does not show the better performance on different road interruptions |
| [45] | GB-RBM | It enhances the segmentation results; it has improved the speed and accuracy | This model is restricted on high spatial resolution thermal infrared images |
| [46] | FCN | The segmentation efficiency is improved while comparing with the conventional approaches; it gets less overhead | The multitask degrades the efficiency of segmentation |
| [47] | FCN | It efficiently discriminates the nonwetland classes from wetland classes; It enhances the accuracy of semantic segmentation | However, processing the restricted availability of ground truth data in large-scale remote-sensing applications is challenging |
| [48] | CNN | This model has enriched the semantic information, which has focused on attaining the representative extracted features | It lacks in performance due to the processing of high-level features |
| [49] | GCN | It has shown superior performance with capturing of complex features; it solves the scarcity problem | It has to enhance the accuracy by adopting different approaches such as optimization and semantic labeling |
| [50] | Modified InceptionV-4 network called DAPN | This technique has robust generalization ability | Although the potsdam dataset has offered consistent performance, there is a considerable reduction in the vaihingen IR-R-G dataset concerning IOU scores |
| [51] | Relation module-equipped FCN | The performance of semantic segmentation is enhanced with the use of a network using aerial scenes | However, the suggested relation modules regarding segmentation are basic one, and thus, it does not offer superior efficiency |
| [52] | Nested dilated residual block (NDRB) | It offers precise object boundaries and labeling for complex scenes | The per-class accuracy is not evaluated which does not estimate the efficiency |
| [53] | CNN | This model has offered the best tradeoff with the fewer number of parameters along with less memory utilization; it gives the suitable mapping of terrain | The considered images do not have a fixed shape or resolution, and thus, the training may be affected |
| [54] | ADSCNet | This model has reduced the network complexity because of the depth-wise convolution; it improves the performance along with better information flow | This model does not evaluate the actual inference speed |
| [55] | FCN-AlexNet model | This model has maintained reasonable accuracy and inference speed | The limited dataset is used for validation, which has to be rectified |

TABLE 2: Continued.

| Citation number | Methodology | Features | Challenges |
|-----------------|--|--|--|
| [56] | HMANet | This model captures the global contextual details for efficient segmentation; it enhances the efficiency of the self-attention scheme and reduces feature redundancy | It takes huge consumption of memory |
| [57] | DCNN | It has shown better smoothing effects; it has extracted the multilevel features | It does not consider the complementary and orthogonal technical progressions |
| [58] | Improved SegNet | It increases the accuracy and speed of sunflower lodging; it efficiently monitors the lodging in equivalent low canopy density crops | The complexity of the identification is increased due to the growth and status of sunflow which is varied through spatial distribution changes |
| [59] | ResUNet-a | It has offered better convergence properties; the superior F1 score is observed | It shows slow operation due to the GPU synchronization that makes it impractical for future processes |
| [60] | Ensemble learning | It extracts multiscale features; the manual dataset offers superior performance with temporal consistency | However, this dataset has different challenges such as number of types in scenes, dataset size and large-scale differentiation for several objects |
| [61] | CNN | It reduces the computational constraints; it provides real-time performance | However, the weak labeling stage is observed that affects the performance |
| [62] | Superpixel-enhanced deep neural forest (SDNF) | It shows superior classification ability with reduced noises; it gives robust results | However, for some of the classes, the accuracy is reduced |
| [63] | MAML-PSO | The misclassification of objects with specific height variance can be effectively minimized by introducing LIDAR data; it increases the testing accuracy | In this model, the overall accuracy is not very good |
| [64] | SegNet | It shows the superior building extraction for medium- and high-sized buildings; it also enhances the classification accuracy | However, the small size buildings are complex for identification |
| [65] | ResNet | It has efficiently extracted the global and local deep features that offer better semantic segmentation results | This study does not consider the digital surface models on both datasets |
| [66] | Ensemble learning | The semantic segmentation is improved due to the extracted features | The suggested model is limited on dataset size |
| [67] | ColorMapGAN | The suggested model has minimized computational complexity and improved accuracy | Though, the results' quality for nonlearning-based approaches is inefficient |
| [68] | DL-DCNN | It has achieved better convergence rate and accelerated network training; it obtains enhanced results with efficient identification of changes | Conversely, it gets overfitting and low accuracy rate |
| [69] | UVid-net | It reduces the computational complexity and provides superior segmentation results on aerial videos | However, it is a laborious and time-consuming task |
| [70] | ALRNet | This model chooses the most nonredundant and representative features to offer outstanding efficiency | ALRNet has higher computational inefficiency |
| [71] | GAN | This model efficiently preserves the edge information and gets a better accuracy rate on segmenting the maps | This model lacks accuracy and also it suffers from extracting the continuous road parts or complex regions |
| [72] | Integrated algorithm (encoder-decoder CNN structures SegNet with index pooling and UNet) | This integrated technique has offered superior features of both CNN and UNet to offer better semantic segmentation of images | For some classes, the suggested integrated model has attained less performance than other algorithms |

TABLE 2: Continued.

| Citation number | Methodology | Features | Challenges |
|-----------------|---------------------------------|--|--|
| [73] | UNet | This approach has the capability of precise segmentation of tree canopies; it also solves complex problems in environments such as agricultural production | Although the designed model shows superior performance on detection, it does not solve the issue of densely merged and located false positives |
| [74] | Hard sampling and soft sampling | This model explores the heterogeneous colour feature and texture feature of the PV panel | On the contrary, the uncertainties have remained |
| [75] | Aerial-BiSeNet | A superior balance among the speed and accuracy is offered; it has shown better efficiency and accuracy on both datasets | It suffers from weak representation ability and high model complexity |
| [76] | Optimized UNet | It solves the computational complexity; it improves the overall performance | Conversely, some of the images attain the worst results because of the estimation problem |
| [77] | Ensemble learning | The accuracy and practical implementation is superior to other existing approaches | The efficiency can be affected due to the noise present in images |
| [78] | DAugNet | The precise maps are generated and have provided life-long adaptation settings | This model does not apply on sentinel and aerial images |
| [79] | DASSN_RSI | It has reduced the training loss and enhanced the convergence rate; it verifies the advancement and efficiency of the suggested method | It lacks in robustness, which does not focus on low-shot learning methods |
| [80] | CNN | The segmentation and overall training time have been reduced; it also improves the overall precision | This model does not consider low-resolution images |
| [81] | InceptionV-4 network | It has attained superior segmentation efficiency and training efficiency | It shows poor generalization ability |
| [82] | GAMNet | The efficiency of the integration module is improved; the accurate results has attained with precise boundaries even for small objects | The confusing problem is occurred and suffered from misclassification problem in shaded areas |
| [83] | GCN | It restores the boundaries of ground objects and reduced the pixel-level noises | However, it does not utilize the spatial correlation details for interpreting remote-sensing images |

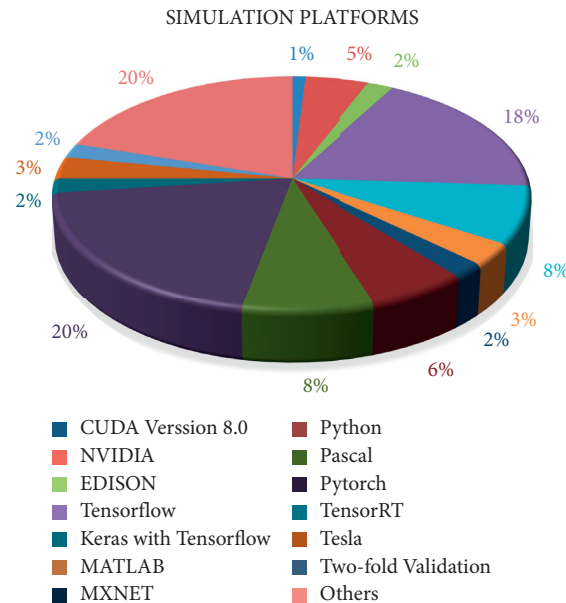


FIGURE 8: Simulation platforms used for implementing the semantic segmentation models.

TABLE 3: Dataset description based on modality of aerial imagery.

| Citation number | Dataset description |
|-----------------|--|
| [38] | Aeriallanes18 dataset |
| [52] | The potsdam dataset and the vaihingen dataset |
| [27] | ISPRS vaihingen and potsdam datasets |
| [24] | Manual dataset that includes 127 aerial images |
| [41] | Mavnet |
| [73] | Manual dataset |
| [53] | The dataset can be accessed at https://github.com/gabrielmtzsoltero/ssegfor_aerial_mapping/ |
| [56] | ISPRS 2D semantic labeling challenging for vaihingen and potsdam |
| [29] | Use the vaihingen dataset |
| [77] | Commercial apple orchard in northeastern melbourne |
| [61] | Vaihingen dataset and potsdam dataset |
| [46] | ISPRS 2D semantic labeling and data fusion contest 2015 |
| [30] | Dataset is composed of 33 orthorectified image tiles acquired by a near infrared (NIR)-green (G)-red (R) aerial camera, over the town of vaihingen (Germany) |
| [48] | ISPRS Benchmarks1, deepglobe contest2, and spacenet competition3 |

TABLE 4: Dataset description based on modality of high-resolution aerial images.

| Citation number | Dataset description |
|-----------------|---|
| [51] | ISPRS vaihingen and potsdam |
| [70] | Potsdam dataset, vaihingen dataset, and whu dataset |
| [71] | Massachusetts dataset |
| [25] | Vaihingen dataset |
| [75] | Potsdam and vaihingen datasets |
| [59] | ISPRS 2D potsdam dataset |
| [31] | ISPRS 2D semantic labeling contest of potsdam and an area of guangzhou in China |
| [33] | ISPRS vaihingen 2D semantic labeling challenge |
| [82] | ISPRS 2D semantic labeling datasets |

TABLE 5: Dataset description based on remote-sensing images.

| Citation number | Dataset description |
|-----------------|--|
| [35] | Vaihingen and potsdam |
| [50] | International society for photogrammetry and remote-sensing (ISPRS) 2D semantic labeling contest potsdam and inria aerial image labeling dataset |
| [28] | ISPRS |
| [42] | Manual dataset on Tokyo |
| [72] | Big data and computing intelligence contest (BDCI) |
| [54] | Cityscapes |
| [58] | The remote-sensing data collected from field 1 |
| [62] | ISPRS 2D semantic labeling benchmark dataset |
| [63] | 2015 igrss data fusion competition |
| [32] | Rit-18 |
| [79] | Gaofen image dataset (GID) datasets |
| [80] | Potsdam and vaihingen datasets |
| [67] | Luxcarta dataset |
| [81] | ISPRS 2D semantic labeling contest vaihingen dataset and Massachusetts building dataset |
| [68] | Ottawa dataset, stone gate dataset, sardinia dataset, yellow river estuary dataset, barbara dataset, and USGS dataset |
| [49] | Landsat-8 satellite and ISPRS vaihingen challenge dataset |
| [83] | UCM dataset and the deepglobe dataset |

4.3. Datasets for Image Segmentation. In this section, we give a synopsis of a portion of the most generally utilized datasets for image segmentation. We combine these datasets into 3 classifications is 2-dimensional images, 2.5-dimensional RGB-D (complexity+ colour) images, and 3-dimensional

images and give subtle ties with regards to the attributes of each dataset. The recorded datasets have pixel-wise marks, which can be utilized for assessing model execution.

It is worth focusing on that a portion of these works, use augmentation of data to expand the quantity of marked

TABLE 6: Dataset description based on modality OF UN-OCCUPIED aerial vehicles (UAVS) images.

| Citation number | Dataset description |
|-----------------|--|
| [37] | UAV images of three regions (Baoxing, Jiaying, and Chengyang) |
| [39] | China southern power grid company |
| [69] | Manual Uavid dataset and cityscape dataset |
| [43] | Worldview-3 |
| [55] | Rice field located in southern China |
| [60] | Manual Uavid dataset |
| [45] | UAV-based thermal infrared imagery named NPU_CS_UAV_IR_DATA that was collected from some streets of China by using FLIR TAU2 |
| [64] | RGB-D UAV dataset |

TABLE 7: Dataset description based on multiscale, multispatial resolution, satellite images, and other high-resolution images.

| Citation number | Dataset description |
|-----------------|---|
| [76] | Lsun dataset |
| [44] | CVPR2018 deepglobe challenge |
| [78] | Dataset consists of pleiades images collected over five cities in Austria |
| [42] | Polarimetric RADARSAT-2 |
| [66] | High-resolution images from LANDSAT-8 datasets of Google Earth engine |
| [34] | Challenging PASCAL VOC2012 database |

samples, uncommonly the ones which manage little datasets such as in the medical domain. Augmentation of data serves to expand the quantity of preparing tests by applying a set of changes either in the information space, or element space, or now and again both to the images, i.e., both the input image and the segmentation map. Some normal changes incorporate interpretation, reflection, pivot, twisting, scaling, colour space shifting, trimming, and projections onto principal components. Augmentation of data has demonstrated to work on the presentation of the models, particularly when gaining from restricted datasets, like those in medical image investigation.

The common image segmentation research has concentrated on 2-dimensional images. From Figure 9 [91], pink, green, and yellow blocks mention semantic occurrence and panoptic segmentation algorithms, respectively. Therefore, several 2-dimensional image segmentation datasets are existing, and they are PASCAL Visual Object Classes (VOC) [92], PASCAL Context [93], Microsoft Common Objects in Context (MS COCO) [94], Cityscapes [95], ADE20K/MIT Scene Parsing (SceneParse150) [96], SiftFlow [97], Stanford background [98], Berkeley Segmentation dataset [99], Youtube-Objects [100], KITTI [101], Semantic Boundaries Dataset (SBD) [102], PASCAL Part [103], SYNTHIA [104], Dobe's Portrait Segmentation [105], etc., With the obtainability of reasonable range scanners, RGB-D images have become standard in both research and industrial applications. Some of the most standard 2.5-dimensional RGB-D datasets are NYU-D V2 [106], SUN-3D [107], SUN RGB-D[108], UW RGB-D Object Dataset [109], ScanNet [110], etc., Three-dimensional image datasets are standard in robotic, medical image analysis, 3D scene analysis, and construction applications. Three-dimensional images are generally provided via meshes or other

volumetric illustrations, such as point clouds. Some of the standard 3-dimensional datasets are Stanford 2D-3D [111], ShapeNet Core [112], Sydney Urban Objects Dataset [113], etc.

4.4. Frameworks and Benchmark Datasets Employed for Different DL Tasks. Several deep learning frameworks and datasets have been developed in the last few years. Various frameworks and libraries have also been used in order to expedite the work with good results. Through their use, the training process has become easier. Tables 8 and 9 [89] list the most utilized frameworks and libraries and Benchmark datasets.

4.5. Algorithms Comparison Based on Different Datasets. Comparison of different algorithmic features and their results obtained based on clustering methods, conditional random field, PASCAL VOC2012 dataset, CamVid dataset, and MS COCO dataset are tabulated (Tables 10–14).

5. Performance Measures and Best Accuracy Rate Attained by the Conventional Semantic Segmentation Models

5.1. Performance Metrics. An exemplary ought to preferably remain assessed in an assortment of ways, including quantitative precision, speed, and capacity necessities. The majority of previous research has concentrated on parameters for assessing model accuracy. The most commonly used parametric for evaluating the accuracy of segmentation algorithms is summarized below [91, 136]. On benchmarks, to analyze various models, quantitative measurements are

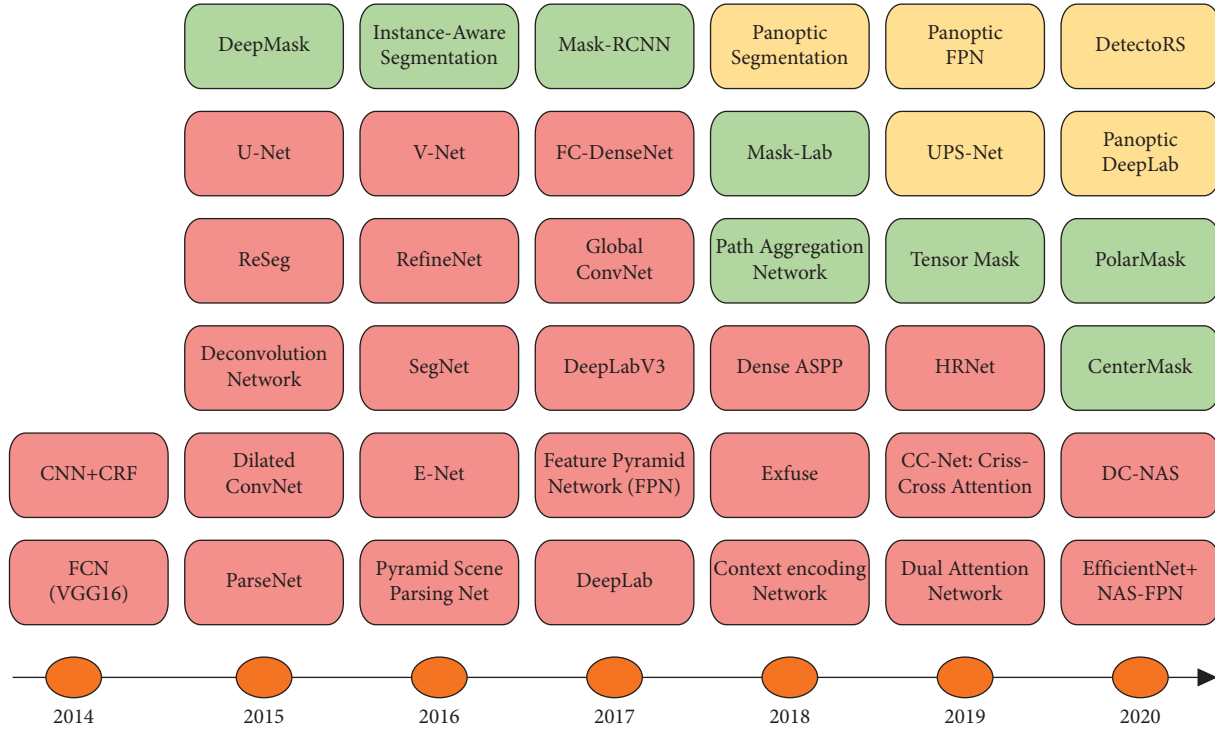


FIGURE 9: The evolution of deep learning-based segmentation algorithms for 2-dimensional images from 2014 to 2020 [91].

TABLE 8: LIST of most common frameworks and libraries.

| Framework | License | Core language | Year of release | Homepages |
|------------|---------------|----------------|-----------------|---|
| TensorFlow | Apache 2.0 | C++ and python | 2015 | https://www.tensorflow.org/ |
| Keras | MIT | Python | 2015 | https://keras.io/ |
| Caffe | BSD | C++ | 2015 | http://caffe.berkeleyvision.org/ |
| MatConvNet | Oxford | MATLAB | 2014 | http://www.vlfeat.org/matconvnet/ |
| MXNet | Apache 2.0 | C++ | 2015 | https://github.com/dmic/mxnet |
| CNTK | MIT | C++ | 2016 | https://github.com/Microsoft/CNTK |
| Theano | BSD | Python | 2008 | http://deeplearning.net/software/theano/ |
| Torch | BSD | C and lua | 2002 | http://torch.ch/ |
| DL4j | Apache 2.0 | Java | 2014 | https://deeplearning4j.org/ |
| Gluon | AWS microsoft | C++ | 2017 | https://github.com/gluon-api/gluon-api/ |
| OpenDeep | MIT | Python | 2017 | http://www.opendeep.org/ |

utilized, and the visual nature of the model yields significance in figuring out.

- (i) Pixel accuracy (PA): basically, pixel accuracy states the ratio of correctly classified pixels to the total quantity of pixels. Pixel accuracy is known for $N + 1$ classes as

$$PA = \frac{\sum_{i=0}^n a_{ii}}{\sum_{i=0}^n \sum_{j=0}^n a_{ij}}, \quad (1)$$

where a_{ij} is the quantity of pixels of class I predicted as belonging to class j .

- (ii) Average/mean pixel accuracy (MPA): mean pixel accuracy has marginally further developed, in

which the ratio of correct pixels is computed in a per-class basis and then averaged over the total number of classes:

$$MPA = \frac{1}{N + 1} \sum_{i=0}^n \frac{a_{ii}}{\sum_{j=0}^n a_{ij}} \quad (2)$$

- (iii) Intersection over union (IoU): this is quite possibly the most generally utilized measurement in semantic segmentation. It is determined as the area of intersection of the predicted division map and the ground truth divided by the area of the union of the predicted segmentation map and the ground truth:

TABLE 9: Benchmark datasets.

| Dataset | No. of classes | Applications | Link to dataset |
|---------------------------|----------------|---|---|
| ImageNet | 1000 | Image classification, object localization, object detection, etc. | http://www.image-net.org/ |
| CIFAR10/100 | 10/100 | Image classification | https://www.cs.toronto.edu/~kriz/cifar.html |
| MNIST | 10 | Classification of handwritten digits | http://yann.lecun.com/exdb/mnist/ |
| Pascal VOC | 20 | Image classification, segmentation, and object detection | http://host.robots.ox.ac.uk/pascal/VOC/voc2012/ |
| Microsoft COCO | 80 | Object detection and semantic segmentation | https://cocodataset.org/#home |
| YFCC100 M | 8M | Video and image understanding | http://projects.dfki.unikl.de/yfcc100m/ |
| YouTube-8M | 4716 | Video classification | https://research.google.com/youtube8m/ |
| UCF-101 | 101 | Human action detection | https://www.crcv.ucf.edu/data/UCF101.php |
| Kinetics | 400 | Human action detection | https://deepmind.com/research/open-source/kinetics |
| Google open images | 350 | Image classification, segmentation, and object detection | https://storage.googleapis.com/openimages/web/index.html |
| CalTech101 | 101 | Classification | http://www.vision.caltech.edu/Image_Datasets/Caltech101/ |
| Labeled faces in the wild | - | Face recognition | http://vis-www.cs.umass.edu/lfw/ |
| MIT-67 scene dataset | 67 | Indoor scene recognition | http://web.mit.edu/torralba/www/indoor.htm |

TABLE 10: Comparison of algorithms based on clustering methods (%).

| Citations | Algorithm features | Datasets | Segmentation results |
|-----------|--|--------------------|----------------------|
| [114] | Weak supervision, spectral clustering, and discriminative clustering | MSRC-21 | 70 (mA) |
| [115] | Weak supervision and double-end clustering | MSRC-21 | 52.9 (mIoU) |
| [116] | FCM algorithm and grouping algorithm | LABLEME | 26 (mA) |
| | | Self-built dataset | 2.2 (mError) |

TABLE 11: Comparison of algorithms based on conditional random field (%).

| Citations | Algorithm features | Datasets | Segmentation results |
|-----------|--|--------------------|----------------------|
| [117] | CRF, dense features, and high-order potential energy | MSRC-21 | 75.8 (mA) |
| [118] | CRF and joint-boosting algorithm | MSRC-21 | 71.6 (mA) |
| [119] | CRF and interactive | Self-built dataset | 95.3 (mA) |
| [120] | CRF and high-order energy items | MSRC-21 | 72.2 (PA) |
| [121] | CRF and maximum flow-minimum cut | MSRC-21 | 0.7 s (time) |

TABLE 12: Comparison of algorithms based on PASCAL VOC2012 dataset (%).

| Citations | Algorithm features | Datasets | Segmentation results |
|-----------|---|----------------|----------------------|
| [122] | Convolution and deconvolution neural networks | PASCAL VOC2012 | 63.6 (mIoU) |
| [123] | Deconvolution networks | PASCAL VOC2012 | 72.5 (mIoU) |
| [124] | PSPNet | PASCAL VOC2012 | 82.6 (mIoU) |
| [125] | RefineNet | PASCAL VOC2012 | 83.4 (mIoU) |
| [126] | Decoupled deep neural networks | PASCAL VOC2012 | 66.6 (mIoU) |

TABLE 13: Comparison of algorithms based on CAMVID dataset (%).

| Citations | Algorithm features | Datasets | Segmentation results |
|-----------|--|----------|----------------------|
| [127] | SegNet | CamVid | 60.1 (mIoU) |
| [128] | Densely connected convolutional networks | CamVid | 66.9 (mIoU) |
| [129] | ENet | CamVid | 51.3 (mIoU) |
| [130] | Gated feedback refinement networks | CamVid | 68.0 (mIoU) |
| [131] | Generative adversarial networks | CamVid | 58.2 (mIoU) |

TABLE 14: Comparison of algorithms based on MS COCO dataset (%).

| Citations | Algorithm features | Datasets | Segmentation results |
|-----------|----------------------------|----------|----------------------|
| [132] | Mask R-CNN | MS COCO | 37.1 (PA) |
| [133] | FCIS | MS COCO | 59.9 (PA) |
| [134] | Multitask network cascades | MS COCO | 51.5 (PA) |
| [135] | Residual networks | MS COCO | 48.4 (PA) |

$$IoU = J(P, Q) = \frac{|P \cap Q|}{|P \cup Q|}, \quad (3)$$

where P = true segmentation map and Q = predicted segmentation maps.

The value of intersection over union lies between 0 and 1.

- (iv) Mean-IoU: mean intersection over union is an alternative standard metric defined by average intersection over union across entire modules. It is commonly used in reporting the performance of contemporary segmentation algorithms [91].
- (v) Precision/recall: for numerous classical image segmentation models, precision and recall are the standard metrics for recording. Definition for precision and recall for every class is as follows:

$$Precision = \frac{TP \text{ Fraction}}{TP \text{ Fraction} + FP \text{ Fraction}}, \quad (4)$$

$$Recall = \frac{TP \text{ Fraction}}{TP \text{ Fraction} + FN \text{ Fraction}},$$

where TP = True Positive, FP = False Positive, and FN = False Negative. Usually, we are attentive in a united form of precision and recall rates.

- (vi) F1 score: F1 score is also the standard metric and defined by the harmonic mean of precision and recall:

$$F1 - Score = \frac{2Precision \times Recall}{Precision + Recall}. \quad (5)$$

- (vii) Dice coefficient: Dice coefficient is an alternative standard metric used in medical image analysis for image segmentation, defined by “twice the overlap area of predicted and ground truth maps, divided by the total number of pixels in both images. The Dice coefficient is very identical to the IoU” [91]:

$$Dice = \frac{2|P \cap Q|}{|P| + |Q|}. \quad (6)$$

While practical to Boolean data, the Dice coefficient is nearly equal to the F1 score:

$$Dice = \frac{2TP}{2TP + FP + FN} = F1score, \quad (7)$$

where TP indicates True Positive Fraction, FP indicates False Positive Fraction, and FN indicates False Negative Fraction.

- (viii) Frequency weighted mIoU: over the raw mIoU, frequency weighted mean intersection over union is an improved which weights each class importance depending on their appearance frequency [136]:

$$FWmIoU = \frac{1}{\sum_{i=0}^K \sum_{j=0}^K a_{ij}} \sum_{i=0}^K \frac{\sum_{j=0}^K a_{ij} a_{ii}}{\sum_{j=0}^K a_{ij} + \sum_{j=0}^K a_{ji} - a_{ii}}. \quad (8)$$

- (ix) Jaccard index: the Jaccard index, commonly known as the Jaccard similarity coefficient, is a statistic used to assess the similarity between sample sets. The measurement stresses similarity between finite sample sets and is officially defined as the intersection size divided by the sample set union size. The mathematical representation of the index is written as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (9)$$

- (x) Confusion matrix: a Confusion matrix is an $N \times N$ matrix that is used to assess the effectiveness of a classification model, where N is the number of target classes. Figure 10 represents the confusion matrix. The matrix compares the actual goal values to the machine learning model’s predictions. This provides us with a comprehensive picture of how well our classification model is working and the kind of errors it is producing. For a binary classification task, we would have a 2×2 matrix with four values, as illustrated in figure [137].

Let us decode the matrix. The target variable has two values: positive or negative. The columns represent the actual values of the target variable. The rows represent the predicted values of the target variable.

- (xi) Kappa coefficient: it is used to assess the level of agreement between two human evaluators or raters (for example, psychologists) when assessing topics (patients). The machine learning community then “appropriated” it to quantify categorization performance. The kappa score, also known as Cohen’s kappa coefficient [138], is named after Jacob Cohen, an American statistician and psychologist who produced the foundational study on the subject. This measure is also known as Cohen’s kappa and the kappa statistic. To compute the kappa score, it is

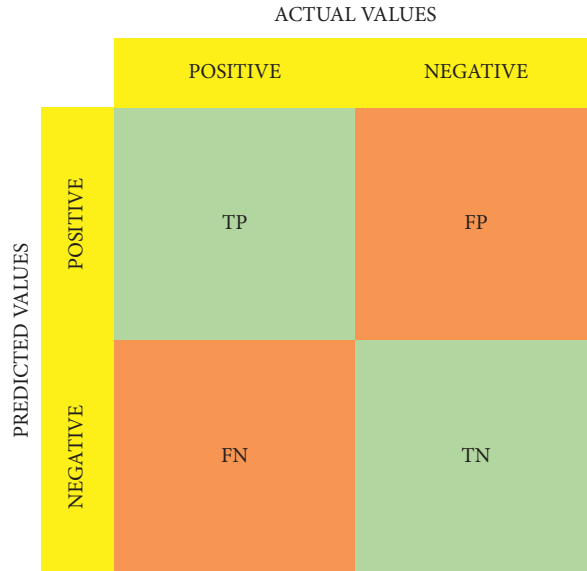


FIGURE 10: Confusion matrix [137].

| | | Professor A | | |
|-------------|--------|-------------|----|--------|
| | | Accept | WL | Reject |
| Professor B | Accept | 4 | 6 | 3 |
| | WL | 1 | 2 | 0 |
| | Reject | 1 | 2 | 6 |

FIGURE 11: Summarization of ratings for kappa coefficient [138].

TABLE 15: The best performances were obtained by diverse semantic segmentation models.

| Citations | Performance metric | Best performance in percentage |
|-----------|-------------------------|--------------------------------|
| [64] | Overall accuracy | 97.00 |
| [68] | F1 score | 99.41 |
| [62] | Intersection over union | 96.50 |
| [24] | Recall | 99.84 |

convenient to first summarize the ratings in a matrix shown in Figure 11.

The columns show the ratings by professor A. The rows show the ratings by Professor B. The value in each cell is the number of candidates with the corresponding ratings by the two professors.

The performance metrics employed for analyzing the diverse semantic segmentation models through deep learning is given in Table 15. From the set of research works, 63.3% of the works use OA, 48.3% of the contributions use F1 score, and 25% of the works consider recall and precision measures, respectively. mIoU metric is taken in 28.3% of the research works, 5% of the papers use Jaccard index, kappa coefficient, and dice coefficient the performance metric,

confusion matrix, and PA are considered in 4% of the research works, respectively, and 23.3% of the contributions consider IoU measure. Furthermore, some of the additional measures are also taken for evaluating the efficiency of semantic segmentation, which are FWIoU, MCC, average accuracy, etc.

5.2. *Best Performance Measures.* The best performance measures obtained by diverse semantic segmentation models are depicted in Figure 12. From this comprehensive survey, Figure 8(a) represents contributions such as [32, 64] to get 97% as the highest accuracy rate than others. Secondly, the work in [50] obtains 94.49%, and the research works

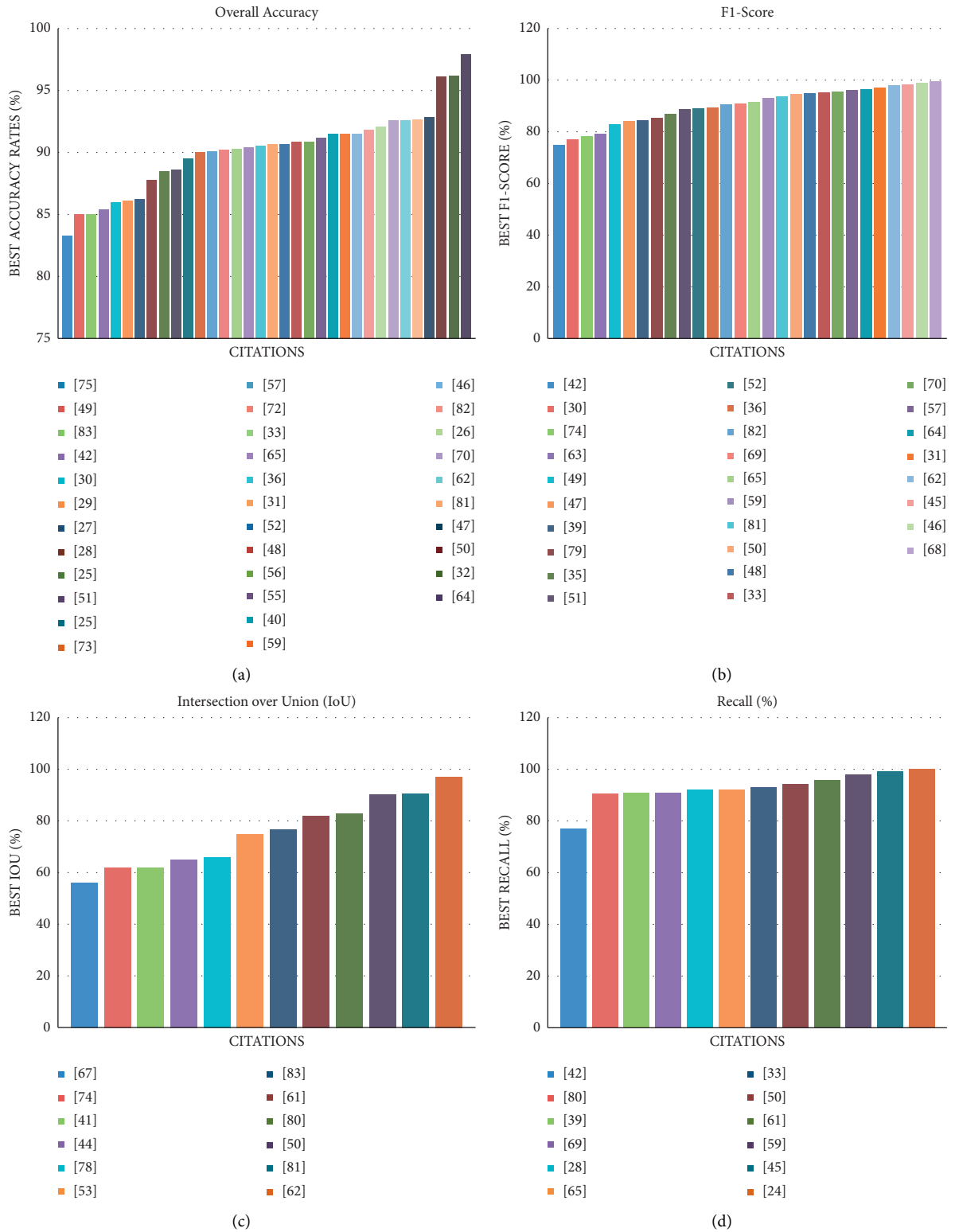


FIGURE 12: Best performance measures obtained by state-of-the-art semantic segmentation models: (a) best accuracy vs. citations; (b) best F1 score vs. citations, (c) best IoU vs. citations, and (d) best recall vs. citations.

TABLE 16: Continued.

| Citations | Overall accuracy (OA) | F1 score | Recall | Precision | Mean intersection over union (mIoU) | Jaccard index | Dice coefficient | Confusion matrix | Intersection over union (IoU) | Kappa coefficient | Pixel accuracy (PA) | Miscellaneous techniques |
|-----------|-----------------------|----------|--------|-----------|-------------------------------------|---------------|------------------|------------------|-------------------------------|-------------------|---------------------|---|
| [60] | — | — | — | — | ✓ | — | — | — | — | — | — | — |
| [61] | — | — | ✓ | — | — | — | — | — | ✓ | — | — | — |
| [62] | ✓ | ✓ | — | — | — | — | — | — | ✓ | — | — | — |
| [63] | ✓ | ✓ | — | — | ✓ | — | — | — | — | — | — | — |
| [64] | ✓ | ✓ | — | — | ✓ | — | — | — | — | — | — | — |
| [65] | ✓ | ✓ | ✓ | — | — | — | — | — | — | — | — | — |
| [66] | — | — | — | — | — | — | — | — | — | — | — | — |
| [67] | — | — | — | — | — | — | — | — | ✓ | — | — | Weighted IoU Running time and execution time Percentage correct classification and overall error rate |
| [68] | — | ✓ | — | — | — | — | — | — | — | ✓ | — | — |
| [69] | — | ✓ | ✓ | — | ✓ | — | — | — | — | — | — | — |
| [70] | ✓ | ✓ | — | — | ✓ | — | — | — | — | — | — | — |
| [71] | — | — | — | — | ✓ | — | — | — | — | — | — | Matthews correlation coefficient (MCC) |
| [72] | ✓ | — | — | — | — | — | — | — | — | — | — | — |
| [73] | ✓ | — | — | — | — | ✓ | — | — | — | — | — | — |
| [74] | — | ✓ | — | — | — | — | — | — | ✓ | — | — | — |
| [75] | ✓ | — | — | — | ✓ | — | — | — | — | — | — | — |
| [76] | — | — | — | — | — | — | — | — | — | — | — | Binary cross entropy (BCE) Binary accuracy, and boundary F1 score Training time |
| [77] | — | — | — | — | ✓ | — | — | — | — | — | — | — |
| [78] | ✓ | — | — | — | — | — | — | — | ✓ | — | — | — |
| [79] | — | ✓ | — | — | ✓ | — | — | ✓ | — | — | — | — |
| [80] | — | — | ✓ | — | — | — | — | — | ✓ | — | — | Training and prediction time |
| [81] | ✓ | ✓ | — | — | — | — | — | — | ✓ | — | — | — |
| [82] | ✓ | ✓ | — | — | — | — | — | — | ✓ | — | — | — |
| [83] | ✓ | — | — | — | — | — | — | — | ✓ | — | — | Frequency weighted IoU (FWIoU) |

such as [26, 43, 47, 56, 59, 81] attain 92.63% accuracy rate when compared with other works. The best performances for some of the metrics such as overall accuracy, F1 score, intersection over union, and recall were noted and tabulated as shown in Table 16.

5.3. Research Gaps and Challenges. In recent decades, several semantic segmentation approaches have been designed for different applications such as surveillance systems, traffic monitoring, and analysis on environmental changes. However, manual segmentation methods are time tedious and complex one. Thus, an automated semantic segmentation of aerial images is emerged as the recent hot topic [139]. On the contrary, the semantic segmentation of aerial images is a complex task due to several constraints such as demand for pixel-level accuracy, nonconventional data, and lack of training examples. Each object in the remote-sensing images specifies important information, which requires to be precisely categorized from the neighboring ones. Numerous works have been proposed for solving this problem, which has been focused on improving regularization and FCN such as object boundary details. More numbers of public datasets have been considered for evaluating the performance of the deep learning approaches. Here, infrared and colour satellite images have gained noteworthy performance that is more equivalent to image sets utilized in the portrait and scenic computer vision tasks. From the comprehensive review, the public datasets such as ISPRS datasets get more importance that has guaranteed the implementation of deep learning approaches for facilitating the semantic segmentation [140]. Though, the semantic segmentation on different data or imaging modality and analysis metrics make evaluation complex. Moreover, handling of different modality of remote-sensing images such as UAV, hyperspectral images, and infrared and RGB images are complex to process. It results in lack of accuracy to estimate the nonconventional data.

Sometimes, a large volume of data and a lack of training examples pose complexities in aerial imaging applications. Conversely, it is much more challenging due to the nonconventional data sources such as LiDAR, hyperspectral images, and synthetic aperture radar images [141]. When the deep learning techniques are utilized for processing the nonconventional remote-sensing datasets with labels, it creates complexities. These deep learning methods suffer from the lack of training dataset. Any deep learning model may need a huge set of training images due to the number of classes and complications of the problem [142]. Moreover, the utilization of deep learning is more complicated while considering the expensive and additional remote-sensing data collection [143]. Thus, different augmentation approaches are mostly employed for increasing the variation and number of the dataset. Consequently, the most common datasets called “ISPRS’s 2D labeling dataset and IEEE’S GRRS dataset” have been attempted for addressing the data inefficiency through offering the very high-resolution remote-sensing images gathered from UAVs [141].

An additional limitation of deep learning-based semantic segmentation is the necessity of a high number of label dataset, which generally requires manual annotation. This issue has also considerably been solved through public datasets through offering the annotations [142]. However, it is still tedious while taking the own or manual datasets. Existing research works have utilized conventional approaches for producing the annotations. Similarly, the label dataset can be created with the feature of pretrained models. From the meta-analysis results, the deep learning provides enhanced efficiency and shows the superior performance when compared to conventional approaches [143]. Many challenges of deep learning-adopted techniques have been solved and reduced in recent decades, which have to increase the performance. The future research areas in the semantic segmentation of aerial images can integrate the well-known deep learning models with hybrid or new variant metaheuristic approaches. As the deep learning-based semantic segmentation models have emerged their future prospects, it has to create a new future scope on different applications using intelligent algorithms for increasing the accuracy rate [144]. In the future, it has to solve the nonconventional data and labeling problems while preparing a new datasets. Thus, this research helps the researchers to understand the semantic segmentation model with several other possibilities for coming up with new future research perspectives.

6. Conclusion

This study has presented a comprehensive review on conventional semantic segmentation models through deep learning approaches. For this purpose, a set of research works has been taken from recent years. This study has given the information regarding different machine learning or deep learning techniques used, simulation tools, performance metrics, features and challenges of conventional semantic segmentation models, different imaging modalities, and the datasets utilized. Finally, the research gaps and limitations were analyzed for exploring a future research perspective of semantic segmentation systems. On the whole, this study has offered the detailed information on semantic segmentation models, which are helpful for assisting the researchers to present a semantic segmentation model in the upcoming years.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the help of the Vellore Institute of Technology, Vellore, India, for giving excellent assets to this work. Also, the authors would like to thank the individual copyright holders for consent conceded to incorporate referred figures in this work.

References

- [1] M. Zhang, G. Xu, K. Chen, M. Yan, and X. Sun, "Triplet-based semantic relation learning for aerial remote sensing image change detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 266–270, 2019.
- [2] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6054–6068, 2017.
- [3] X. Zhang, W. Ma, C. Li, J. Wu, X. Tang, and L. Jiao, "Fully convolutional network-based ensemble method for road extraction from aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 10, pp. 1777–1781, 2020.
- [4] R. Manickam, S. Kumar Rajan, C. Subramanian, A. Xavi, G. J. Eanoch, and H. R. Yesudhas, "Person identification with aerial imagery using SegNet based semantic segmentation," *Earth Science Informatics*, vol. 13, no. 4, pp. 1293–1304, 2020.
- [5] B. Du, Z. Zhao, X. Hu et al., "Landslide susceptibility prediction based on image semantic segmentation," *Computers & Geosciences*, vol. 155, 2021.
- [6] G. Can, D. Mantegazza, G. Abbate, S. Chappuis, and A. Giusti, "Semantic segmentation on Swiss3DCities: A benchmark study on aerial photogrammetric 3D pointcloud dataset," *Pattern Recognition Letters*, vol. 150, pp. 108–114, 2021.
- [7] G. Mandlbürger, M. Kölle, H. Nübel, and U. Soergel, "BathyNet: A deep neural network for water depth mapping from multispectral aerial images," *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 89, pp. 71–89, 2021.
- [8] Bo Li, C. Chen, S. Dong, and J. Qiao, "Transmission line detection in aerial images: An instance segmentation approach based on multitask neural networks," *Signal Processing: Image Communication*, vol. 96, August 2021.
- [9] P. K. R. Maddikunta, S. Hakak, M. Alazab et al., "Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17608–17619, 2021.
- [10] Q. Gao and X. Shen, "ThickSeg: Efficient semantic segmentation of large-scale 3D point clouds using multi-layer projection," *Image and Vision Computing*, vol. 108, Article ID 104161, 2021.
- [11] A. S. Edun, K. Perry, J. B. Harley, and C. Deline, "Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough transform," *Applied Energy*, vol. 298, Article ID 117273, 2021.
- [12] P. Wei, D. Chai, T. Lin, C. Tang, M. Du, and J. Huang, "Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 174, pp. 198–214, 2021.
- [13] R. Ch, G. Srivastava, T. R. Gadekallu, P. K. R. Maddikunta, and S. Bhattacharya, "Security and privacy of UAV data using blockchain technology," *Journal of Information Security and Applications*, vol. 55, Article ID 102670, 2020.
- [14] G. Bhattacharjee and S. K. Pujari, "Semantic segmentation of aerial images survey," *International Journal of Applied Information Systems (IJ AIS)*, vol. 12, no. No. 5, pp. 28–34, 2017.
- [15] H. Xiu, P. Vinayaraj, K.-S. Kim, R. Nakamura, and W. Yan, "3D semantic segmentation for high-resolution aerial survey derived point clouds using deep learning (demonstration)," in *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 588–591, Washington, Seattle, November 2018.
- [16] T. P. Singh, R. R. Singh, Himanshu, A. Mishra, and N. Sharma, "Semantic segmentation of satellite images: A survey," *International Research Journal of Engineering and Technology (IRJET)*, vol. 07, no. Issue. 12, 2020.
- [17] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Systems with Applications*, vol. 169, 2021 [18] Juhong Wang, Bin Liu & Kun Xu "Semantic segmentation of high-resolution images," *Science China Information Sciences*, vol. 60, 2017, Article ID 114417.
- [18] M. Chouai, M. Merah, and M. Mimi, "Correction to: CH-Net: deep adversarial autoencoders for semantic segmentation in X-ray images of cabin baggage screening at airports," *Journal of Transportation Security*, vol. 13, no. 1-2, p. 91, 2020.
- [19] J. Wang, B. Liu, and K. Xu, "Semantic segmentation of high-resolution images," *Science China Information Sciences*, vol. 60, Article ID 123101, 2017.
- [20] S. Gupta, P. Arbeláez, R. Girshick, and J. Malik, "Indoor scene understanding with RGB-D images: Bottom-up segmentation, object detection and semantic segmentation," *International Journal of Computer Vision*, vol. 112, no. 2, pp. 133–149, 2015.
- [21] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, T. R. Gadekallu, and G. Srivastava, "SP2F: A secured privacy-preserving framework for smart agricultural Unmanned Aerial Vehicles," *Computer Networks*, vol. 187, p. 1, Article ID 07819, 2021.
- [22] N. M. Balamurugan, S. Mohan, M. Adimoolam, A. John, and W. Wang, "DOA tracking for seamless connectivity in beamformed IoT-based drones," *Computer Standards & Interfaces*, vol. 79, Article ID 103564, 2022.
- [23] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 137–178, 2021.
- [24] S. Saito, R. Arai, and Y. Aoki, "Seamline determination based on semantic segmentation for aerial image mosaicking," *IEEE Access*, vol. 3, pp. 2847–2856, 2015.
- [25] D. Marmanis, J. D. Wegner, S. Galliani, K. Schindler, M. Datcu, and U. Stilla, "Semantic segmentation OF aerial images with an ensemble OF CNNs," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 473–480, 2016.
- [26] A. Holliday, M. Barekatin, J. Laurmaa, C. Kandaswamy, and H. Prendinger, "Speedup of deep learning ensembles for semantic segmentation using a model compression technique," *Computer Vision and Image Understanding*, vol. 164, pp. 16–26, 2017.
- [27] K. Chen, K. Fu, M. Yan, X. Gao, X. Sun, and X. Wei, "Semantic segmentation of aerial images with shuffling convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 173–177, Feb2018.
- [28] B. Yu, L. Yang, and F. Chen, "Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 9, pp. 3252–3261, 2018.
- [29] K. Chen, M. Weinmann, X. Sun et al., "Semantic segmentation OF aerial imagery via multi-scale shuffling convolutional neural networks with deep supervision," *ISPRS Annals*

- of the *Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-1, pp. 29–36, 2018.
- [30] M. Volpia and D. Tuia, “Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 144, pp. 48–60, 2018.
- [31] Y. Sun, X. Zhang, Q. Xin, and J. Huang, “Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 143, pp. 3–14, 2018.
- [32] R. Kemker, C. Salvaggio, and C. Kanan, “Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, no. Part A, pp. 60–77, 2018.
- [33] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, “Classification with an edge: Improving semantic image segmentation with boundary detection,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 135, pp. 158–172, 2018.
- [34] D. M. Vo and S.-W. Lee, “Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions,” *Multimedia Tools and Applications*, vol. 77, no. 14, pp. 18689–18707, 2018.
- [35] C. Peng, Y. Li, L. Jiao, Y. Chen, and R. Shang, “Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 2612–2626, 2019.
- [36] H. Luo, C. Chen, L. Fang, X. Zhu, and L. Lu, “High-resolution aerial images semantic segmentation using deep fully convolutional network with Channel attention mechanism,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3492–3507, 2019.
- [37] Y. Li, B. Peng, L. He, K. Fan, and L. Tong, “Road segmentation of unmanned aerial vehicle remote sensing images using adversarial network with multiscale context aggregation,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2279–2287, 2019.
- [38] S. M. Azimi, P. Fischer, M. Korner, and P. Reinartz, “Aerial LaneNet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2920–2938, 2019.
- [39] L. Wang, Z. Chen, D. Hua, and Z. Zheng, “Semantic segmentation of transmission lines and their accessories based on UAV-taken images,” *IEEE Access*, vol. 7, pp. 80829–80839, 2019.
- [40] Z. Cao, K. Fu, X. Lu et al., “End-to-End DSM fusion networks for semantic segmentation in high-resolution aerial images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 11, pp. 1766–1770, 2019.
- [41] T. Nguyen, J. Wozencraft, C. J. Taylor et al., “MAVNet: An effective semantic segmentation micro-network for MAV-based tasks,” *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3908–3915, 2019.
- [42] Z. Guo, G. Wu, X. Song et al., “Super-resolution integrated building semantic segmentation for multi-source remote sensing imagery,” *IEEE Access*, vol. 7, pp. 99381–99397, 2019.
- [43] D. M. Igonin and Yu. V. Tiumentseva, “Comparative efficiency analysis for various neuroarchitectures for semantic segmentation of images in remote sensing applications,” *Optical Memory & Neural Networks*, vol. 28, no. 4, pp. 306–320, 2019.
- [44] M. Wu, C. Zhang, J. Liu, L. Zhou, and X. Li, “Towards accurate high resolution satellite image semantic segmentation,” *IEEE Access*, vol. 7, pp. 55609–55619, 2019.
- [45] M. K. Masouleh and R. Shah-Hosseini, “Development and evaluation of a deep learning model for real-time ground vehicle semantic segmentation from UAV-based thermal infrared imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 155, pp. 172–186, 2019.
- [46] N. Audebert, A. Boulch, B. Le Saux, and S. Lefèvre, “Distance transform regression for spatially-aware deep semantic segmentation,” *Computer Vision and Image Understanding*, vol. 189, Article ID 102809, 2019.
- [47] F. Mohammadimanesha, B. Salehic, M. Mahdianparia, E. Gill, and M. Molinier, “A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 151, pp. 223–236, 2019.
- [48] L. Ding, T. Tang, and L. Bruzzone, “Improving semantic segmentation of aerial images using patch-based attention,” 2019, <https://arxiv.org/abs/1911.08877>.
- [49] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathien, and P. Vateekul, “Semantic segmentation on remotely sensed images using an enhanced global convolutional network with Channel attention and domain specific transfer learning,” *Remote Sensing*, vol. 11, no. Issue. 1, p. 83, 2019.
- [50] W. Liu, Y. Zhang, H. Fan, Y. Zou, and Z. Cui, “A new multi-channel deep convolutional neural network for semantic segmentation of remote sensing image,” *IEEE Access*, vol. 8, pp. 131814–131825, 2020.
- [51] L. Mou, Y. Hua, and X. X. Zhu, “Relation matters: Relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7557–7569, 2020.
- [52] F. Wang, S. Piao, and J. Xie, “CSE-HRNet: A context and semantic enhanced high-resolution network for semantic segmentation of aerial imagery,” *IEEE Access*, vol. 8, pp. 182475–182489, 2020.
- [53] G. Martinez-Soltero, A. Y. Alanis, N. Arana-Daniel, and C. Lopez-Franco, “Semantic segmentation for aerial mapping,” *Mathematics*, vol. 8, no. 9, p. 1456, 2020.
- [54] W. Jiawe, H. Xiong, H. Wang, and X. Nian, “ADSCNet: Asymmetric depthwise separable convolution for semantic segmentation in real-time,” *Applied Intelligence*, vol. 50, no. issue. 12, pp. 1045–1056, 2020.
- [55] J. Deng, Z. Zhong, H. Huang, Y. Lan, Y. Han, and Y. Zhang, “Lightweight semantic segmentation network for real-time weed mapping using unmanned aerial vehicles,” *Applied Sciences*, vol. 10, no. 20, p. 7132, 2020.
- [56] R. Niu, X. Sun, Y. Tian, W. Diao, K. Chen, and K. Fu, “Hybrid multiple attention network for semantic segmentation in aerial images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2021.
- [57] D. Chai, S. Newsam, and J. Huang, “Aerial image semantic segmentation using DCNN predicted distance maps,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 309–322, 2020.

- [58] Z. Song, Z. Zhang, S. Yang, D. Ding, and J. Ning, "Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging," *Computers and Electronics in Agriculture*, vol. 179, Article ID 105812, 2020.
- [59] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.
- [60] L. Ye, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "UAVid: A semantic segmentation dataset for UAV imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 165, pp. 108–119, 2020.
- [61] L. C. L. Bianco, J. Beltrán, G. F. López, F. García, and A. Al-Kaff, "Joint semantic segmentation of road objects and lanes using Convolutional Neural Networks," *Robotics and Autonomous Systems*, vol. 133, Article ID 103623, 2020.
- [62] Li Mi and Z. Chen, "Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 140–152, 2020.
- [63] K. Zhang, Yu Han, J. Chen, Z. Zhang, and S. Wang, "Semantic segmentation for remote sensing based on RGB images and lidar data using model-agnostic meta-learning and partial Swarm optimization," *IFAC-papersOnLine*, vol. 53, no. Issue 5, pp. 397–402, 2020.
- [64] W. Boonpook, Y. Tan, and Bo Xu, "Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry," *International Journal of Remote Sensing*, vol. 42, no. Issue 1, pp. 1–19, 2021.
- [65] H. Yang, Bo Yu, J. Luo, and F. Chen, "Semantic segmentation of high spatial resolution images with deep neural networks," *GIScience and Remote Sensing*, vol. 56, no. Issue. 5, 2019.
- [66] A. Mehra, N. Jain, and H. S. Srivastava, "A novel approach to use semantic segmentation based deep learning networks to classify multitemporal SAR data," *Geocarto International*, vol. 37, no. 1, pp. 163–178, 2020.
- [67] O. Tasar, S. L. Happy, Y. Tarabalka, and P. Alliez, "Color-MapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7178–7193, 2020.
- [68] N. Venugopal, "Automatic semantic segmentation with DeepLab dilated learning network for change detection in remote sensing images," *Neural Processing Letters*, vol. 51, no. 3, pp. 2355–2377, 2020.
- [69] S. Girisha, U. Verma, M. M. Manohara Pai, and R. M. Pai, "UVid-net: Enhanced semantic segmentation of UAV aerial videos by embedding temporal information," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4115–4127, 2021.
- [70] J. Huang, X. Zhang, Y. Sun, and Q. Xin, "Attention-guided label refinement network for semantic segmentation of very high resolution aerial orthoimages," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4490–4503, 2021.
- [71] A. Abdollahi, B. Pradhan, G. Sharma, K. N. A. Maulud, and A. Alamri, "Improving road semantic segmentation using generative adversarial network," *IEEE Access*, vol. 9, pp. 64381–64392, 2021.
- [72] M. Alam, J.-F. Wang, C. Guangpei, L. Yunrong, and Y. Chen, "Convolutional neural network for the semantic segmentation of remote sensing images," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 200–215, 2021.
- [73] A. Anagnostis, A. C. Tagarakis, D. Kateris et al., "Orchard mapping with deep learning semantic segmentation," *Sensors*, vol. 21, no. 11, 2021.
- [74] P. Li, H. Zhang, Z. Guo et al., "Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning," *Advances in applied energy*, vol. 4, Article ID 100057, 2021.
- [75] F. Wang, X. Luo, Q. Wang, and Lu Li, "Aerial-BiSeNet: A real-time semantic segmentation network for high resolution aerial imagery Author links open overlay," *Chinese Journal of Aeronautics*, vol. 34, no. 9, pp. 47–59, 2021, Available online.
- [76] L. Vasquez-Espinoza, M. Castillo-Cara, and L. Orozco-Barbosa, "On the relevance of the metadata used in the semantic segmentation of indoor image spaces," *Expert Systems with Applications*, vol. 184, Article ID 115486, 2021.
- [77] Z. Chen, D. Ting, R. Newbury, and C. Chen, "Semantic segmentation for partially occluded apple trees based on deep learning," *Computers and Electronics in Agriculture*, vol. 181, Article ID 105952, 2021.
- [78] O. Tasar, A. Giros, Y. Tarabalka, P. Alliez, and S. Clerc, "DAugNet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1067–1081, 2021.
- [79] F. X. Li, X. Lyu, H. Gao et al., "Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images," *International Journal of Remote Sensing*, vol. 42, no. 9, pp. 3583–3610, 2021.
- [80] Yi-Z. Jiang, "Semantic segmentation of remote sensing image based on convolutional neural network and mask generation," *Mathematical Problems in Engineering*, vol. 2021, 2021.
- [81] W. Liu, Y. Zhang, J. Yan, Y. Zou, and Z. Cui, "Semantic segmentation network of remote sensing images with dynamic loss fusion strategy," *IEEE Access*, vol. 9, pp. 70406–70418, 2021.
- [82] Z. Zheng, X. Zhang, P. Xiao, and Z. Li, "Integrating gate and attention modules for high-resolution image semantic segmentation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4530–4546, 2021.
- [83] S. Ouyang and Y. Li, "Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery," *Remote Sensing*, vol. 13, no. Issue. 1, p. 119, 2021.
- [84] S. Mouakket and A. M. Bettayeb, "Investigating the factors influencing continuance usage intention of Learning management systems by university instructors: The Blackboard system case," *International Journal of Web Information Systems*, vol. 11, no. 4, pp. 491–509, 2015.
- [85] F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel, "Ensemble adversarial training: Attacks and defenses," arXiv preprint arXiv:1705.07204, 2017.
- [86] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, 2019.
- [87] Z. Ji, Z. C. Lipton, and C. Elkan, "Differential privacy and machine learning: A survey and review," arXiv preprint arXiv:1412.7584, 2014.
- [88] D. Zhang, X. Chen, D. Wang, and J. Shi, "A survey on collaborative deep learning and privacy-preserving," in *Proceedings of the 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, pp. 652–658, IEEE, Guangzhou, China, June 2018.

- [89] L. Alzubaidi, J. Zhang, A. J. Humaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, no. 1, pp. 53–74, 2021.
- [90] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, Boston, MA, USA, June 2015.
- [91] M. Shervin, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [92] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [93] R. Mottaghi, X. Chen, X. Liu et al., "The role of context for object detection and semantic segmentation in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 891–898, Columbus, OH, USA, June 2014.
- [94] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft coco: Common objects in context," in *Proceedings of the European conference on computer vision*, Springer, Cham, pp. 740–755, 2014.
- [95] M. Cordts, O. Mohamed, S. Ramos et al., "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, Las Vegas, NV, USA, June 2016.
- [96] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 633–641, Honolulu, HI, USA, July 2017.
- [97] Ce Liu, J. Yuen, and A. Torralba, "Nonparametric scene parsing: Label transfer via dense scene alignment," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1972–1979, IEEE, Miami, FL, USA, August 2009.
- [98] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *Proceedings of the 2009 IEEE 12th international conference on computer vision*, pp. 1–8, IEEE, Kyoto, Japan, 2009, September.
- [99] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," vol. 2, pp. 416–423, in *Proceedings of the Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, pp. 416–423, IEEE, Vancouver, BC, Canada, 2001, July.
- [100] A. Prest, C. Leistner, J. Civera, C. Schmid, and V. Ferrari, "Learning object class detectors from weakly annotated video," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3282–3289, IEEE, Providence, RI, USA, 2012, June.
- [101] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [102] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik, "Semantic contours from inverse detectors," in *Proceedings of the 2011 International Conference on Computer Vision*, pp. 991–998, IEEE, Barcelona, Spain, November 2011.
- [103] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille, "Detect what you can: Detecting and representing objects using holistic models and body parts," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1971–1978, Columbus, OH, USA, June 2014.
- [104] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3234–3243, Las Vegas, NV, USA, June 2016.
- [105] X. Shen, A. Hertzmann, J. Jia et al., "Automatic portrait segmentation for image stylization," *Computer Graphics Forum*, vol. 35, no. 2, pp. 93–102, 2016.
- [106] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgb-d images," in *Proceedings of the European conference on computer vision*, Springer, Berlin, Heidelberg, pp. 746–760, 2012.
- [107] J. Xiao, A. Owens, and A. Torralba, "Sun3d: A database of big spaces reconstructed using sfm and object labels," in *Proceedings of the IEEE international conference on computer vision*, pp. 1625–1632, Sydney, NSW, Australia, December 2013.
- [108] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 567–576, Boston, MA, USA, June 2015.
- [109] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *Proceedings of the 2011 IEEE international conference on robotics and automation*, pp. 1817–1824, IEEE, Shanghai, China, May 2011.
- [110] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3d reconstructions of indoor scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5828–5839, Honolulu, HI, USA, July 2017.
- [111] I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, "Joint 2d-3d-semantic data for indoor scene understanding," arXiv preprint arXiv:1702.01105, 2017.
- [112] A. X. Chang, T. Funkhouser, L. Guibas et al., "Shapenet: An information-rich 3d model repository," arXiv preprint arXiv:1512.03012, 2015.
- [113] M. De Deuge, A. Quadros, C. Hung, and D. Bertrand, "Unsupervised feature learning for classification of outdoor 3d scans," *Australasian Conference on Robotics and Automation*, vol. 2, p. 1, 2013.
- [114] W. Song, N. Zheng, R. Zheng, X. Zhao, and A. Wang, "Digital image semantic segmentation algorithms: A survey," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 10, no. 1, pp. 196–211, 2019.
- [115] Y. Liu, J. Liu, Z. Li, J. Tang, and H. Lu, "Weakly-supervised dual clustering for image semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2075–2082, Portland, OR, USA, June 2013.
- [116] R. Guo, X. P. Yang, and J. Wang, "The optimization analysis of the image segmentation and denoising based on the improved FCM clustering algorithm [J]," *CAAI Transactions on Intelligent Systems*, vol. 11, no. 6, pp. 227–233, 2016.
- [117] X. X. Zhang, "Research on image semantic segmentation based on probability Graph Model [D]," Dissertation, Xiamen University, Xiamen, Fujian, China, 2014.

- [118] C. F. Zhang, "Image semantic segmentation based on conditional random _led [J]," *Computer CD software and applications*, no. 9, pp. 21–23, 2012.
- [119] X. M. Zuo, Z. Zhao, and T. T. Gou, "RGB-D image segmentation method based on interactive conditional random _elds [J]," *Computer applications and software*, vol. 34, no. 3, pp. 174–180, 2017.
- [120] L. Mao and M. Xie, "Image semantic segmentation based on higher-order CRF model [J]," *Application research of compute*, vol. 30, no. 11, pp. 3514–3517, 2013.
- [121] L. J. Wang, Y. Q. Zhong, and H. Guo, "Improved image segmentation algorithm based on order conditional random _eld model [J]," *Computer Engineering*, vol. 42, no. 6, pp. 241–246, 2016.
- [122] H. X. Chen, "Semantic segmentation based on convolutional neural networks [D]," Dissertation, Zhejiang University, Hangzhou, China, 2016.
- [123] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, pp. 1520–1528, Santiago, Chile, December 2015.
- [124] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, Honolulu, HI, USA, July 2017.
- [125] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1925–1934, Honolulu, HI, USA, July 2017.
- [126] S. Hong, H. Noh, and B. Han, "Decoupled deep neural network for semi-supervised semantic segmentation," arXiv preprint arXiv:1506.04924, 2015.
- [127] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [128] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 11–19, Honolulu, HI, USA, July 2017.
- [129] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," arXiv preprint arXiv:1606.02147, 2016.
- [130] Md A. Islam, M. Rochan, N. D. B. Bruce, and Y. Wang, "Gated feedback refinement network for dense image labeling," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3751–3759, Honolulu, HI, USA, July 2017.
- [131] N. Souly, C. Spampinato, and M. Shah, "Semi and weakly supervised semantic segmentation using generative adversarial network," arXiv preprint arXiv:1703.09695, 2017.
- [132] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, Venice, Italy, October 2017.
- [133] Yi Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2359–2367, Honolulu, HI, USA, July 2017.
- [134] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3150–3158, Las Vegas, NV, USA, June 2016.
- [135] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [136] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Applied Soft Computing*, vol. 70, pp. 41–65, 2018.
- [137] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*, Springer Science & Business Media, Boston, MA, 2011.
- [138] <https://towardsdatascience.com/multi-class-metrics-made-simple-the-kappa-score-aka-cohens-kappa-coefficient-bdea137af09c> [Multi-Class Metrics Made Simple, Part III: the Kappa Score (aka Cohen's Kappa Coefficient)].
- [139] M. Zhang, Y. Zhou, J. Zhao, Y. Man, B. Liu, and R. Yao, "A survey of semi- and weakly supervised semantic segmentation of images," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4259–4288, 2020.
- [140] Q. Geng, Z. Zhou, and X. Cao, "Survey of recent progress in semantic image segmentation with CNNs," *Science China Information Sciences*, vol. 61, Article ID 051101, 2018.
- [141] B. Zhao, J. Feng, X. Wu, and S. Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," *International Journal of Automation and Computing*, vol. 14, no. 2, pp. 119–135, 2017.
- [142] U. Srinivasan, S. Pfeiffer, S. Nepal, M. Lee, L. Gu, and S. Barras, "A survey of MPEG-1 audio, video and semantic analysis techniques," *Multimedia Tools and Applications*, vol. 27, no. 1, pp. 105–141, 2005.
- [143] S. I. Anishchenko and M. V. Petrushan, "Optimal feature space for semantic image segmentation," *Pattern Recognition and Image Analysis*, vol. 24, no. 4, pp. 502–505, 2014.
- [144] Q. Ning, J. Zhu, and C. Chen, "Very fast semantic image segmentation using hierarchical dilation and feature refining," *Cognitive Computation*, vol. 10, no. 1, pp. 62–72, 2018.

Review Article

The Systematic Literature Review of Privacy-Preserving Solutions in Smart Healthcare Environment

Driss El Majdoubi , **Hanan El Bakkali** , **Souad Sadki** , **Zaina Maqour,**
and Asmae Leghmid

Rabat IT Center, Smart Systems Laboratory (SSL), ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

Correspondence should be addressed to Driss El Majdoubi; driss.elmajdoubi@um5s.net.ma

Received 1 November 2021; Revised 9 February 2022; Accepted 10 February 2022; Published 16 March 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Driss El Majdoubi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rapid development of the Internet of Medical Things (IoMT) technology has resulted in various advances in the smart healthcare field; it improves healthcare systems to offer more complicated real-time services and provides an efficient patient motioning system. However, despite the brilliant side of IoMT, several concerns continue to undercut its adoption. In fact, collecting, transmitting, storing, and using data in IoMT applications raises issues regarding privacy and data protection, especially with the multitude of stakeholders involved during the whole data life cycle. Motivated from these facts, this article is devoted to perform a Systematic Literature Review (SLR) of privacy-preserving solutions used in the smart healthcare ecosystem. The recent research papers disseminated between 2017 and 2021 are selected from multiple databases and a standardized SLR method is conducted. A total of 100 papers were reviewed and a critical analysis was conducted on the selected papers. Moreover, this review study attempts to highlight the limitation of the current approaches and aims to find possible solutions to them. Thus, a detailed analysis was carried out on the selected papers in terms of the privacy techniques they deployed, the data life cycle phase they addressed, the stakeholders needs they met, and the privacy principles they covered according to privacy laws and regulations. Finally, we summarize our results showing privacy-preserving trends and identifying recommendations to involve privacy principles coverage in smart healthcare applications.

1. Introduction

In recent years, smart healthcare is one of the fastest-growing technologies that provide an opportunity for accurate and efficient prevention of several diseases. The Internet of Medical Things (IoMT) is a connected infrastructure of medical devices, health systems, and services. The IoMT [1] enables the connection, communication, capture, and exchange of Electronic Medical Records (EMR) between entities. The EMR includes sensitive health data, whereas the implementation of any Internet of Things Technology usually comes with various concerns about privacy and data protection. When it comes to patient privacy, the things to consider are even more. Hence, data security and privacy issues have become the biggest concerns of people in smart healthcare field. For example, a patient

usually expects that his or her EMRs, such as blood pressure and pulse rate, can only be accessed by authorized professional health caregivers and with his or her consent and control.

Recently, several researchers have shown interest in security and privacy preservation in a smart healthcare environment. Yet, understanding the current security and privacy issues of the IoMT system is essential. Moreover, it is significant to know the effectiveness of the offered solutions. We found that little attention has been paid in the literature to elaborate on these issues. Therefore, in this work, a Systematic Literature Review (SLR) [2] is presented.

1.1. Scope. Recently, Many surveys have been conducted which highlighted the privacy-preserving issues in healthcare environments. Most of these surveys have given an

insight into the privacy issues and their solutions in different areas of the healthcare field. In the proposed survey, we have given a comprehensive overview of different privacy-preserving approaches in the smart healthcare ecosystem which use many smart technologies (Cloud Computing, Fog Computing, Internet of Things, and telehealthcare technologies) to share data between various stakeholders. To this end, the current systematic literature review is intended to address privacy-preserving solutions in IoMT considering different needs of stakeholders, the whole data life cycle, and limitations in terms of privacy criteria coverage view.

One of the most recent survey papers of privacy-preserving in healthcare environments was performed by Hameed et al. [3]. In this paper, the authors highlighted a systematic literature review around the IoMT security and privacy issues and how machine learning techniques are applied to solve these problems.

Within the scope of another study, performed by Tanriverdi [4], blockchain-based studies on the preservation of medical data sharing privacy were analyzed. In this study, information about the research publications in the literature and possible issues that can be examined in the future were discussed. In another study, Iwaya et al. [5] reviewed, analyzed, and synthesized the related literature on the security and privacy of m/uHealth systems using an evidence-based software engineering methodology, a Systematic Mapping Study (SMS).

Another exhaustive survey on security and privacy issues in Healthcare 4.0 was carried out by Hathaliya and Tanwar [6]. The authors explored the blockchain-based solution to give insights to researcher communities. The technology used, the problem formulation, the parameters to handle the security, and privacy issues were implemented in a comparative analysis of the existing survey on security and privacy in Healthcare 4.0.

A review of security and privacy in the medical Internet of things was conducted by Sun et al. [7]. The authors survey the existing solutions for security and privacy in the IoMT; the proposed solutions are focusing on data encryption, access control, trusted third party auditing, data search, and data anonymization. It had also highlighted the future challenges of security and privacy in IoMT.

1.2. Motivation. The motivation of this paper was as follows:

- (i) Importance of privacy preserving in the smart healthcare field is one of the key criteria to explore this area.
- (ii) The existing literature mainly discussed some privacy aspects of smart healthcare such as technical aspects; IoT-based and machine learning-based solutions. Many other emerging areas of privacy in smart healthcare, such as compliance with privacy laws, in accordance with patient's preferences and privacy preserving in the whole data life cycle were not explored to their full potential. So, there is a need to write a survey that considers the integration of all these aspects as mentioned above.

- (iii) This systematic literature review is intended for new researchers in the field, and for those who are keen to know about recent advances and limitations of privacy-preserving in a smart healthcare environment. In addition, this kind of study enables the identification of research trends, raising the most discussed aspects and open issues, indicating possibilities of research in less discussed aspects.

1.3. Contributions. In this paper, we make the following contributions:

- (i) We discuss the background and the importance of privacy in smart healthcare.
- (ii) We identify several aspects that should be considered while treating privacy issues on smart healthcare systems.
- (iii) We identify 3425 primary studies that present privacy-preserving solutions in the smart healthcare sector.
- (iv) We further select 100 primary studies that meet the inclusion and exclusion criteria we set for the paper screening phase.
- (v) We conduct an in-depth assessment through a critical analysis of the 100 selected papers and present the research ideas, techniques, and the adopted aspects and considerations in the field of privacy in IoMT.
- (vi) We make different combinations of the prespecified privacy aspects in the form of bubble charts to conclude the state of privacy principles coverage and security requirement fulfillment by the primary studies.
- (vii) Finally, we summarize the lessons learned from this SLR and identify the recommendations that lead toward a holistic approach to preserve privacy in smart healthcare applications.

1.4. Organization of the Paper. The remainder of the paper is structured as follows: Section 2 presents a general background about IoMT, state of art, and terminologies of privacy in smart healthcare. Section 3 describes the methodology with which the primary studies were systematically selected for analysis. The findings of all the primary studies selected are presented in section 4, followed by the discussion of the results in Section 5. Section 6 presents recommendations to upgrade toward a holistic approach to preserve privacy. Section 7 concludes the paper and presents some future research directions.

2. Background

In this section, we present background knowledge about privacy, security, and smart healthcare, by defining each concept, and presenting the existing privacy-preserving techniques, laws, and criteria; we also provide an overview of

smart healthcare, and the existing IoMT categories and architectures.

2.1. Privacy Definitions and Techniques

2.1.1. Definitions: Privacy VS. Security. Privacy and security are two different concepts, yet they are frequently misunderstood or conflated by the concerned users and organizations while dealing with Internet services and personal data. Thus, it is mandatory to rectify the meaning of each and discuss their differences.

Privacy is related to the right to have control over information, identity, and activity of oneself, and take part in data processing decisions, such as disclosure, retention, and erasure, whilst security is related to how the data is protected, and the measures to follow against the different threats.

In other words and according to Ref. [8], the difference between security and privacy can be represented as: “Information security means protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction so that the confidentiality, integrity, and availability of information are maintained. In contrast, privacy ensures that user’s data are stored, used and disclosed fairly according to the data owner’s preferences.”

2.1.2. Privacy-Preserving Techniques. Not any privacy-preserving techniques are suitable for all sorts of applications [9]; Herein, we present a few techniques that are effectively used to preserve privacy in the smart healthcare sector:

- (i) **Access Control:** The intent behind these techniques is to restrict access to only authorized parties [10]. Its mechanisms can take many forms depending on the adopted approach while granting permissions; assigning permissions based on roles designated as Role-Based Access Control (RBAC), and based on attributes designated as Attribute-Based Access Control (ABAC).
- (ii) **Cryptography:** Various cryptographic techniques are being applied in order to preserve privacy, They can be classified into three main collections: *Secret Key Cryptography (SKC)* which uses the same key for encryption and decryption, i.e., DES, *Public Key Cryptography (PKC)*, a system in which two different keys are used, i.e., RSA, *Hash Function*, which is an irreversible function that generates an output data with fixed size from an unfixed input size [11].
- (iii) **Anonymization:** This technique is commonly performed before the distribution and analysis processes with the aim of data sanitization, also known as de-identification [11]; it makes the data less precise and hides the identity of patients.
- (iv) **Blockchain:** Recently, blockchain has extended beyond the financial sector and has become a trending solution for decentralization, and privacy issues in the smart healthcare domain, due to its

numerous features [12], namely, Decentralization; Transparency; Open-source; Autonomy; Immutability; and Anonymity.

2.1.3. Data Access Management. The use of smart health has become the key source of data breaches since medical data are more sensitive than the other types. The 2021 Mid-Year Data Breach Quick View Report published by Risk-Based Security affirmed that 238 healthcare data breaches were reported in the first 6 months of 2021, which makes the healthcare sector in the top position as the most breached economic sector [13] (Figure 1); moreover, “Hacking” or Unauthorized Access is considered the number one breach type (Figure 2), which points to the importance of data access management.

Many techniques are being used for this purpose, including the aforementioned techniques, Access Control and Cryptography; authentication process is also used as a solution to provide secure access to the medical data.

With the emergence of blockchain, new technologies are added to this block, namely, permissioned blockchain and smart contracts. Permissioned blockchain requires an access control layer, which makes it provide an additional level of security over the typical blockchain, while smart contracts are being applied to manage the permissions to a patient’s HER [14].

2.2. Smart Healthcare Overview. In our SLR, we are interested in privacy-preserving solutions, particularly in the smart healthcare sector. With that being said, in this section, we will clarify the status of this sector in the field of health as a whole, and its different actors.

2.2.1. Electronic Health (e-Health). E-health is an emerging field at the intersection of classical health, and Information and Communication Technologies (ICT), for instance, the use of Electronic Health Records (EHR) or databases that store medical information of patients.

Nevertheless, an article published in the Journal of Medical Internet Research [15] claims that the definition of e-health has a broader sense, as the “e” does not simply mean electronic, but implies several other “e’s,” which combine to provide a full definition of e-health; these “10 e’s,” are the following: Efficiency, Enhancing Quality, Evidence-Based, Empowerment, Encouragement, Education, Enabling, Extending, Ethics, and Equity.

2.2.2. Mobile Health (m-Health). More recently the emergence of smartphones has led to their recognition as a great help in the healthcare sector, hence the emergence of mobile healthcare. Known as m-health.

M-health is a subsection of e-health and defined by The World Health Organization (WHO) in collaboration with the Global Observatory for eHealth as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices” [16]. M-health

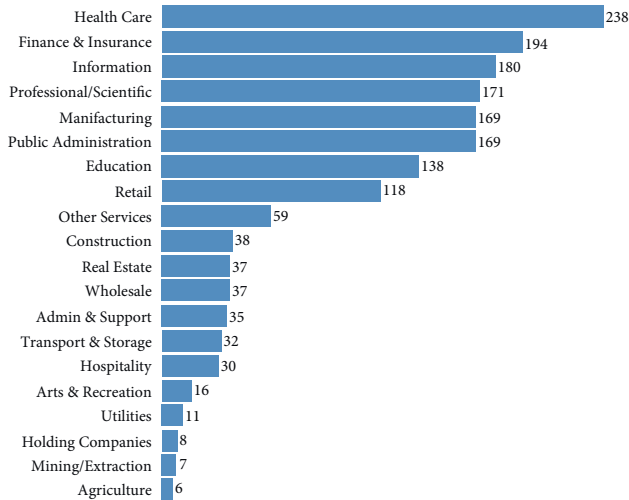


FIGURE 1: Number of breaches by economic sector, reported by Q2 2021 [12].

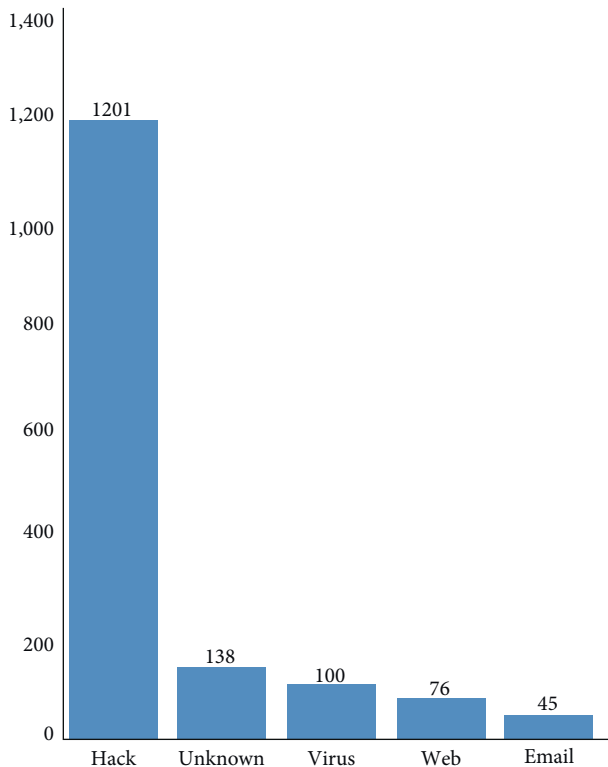


FIGURE 2: Number of breaches by breach type, reported by Q2 2021 [12].

applications facilitate the communication between patients and caregivers. Moreover, they allow remote and real-time monitoring.

2.2.3. Smart Health (s-Health). The abiding progression of ICT has led to a new concept named context-aware environments, such as smart homes and smart cities, which are built with high sensing, analyzing, and decision-making

capabilities. The implementation of these context-aware environments along with both e-health and m-health has engendered the term smart health (or s-health).

2.2.4. S-Health VS. m-Health. S-health and m-health are distinct in terms of the underlying infrastructure. Ref. [17] stated that “the concept of s-health could be considered an augmentation of m-health with the sensing capabilities of smart cities,” and summarizes the differences in two points; differences in information sources as the data may come from different sources not only from patients which exceed m-health, and differences in information flows as the collected data may be processed by several parties, hence it is beyond the user-centric approach and beyond m-health.

2.2.5. Key Stakeholders of s-Health. To ensure full coverage of privacy in s-health, it is mandatory to define the different actors as well as the needs of each of them:

- (i) **Patients:** are the data owners; therefore, they have the complete right to take control over their data, in terms of access, modification, retention, erasure, disclosure. . . In other words, the patients’ preferences must be considered by the smart healthcare systems.
- (ii) **Services providers:** are the actors who provide smart healthcare and well-being services, such as doctors, service developers, and cloud providers.
- (iii) **Governmental bodies:** are either the organizations that define privacy regulations and laws, such as the European Data Protection Law, or the organizations that supervise the legitimacy of data processing in the smart healthcare sector.

To summarize, a solution that preserves privacy in a way that satisfies the various stakeholders’ needs is a solution that respects both the patient’s preferences and the service provider’s privacy policy, while complying with the privacy laws and regulations.

2.3. IoMT: Data Life Cycle, Categories, and Architectures

2.3.1. Data Life Cycle. Wireless body area networks (WBAN) [18] and IoMT collect data streams such as heart rate, blood pressure, ECG from sensors, actuators, which are then transmitted to different units, i.e., mobile devices, hospital data centers, etc. These data streams are then stored in cloud servers, databases, ready for any further processing or use. That being said, the data life cycle can be assembled in 4 global stages, namely, **collection, transmission, storage, and process.**

In our SLR, we are targeting IoMT data protection and patient’s privacy-preserving solutions in s-health, and since privacy must be protected in each data phase, we are joining the data life cycle to the adopted aspects to assess the proposed studies.

2.3.2. *Categories of IoMT.* Different categories of IoMT are being adopted by the privacy-preserving solutions in the s-health. We present them as follows:

- (i) **Fog/Edge/Cloud-based:** Cloud computing allows the data to be stored on multiple servers and accessed from different locations. Yet, despite the recent efforts to make data closer to the user, fog and edge computing have overlapped to enhance the velocity of data processing. In edge computing, the data are stored in the device itself or closer to the device and not sent to the cloud [19]; similarly, fog computing provides an additional intermediate layer where the data are processed within a node, gateway, or router and then transferred to the proper devices [20].
- (ii) **Blockchain-based:** As we already mentioned in the privacy techniques section, blockchain technology becomes hugely implemented in the s-health domain in order to address privacy issues and maintain seamless accessibility of data by the different stakeholders; it is built on public-key cryptography which is used to conduct transactions among nodes; these transactions are then stored on a shared ledger [1]. Once the data are recorded in the blockchain, they cannot be modified or removed.
- (iii) **Policy-based:** A privacy policy is another facet that should be concerned about while dealing with privacy issues in s-health since they are the main intersection point between the multiple actors on the patients' data, i.e., service providers, governmental organizations, etc., wherein every actor expresses his needs on how the data should be used, when, and how. Many privacy-preserving solutions have followed privacy by design approach, yet they fail to cover the needs of all the stakeholders.

2.3.3. *Architectures of IoMT.* Based on the aforementioned categories, we can distinguish 4 architectures for the IoT in the healthcare domain, namely, **centralized architecture, decentralized architecture, hybrid architecture, and third parties architecture**, as illustrated in Table 1.

2.4. Privacy-Preserving in s-Health: Laws, Policies, and Preferences

2.4.1. *Privacy Legislation.* Privacy preservation is a common responsibility among the different stakeholders; hence, they are attempting to manage it by applying many mechanisms; however, they only rely on the technical perspective while neglecting the perspective of legitimacy and law compliance. In this section, we will promote the existing and relevant laws and regulations of protecting data and personal health information (PHI).

As shown in Figure 3, since 1988, countries are attempting to put boundaries on data usage, and allow citizens to have control over their data, furthermore setting

penalties on any violation behavior of personal privacy. A comparative study of major privacy laws and regulations is surveyed in Ref. [21].

The most relevant data protection legislation enacted to date is the EU law General Data Protection Regulation (GDPR), which is not only restricted to European-based companies and service providers but also deals with international parties that are involved in processing the data of the EU citizens. For that reason, many recent privacy laws consider the adoption of GDPR, for instance, the Consumer Privacy Act (CCPA) in California, enacted in June 2018 and took effect in January 2020, the General Data Protection Law (LGPD) in Brazil passed in 2018 and goes into effect in February 2020 [22], and the Consumer Data Protection Act (CDPA) in Virginia enacted on March 2, 2021.

Besides the EU GDPR, the International Standard ISO/IEC 29100 defines 11 privacy principles to help organizations define their privacy safeguarding requirements related to personally identifiable information (PII) [23], namely: Consent and choice; Purpose legitimacy and specification; Collection limitation; Data minimization; Use, retention, and disclosure limitation; Accuracy and quality; Openness, transparency, and notice; Individual participation and access; Accountability; Information security; and Privacy compliance. The principles are described in Ref. [23].

2.4.2. *Privacy Policies and Patients' Preferences.* As stated in section 2.3.2, a Privacy Policy is a statement wherein an organization clarifies how it will handle the collected Personal Health Information (PHI); however, it may not necessarily satisfy the patients' preferences. In order to prevent this kind of conflict, the Privacy Policy must be written understandably and mostly natural languages are used for this purpose, while at the same time allowing patients to express their preferences beforehand. In fact, this approach is proceeding toward a mutual agreement among the s-health stakeholders regarding privacy preservation, yet it provokes many challenges particularly the agreement process and the conflicting policies.

3. Systematic Literature Review

To choose and subsequently analyze a series of scientific articles, the methodology used to conduct the literature search and the selection of the studies to be included in our analysis has been presented in this section. An SLR is composed of five phases, namely: (i) *definition of Research Scope*, (ii) *Selection of primary studies*, (iii) *Inclusion and exclusion criteria*, (iv) *Selection results*, and (v) *Data analysis*.

3.1. *Step1: Definition of the Research Scope.* The first step to perform an SLR is identifying the need to uncover gaps and trends related to the privacy-preserving aspects addressed in this study. Therefore, it is necessary to identify some research questions (RQ) to be answered from the inputs provided by the analysis of relevant studies, which will constitute the primary studies.

TABLE 1: Architectures of IoMT.

| Architectures | Entities participation | Storage type | Pros | Cons |
|---------------|------------------------|-------------------------|---|---|
| Centralized | N | Centric | Efficiency: All data are managed in one place; affordable to maintain | Single point of failure; no control over data usage |
| Decentralized | Y | Distributed | Secure storage; control over data usage | Higher maintenance costs |
| Third-party | N | Centric | Cost-effective | Data disclosed to untrusted third party; no control over data usage |
| Hybrid | N/Y | Centric/ Distributed | Takes advantage of the pros of each architecture | Difficulty in management |

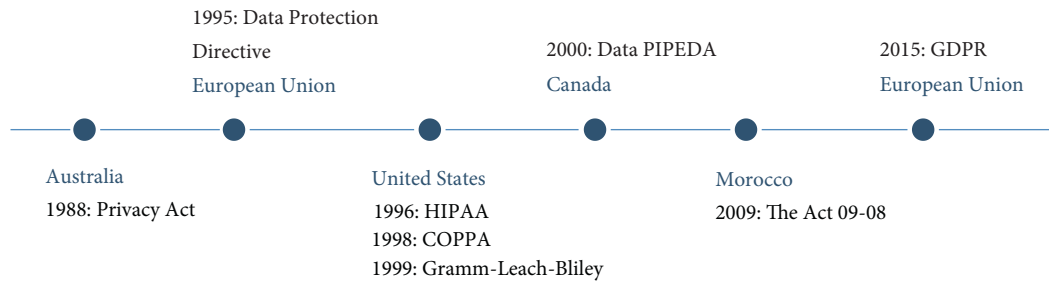


FIGURE 3: Privacy laws and regulations.

The purpose of this research was to analyze existing studies and their solutions, to summarize the efforts of research on privacy-preserving in smart healthcare applications from an end-to-end view (Different needs of stakeholders and whole data life cycle), and to discover limitations in terms of privacy criteria coverage view.

Therefore, to achieve these objectives, we have chosen three research questions as listed in Table 2.

3.2. Step2: Selection of Primary Studies. In this phase, we will identify the source bases and the source strings used to select the primary studies for our study. To form our research query, we used the Boolean operators AND and OR to combine the multiple keywords describing our research subject, the final result is the following:

(Privacy OR Cybersecurity OR “Cyber security” OR security) AND (“smart healthcare” OR “smart health” OR “digital healthcare” OR “medical Internet of things” OR “medical IoT” OR “Internet of medical things”)

We have submitted this query in various relevant databases, namely: Science Direct, Scopus, Web of Science, and Springer Link. The obtained results were then filtered through the inclusion and exclusion criteria defined in section 2.3; afterward, we conducted the snowballing technique to the new set of results, including both forward and backward processes.

3.3. Step3: Inclusion and Exclusion Criteria. After the initial selection from the previously mentioned databases, the next step is the paper screening, which consists of checking the eligibility of each article according to many criteria for inclusion and exclusion, as presented in Table 3, to retrieve

only the most relevant studies that present a privacy-preserving solution in the smart healthcare environment.

3.4. Step4: Selection Results. The initial query search in the selected databases provided us with an amount of **3425** articles as shown in Figure 4; after removing the duplicated studies, the number was reduced to **3391**; these studies are then examined through the inclusion/exclusion criteria and reduced to **59**. An additional **8** and **33** studies were identified by forward and backward snowballing, respectively, making the outcome of the papers to be included in our systematic literature review equal to **100** papers.

4. Result and Finding

After selecting the primary studies, in this section, two types of analysis were performed to evaluate and synthesize the primary studies—bibliometric analysis and technical analysis, as discussed in the following subsections.

4.1. Bibliometric Analysis. After the paper screening phase, in this section, we focused on the evaluation of the primary studies in terms of their publishers and publication year.

According to Figure 5, an important number of primary studies (51%) was published by Science Direct, and 33% was published by Scopus; moreover, a percentage of 21% was published by Web of Science, while the least number of papers was found in SpringerLink (4%).

It is worth mentioning that the researchers’ concern about security in smart healthcare and privacy of medical data is constantly growing, as seen in Figure 6; the published primary studies went from 9 percent in 2017 to 29 percent in 2021. The increasing adoption of telemedicine, usage of the

TABLE 2: Research questions for our systematic literature review.

| Research questions | Goals |
|---|---|
| RQ1: What are the proposed solutions to preserve privacy according to the different stakeholders' needs (patients, providers, government bodies) while considering data-access management? | This question aims at identifying the existing solutions to preserve privacy from different stakeholders' points of view; therefore, it will help to know the missing stakeholders' needs that require more interest in the future. |
| RQ2: What are the privacy criteria that have been considered by the proposed solutions, and in which data life cycle phase? | This question aims at identifying the privacy criteria coverage stated by the existing solutions, as well as the phase of the data life cycle that should be more enhanced. |
| RQ3: How and what are the techniques used by published papers to preserve privacy in smart healthcare, and in which architecture and category? | This question aims to identify the privacy-preserving techniques and their impact on the architecture choice. |

TABLE 3: Inclusion and exclusion criteria.

| Inclusion criteria | Exclusion criteria |
|---|--|
| The paper must present a privacy-preserving solution related to smart healthcare | Duplicated articles Articles published before 2017 |
| The paper must be a research article published in a peer-reviewed journal or conference | Articles that are written in a language other than English |
| We suppose that relevant articles should have at least one citation per year if the publication year is between 2017 and 2019. For example, if an article is published in 2017, it should have at least 4 citations | Articles that present a previous version of a complete research study The study is a survey |

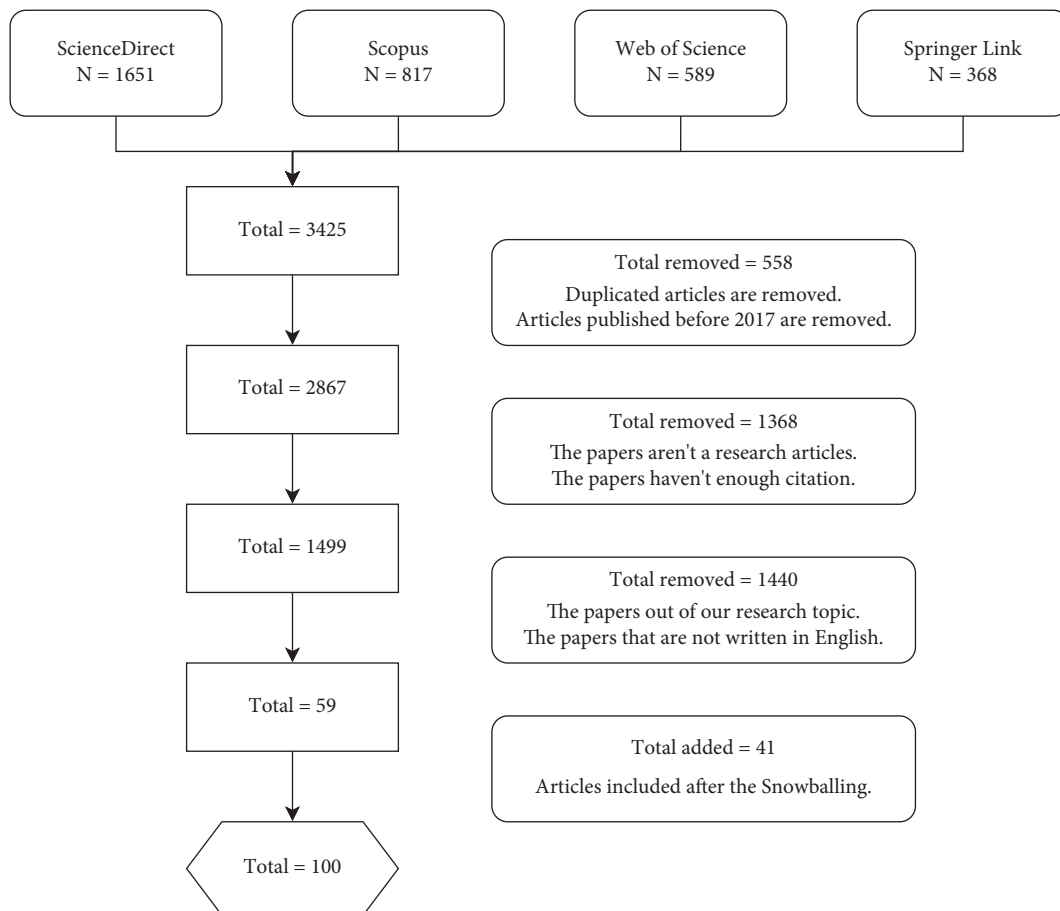


FIGURE 4: Number of articles included and excluded through paper screening.

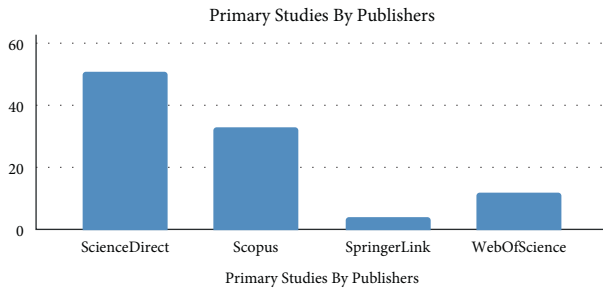


FIGURE 5: Number of primary studies according to their publishers.

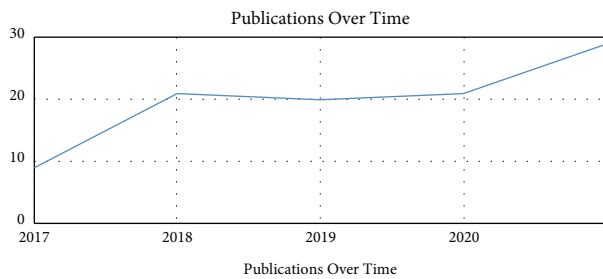


FIGURE 6: Number of primary studies published over time.

Internet of medical things by different stakeholders, and cyber awareness of patients, imply that in the next few years, the research will be further enriched by studies that arise in the context of privacy-preserving in digital healthcare.

4.2. Technical Analysis. This step consists of defining a classification scheme composed of different privacy aspects in order to compare the different solutions based on a comprehensive analysis. We define *seven privacy aspects* for our study as follows:

- (i) **Architecture:** Centralized, Decentralized, Third party, and Hybrid.
- (ii) **Category:** Fog-/Edge-/Cloud-based, Blockchain-based and Policy-based.
- (iii) **Data access management and privacy-preserving technique:** Access Control, Blockchain, Authentication, Cryptography, and Anonymization.
- (iv) **Stakeholders' needs:** Patient Preferences, Privacy Policy, and Privacy laws.
- (v) **Data Life Cycle:** Collection, Transmission, Storage, and Process/Use.
- (vi) **ISO Privacy Principle:** P_1 : Consent and choice; P_2 : Purpose legitimacy and specification; P_3 : Collection limitation; P_4 : Data minimization; P_5 : Use, retention, and disclosure limitation; P_6 : Accuracy and quality; P_7 : Openness, transparency, and notice; P_8 : Individual participation and access; P_9 : Accountability; P_{10} : Information security; P_{11} : Privacy compliance.

In this phase, we will classify the papers into four classes (centralized, decentralized, third party, and

hybrid) based on the architecture privacy aspect. Furthermore, an in-depth assessment will be accomplished through a critical analysis of the selected papers based on the other privacy aspects: category, data access management and privacy-preserving technique, stakeholders' needs, data life cycle, and ISO privacy principle. In the following subsections, we present the findings of the aforementioned analysis.

4.2.1. Centralized Architecture. In this subsection, we present the main features and limitations of the relevant papers in a centralized architecture, while the detailed results are summarized in Table 4.

Kumar et al. [24], proposed a Secure Addressing and Mutual Authentication protocol (SAMA) scheme to protect the network from multiple attacks by modifying the standard IPv6 protocol, and by establishing a secure session key and mutual authentication.

In another work carried out by de Oliveira et al. [25], a dynamic revocable data access control protocol for Acute Care teams (AC-AC) was proposed, by adding a security mechanism that enables break-glass access to the Electronic Medical Records (EMR) with dynamic revocation to provide access to a patient's encrypted EMR during acute care.

In the same context, and using Hyperelliptic Curve Cryptosystem (HECC), authors proposed in Red. [81] a secure and efficient software-defined healthcare-enabled WBANs architecture. More explicitly, authors integrated the SDN technology into the proposed solution while separating the control and data planes in an efferent manner. Hence, convenient results were obtained in terms of security, computation, communication, and storage costs.

On the other hand, Zhong et al. [26] proposed an efficient attribute-based encryption (ABE) scheme that outsources part of the encryption and decryption to the edge nodes and supports attribute updates, enabling flexible right control. This scheme is tested and evaluated in different security levels and proves that it is more efficient for resource-constrained devices than the traditional ABE schema. Furthermore, in a study reported by Onesimu et al. [27], a privacy-preserving data collection scheme was implemented based on the clustering-based anonymity mode for IoT-based healthcare services and formulates the threat model as client-server-to-user to ensure privacy on both ends.

In the same context, an integrated privacy-preserving framework in IoT-based smart healthcare was suggested in Ref. [29]; the particularity of this solution is its ability to allow patients making pragmatic data sharing deals with smart services by indicating the data items that can be shared or used along with their precision.

Izza et al. [28] proposed an IoT-based Radio Frequency Identification (RFID) authentication scheme for Wireless Body Area Networks (WBAN), which is an improved version of the RFID authentication scheme for IoT proposed by Naeem et al. This study focused on solving the remaining security challenges of the previous protocol during the transmission phase.

TABLE 4: Summary of the studies reported on centralized architecture.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|------------------|---|--|------------------------------------|---|----------------------------------|
| [24] | Cloud-based | Cryptography anonymization | Cryptography authentication process | No | Transmission | P6; P10; P11 |
| [25] | Cloud-based | Cryptography | Cryptography access control | No | Transmission storage process | P1; P2; P5; P6; P8; P10 |
| [26] | Edge/Cloud-based | Cryptography access control | Access control | No | Transmission storage | P6; P10; P11 |
| [27] | Cloud-based | Anonymization | No | No | Collection | P5; P6; P7; P10; P11 |
| [28] | Cloud-based | Cryptography \Anonymization | Cryptography | No | Transmission | P6; P10; P11 |
| [29] | Cloud-based | Anonymization | Access control | Patient preferences privacy policy | Process | P1; P3; P5; P7; P8; P11 |
| [30] | Edge-based | Cryptography | Access control | Patient preferences privacy policy | Collection transmission storage | P1; P3; P4; P6; P8; P10; P11 |
| [31] | Cloud-based | Cryptography access control | Access control | Privacy policy | Transmission storage process | P2; P8; P6; P9; P10; P11 |
| [32] | Cloud-based | Cryptography access control | Cryptography access control | Privacy policy | Transmission storage | P1; P2; P5; P8; P9; P11 |
| [33] | Cloud-based | Cryptography anonymization access control | Cryptography access control authentication process | No | Storage process | P6; P7; P10; P11 |
| [34] | Cloud-based | Cryptography anonymization | Cryptography authentication process | No | Transmission storage process | P6; P7; P11 |
| [35] | Cloud-based | Access control | Cryptography | Privacy policy | Transmission storage process | P2; P5; P10; P11 |
| [36] | Privacy-based | Anonymization | No | No | Process | P2; P5 |
| [37] | Cloud-based | | Authentication process | No | Process | P5; P9; P10 |
| [38] | Cloud-based | Cryptography | Authentication process | Patient preferences | Process | P2; P5; P6; P8 |
| [39] | Cloud-based | Cryptography anonymization | Cryptography authentication | No | Transmission process | P6; P10; P11 |
| [40] | Cloud-based | Cryptography access control | Access control authentication process | No | Collection transmission storage | P2; P3; P4; P10; P11 |
| [41] | Cloud-based | Cryptography | No | No | Transmission | P6; P10; P11 |
| [42] | Edge/Cloud-based | Cryptography | No | No | Transmission storage | P6; P9; P10; P11 |
| [43] | Cloud-based | Cryptography access control | Access control | No | Storage process | P1; P6; P8; P10; P11 |
| [44] | Privacy-based | Anonymization access control | No | No | Storage process | P6; P10; P11 |
| [45] | Cloud-based | Cryptography | Cryptography | Patient preferences privacy policy | Collection transmission storage process | P1; P4; P7; P5; P3; P6; P9; P10 |
| [46] | Cloud-based | Cryptography access control | Access control authentication process | Patient preferences | Process | P2; P5; P8; P9; P10 |
| [47] | Cloud-based | Cryptography | Cryptography | No | Transmission storage process | Purpose legitimacy; P5; P10 |
| [48] | Cloud-based | Cryptography anonymization | Cryptography authentication process | No | Process | P6; P7; P10; P11. |
| [49] | Cloud-based | Cryptography anonymization | Cryptography authentication process | No | Process | P6; P7; P10; P11 |
| [50] | Cloud-based | Cryptography | Cryptography | No | Collection transmission storage process | P6; P10; P11 |

TABLE 4: Continued.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|-------------------------------|---|--|---|------------------------------|----------------------------------|
| [51] | Cloud-based | Cryptography access control | Cryptography access control | Privacy policy | Collection storage process | P1; P4; P7; P2 |
| [52] | Cloud-based | Cryptography | Cryptography authentication process | No | Transmission storage | P7; P8; P10 |
| [53] | Cloud-based | Cryptography access control | Cryptography access control | Privacy policy | Storage process | P9; P10; P11 |
| [54] | Cloud-based | Cryptography anonymization access control | Cryptography access control authentication | Privacy policy | Transmission storage process | P6; P10; P11 |
| [55] | Cloud-based | Cryptography anonymization | Cryptography authentication process | No | Transmission | P5; P7; P10; P11 |
| [56] | Cloud-based | Cryptography anonymization access control | Cryptography access control | Patient preferences | Storage process | P1; P5; P8; P10; P11 |
| [57] | Edge/Cloud-based | Cryptography | Cryptography authentication process | Patient preferences | Transmission storage process | P1; P2; P5; P8; P9 |
| [58] | Cloud-based | Cryptography access control | Access control | No | Collection process | P1; P2; P8; P11 |
| [59] | Privacy-based | Access control | Authentication process | No | Transmission process | P6; P10; P11 |
| [60] | Cloud-based | Cryptography | Cryptography | Privacy policy | Process | P2; P5; P9; P10 |
| [61] | Fog/Edge/Cloud-based | Access control | Access control | Patient preferences | Collection transmission | P1; P3; P4; P7; P8; P11 |
| [62] | Cloud-based | Cryptography anonymization | Smart contracts | No | Transmission process | P2; P5; P10 |
| [63] | Cloud-based | Cryptography | Cryptography | Patient preferences | Transmission/Storage | P7; P5; P8; P9; P10 |
| [64] | Policy-based | Anonymization access control | Access control | Privacy policy | Transmission storage process | P1; P6; P7; P8; P10; P11 |
| [65] | Cloud-based, distributed data | Anonymization | No | Privacy policy | Collection storage | P3; P4; P7; P10 |
| [66] | Cloud-based | Cryptography | No | No | Transmission | P6; P10; P11 |
| [67] | Privacy-based | Cryptography access control | Access control | Patient preferences/Privacy policy | Transmission | P2; P10; P11 |
| [68] | Cloud-based | Cryptography access control | Access control | No | Transmission storage | P6; P9; P10; P11 |
| [69] | Cloud-based | Cryptography Anonymization Access control | Cryptography access control | No | Storage process | P6; P10; P11 |
| [70] | Cloud-based | Cryptography anonymization access control | Cryptography access control | No | Storage process | P7; P8; P10; P11 |
| [71] | Privacy-based | Cryptography anonymization | Cryptography authentication process | No | Process | P5; P9; P10 |
| [72] | Cloud-based | Anonymization | No | No | Collection | P3; P4; P9; P10; P11 |
| [73] | Cloud-based | Cryptography | No | Privacy policy Compliant with laws privacy policy | Storage process | P2; P5; P10; P11 |
| [74] | Privacy-based | Cryptography | Access control | Privacy policy Compliant with laws privacy policy | Collection process | P1; P2; P3; P7; P8; P11 |
| [75] | Cloud-based | Cryptography | Cryptography | No | Transmission storage | P7; P8; P10 |
| [76] | Edge/Cloud-based | Cryptography | No | Privacy policy | Collection | P5; P6; P7; P10; P11 |
| [77] | Cloud-based | Cryptography anonymization access control | Cryptography access control authentication | No | Storage process | P6; P9; P11 |

TABLE 4: Continued.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|---------------|--|---------------------------------------|---------------------|----------------------|----------------------------------|
| [78] | Cloud-based | Cryptography | No | Compliant with laws | Transmission storage | P9; P10; P11 |
| [79] | Privacy-based | Cryptography anonymization access control | Access control authentication process | Privacy policy | Storage process | P2; P5; P8; P10 |
| [80] | Cloud-based | Cryptography access control | Access control | No | Storage process | P2; P5; P8 |

Moreover, Alraja et al. [29] focused on protecting the privacy of the IoT users, and helping them make pragmatic data-sharing deals with smart services and data consumers by determining the existing privacy risks concerning each data sharing.

Singh and Chatterjee [30] designed a smart healthcare system based on edge computing architecture which consists of an intermediary layer called an edge computing layer responsible for maintaining the network latency and preserving the privacy of the patient data.

The emerging healthcare Industrial Internet of Things (HealthIIoT) faces several fundamental security and privacy challenges, such as secure fine-grained data delivery, privacy-preserving keyword-based ciphertext retrieval, and malicious key delegation. For these challenges, Sun et al. [31] proposed a Privacy-aware and Traceable Fine-grained System (PTFS) in cloud-assisted HealthIIoT, which enables secure fine-grained data delivery, privacy-preserving data retrieval, efficient encryption, and decryption operations.

Sathya and Raja [32] proposed a Euclidean L3P-based Multiobjective Successive Approximation (EMSA) algorithm, a powerful measure of privacy in the smart healthcare environment. Based on the critical foundation for the storage of sensitive data in cloud environments, the role-based encryption keys.

Furthermore, other research groups and Ogundoyin et al. [33] proposed a lightweight privacy-preserving authentication and fine-grained access control scheme (PAASH) for smart health. This study addresses the security, efficiency, and privacy challenges of smart healthcare in smart cities.

Furthermore, Vineela et al. [34] proposed an authentication scheme for preserving the security and privacy of the big data in a cloud environment; this schema follows a mutual authentication and performs encryption operation between user and cloud environment.

A human-in-the-loop-aided (HitL-aided) scheme was designed by Zhou et al. [35] to preserve privacy in smart healthcare. They employed a block design technique to obfuscate various health indicators from the hospitals and the smart devices. They also introduced a human-in-the-loop (HitL) to enable privacy access of the health reports from the smart healthcare platform.

In addition, in a study made by Krall et al. [36], a new approach for preserving privacy in the framework of predictive modeling was proposed. This solution meets the

requirement of differential privacy while mitigating the risk of model inversion.

Based on machine learning techniques to detect deviated user access against Electronic Health Records (EHR), and to maintain the privacy of healthcare data, Hussain Seh et al. [37] defined an efficient framework for securing the privacy and confidentiality of healthcare data proactively. On the other hand, He et al. [38] presented a password strength meter that takes into account users' personal information. It helps users to select passwords with a higher degree of security.

Furthermore, in another recent work performed by Ibaida et al. [39], a novel privacy-preserving and efficient technique was proposed, that implements a lightweight shallow neural network to reduce the burden on the network while ensuring the privacy of the Electrocardiogram signals (ECG).

For the same purpose, another recent work carried out by El Zouka et al. [40] defined a secured healthcare monitoring system using fuzzy logic-based decision support (FBIS) systems to get the status of the patient. The proposed model consists of a trusted environment that is responsible for collecting authenticated physiological data.

Furthermore, a secure certificateless searchable public-key encryption (SPE) scheme for SHS was defined by Ma et al. [41], named SCF-CLSPE scheme, and it can resist keyword guessing attacks (KGA) and chosen keyword attacks (CKA) under the standard model. This scheme was also tested and proved that it has lower computation and communication costs.

Jayaram and Prabakaran [42] presented an edge-level privacy-preserving additive homomorphic encryption for secure data processing and filtering nonsensitive data in the edge layer. Also, an adaptive weighted probabilistic classifier model is proposed in the cloud layer for onboard disease prediction and rehabilitation of remote patients.

Also, many healthcare-based solutions, including Refs. [82–84], focus on predicting serious disease using deep learning, machine learning, or a combination of the two. These works aim to analyze and monitor patient health to prevent severe health complications. Yet, these works focus more on patients' data while little attention is given to patient's privacy. Contrary to these propositions, the work performed by Ge et al. [43] while aiming to predict disease by using deep learning also assured the data deletion approach by the data owner to limit access to their health data.

Toward the identification of anomalous behaviors within electronic patient record (EPR) datasets, researchers Hurst et al. [44] presented an investigation methodology. The proposed framework uses the LOF algorithm to detect unusual data patterns, labeling points as normal or anomalous, under the consideration of an HIL approach.

Moreover, Abdo et al. [45] used machine learning techniques for classifying a user's health state and crowdsensing for collecting information about a person's privacy preferences. They proposed a novel cloud-based secure location privacy-preserving mobile healthcare framework with securely storing, processing, and decision-making capabilities.

Furthermore, to provide users with secure and efficient access to their data, a lightweight user authentication system was designed by Kaul et al. [46]. In order to prevent unauthorized users from accessing the data, a proposed authentication describes a lightweight data access control process.

Moreover, with the intent to protect a patient's images from a compromised broker, Hamza et al. [47] proposed a privacy-preserving chaos-based encryption cryptosystem. They proposed a fast probabilistic cryptosystem to secure medical keyframes that are extracted from wireless capsule endoscopy procedures using a prioritization method.

Hathaliya et al. [48] proposed a mobile-based healthcare system with a biometric authentication approach, to ensure the security and privacy of electronic healthcare records in the Healthcare 4.0 era. For the same purpose, researchers Hathaliya et al. [49] previously proposed a biometric-based authentication scheme to ensure secure access of the patient's EHR from any location. The proposed scheme is tested and validated by the Automated Validation of Internet Security Protocols and Applications (AVISPA) tool.

Furthermore, in a study reported by Xie et al. [50], iCLAS was presented, which is an improved certificateless aggregate signature scheme that can resist all kinds of security attacks and can ensure patient privacy protection.

A privacy preservation framework was presented by Azad et al. [51] within smart context-aware healthcare emphasizing privacy assurance challenges within Electronic Transfer of Prescription. They proposed an enhancement to the widely used Salford model to achieve privacy preservation against masquerading and impersonation threats.

An anonymity-based user authentication protocol is preferred to resolve the privacy preservation issues in the IoMT. For this purpose, Deebak et al. [52] proposed a Secure and Anonymous Biometric Based User Authentication Scheme (SAB-UAS) to ensure secure communication in healthcare applications.

Moreover, in Yang et al. [53], a privacy-preserving smart IoT-based healthcare big data storage system with self-adaptive access control was defined. This solution aims at solving the following challenges: privacy of patients' medical data, access control in emergency scenarios, and optimization of data storage in big data systems. In another recent work by the same group, Yang et al. [54] proposed a privacy-preserving e-health system, where it defines a noninteractive and authenticated key distribution procedure for the

medical IoT network, as well as a novel keyword match-based policy update mechanism. Also, note that this system is a fusion of Internet-of-things (IoT), big data, and cloud storage.

Furthermore, in a study reported by Aghili et al. [55], the limitations of the previously proposed lightweight RFID mutual authentication (LRMI) protocol were presented, to eventually propose a new secure and lightweight mutual RFID authentication (SecLAP) protocol that provides secure communication and preserves privacy in the smart medical systems. The proposed security features are verified using the BAN logic.

Greene et al. [56] proposed ShareHealth, an end-to-end system for secure sharing of the collected medical data, by allowing the data owners to specify access-control policies and to cryptographically enforce those policies.

In addition, with the aim of solving the problem of limited computation ability of sensors on a patient in a smart healthcare system, Ding et al. [57] proposed a lightweight secure smart healthcare storage system that employs edge servers to compute data authenticators and verify data integrity.

MPPDS, a novel collaborative eHealth system that supports Multilevel Privacy-Preserving Data Sharing, was developed by Kim et al. [58]. The proposed system gives the data owner the possibility to share his or her health data with several data users within a collaborative eHealth system, under different levels of privacy protection.

Huang et al. [59] presented a practical scheme that can reliably authenticate patients with biometric authentication; electrocardiogram (ECG) signals, and provide differentially private protection simultaneously.

Furthermore, intending to protect patients' sensitive data while the smart health platform needs to do some analysis over these data, researchers Wang et al. [60] proposed a privacy-preserving outsourced computation scheme in the healthcare system. They enhanced the security of this scheme by splitting the decryption permissions into both servers.

A fog-based access control model was proposed by Wang et al. [61], to ensure high-level privacy protection without reducing the efficiency in cloud/fog computing, especially on the Internet of medical things IoMT.

Zhang et al. [62] defined a secure smart healthcare system based on a leakage-resilient anonymous HIBE scheme in the bounded leakage model. It can protect the patient's privacy well, even when the private key is partially leaked. It also achieves the safe transmission of the patient's electronic health records (EHR) in the case of leakage attacks.

Based on multi-party random masking and polynomial aggregation techniques, Kaur et al. [63] proposed a PPCF scheme. Privacy-Preserving Collaborative Filtering scheme on Arbitrary Distributed Data (ADD), where two phases are considered namely: off-line model generation and online prediction generation.

Vora et al. [64] presented an approach to preserve the identity and to protect the privacy of clinical data using an ARCANA encryption scheme. They also discussed an

authorization framework using access of varying degrees. Moreover, they had implemented the AT&T scheme for managing the access control mechanism of patients' data.

A practical framework called PrivacyProtector was defined by Luo et al. [65]. The proposed framework is a patient privacy-protected data collection, intending to prevent collision attacks and data leakage. PrivacyProtector includes the ideas of secret sharing and shares repairing for patients' data privacy.

In another study made by Elhoseny et al. [66], a hybrid security model was proposed. This model aims to secure the diagnostic text data in medical images. They also developed through integrating 2-D discrete wavelet transform 1 level (2D-DWT-1L) steganography technique with a proposed hybrid encryption scheme.

An attribute-based credential (ABC) was presented by Maria de Fuentes et al. [67], to cope with smart health privacy issues and to set the stage for the further adoption in other privacy-aware IoT-based smart cities' services.

Zhang et al. [68] introduced PASH, a privacy-aware s-health access control system, in which the key ingredient is a large universe ciphertext-policy attribute-based encryption (CP-ABE) with access policies partially hidden.

In another work, Zhang et al. [69], the authors, have introduced SSH, a Secure Smart Health system with privacy-aware aggregate authentication and access control in IoT. This solution is built on an anonymous certificateless aggregate signature and an anonymous CP-ABE scheme.

An efficient work carried out by Zheng et al. [70] presented a medical data sharing scheme in cloud storage. To solve the privacy issues in users' data sharing, they utilize attribute-based encryption to enable data sharing. And, they use the attribute bloom filter to hide all the attributes in the access control structure.

Zhang et al. [71] provided a privacy protection mechanism offering biometric authentication that allows the server to authenticate users with a biometric template. The user's anonymity is maintained during the authentication and key negotiation process.

A novel method for preserving the privacy of the collected data in the healthcare environment was developed by Kim et al. [72]. The proposed method is characterized as temporal data collected at fixed intervals by leveraging local differential privacy.

Practical privacy-preserving analytics in healthcare information systems was developed by Sharma et al. [73]. The study is based on kHealth, a personalized digital healthcare information system that is being developed and tested for disease monitoring.

Furthermore, with the aim of proposing the requirements and the practical approaches that should be considered when designing and developing IoT for data collection and data sharing within the healthcare domain, O'Connor et al. [74] define a "Privacy by Design approach."

A methodology to secure patients' medical big data MBD in the healthcare cloud was proposed by Al Hamid et al. [75], using the decoy technique with a fog computing facility. It is based on the bilinear pairing cryptography that can generate

a session key among the participants and communicate among them securely.

A pioneer work carried out by Bhuiyan et al. [76] investigated the concerns with privacy-protected data collection. For this purpose, a new secret sharing scheme and a share reconstruction scheme were defined for patient data privacy. They consider a distributed database consisting of multiple edge servers and each server receives a share of the patient data.

A new schema named OOABS was defined by Liu et al. [77] to replace the traditional Mobile Internet Devices (MIDs), and embedded Devices (EDs) of the electronic Health systems, to overcome their limitations in terms of storage space, power supply, and computational capacity.

Yang et al. [78] proposed a new solution to preserve privacy in e-health. This solution is based on the dynamic searchable symmetric encryption scheme with forwarding privacy and delegated verifiability for periodically generated healthcare data.

Rahman et al. [79] designed a security framework named PriSens-HSAC. The proposed framework is the first framework that provides increased privacy for Radio Frequency Identification (RFID) based healthcare systems, using RFID authentication along with access control techniques.

Moreover, Zhang et al. [80] defined a secure smart healthcare system that fulfills fine-grained access control on smart healthcare cloud data and hence ensures users' privacy protection. The key technique is a promising cryptographic primitive called ciphertext-policy attribute-based encryption.

4.2.2. Decentralized Architecture. Each primary research paper of this class was read in full, and relevant data were extracted and summarized in Table 5. The main idea of each paper is also recorded below in the following section. Chelladurai et al. [85] proposed a Patient-Centric secure EHR Management system using blockchain technology, to provide a regulated solution to the requirements of patients, doctors, and health service providers with integrity management. The proposed system provides high security and integrity through cryptographic hash functions. Lee et al. [86] proposed a blockchain-based medical data preservation scheme for telecare medical information systems (TMISs), which consist of a medical sensor area authentication protocol (WBAN) and a social network information transfer protocol. A Double Blockchain Telemedicine Diagnosis (DBTMD) scheme was proposed by Wang et al. [87] for privacy protection, which constructs a public chain Userchain and a consortium Medicalchain. It also develops an identity authentication chain to ensure the real-time accuracy of the doctor's identity information. This study reduces the communication costs of keys' transactions.

Furthermore, Wang et al. [88] proposed a data privacy protection, efficient retrieval, and analysis service scheme of IoMT based on low-cost fog computing. The fog computing system is set between the IoMT and cloud services, and

TABLE 5: Summary of studies reported on decentralized architecture.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|------------------|---------------------------------------|---|---|---|--------------------------------------|
| [85] | Blockchain-based | Cryptography blockchain | Access control smart contracts | No | Collection transmission storage process | P1; P6; P7; P8; P10 |
| [86] | Blockchain-based | Cryptography blockchain anonymization | Cryptograph access control | No | Collection transmission storage | P1; P6; P10 |
| [87] | Blockchain-based | Cryptography blockchain | Cryptography access control authentication process | No | Transmission storage process | P1; P4; P5; P6; P7; P8; P10; P11 |
| [88] | Fog/Cloud-based | Cryptography access control | Cryptography access control | No | Transmission storage process | P5; P6; P10; P11 |
| [89] | Blockchain-based | Blockchain access control | Cryptography smart contracts authentication process | No | Transmission storage process | P4; P5; P6; P7; P8; P10; P11 |
| [90] | Blockchain-based | Cryptography blockchain anonymization | Cryptograph access control | No | Collection transmission storage | P1; P5; P6; P7; P10 |
| [91] | Fog/Edge-based | Cryptography | No | No | Transmission storage | P6; P10 |
| [92] | Blockchain-based | Cryptography blockchain | Cryptography | No | Transmission storage process | P5; P6; P7; P8; P10; P11 |
| [93] | Blockchain-based | Cryptography blockchain | Smart contracts | Patient preferences privacy policy | Transmission storage process | P1; P3; P4; P5; P7; P11 |
| [94] | Blockchain-based | Cryptography blockchain | Cryptography smart contracts | Patient preferences | Storage process | P7; P8; P10; P11 |
| [95] | Blockchain-based | Cryptography blockchain | Cryptography smart contracts | Patient preferences | Storage process | P1; P3; P7; P8; P10; P11 |
| [96] | Blockchain-based | Blockchain access control | Access control smart contracts | Compliant with laws patient preferences | Collection storage process | P1; P2; P3; P5; P7; P8; P9; P10; P11 |
| [97] | Blockchain-based | Cryptography blockchain | Smart contracts | Patient preferences privacy policy | Transmission storage process | P1; P3; P4; P5; P7; P8; P10; P11 |
| [98] | Blockchain-based | Cryptography blockchain | Cryptography | Patient preferences | Storage process | P1; P3; P7; P8; P10; P11 |
| [99] | Blockchain-based | Cryptography blockchain | No | No | Transmission storage | P7; P10; P11 |
| [100] | Blockchain-based | Cryptography blockchain | Cryptography | No | Storage | P6; P10; P11 |
| [101] | Blockchain-based | Cryptography blockchain | Cryptography smart contracts authentication process permissioned blockchain | Patient preferences | Storage process | P1; P5; P7; P8; P10; P11 |
| [102] | Blockchain-based | Cryptography blockchain | Cryptography smart contracts | Compliant with laws | Transmission storage | P1; P2; P6; P7; P8; P9; P10; P11 |
| [103] | Blockchain-based | Cryptography blockchain | Smart contracts | Compliant with laws privacy policy | Transmission storage process | P2; P5; P6; P7; P8 |
| [104] | Blockchain-based | Cryptography blockchain | Authentication process | No | Storage | P5; P6; P7; P10; P11 |
| [105] | Blockchain-based | Cryptography blockchain | Smart contracts authentication process | Compliant with laws patient preferences | Transmission storage process | P1; P2; P5; P6; P7; P8; P10; P11 |
| [106] | IOTA-based | Cryptography | Access control authentication process | Patient preferences | Transmission storage process | P1; P5; P6; P7; P8; P10; P11 |

TABLE 5: Continued.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|------------------|--|--|---|------------------------------|----------------------------------|
| [107] | Blockchain-based | Cryptography blockchain | Smart contracts authentication process | Compliant with laws patient preferences | Transmission storage process | P1; P2; P5; P6; P7; P8; P10; P11 |
| [108] | Blockchain-based | Blockchain cryptography anonymization access control | Cryptography smart contracts | No | Transmission storage process | P5; P6; P10; P11 |

provides low latency, high computing efficiency, and decentralization.

Moreover, in a research paper made by Kumar et al. [89], a smart contract-enabled consortium blockchain network was defined, which is built on the interplanetary file systems (IPFS) cluster node and smart contracts for authentication of patients and medical devices.

Furthermore, in a study reported by Zhang et al. [90], the PTBM scheme was proposed, a contact tracing scheme in 5G-integrated and Blockchain-based Medical applications, which enables patients' location tracking and checking in a privacy-preserving manner.

Wang et al. [91] define a computation transferable authenticated key agreement protocol without an online registration center for smart healthcare. The proposed scheme adopts certificateless public-key cryptography, which can solve the problems of certificate management and key escrow. For the same intent, researchers Wang et al. [92] previously proposed GuardHealth, a decentralized blockchain system for data privacy-preserving and sharing. The proposed system manages confidentiality, authentication, data preserving, and data sharing when handling sensitive information.

A blockchain-based knapsack system has been proposed by Ranjith and Mahantesh [93]. The proposed blockchain method was evaluated on medical data to analyze the performance. The results show that the proposed method has less computation time and memory use compared to the existing methods. For the same purpose, in a pioneer work carried out by Dai et al. [94], a blockchain-enabled IoMT was proposed to increase the security and privacy concerns of IoMT systems. They also discuss the solutions brought by blockchain-enabled IoMT to COVID-19 from five different perspectives.

Moreover, with the aim to access control over individual health data, Jaiman and Urovi [95] presented a blockchain-based data-sharing consent model by using smart contracts. The dynamic consent model extends to two ontologies: The Data Use Ontology (DUO), which models the individual consent of users, and the Automatable Discovery and Access Matrix (ADA-M), which describes queries from data requesters.

Zhuang et al. [96] presented a blockchain model that achieves patient-centric HIE to protect data security and patients' privacy, ensure data provenance, and provide

patients full control of their health records, by personalizing data segmentation and an "allowed list" for clinicians to access their data.

Uddin et al. [97] proposed a blockchain leveraged decentralized architecture for eHealth. This architecture comprises three layers, the sensing layer (Body Area Sensor Network), the NEAR processing layer (the Fog), and the FAR processing layer (the Cloud).

Furthermore, a study was performed by Aruna Sri and Lalitha Bhaskari [98] aiming at analyzing blockchain-based encryption for patients' data and proposes a consensus mechanism to validate Proof of Word and Interoperability for data discovery and access.

Sun et al. [99] defined an attribute-based encryption scheme for secure storage and efficient sharing of electronic medical records in an InterPlanetary File System (IPFS) storage environment. The proposed model includes fine-grained and flexible access control, revocability of consent, auditability, and tamper resistance.

Tripathi et al. [100] proposed the Smart and Secured Healthcare System (S2HS), which is a two-level blockchain-based smart healthcare systems (SHS) framework to provide intrinsic security and integrity of the system.

Moreover, Usman and Qamar [101] presented a blockchain-based records management system that implements permissioned blockchain platform "Hyperledger" for efficient management and sharing of electronic medical records (EMRs).

In a research paper made by Hylock and Zeng [102], HealthChain was presented, a novel patient-centered blockchain framework to support immutable logging, promote patient engagement, and facilitate secure mediated information exchange between patients and providers.

Existing solutions on retrieval of electronic medical records either fail to protect sensitive data or are limited to a single image data provider. To resolve these challenges, Shen et al. [103] proposed a medical encrypted image retrieval scheme based on blockchain for privacy protection. They presented the layered architecture and threat model of the proposed scheme, using the emerging blockchain techniques.

In another study made by Xu et al. [104], the Healthchain scheme was proposed, a large-scale health data privacy-preserving scheme based on blockchain technology, where health data are encrypted to conduct fine-grained access control.

Daraghmi et al. [105] designed a MedChain system for medical records access management, where the timed-based smart contracts can interact with the various demands of health providers, patients, and third parties.

To provide a storage solution while preserving privacy for users, Li et al. [106] proposed a novel blockchain-based data preservation system (DPS) for medical data. With the proposed system, users can preserve important data in perpetuity, and the originality of the data can be verified if tampering is suspected.

Furthermore, Dagher et al. [107] proposed a framework named Ancile, which utilizes smart contracts in an Ethereum-based blockchain to define heightened access to medical records by the different stakeholders while preserving the privacy of patients' sensitive information.

In another work carried out by Brogan et al. [108], the role of distributed ledger technologies in ensuring security within electronic health was highlighted and proposes a Masked Authenticated Messaging (MAM) module of the IOTA protocol, which focuses on the transport of health activity data generated by wearable and embedded devices to a distributed ledger.

In order to handle the aim of protecting health information (PHI) generated by IoMT devices, Griggs et al. [109] proposed utilizing blockchain-based smart contracts to facilitate secure analysis and management of medical sensors. Using a private blockchain based on the Ethereum protocol.

Zhang et al. [110] proposed a BSPP scheme, blockchain-based secure and privacy-preserving personal health information (PHI) sharing scheme for diagnosis improvements in eHealth systems. The scheme is constructed using two blockchains, private blockchain for storing the PHI and consortium blockchain for maintaining the records of its secure indexes.

In a pioneer work carried out by Al Omar et al. [111], a MediBchain was presented, a patient-centric healthcare data management system by using blockchain as storage to attain privacy. Pseudonymity is ensured by using cryptographic functions to protect patients' data.

While a lot of blockchain-based solutions for smart healthcare focusing on the nature of the network architecture as a first step toward ensuring patients' privacy, some recent contributions give a particular attention to the protection of the communication contents and real identities of the nodes in a blockchain-based environment.

In this context, an interesting work was proposed in Ref. [112] where authors suggest a large-scale and efficient batch verification scheme based Elliptic Curve Digital Signature Algorithm (ECDSA) and group testing technology. Indeed, contrary to many recent propositions, this research paper does not only focus on improving the efficiency of batch verification algorithms but also considers the problem of invalid signatures identification. By doing so, this paper resolves the problem of performance degradation in case the batch verification fails. Another research work [113] emphasizes the importance of considering the open communication channel between patients and healthcare professionals in an (IoMT)-based environment. In particular, and based on the blockchain technology, authors

propose a lightweight and reliable authentication protocol while trying to address the problem of physical layer security and over-centralized server in wireless medical sensor networks.

4.2.3. Third-Party Architecture. Herein, the main results of the papers where a public institution or a private corporation is responsible for the data management are given followed by a summarized illustration of the studies, as shown in Table 6.

In a work performed by Larrucea et al. [114], the Healthcare Industry architecture reference model was extended, with a set of tools dealing with consent management and data hiding tools, while considering the legal aspects such as general data protection regulation (GDPR).

In addition, a done made by Zhang et al. [115] proposes an efficient and privacy-preserving disease prediction system, called PPDP. In PPDP, patients' historical medical data are encrypted and outsourced to the cloud server, which can be further utilized to train prediction models by using the Single-Layer Perceptron learning algorithm in a privacy-preserving way.

Moreover, the CP-ABSC scheme was proposed by Rao [116], a Ciphertext-Policy Attribute-Based Signcryption, with a public ciphertext verifiability framework, that achieves essential security goals of an attribute-Based Encryption (ABE) and Attribute-Based Signatures (ABS) schemes such as data confidentiality, unforgeability, and signcryption privacy.

4.2.4. Hybrid Architecture. In this section, results and analysis of the relevant papers that combine several architectures are given, while the main points and limitations of the related papers are summarized in Table 7.

In the research made by Chen et al. [117], a medical data information system model was proposed. The proposed model is based on blockchain, the Internet of Things, cloud storage, and proxy re-encryption algorithm to realize the reliable collection, safe storage, and sharing of medical data.

In another recent work done by Wang et al. [118], a novel handover authentication model of ITS with multi-server edge computing architecture was defined, a handover authentication scheme that allows the authenticated server to assist users to subsequently authenticate with another server and blockchain technology to preserve user's privacy.

Furthermore, Ngabo et al. [119] proposed a public-permissioned blockchain security mechanism using the elliptic curve crypto (ECC) digital signature that supports a distributed ledger database to provide an immutable security solution, transaction transparency and prevent the patient electronic health records from tampering at the IoMT fog layer.

Healthcare facilities and insurance companies ought to guarantee authenticity before offering any assistance to an individual. Therefore, Al Omar et al. [120] had implemented a blockchain framework to safeguard patients' sensitive data and insurance policy. They defined a solution for the healthcare system that provides data privacy and transparency. Data privacy is shielded with cryptographic

TABLE 6: Summary of studies reported on third party architecture.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|-------------|-----------------------------|------------------------|------------------------------------|------------------------------|----------------------------------|
| [114] | Cloud-based | Cryptography | Access control | Compliant with laws privacy policy | Process | P1; P2; P5; P7; P8; P11 |
| [115] | Cloud-based | Cryptography | No | No | Transmission storage process | P2; P5; P8; P10 |
| [116] | Cloud-based | Cryptography access control | Cryptography | Privacy policy | Transmission storage process | P5; P8; P6; P10 |

TABLE 7: Summary of the studies reported on hybrid architecture.

| Primary studies | Category | Privacy techniques | Data-access management | Stakeholders needs | Data life cycle | Privacy criteria (ISO/IEC 29100) |
|-----------------|---|---------------------------------------|---|--|---|--|
| [117] | Blockchain-based, cloud-based | Cryptography blockchain anonymization | Permissioned blockchain | Compliant with laws | Collection transmission storage process | P1; P2; P5; P6; P7; P8; P10; P11 |
| [118] | Edge/Cloud-based, blockchain-based | Cryptography blockchain anonymization | Authentication process | No | Transmission storage process | P1; P6; P7; P10; P11 |
| [119] | Blockchain-based, Edge/Cloud-based | Blockchain cryptography | Smart contracts permissioned blockchain | Patient preferences | Collection transmission storage process | P1; P6; P7; P8; P10 |
| [120] | Blockchain-based cloud storage | Cryptography blockchain | Smart contracts | Privacy policy | Transmission storage process | P1; P3; P5; P7; P10, P11 |
| [121] | Blockchain-based, distributed data | Cryptography blockchain anonymization | Access control smart contracts | Privacy policy | Transmission storage process | P1; P7; P2; P5; P8; P6; P9; P10; P11 |
| [122] | Blockchain-based, privacy-based, Edge/Cloud-based | Blockchain anonymization | Cryptography smart contracts authentication process | Compliant with laws patient preferences privacy policy | Collection transmission storage process | P1; P2; P3; P4; P5; P6; P7; P8; P9; P10; P11 |
| [123] | Cloud-based, privacy-based | Cryptography anonymization | Cryptography access control | Privacy policy | Transmission | P2; P5; P7; P8; P9; P10; P11 |
| [124] | Cloud-based | Access control | Access control | No | Transmission storage | P2; P5; P8; P10 |
| [125] | Blockchain, cloud storage | Cryptography blockchain | Cryptography | No | Transmission storage | P1; P3; P5; P7; P10; P11 |
| [126] | Blockchain-based, cloud-based | Cryptography blockchain | Access control | No | Transmission storage process | P1; P3; P4; P5; P6; P7; P8; P10; P11 |
| [127] | Cloud-based | Cryptography anonymization | Cryptography | No | Storage | P2; P8; P5; P10 |
| [128] | Blockchain-based/ Cloud storage | Cryptography blockchain | Smart contracts permissioned blockchain | No | Transmission storage process | P2; P5; P8; P11 |

mechanisms, and insurance policies are included in blockchain via the Ethereum platform.

Moreover, Egala et al. [121] had introduced a blockchain-based novel architecture that affords a decentralized EHR and smart-contract-based service automation without settling with the system's security and privacy. In the proposed architecture, they had proposed the hybrid computing model with the blockchain-based distributed data storage system to overcome blockchain-based cloud-centric IoMT healthcare system disadvantages.

In another work carried out by Robles et al. [122], a new trustworthy personal data protection mechanism was

presented for well-being services, based on privacy-by-design technologies. This mechanism is based on Blockchain networks and indirection functions and tokens.

Besides, in a study reported by Sun et al. [123], lightweight policy-hiding ciphertext-policy attribute-based encryption CP-ABE scheme was defined, for the IoT-oriented smart health application.

Liu et al. [124] designed a cooperative privacy preservation scheme for wearable devices with identity authentication and data access control considerations in the time and space contexts. In the time-aware cloud computing mode, ciphertext policy attribute-based encryption is applied for

fine-grained access control, and a bloom filter is used to achieve an efficient data structure without privacy exposure. In the space-aware edge computing mode, secret sharing and MinHash-based authentication is designed to enhance privacy preservation along with similarity computing without revealing sensitive data.

Moreover, another research group Dwivedi et al. [125] proposed a modified blockchain model suitable for IoT devices that rely on their distributed nature and other additional privacy and security properties of the network to provide secure management and analysis of healthcare big data.

Furthermore, with the intent to support a tamper resistance feature, Thwin and Vasupongayya [126] proposed a blockchain-based personal health record system (PHR system). The proposed model employed proxy re-encryption and other cryptographic techniques to preserve privacy.

Natgunanathan et al. [127] proposed a location privacy protection mechanism in which location privacy is protected while maintaining the utility of the location data. The MPU has the necessary data and processing ability to decide whether the patient is in a critical state or not.

Several solutions have been introduced to control the consequence of attacks using the decentralized method, but these solutions somehow failed to assure the overall privacy of patient-centric systems. Therefore, Omar et al. [128] presented a patient-centric healthcare data management system that uses blockchain technology to store and protect their data. They use cryptographic functions to encrypt patients' data and to ensure pseudonymity.

5. Discussion

This section aims at answering the research questions, by combining the different aforementioned privacy aspects. It should be noted that the systematic review results may have been impacted by multiple parts such as selection of databases, researchers' ideas, and time restrictions.

RQ1: What are the proposed solutions to preserve privacy according to the different stakeholders' needs while considering data-access management?

To answer this question, we combine stakeholders' needs with data-access management. According to our study, few solutions (35%) are found to cover at least one stakeholder's needs. Figure 7 shows that the most addressed stakeholder need by the studied papers is the patient preferences (35 publications), followed by the privacy policy (31 publications). The compliance with laws is the least addressed by papers (22 publications). Access control (28 publications) is the most used data access management technique in the currently proposed solutions for each stakeholder's needs. Furthermore, smart contracts (20 publications) and cryptography (19 publications) have also emerged in the stakeholders' needs.

RQ2: What are the privacy criteria that have been considered by the proposed solutions, and in which data life cycle phase?

The result of combining the privacy criteria and the data life cycle facets is shown in Figure 8. According to our

analysis, we can observe the privacy criteria coverage by the different studies. Thus, the most covered criterion is the "information security (16.23%) followed by the 'Privacy compliance' criteria (12.52%). Moreover, Accuracy and quality (11.99%); Openness, transparency, and notice (11.52%); and Individual participation and access (11.4%) have also emerged in the whole data life cycle. This can be explained by the relationship between these principles and the most addressed data life cycle phases, such as storage and process. However, "Data minimization" (3.59%) is the least considered criterion.

It seems that few of the existing solutions (7%) address the whole data life cycle. Figure 9 shows that the most addressed data phase by the studied publications is the storage phase (75 publications), followed by the process phase (66 publications). The collection phase is the least enhanced by publications (20 publications).

RQ3: How and what are the techniques used by published papers to preserve privacy in smart healthcare, and in which architecture and category?

To answer the above research question, privacy-preserving techniques, category, and architecture used and reported in the reviewed studies are analyzed. As shown in Figure 10, cryptography is the dominant technique in most of the proposed solutions that helps to preserve privacy in several architectures. This blockchain-based category is used with 37% of the total publications.

Figure 11 shows that the most addressed architecture by the studied papers is the centralized architecture (58 publications), followed by the decentralized architecture (27 publications). Third-party architecture gives full trust to the third party for the whole data management and does not control data usage. Therefore, it is the least architecture (3 publications) addressed by the proposed solutions.

6. Toward a Holistic Approach to Preserve Privacy: Discussion and Future Directions

By integrating new and innovating technologies such as mobile and smart IoT solutions or favorizing decentralized healthcare systems over traditional health infrastructures, we guarantee that healthcare services are delivered faster and in an efficient way to patients. Still, as demonstrated throughout the proposed SLR, there are challenges to address, mainly the lack of a holistic approach that considers privacy all along the data lifecycle and the stakeholders' needs.

Protecting patients from any privacy violation becomes even harder since privacy has to be taken into ground during the collection, transmission, usage, and storage of patient's sensitive data. Nevertheless, what if we consider privacy even before developing such smart and innovative solutions? What if instead of trying harder to enhance actual smart IoT-based solutions for healthcare we emphasize the importance of privacy before even these systems come into existence?

We believe that privacy together with possible best practices to preserve it have to be designed before the development of smart solutions. To illustrate this, we consider

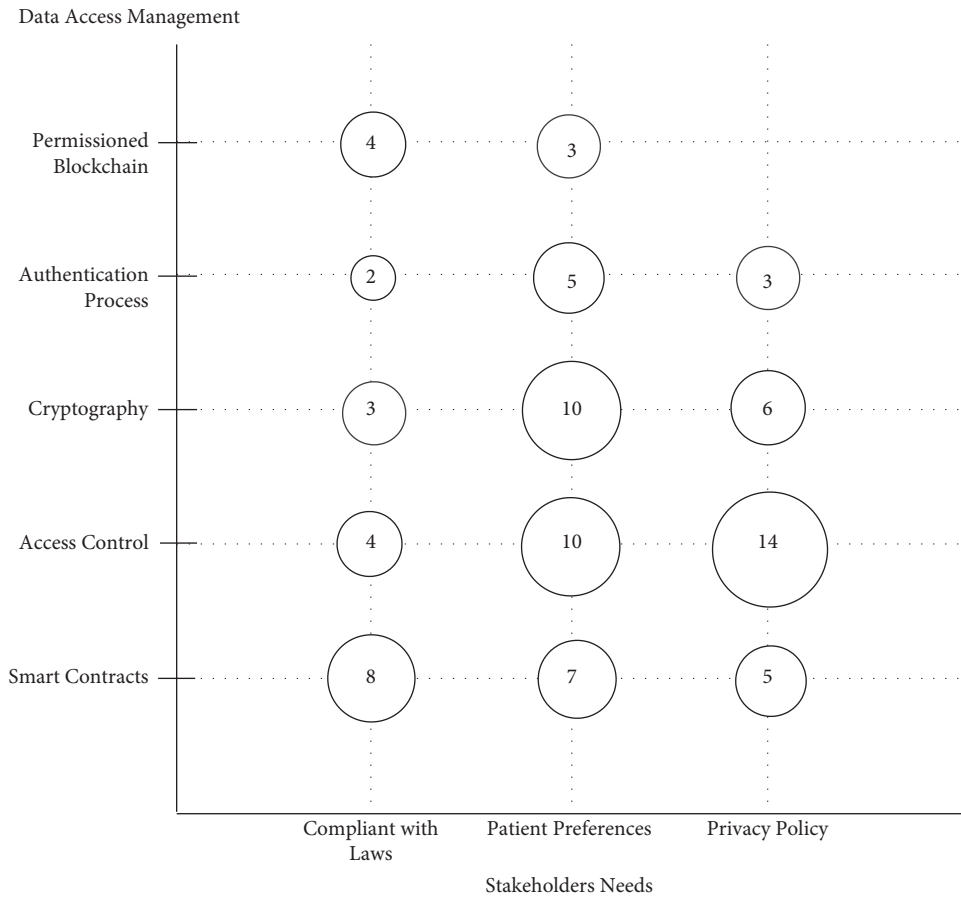


FIGURE 7: Papers distribution by the data access management and the stakeholders’ needs.

the classification previously suggested in section 2 (Fog/Edge/Cloud-based, blockchain-based, and policy-based) and we compare them according to the privacy-by-design principles. As stated in Ref. [129], the seven privacy-by-design principles are summarized in Table 8.

6.1. Discussion. Table 9 gives a clear picture of how the development in technologies for healthcare is significantly helping in protecting patients’ privacy. At the same time, these same technologies and new architectures need to be enhanced in order to meet patients’ privacy needs as well as the different stakeholders’ needs. We believe that it is a difficult equation to solve. Still, by encouraging organizations and project holders to follow best practices and raising their awareness about the different privacy risks, we guarantee that privacy risks and invasions can be reduced or prevented even before they happen. Therefore, the seven privacy-by-design principles were adopted as criteria or requirements that when met by a specific type of architecture (one of three proposed classes) reflects nothing but how much this architecture or solution considers the user’s privacy and also the other actors’ privacy needs. For instance, in cloud-based solutions by applying practices such as “least privilege,” which consists on granting only the required permissions to complete a task, we prevent unauthorized access which then leads to a proactive approach (P1). From another line, Cloud

providers such as Amazon Web Services (AWS) make security a shared responsibility between the service provider and the service consumer. This makes the design of privacy aligned with a win-win approach (P3).

In blockchain-based solutions, the fidelity and security of data are guaranteed, and trust is generated without the need for a trusted third party. This makes the exchange of private data performed in a transparent manner (P6). In addition to that, the architecture of a blockchain network and the adoption of schemes such as Zero-knowledge proof make this kind of solution highly secure while preserving user privacy (P3). To achieve this, data owners’ privacy is protected by the use of encrypted keys instead of divulging the real identity of users (P7).

In policy-based solutions, the way patients’ data are collected, stored, and shared is detailed by the means of formalized policies/preferences. Hence, patients can be reassured about who, when, and how these data are being used, stored, and shared (P6). Notably, the most important parameter in the third class (policy-based solution) is policy/preferences. That said, by involving the patient in the decision-making, we are indirectly protecting him or her from unauthorized use or disclosure (P7).

6.2. Recommendations and Future Directions. Health data are very private and can have a huge impact on patients’

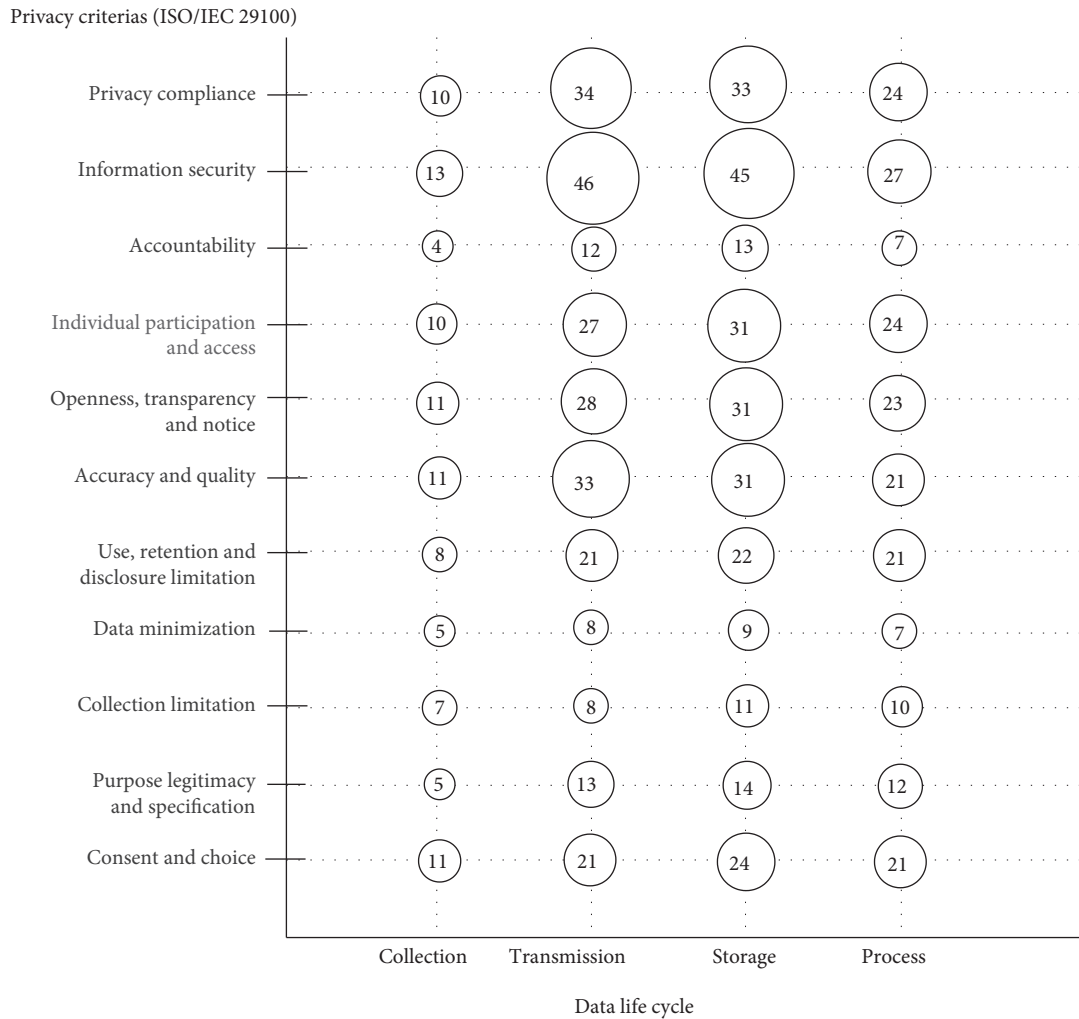


FIGURE 8: Papers distribution by the privacy criteria (ISO/IEC 29100) and the data life cycle.

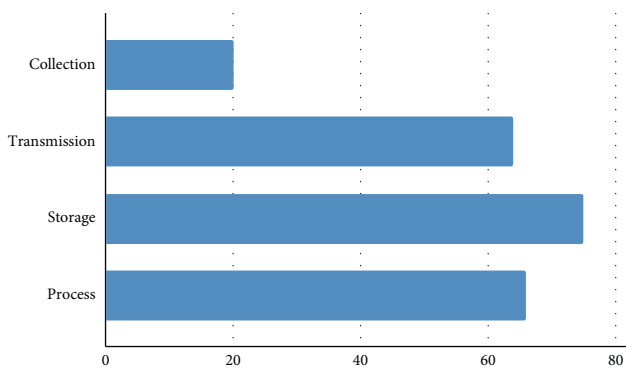


FIGURE 9: Data life cycle vs. number of publications.

health and outcomes. Hence, and in order to be reassured about the safety of their data, patients have the right to know what and how much data are collected, why are collected, and for what purposes these data are going to be used for. As a first step toward this aim, patients’ privacy preferences have to be well explained, formalized, and then shared across all the involved parties. Notably, both patients’ and third parties’ policies have to be accessible, clear, and regularly updated.

Next, as a second-best recommendation, it is of utmost importance to identify third parties that have access to patients’ data, especially when data are gathered from different sources. For instance, in IoT-based health solutions, medical data are generated from different wearable IoT devices and are then sent to different sources for processing and storage. Therefore, and since it is directly related to a patient’s life, these data have to be carefully used and shared [119]. Indeed, as stated in Ref. [130], Blockchain is a good solution to safeguard sensitive medical data generated by these IoT devices. First, thanks to its nature of decentralization, data are not owned nor managed by one single entity. Second, and particularly in private blockchains, it is possible to control who can access the blockchain network and may also control the type of access rights each entity has. Above this, personal data have to be kept anonymous so that they cannot be divulged—by mistake—to a specific person.

Apart from restricting access to patients’ data and respecting patients’ preferences when using and sharing data, it is highly recommended not to ask for more information than necessary from patients. These kinds of practices are strongly considered in healthcare cloud-based solutions

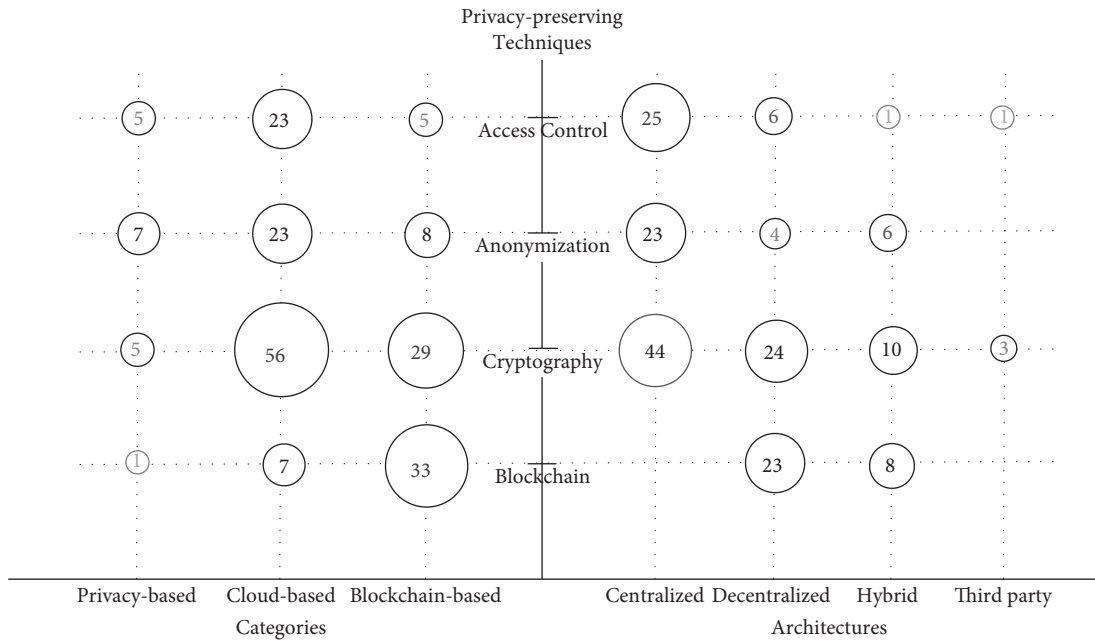


FIGURE 10: Papers distribution by the privacy-preserving techniques, categories, and architectures.

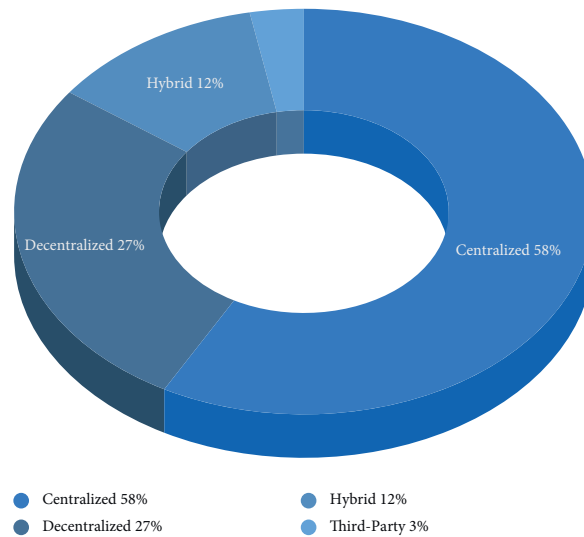


FIGURE 11: Storage architecture VS Number of Publications.

TABLE 8: Summary of the seven privacy-by-design principles.

| The principal | Definition |
|--------------------------------|---|
| Proactive not reactive | It refers to the ability of a system to anticipate and prevent privacy risks before they happen |
| Privacy as the default setting | This principle seeks to ensure a higher degree of personal data protection by enforcing privacy default settings. This way, privacy remains intact even if the user does not follow any privacy practices |
| Privacy embedded into design | This principle means that privacy has to be considered as an essential component or functionality of any system |
| Positive-sum, not zero-sum | Designing privacy follows a “win-win” approach where privacy can be reached by compromising other concepts. (e.g., security) |
| End-to-end security | This principle means that security, considered as an important concept for protecting patients’ data, has to be taken into grounded during the entire lifecycle of the data, from start to finish. |
| Visibility and transparency | By respecting this principle, we make sure that all stakeholders are able to see how information moves through the system |
| Respect for user privacy | This principle encourages organizations and solution architects to design user-centric solutions where both the owner of data as well as the other stakeholders play an active role |

TABLE 9: Comparison of the categories of MIoT and the privacy-by-design principles.

| Class | Principle | | | | | | |
|------------------|-----------|-----|----|-----|-----|----|-----|
| | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
| Cloud-based | + | - | - | +/- | + | - | +/- |
| Blockchain-based | + | +/- | + | + | + | + | + |
| Policy-based | +/- | - | - | +/- | +/- | - | + |

(e.g., AWS) where principles such as “the least privilege” are integrated and implemented in secure cloud services. This way, users have only access to resources they are allowed to. Also, it is worth nothing that the emplacement of data is another factor affecting patients’ privacy. In fact, it is highly recommended to carefully choose the location for storing sensitive data and ask questions such as: does this storage service follow the Security privacy by Design approach? Is the storage service provider compliant with local laws and regulations? Finally, more efforts have to be made to develop procedures to automatically detect, report, and investigate data breaches [131].

7. Conclusion and Future Work

In this research, we conducted a Systematic Literature Review (SLR), where we presented an exhaustive study of the existing privacy-preserving solutions in the smart healthcare environment. By analyzing the content of the primary studies based on several aspects, i.e., the implemented privacy techniques, privacy principles, IoMT architecture, stakeholders’ needs, etc., the designated research questions were answered. The findings of this study revealed that among one hundred primary studies, more than 70 percent had neglected the stakeholders’ needs, especially the compliance with patient privacy preferences and privacy laws; moreover, the data collection is the least considered phase, which translates the neglecting of the two criteria: data collection limitation and data minimization. Many other limitations are cited in the previous sections, which all point out the lack of a holistic approach that aims to preserve privacy all along the data life cycle and according to the different stakeholders’ needs. We believe that our systematic literature review will be of great help to researchers targeting privacy preservation in the IoMT field, as they can rely on the aforementioned limitations, to propose further scientific contributions.

In our ongoing research, we aim to propose a blockchain-based solution for privacy-preserving in IoMT, which overcomes the limitations of the prior solutions and considers the whole privacy aspects adopted in the previous analysis, to ensure full coverage of privacy and security.

Data Availability

The data used to support the findings of this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] N. Dilawar, M. Rizwan, and F. Ahmad, “Blockchain: securing internet of medical things (IoMT),” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, p. 31, 2019.
- [2] Y. Xiao and M. Watson, “Guidance on conducting a systematic literature review,” *Journal of Planning Education and Research*, vol. 39, no. 1, pp. 93–112, 2019.
- [3] S. S. Hameed, W. H. Hassan, L. A. Latiff, and F. Ghabban, “A systematic review of security and privacy issues in the internet of medical things; the role of machine learning approaches,” *Peer J Comput. Sci.* vol. 7, p. e414, 2021.
- [4] M. Tanriverdi, “A systematic review of privacy preserving healthcare data sharing on blockchain,” *Journal of Cybersecurity and Information Management*, vol. 5, pp. 31–37, 2020.
- [5] L. H. Iwaya, A. Ahmad, and M. A. Babar, “Security and privacy for mHealth and uHealth systems: a systematic mapping study,” *IEEE Access*, vol. 8, pp. 150081–150112, 2020.
- [6] J. J. Hathaliya and S. Tanwar, “An exhaustive survey on security and privacy issues in Healthcare 4.0,” *Computer Communications*, vol. 153, pp. 311–335, 2020.
- [7] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, and G. Wang, “Security and privacy in the medical internet of things: a review,” *Security and Communication Networks*, vol. 2018, Article ID e5978636, 9 pages, 2018.
- [8] Z. El Ouazzani, H. El Bakkali, and S. Sadki, “Privacy preserving in digital health: main issues, technologies, and solutions,” *Research Anthology on Privatizing and Securing Data*, IGI Global, Hershey, Pennsylvania, pp. 1503–1526, 2021.
- [9] W. Fang, X. Z. Wen, Y. Zheng, and M. Zhou, “A survey of big data security and privacy preserving,” *IETE Technical Review*, vol. 34, no. 5, pp. 544–560, 2017.
- [10] C. S. Kruse, B. Smith, H. Vanderlinden, and A. Nealand, “Security techniques for the electronic health records,” *Journal of Medical Systems*, vol. 41, no. 8, p. 127, 2017.
- [11] Z. El Ouazzani and H. El Bakkali, “A classification of non-cryptographic anonymization techniques ensuring privacy in big data,” *International Journal of Communication Networks and Information Security*, vol. 12, pp. 142–152, 2020.
- [12] D. Preethi, N. Khare, and B. K. Tripathy, “Security and privacy issues in blockchain technology,” in *Blockchain Technology and the Internet of Things*, Apple Academic Press, New Jersey, NJ, USA, 2020.
- [13] Risk Based Security, “2021 Mid year data breach QuickView report,” Risk Based Security, Richmond, VA, USA, <https://pages.riskbasedsecurity.com/hubfs/Reports/2021/2021%20Mid%20Year%20Data%20Breach%20QuickView%20Report.pdf>, 2021.
- [14] M. Verdonck and G. Poels, “Decentralized Data Access with IPFS and Smart Contract Permission Management for Electronic Health Records,” in *Proceedings of the*

- International Conference on Business Process Management*, Rome, Italy, September 2021.
- [15] G. Eysenbach, "What is e-health?" *Journal of Medical Internet Research*, vol. 3, no. 2, p. e833, 2001.
 - [16] Indicator Metadata Registry Details, "Indicator Metadata Registry Details," 2021, <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4774>.
 - [17] A. Solanaset, C. Patsakis, M. Conti et al., "Smart health: a context-aware health paradigm within smart cities," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 74–81, 2014.
 - [18] T. Francis, M. Madijagan, and V. Kumar, "Privacy issues and techniques in E-health systems," in *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research*, pp. 113–115, New York, NY, USA, June 2015.
 - [19] P. Shalin, D. Dharmin, P. Reema, and D. Nishant, "Security and privacy issues in cloud, fog and edge computing," *Procedia Computer Science*, vol. 160, pp. 734–739, 2019.
 - [20] D. C. Klonoff, "Fog computing and edge computing architectures for processing data from diabetes devices connected to the medical internet of things," *J Diabetes Sci Technol*, vol. 11, no. 4, pp. 647–652, 2017.
 - [21] D. E. Majdoubi and H. E. Bakkali, "A survey of major data privacy laws, languages and approaches in smart cities environments," in *Proceedings of the 4th International Conference on Smart City Applications*, pp. 1–8, New York, NY, USA, October 2019.
 - [22] The four most internationally significant recent privacy laws, "The four most internationally significant recent privacy laws," *TechGenix*, vol. 26, 2019.
 - [23] Online Browsing Platform (OBP)," 2021, <https://www.iso.org/obp/ui>.
 - [24] P. Kumar and L. Chouhan, "A privacy and session key based authentication scheme for medical IoT networks," *Computer Communications*, vol. 166, pp. 154–164, 2021.
 - [25] M. T. de Oliveira, H.-V. Dang, L. H. A. Reis, H. A. Marquering, and S. D. Olabbarriaga, "AC.-AC: Dynamic revocable access control for acute care teams to access medical records," *Smart Health*, vol. 20, Article ID 100190, 2021.
 - [26] H. Zhong, Y. Zhou, Q. Zhang, and Y. Xu, J. Cui, "An efficient and outsourcing-supported attribute-based access control scheme for edge-enabled smart healthcare," *Future Generation Computer Systems*, vol. 115, pp. 486–496, 2021.
 - [27] J. A. Onesimu, J. Karthikeyan, and Y. Sei, "An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services », Peer-to-Peer Netw, " *Peer-to-Peer Netw Appl*, vol. 14, no. 3, pp. 1629–1649, 2021.
 - [28] S. Izza, M. Benssalah, and K. Drouiche, "An enhanced scalable and secure RFID authentication protocol for WBAN within an IoT environment," *Journal of Information Security and Applications*, vol. 58, Article ID 102705, 2021.
 - [29] M. N. Alraja, H. Barhamgi, A. Rattrout, and M. Barhamgi, "An integrated framework for privacy protection in IoT — applied to smart healthcare," *Computers & Electrical Engineering*, vol. 91, Article ID 107060, 2021.
 - [30] A. Singh and K. Chatterjee, "Securing smart healthcare system with edge computing," *Computers & Security*, vol. 108, Article ID 102353, 2021.
 - [31] J. Sun, D. Chen, N. Zhang et al., "A privacy-aware and traceable fine-grained data delivery system in cloud-assisted healthcare IIoT," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 10034–10046, 2021.
 - [32] A. Sathya and S. K. S. Raja, "Privacy preservation-based access control intelligence for cloud data storage in smart healthcare infrastructure," *Wireless Personal Communications*, vol. 118, no. 4, pp. 3595–3614, 2021.
 - [33] S. O. Ogundoyin and I. A. Kamil, "PAASH: a privacy-preserving authentication and fine-grained access control of outsourced data for secure smart health in smart cities," *Journal of Parallel and Distributed Computing*, vol. 155, pp. 101–119, 2021.
 - [34] A. Vineela, N. Kasiviswanath, and C. ShobaBindu, "Theoretical analysis on applications aspects of smart materials preserving the security and privacy in medical big data and cloud," *Materials Today Proceedings*, 2021.
 - [35] T. Zhou, J. Shen, D. He, P. Vijayakumar, and N. Kumar, "Human-in-the-Loop-Aided privacy-preserving scheme for smart healthcare," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, pp. 1–10, 2020.
 - [36] A. Krall, D. Finke, and H. Yang, "Mosaic privacy-preserving mechanisms for healthcare analytics," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2184–2192, 2021.
 - [37] A. Hussain Seh, J. F. Al-Amri, A. F. Subahi, A. Agrawal, R. Kumar, and R. Ahmad Khan, "Machine learning based framework for maintaining privacy of healthcare data," *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 697–712, 2021.
 - [38] D. He, R. Ye, S. Chan, M. Guizani, and Y. Xu, "Privacy in the internet of things for smart healthcare," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 38–44, 2018.
 - [39] A. Ibaida, A. Abuadba, and N. Chilamkurti, "Privacy-preserving compression model for efficient IoMT ECG sharing," *Computer Communications*, vol. 166, pp. 1–8, 2021.
 - [40] H. A. El Zouka, M. M. Hosni, "Secure IoT communications for smart healthcare monitoring system," *Internet of Things*, vol. 13, Article ID 100036, 2021.
 - [41] M. Ma, D. He, S. Fan, and D. Feng, "Certificateless searchable public key encryption scheme secure against keyword guessing attacks for smart healthcare," *Journal of Information Security and Applications*, vol. 50, Article ID 102429, 2020.
 - [42] R. Jayaram, S. Prabakaran, "Onboard disease prediction and rehabilitation monitoring on secure edge-cloud integrated privacy preserving healthcare system," *Egyptian Informatics Journal*, vol. 22, 2020.
 - [43] C. Ge, C. Yin, Z. Liu, L. Fang, J. Zhu, and H. Ling, "A privacy preserve big data analysis system for wearable wireless sensor network," *Computers & Security*, vol. 96, Article ID 101887, 2020.
 - [44] W. Hurst, A. Boddy, M. Merabti, and N. Shone, "Patient privacy violation detection in healthcare critical infrastructures: an investigation using density-based benchmarking," *Future Internet*, vol. 12, p. 100, 2020.
 - [45] M. A. Abdo, A. A. Abdel-Hamid, and H. A. Elzouka, "A cloud-based mobile healthcare monitoring framework with location privacy preservation," in *Proceedings of the 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, pp. 1–8, Sakheer, Bahrain, December 2020.
 - [46] S. D. Kaul, V. K. Murty, and D. Hatzinakos, "Secure and privacy preserving biometric based user authentication with data access control system in the healthcare environment," in *Proceedings of the 2020 International Conference on Cyberworlds (CW)*, pp. 249–256, Caen, France, September 2020.

- [47] R. Hamza, Z. Yan, K. Muhammad, P. Bellavista, and F. Titouna, "A privacy-preserving cryptosystem for IoT E-healthcare," *Information Sciences*, vol. 527, pp. 493–510, 2020.
- [48] J. J. Hathaliya, S. Tanwar, and R. Evans, "Securing electronic healthcare records: a mobile-based biometric authentication approach," *Journal of Information Security and Applications*, vol. 53, Article ID 102528, 2020.
- [49] J. J. Hathaliya, S. Tanwar, S. Tyagi, and N. Kumar, "Securing electronics healthcare records in Healthcare 4.0: a biometric-based approach," *Computers & Electrical Engineering*, vol. 76, pp. 398–410, 2019.
- [50] Y. Xie, X. Li, S. Zhang, and Y. Li, "ICLASS: an improved certificateless aggregate signature scheme for healthcare wireless sensor networks," *IEEE Access*, vol. 7, pp. 15170–15182, 2019.
- [51] M. A. Azad, J. Arshad, S. Mahmoud, K. Salah, and M. Imran, "A privacy-preserving framework for smart context-aware healthcare applications," *Transactions on Emerging Telecommunications Technologies*, Article ID e3634, 2019.
- [52] B. D. Deebak, F. Al-Turjman, M. Aloqaily, and O. Alfandi, "An authentic-based privacy preservation protocol for smart e-healthcare systems in IoT," *IEEE Access*, vol. 7, pp. 135632–135649, 2019.
- [53] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system," *Information Sciences*, vol. 479, pp. 567–592, 2019.
- [54] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving fusion of IoT and big data for e-health," *Future Generation Computer Systems*, vol. 86, pp. 1437–1455, 2018.
- [55] S. F. Aghili, H. Mala, P. Kaliyar, and M. Conti, "SecLAP: Secure and lightweight RFID authentication protocol for Medical IoT," *Future Generation Computer Systems*, vol. 101, pp. 621–634, 2019.
- [56] E. Greene, P. Proctor, and D. Kotz, "Secure sharing of mHealth data streams through cryptographically-enforced access control," *Smart Health*, vol. 12, pp. 49–65, 2019.
- [57] R. Ding, H. Zhong, J. Ma, X. Liu, and J. Ning, "Lightweight privacy-preserving identity-based verifiable IoT-based health storage system," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8393–8405, 2019.
- [58] J. W. Kim, K. Edemacu, and B. Jang, "MPPDS: Multilevel privacy-preserving data sharing in a collaborative eHealth system," *IEEE Access*, vol. 7, pp. 109910–109923, 2019.
- [59] P. Huang, L. Guo, M. Li, and Y. Fang, "Practical privacy-preserving ECG-based authentication for IoT-based healthcare," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9200–9210, 2019.
- [60] Q. Wang, D. Zhou, S. Yang, P. Li, C. Wang, and Q. Guan, "Privacy preserving computations over healthcare data," in *Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 635–640, Atlanta, GA, USA, July 2019.
- [61] X. Wang, L. Wang, Y. Li, and K. Gai, "Privacy-aware efficient fine-grained data access control in internet of medical things based fog computing," *IEEE Access*, vol. 6, pp. 47657–47665, 2018.
- [62] Y. Zhang, P. Lang, D. Zheng, M. Yang, and R. Guo, "A secure and privacy-aware smart health system with secret key leakage resilience," *Security and Communication Networks*, vol. 2018, Article ID e7202598, 13 pages, 2018.
- [63] H. Kaur, N. Kumar, and S. Batra, "An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system," *Future Generation Computer Systems*, vol. 86, pp. 297–307, 2018.
- [64] J. Vora, I. Prit, S. Tyagi, N. Kumar, M. Obaidat, and K. F. Hsiao, "Ensuring privacy and security in E-health records," in *Proceedings of the 2018 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1–5, Alsace, Colmar, France, July 2018.
- [65] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiquzzaman, "Privacy-protected patient data collection in IoT-based healthcare systems," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 163–168, 2018.
- [66] M. Elhoseny, G. Ramirez-González, O. M. Abu-Elnasr, S. A. Shawkat, N. Arunkumar, and A. Farouk, "Secure medical data transmission model for IoT-based healthcare systems," *IEEE Access*, vol. 6, pp. 20596–20608, 2018.
- [67] J. Maria de Fuentes, L. Gonzalez-Manzano, A. Solanas, and F. Veseli, "Attribute-based credentials for privacy-aware smart health services in IoT-based smart cities," *Computer*, vol. 51, no. 7, pp. 44–53, 2018.
- [68] Y. Zhang, D. Zheng, and R. H. Deng, "Security and privacy in smart health: efficient policy-hiding attribute-based access control," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2130–2145, 2018.
- [69] Y. Zhang, R. H. Deng, G. Han, and D. Zheng, "Secure smart health with privacy-aware aggregate authentication and access control in Internet of Things," *Journal of Network and Computer Applications*, vol. 123, pp. 89–100, 2018.
- [70] D. Zheng, A. Wu, Y. Zhang, and Q. Zhao, "Efficient and privacy-preserving medical data sharing in internet of things with limited computing power," *IEEE Access*, vol. 6, pp. 28019–28027, 2018.
- [71] L. Zhang, Y. Zhang, S. Tang, and H. Luo, "Privacy protection for E-health systems by means of dynamic authentication and three-factor key agreement," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2795–2805, 2018.
- [72] J. W. Kim, B. Jang, and H. Yoo, "Privacy-preserving aggregation of personal health data streams," *PLOS ONE*, vol. 13, no. 11, Article ID e0207639, 2018.
- [73] S. Sharma, K. Chen, and A. Sheth, "Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems," *IEEE Internet Computing*, vol. 22, no. 2, pp. 42–51, 2018.
- [74] Y. O'Connor, W. Rowan, L. Lynch, and C. Heavin, "Privacy by design: informed consent and internet of things for smart health," *Procedia Computer Science*, vol. 113, pp. 653–658, 2017.
- [75] H. A. Al Hamid, S. M. M. Rahman, M. S. Hossain, A. Almogren, and A. Alamri, "A security model for preserving the privacy of medical big data in a healthcare cloud using a fog computing facility with pairing-based cryptography," *IEEE Access*, vol. 5, pp. 22313–22328, 2017.
- [76] M. Z. A. Bhuiyan, M. Zaman, G. Wang, T. Wang, and J. Wu, "Privacy-protected data collection in wireless medical sensor networks," in *Proceedings of the 2017 International Conference on Networking, Architecture, and Storage (NAS)*, pp. 1–2, Shenzhen, China, August 2017.
- [77] J. Liu, J. Ma, W. Wu, X. Chen, X. Huang, and L. Xu, "Protecting mobile health records in cloud computing: a secure, efficient, and anonymous design," *ACM Transactions on Embedded Computing Systems*, vol. 16, no. 2, pp. 1–20, 2017.

- [78] L. Yang, Q. Zheng, and X. Fan, "RSPP: A reliable, searchable and privacy-preserving e-healthcare system for cloud-assisted body area networks," in *Proceedings of the IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, May 2017.
- [79] F. Rahman, M. Z. A. Bhuiyan, and S. I. Ahamed, "A privacy preserving framework for RFID based healthcare systems," *Future Generation Computer Systems*, vol. 72, pp. 339–352, 2017.
- [80] Y. Zhang, J. Li, D. Zheng, X. Chen, and H. Li, "Towards privacy protection and malicious behavior traceability in smart health," *Personal and Ubiquitous Computing*, vol. 21, no. 5, pp. 815–830, 2017.
- [81] J. Iqbal, M. Adnan, Y. Khan et al., "Designing a healthcare-enabled software-defined wireless body area network architecture for secure medical data and efficient diagnosis," *Journal of Healthcare Engineering*, vol. 2022, Article ID 9210761, 19 pages, 2022.
- [82] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 11 pages, 2021.
- [83] M. D. Amzad Hossen, T. Tazin, S. Khan, and E. Alam, "Supervised machine learning-based cardiovascular disease analysis and prediction," *Mathematical Problems in Engineering*, vol. 2021, Article ID 1792201, 10 pages, 2021.
- [84] R. Krishnamoorthi, S. Joshi, H. Z. Almarzouki et al., "A novel diabetes healthcare disease prediction framework using machine learning techniques," *Journal of Healthcare Engineering*, vol. 2022, Article ID 1684017, 10 pages, 2022.
- [85] M. U. Chelladurai, D. S. Pandian, and D. K. Ramasamy, "A Blockchain Based Patient Centric EHR Storage and Integrity Management for E-Health Systems," *Health Policy and Technology*, vol. 10, Article ID 100513, 2021.
- [86] T.-F. Lee, H.-Z. Li, and Y.-P. Hsieh, "A blockchain-based medical data preservation scheme for telecare medical information systems," *International Journal of Information Security*, vol. 20, no. 4, pp. 589–601, 2021.
- [87] W. Wang, L. Wang, P. Zhang et al., "A privacy protection scheme for telemedicine diagnosis based on double blockchain," *Journal of Information Security and Applications*, vol. 61, Article ID 102845, 2021.
- [88] N. Wang, Y. Cai, J. Fu, and J. Xu, "Privacy-preserving efficient data retrieval in IoMT based on low-cost fog computing," *Complexity*, vol. 2021, Article ID 6211475, 13 pages, 2021.
- [89] R. Kumar and R. Tripathi, "Towards design and implementation of security and privacy framework for Internet of Medical Things (IoMT) by leveraging blockchain and IPFS technology," *The Journal of Supercomputing*, vol. 77, no. 8, pp. 7916–7955, 2021.
- [90] C. Zhang, C. Xu, K. Sharif, and L. Zhu, "Privacy-preserving contact tracing in 5G-integrated and blockchain-based medical applications," *Computer Standards & Interfaces*, vol. 77, Article ID 103520, 2021.
- [91] W. Wang, H. Huang, F. Xiao, Q. Li, L. Xue, and J. Jiang, "Computation-transferable authenticated key agreement protocol for smart healthcare," *Journal of Systems Architecture*, vol. 118, Article ID 102215, 2021.
- [92] Z. Wang, N. Luo, and P. Zhou, "GuardHealth: blockchain empowered secure data management and Graph Convolutional Network enabled anomaly detection in smart healthcare," *Journal of Parallel and Distributed Computing*, vol. 142, pp. 1–12, 2020.
- [93] J. Ranjith and K. Mahantesh, "Blockchain-based knapsack system for security and privacy preserving to medical data," *SN COMPUT. SCI.* vol. 2, no. 4, pp. 1–7, 2021.
- [94] H.-N. Dai, M. Imran, and N. Haider, "Blockchain-enabled internet of medical things to combat COVID-19," *IEEE Internet of Things Magazine*, vol. 3, no. 3, pp. 52–57, 2020.
- [95] V. Jaiman, V. Urovi, A consent model for blockchain-based health data sharing platforms," *IEEE Access*, vol. 8, pp. 143734–143745, 2020.
- [96] Y. Zhuang, L. R. Sheets, Y.-W. Chen, Z.-Y. Shae, J. J. P. Tsai, and C.-R. Shyu, "A patient-centric health information exchange framework using blockchain technology," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2169–2176, 2020.
- [97] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "Blockchain leveraged decentralized IoT eHealth framework," *Internet of Things*, vol. 9, Article ID 100159, 2020.
- [98] P. S. G. Aruna Sri and D. Lalitha Bhaskari, "Blockchain technology for secure medical data sharing using consensus mechanism," *Materials Today: Proceedings*, 2020.
- [99] J. Sun, X. Yao, S. Wang, and Y. Wu, "Blockchain-based secure storage and access scheme for electronic medical records in IPFS," *IEEE Access*, vol. 8, pp. 59389–59401, 2020.
- [100] G. Tripathi, M. A. Ahad, and S. Paiva, "S2HS- A blockchain based approach for smart healthcare system," *Healthcare*, vol. 8, no. 1, Article ID 100391, 2020.
- [101] M. Usman and U. Qamar, "Secure electronic medical records storage and sharing using blockchain technology," *Procedia Computer Science*, vol. 174, pp. 321–327, 2020.
- [102] R. H. Hylock and X. Zeng, "A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study," *Journal of Medical Internet Research*, vol. 30, 2019.
- [103] M. Shen, Y. Deng, L. Zhu, X. Du, and N. Guizani, "Privacy-preserving image retrieval for medical IoT systems: a blockchain-based approach," *IEEE Network*, vol. 33, no. 5, pp. 27–33, 2019.
- [104] J. Xu, K. Xue, S. Li et al., "Healthchain: a blockchain-based privacy preserving scheme for large-scale health data," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8770–8781, 2019.
- [105] E.-Y. Daraghmi, Y.-A. Daraghmi, and S.-M. Yuan, "MedChain: a design of blockchain-based system for medical records access and permissions management," *IEEE Access*, vol. 7, pp. 164595–164613, 2019.
- [106] H. Li, L. Zhu, M. Shen, F. Gao, X. Tao, and S. Liu, "Blockchain-based data preservation system for medical data," *Journal of Medical Systems*, vol. 42, no. 8, pp. 1–13, 2018.
- [107] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustainable Cities and Society*, vol. 39, pp. 283–297, 2018.
- [108] J. Brogan, I. Baskaran, and N. Ramachandran, "Authenticating health activity data using distributed ledger technologies," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 257–266, 2018.
- [109] K. N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson, and T. Hayajneh, "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring," *Journal of Medical Systems*, vol. 42, no. 7, pp. 1–7, 2018.

- [110] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain," *Journal of Medical Systems*, vol. 42, no. 8, pp. 1–18, 2018.
- [111] A. Al Omar, M. S. Rahman, A. Basu, and S. Kiyomoto, "MediBchain: a blockchain based privacy preserving platform for healthcare data," in *Proceedings of the International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, pp. 534–543, Guangzhou, China, December 2017.
- [112] H. Xiong, C. Jin, M. Alazab et al., "On the design of blockchain-based ECDSA with fault-tolerant batch verification protocol for blockchain-enabled IoMT," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [113] W. Wang, C. Qiu, Z. Yin, G. Srivastava, T. R. Gadekallu, and F. Alsol, "Blockchain and PUF-based lightweight Authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, 2021.
- [114] X. Larrucea, M. Moffie, S. Asaf, and I. Santamaria, "Towards a GDPR compliant way to secure European cross border Healthcare Industry 4.0," *Computer Standards & Interfaces*, vol. 69, Article ID 103408, 2020.
- [115] C. Zhang, L. Zhu, C. Xu, and R. Lu, "PPDP: An efficient and privacy-preserving disease prediction scheme in cloud-based e-Healthcare system," *Future Generation Computer Systems*, vol. 79, pp. 16–25, 2018.
- [116] Y. S. Rao, "A secure and efficient ciphertext-policy attribute-based signcryption for personal health records sharing in cloud computing," *Future Generation Computer Systems*, vol. 67, pp. 133–151, 2017.
- [117] Z. Chen, W. Xu, B. Wang, and H. Yu, "A blockchain-based preserving and sharing system for medical data privacy," *Future Generation Computer Systems*, vol. 124, pp. 338–350, 2021.
- [118] W. Wang, H. Huang, L. Xue, Q. Li, R. Malekian, and Y. Zhang, "Blockchain-assisted handover authentication for intelligent telehealth in multi-server edge computing environment," *Journal of Systems Architecture*, vol. 115, Article ID 102024, 2021.
- [119] D. Ngabo, D. Wang, C. Iwendi, J. Henry Anajemba, L. Adewale Ajao, and C. Biamba, "Blockchain-Based Security Mechanism for the Medical Data at Fog Computing Architecture of Internet of Things," 2021, <https://www.mdpi.com/2079-9292/10/17/2110/htm>.
- [120] A. Al Omar, A. Kaisar Jamil, A. Khandakar, A. Razzak Uzzal, and R. Bosri, "A transparent and privacy-preserving healthcare platform with novel smart contract for smart cities," *IEEE Access*, vol. 9, pp. 90738–90749, 2021.
- [121] B. S. Egala, A. K. Pradhan, V. Badarla, and S. P. Mohanty, "Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11717–11731, 2021.
- [122] T. Robles, B. Bordel, R. Alcarria, and D. Sánchez-de-Rivera, "Enabling trustworthy personal data protection in eHealth and well-being services through privacy-by-design," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, Article ID 1550147720912110, 2020.
- [123] J. Sun, H. Xiong, X. Liu, Y. Zhang, X. Nie, and R. H. Deng, "Lightweight and privacy-aware fine-grained access control for IoT-oriented smart health," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6566–6575, 2020.
- [124] H. Liu, X. Yao, T. Yang, and H. Ning, "Cooperative privacy preservation for wearable devices in hybrid computing-based smart health," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1352–1362, 2019.
- [125] A. D. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors*, vol. 19, no. 2, Article ID 2, 2019.
- [126] T. T. Thwin and S. Vasupongayya, "Blockchain-based access control model to preserve privacy for personal health record systems," *Security and Communication Networks*, vol. 2019, Article ID e8315614, 15 pages, 2019.
- [127] I. Natgunanathan, A. Mehmood, Y. Xiang, L. Gao, and S. Yu, "Location privacy protection in smart health care system," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3055–3069, 2019.
- [128] A. A. Omar, M. Z. A. Bhuiyan, A. Basu, S. Kiyomoto, and M. S. Rahman, "Privacy-friendly platform for healthcare data in cloud based on blockchain environment," *Future Generation Computer Systems*, vol. 95, pp. 511–521, 2019.
- [129] A. Cavoukian, "The 7 Foundational Principles," 2010, <https://www.ryerson.ca/content/dam/pbdce/seven-foundational-principles>.
- [130] S. Sabu, H. M. Ramalingam, M. Vishaka, H. R. Swapna, and S. Hegde, "Implementation of a secure and privacy-aware E-Health record and IoT data sharing using blockchain," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 429–433, 2021.
- [131] F. Khan, J. H. Kim, L. Mathiassen, and R. Moore, "Data breach management: an integrated risk model," *Information & Management*, vol. 58, no. 1, Article ID 103392, 2021.

Research Article

Creative Destruction Path Selection for Industrial Park Transformation and Upgrading under the Concept of Character Town in the Era of Big Data

Yue Bai  and Xuewen Li 

School of Public Affairs, Zhejiang University, Hangzhou, 310000, China

Correspondence should be addressed to Xuewen Li; lixuewen@zju.edu.cn

Received 21 January 2022; Revised 9 February 2022; Accepted 10 February 2022; Published 11 March 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Yue Bai and Xuewen Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the reshuffle of the global industrial structure, some industrial parks in China are facing problems such as unclear industrial positioning, unclear agglomeration effect, and lack of impetus for independent innovation. It is urgent to seek a way of transformation and upgrading. The arrival of the era of big data also provides new opportunities for the transformation of China's industrial parks. In this context, the model of "character towns" proposed by Zhejiang Province provides a new concept for breaking the bottleneck of industrial park transformation and also provides a platform and space for the application of data technology and digital governance in the era of big data. Based on the theory of creative destruction and self-organization, this paper holds that the essence of industrial park transformation under the concept of character towns is the self-organization process of creative destruction and high-quality reconstruction. After sorting out the development process of parks in China, this paper divides industrial parks into production-oriented parks, consumption-oriented parks, and trade-oriented parks according to the economic activities classified by classical economics. By constructing a six-stage creative destruction model, this paper analyzes the path selection of transformation and upgrading of three types of parks under the concept of character towns and proposes that industrial structure replacement for the production-oriented parks, upgrading the consumption structure for the consumption-oriented parks, and change of trade control points for the trade-oriented parks are the possible transformation paths for the three types of parks.

1. Introduction

In recent years, China's economy has faced problems such as slow growth in demand and severe overcapacity in some industries, and the competitiveness of traditional industries has weakened. It has become the consensus of the entire society to accelerate industrial transformation and upgrading [1]. In China's economic development over the past three decades, industrial parks have flourished, effectively gathering economic development factors, creating huge economic and social benefits, and promoting the rapid development of China's industrialization and urbanization [2]. However, with the reshuffling of global industries, the price of labor, land, and other factors has risen worldwide, the cost of traditional industries has increased, and the

unsustainability in the development of industrial parks has come out. The problems of single industrial structure, lack of characteristics, lack of clustering, and insufficient capacity for independent innovation make industrial parks enter a bottleneck period of development. They need to seek a path of transformation and upgrading urgently [3]. In 2014, the concept of "character towns" proposed by Zhejiang Province was a spatial development platform located within 3.5 square kilometers, relatively independent of urban areas, with clear industrial positioning, cultural connotation, tourism, and certain community functions [4].

Character town gives a new direction to industrial upgrading and park transformation. Schumpeter's innovation theory has been put forward for a century, and there are many pieces of literature discussing this theory. This paper

intends to further explore this theory from the perspective of character towns and traditional parks. Along with the benefits of the big data era, Zhejiang Province supports all regions to promote industrial transformation and upgrading with the concept of character towns; the current development of artificial intelligence and the application of big data provide the possibility for this study. The research results of this paper provide new ideas for enriching urban governance in the era of big data. The character towns can solve the problems of economic innovation and lead to the upgrading of traditional industrial parks [5].

This paper will start from the theory of creative destruction and self-organization and classify parks according to the division of economic activities in classical economics. By analyzing the functional characteristics and existing problems of the three types of parks, a six-stage creative destruction model is constructed to analyze the creative destruction path of each type of industrial park by character towns in the era of big data and provide ideas for the transformation and upgrading of industrial parks with the concept of character towns.

2. Analysis of the Pressure of Industrial Park Transformation and Upgrading

According to the documents issued by the Ministry of Commerce, Ministry of Science and Technology, Ministry of Economics and Information, General Administration of Customs, and other ministries and commissions, the industrial parks approved by the state mainly include economic and technological development zones, high-tech industrial development zones, free trade zones, special customs supervision zones, and cross-border economic cooperation areas, tourist resorts, national independent innovation demonstration areas, and future technology cities.

2.1. The Evolution of China's Industrial Parks. Since the establishment of the Shenzhen Shekou Industrial Zone in 1979, the industrial park has undergone four steps of change [6]. The period from the beginning of reform and opening to the end of the 1980s was the start of industrial park entrepreneurship. The number of industrial parks approved by the state increased year by year, but at this time, the layout of the park mainly considered that the location and transportation conditions, the technology, and capital foundation were weak. It is dominated by labor-intensive industries, with various types of enterprises and unclear industrial position. The park can only provide physical space, and the urban function has not yet been formed in the park. From the 1980s to the end of the 1990s, China approved 32 economic and technological development zones and 53 high-tech industrial development zones. The park entered a rapid development era of capital-intensive and investment-driven, with increased production efficiency, enhanced industrial concentration, and property levels enhancement, but industrial supporting services are still in their infancy.

In the twenty-first century, the park has entered a technology-intensive era. Economic and technological development zones and high-tech development zones have gradually developed and stabilized, the industrial division of labor has been more refined, the industrial chain has begun to extend, and the park's living and commercial facilities have gradually improved. In the second decade of the twenty-first century, innovation has become an important production factor. The park has changed from technology-intensive to innovation-driven. The previous 40 years of development have exposed the problems of industrial-urban separation and single management model [7]. In the new international competitive situation, higher requirements have been put forward for the improvement of the industrial ecological chain, supporting services, and management level. In 2014, Zhejiang Province first proposed the concept of "character town." As an interior part of the park, the character town is an industrial space organization that helps parks to build a good industrial niche. Innovation is its core element [8], and its top-level design concept will satisfy the requirements of industrial-urban integration and management mode innovation through the process of park transformation [9]. In the era of big data, character towns can become the command center of regional governance using big data. Represented by Yunqi Character Town in Xihu District of Hangzhou, the establishment of "Urban Brain" in Yunqi Town provides a digital governance mode for Zhuantang Science Park and the whole city of Hangzhou. Promoting industrial transformation and upgrading through the construction of character towns is a new direction for the development of industrial parks in the new era [10].

2.2. Classification of Industrial Park Types. Marshall's book "Principles of Economics" first mentioned the concept of industrial parks. Industrial park is a concentrated platform of specialized industries in specific places [11]. The current research classifies industrial parks from different angles. According to the proportion of investment in science and technology, it is divided into industrial parks, characteristic industrial parks, comprehensive science and technology parks, and high-tech parks [12]. According to the origin of development and the status in the urban system, it can be summarized as a comprehensive park that starts with attracting foreign investment and contains various investment models. Agricultural industrialization, which developed on original towns, formed by the cluster of agricultural products, gradually matured through the aggregation of similar enterprises and in a subordinate position of the urban system [13]. No matter from which angle the park is classified, the fundamental purpose of the park is to engage in economic activities. The concentration and development of the industry are always the essence of the park. The development style and transformation methods of the parks are different as long as the major industries of the parks are different. The core of the park's transformation and upgrading is industrial transformation and upgrading. Therefore, according to the research paradigm of classical economics and neoclassical economics,

this paper divides economic activities into three categories: production activities, consumption activities, and trading activities. Accordingly, the parks are divided into product-oriented parks, consumer-oriented parks, and trade-oriented parks. Among them, product-oriented parks are divided into production parks for tangible products and production parks for intangible products based on different product types. High-tech parks, economic and technological development zones, and other parks that produce physical products are tangible product-oriented parks. Cultural and creative industrial parks, financial industrial parks, Internet information economy industrial parks, and other parks that produce ideas are intangible product-oriented parks. Consumer-oriented parks mainly refer to parks that provide consumers with goods and experience scenes, including tourism and leisure scenic spots, health and wellness areas, and urban agricultural parks. Trade-oriented park refers to a pivotal zone that provides modern logistics facilities, trade, and exhibition places benefit from transportation advantages. The major industries are bonded warehousing, logistics and distribution, and trade and exhibition, including airport economic demonstration zone, bonded port area, and logistics park.

As shown in Figure 1, the main functions of product-oriented parks are to provide physical products and virtual products, gather various resources, incubate scientific and technological achievements, transform ideas and R&D achievements, and provide high-value-added knowledge and information to society [14, 15]. The main functions of the consumer-oriented park are providing experience-based services to the customer and satisfy China's current transition demand from a production-oriented society to a consumer-oriented society [16]. The main functions of the trade-oriented park are to integrate regional logistics resources, realize the effective flow of raw material intermediate processes, final products, and related information, realize the informatization of the logistics market carrier, and optimize the regional material network [17].

2.3. Different Types of Industrial Parks Face Transformation and Upgrading Pressures and Dilemmas. At present, most production parks are in the middle and later stages of growth, with low levels of industrial development and irrational industrial structure. The park's function positioning and industry selection are facing strategic adjustments, and its driving effect on urban development is also very limited [18]. Problems such as unclear function positioning, obvious industrial homogeneity, incomplete service platform, lack of high-end talents, and incomplete industrial chain have made many industrial enterprises in the development zones with poor benefits, poor economic conditions, and unsatisfactory land use [19–23].

The main problems faced by consumer-oriented parks are unclear business models, high prices, inconvenient transportation, insufficient public facilities, and low service quality of the staff [24]. The lack of cultural connotation leads to poor local rooting; the low degree of scale effect leads to weak regional driving force; the lack of innovation leads to poor heterogeneity and complementarity of leisure products, and the value chain is difficult to extend effectively [25].

Trade-oriented parks mainly suffer from inadequate third-party logistics supply, high vacancy rates in the park, lack of overall planning, excessive government intervention, and low efficiency of land resource utilization [26]. It is urgent to transform into an efficient operating platform.

The three types of parks face different problems in their development. But the homogeneity of products, incomplete industrial chain, lack of innovation, and high-end elements are common problems faced by all types of parks. The emergence of character towns provides three creative destruction paths for the transformation and upgrading of three types of parks.

3. The Theoretical Basis for the Character Towns to Promote the Transformation and Upgrading of the Park: From Creative Destruction to Self-Organization

3.1. Creative Destruction Theory. Evolutionary thoughts began with Thorstein B Veblen. Taking Darwin's theory of evolution and Lamarck's theory of genetics as the ideological basis and taking the evolutionary laws of nature as a reference formed a core category of "genetic mechanism-variation and innovation mechanism-selection mechanism" [27]. Evolutionary economics simulates the dynamic evolution of human economics. In the modern development of evolutionary economics, Schumpeter's economic theory has become its important thought branch. In <Theory of Economic Development>, Schumpeter found the forces that led to an economic change from the internal of the economic system, namely, innovation, summarized as the "new combination" of supply implemented by entrepreneurs. Schumpeter regarded innovation as the essence of the process of economic development, emphasized the important role of nonequilibrium and qualitative change in the economic system, and proposed the concept of "creative destruction;" innovation constantly destroys the internal old economic structure and creates a new economic structure [28, 29]. "Creative destruction" is the fundamental driving force of economic growth. Entrepreneurs and technological innovation play a central role in the process of "creative destruction." On the basis of Schumpeter's theory, Nelson and Winter [30] linked innovation and organization theory to form an evolutionary growth theory in which technology, institutions, and industrial structures coexist. It is believed that the diffusion of innovation requires the cooperation of social institutions and political factors. The creative destruction will inevitably be accompanied by changes in organizational systems.

The connotation of innovation has been continuously expanded on the basis of Schumpeter's ideas, and it has been used to express many creative behaviors: technological innovation, institutional innovation, industrial innovation, cultural innovation, management innovation, and so on [31]. Character towns are embedded in the park, introducing high-end elements to creatively destroy the park industry and introducing a new institution of "government guidance, market leadership, and enterprise entities" to creatively

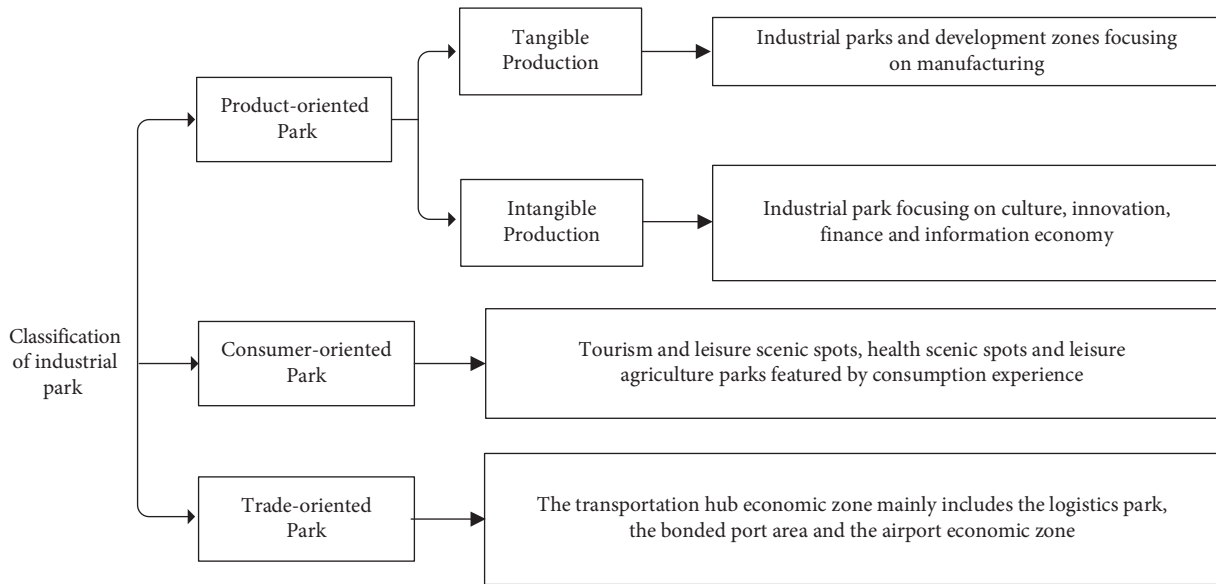


FIGURE 1: Classification of industrial park.

destroy the traditional institutional mechanisms of the park. Combining industrial innovation and institutional innovation, the character towns are positioned as a “nontown and nonzone” area. With the “3.5 square kilometers” of spatial innovation and a “scenic mode” space innovation, character towns creatively destroy the park from different aspects. Taking advantage of the character towns, the parks developed from extensive growth to refinement growth, from industrial decentralization to industrial agglomeration, and from industry-city separation to industry-city integration.

3.2. Self-Organization Theory. With Jacob and Monod revealing the order of self-organization of genetic arrangement in the gene bank, taking dissipative structure theory as a guide, self-organization theory has provided power for evolutionary economics. Self-organization theory is represented by synergetics, catastrophe theory, chaos, and fractal theory. The theory of self-organization believes that the fundamental power of the evolution of the social economic system lies in the material, energy, and information exchange between self-organization power within the system and the outside power, which is far away from the stable state. The exchange created a new structure [32]. Self-organization refers to the evolution process of a system, which is produced by the interaction of various elements in subsystems rather than external effects [33]. The fundamental power of the evolution of the social economic system lies in the self-organizing power within the system [30, 34]. Once a self-organizing system or subsystem emerges, natural selection will distinguish the adaptability of different organizations [35]. Self-replication is a form of self-organization. It produces offspring with the same structure of subsystem in the overall system so that the overall system gradually forms an orderly state and is maintained. In the initial stage of innovation diffusion, old

thinking and habits may stifle innovation in the cradle, but if the system is open and away from equilibrium, innovation will be amplified by self-replication and exceed an unstable threshold to enter a new organizational structure. After the formation of the new structure, the self-reinforcing mechanism will allow new ideas and new methods to enter the stage of rapid diffusion and finally evolve into a popular state of society, completing the routine process of evolutionary economics [36, 37].

The industrial park is a whole system, and the character town is the subsystem of the park. After the industry-city integration, smart growth, and high-quality development, character towns have become a subsystem with a self-blood production function. Its industrial model, governance model, and spatial form will continue to exchange material and energy from the outside world (park) through self-replication and eventually spread to the entire park. Through the combined action of creative destruction and self-organization, the overall system (park) transformation and upgrading will be realized. In the process of transformation and upgrading, since the self-organizing system has the ability of supercircular evolution, it will evolve to a more complicated level. Therefore, the interaction of character towns and parks will drive the industry to develop in a more sophisticated direction, form an industrial culture, and realize character towns and parks coexisting in symbiosis, as shown in Figure 2.

4. The Creative Destruction Path Design of Different Parks under the Concept of Character Towns

4.1. Creative Destruction Path of Product-Oriented Park. The main function of creating character towns in the product-oriented park is to help the park gather high-end production factors [38, 39]. The process of industrial development with

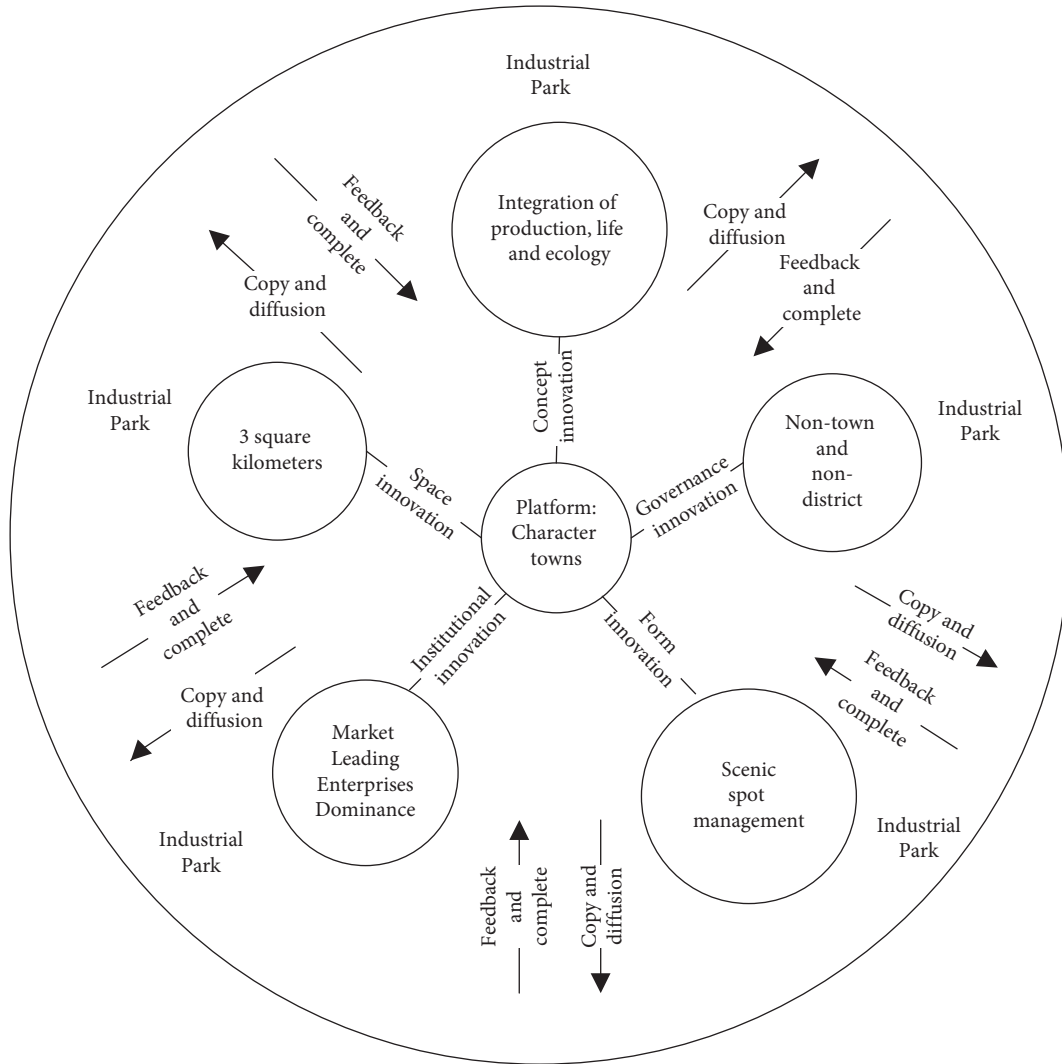


FIGURE 2: The creative destruction mechanism and self-organization evolution process of the character towns to the traditional industry park.

character towns can be divided into six stages. In the pre-aggregation stage, the character towns in the parks entered the provincial establishment list, and the first batch of R&D enterprises was attracted to settle in through preferential policies. The deepening of preferential policies has increased the number of high-end enterprises, professional enterprises have gathered in character towns, and the industrial space has begun to develop and entered the stage of primary agglomeration [40]. The spatial agglomeration of enterprises produces knowledge spillovers [41], the innovative activities brought by industrial agglomeration, knowledge spillovers, and collective actions promote collaborative development among enterprises, production efficiency improved, the value chains formed [42], and the parks entered the advanced agglomeration stage. Trade and cooperation with comparative advantages have been formed between enterprises, and industrial specialization has gradually formed, and the character town formed a clear industrial positioning, breaking through the original lack of organic correlation between enterprises within the park and the unformed industry chain. The old

industrial structure gradually disintegrated, and a new industrial structure was formed inside the character town, entering the stage of primary destruction. At this time, the destruction only happened to the interior of the town. The agglomeration of enterprises has reduced transaction costs. With the strengthening of vertical industrial linkages, horizontal industrial linkages have also gradually expanded. Enterprise service supporting departments such as financial institutions, accounting institutions, and legal institutions have also been attracted to character towns. High-end talents brought by high-end industries have put forward new requirements for the construction of township living facilities. Township infrastructure is gradually improved. On the basis of industrial specialization, a town brand of “one town, one product” is formed [43]. The completeness of the industrial chain has formed an industrial spillover effect, accelerating the connection between the character town and the external overall park, the internal and external networking and innovation of the town has formed, and the town model began to spread into the park and entered the stage of advanced

destruction. As the character town interacts with the park where it is located, the town model is constantly being copied, and the entire park eventually achieves a “dragon-for-bird” exchange. Traditional manufacturing enterprises with high pollution, high energy consumption, and low output are gradually replaced by high-end enterprises with R&D, innovation, and high added value [44, 45]. A complete industrial chain and value chain from R&D to consumer experience are formed in the park. So far, the park has changed from a traditional industrial park with low added value to a new industrial park with high added value, and a character town has become an important tool for the establishment and expansion of the new pattern of the park [40].

The upgrading process of product-oriented parks and the growth process of character towns are mutually reinforcing. The overall improvement of the park as a result of the construction of character towns will deepen the brand effect of them. When agglomeration develops to a certain extent, the business invitation will become an important form. Relaxation or cancellation of preferential policies will still attract enterprises to enter the town. The rooting and expansion of the park have promoted the interaction between the character town and the area, the transaction cost has been further reduced, and the collaborative innovation capability has been further improved. The division of labor in the town is more detailed, internal R&D is deeper, and new industries may be born. The transformation and upgrading of the park in turn also promote the growth of character towns. The upgrading process of product-oriented parks is illustrated in Figure 3.

For example, in this kind of park, the character towns with the health industry as the leading industry are a kind of product-oriented town, which are divided into characteristic towns providing medical health services and nonmedical health services. Among them, most of the characteristic towns providing medical health services build smart hospitals and apply unmanned treatment technology to provide remote diagnosis services for patients. Through big data technology, it can save time and money of communication between doctors and patients. At the same time, it can protect the privacy of patients as far as possible and can also put forward a new method to solve the contradiction between doctors and patients. In addition, many Chinese and foreign joint medical research teams and pharmaceutical manufacturing enterprises have been introduced into these character towns, carry out research on local cases and strengthen R&D capacity, try hard to achieve transformation of R&D in the biomedical field, and promote the transformation and upgrading of the local health industry.

4.2. The Creative Destruction Path of Consumer Parks. Consumer-oriented parks are represented by tourist attractions. The six-stage creative destruction model was first proposed by Mitchell based on the development of the Canadian heritage community St. Jacobs. Mitchell used St. Jacobs to explain the commercialization of historical villages and the evolution of tourism development [46, 47].

Character towns use the concept of global tourism and create an innovative integration model of “industry + tourism” to enhance traditional professional towns and traditional scenic spots [48]. At the birth of character towns, consumer-oriented parks were still dominated by agriculture or industrial production, the consumption structure was relatively simple, commercial investment was not initiated, and the landscape still retained the naturalness of the production-oriented rural landscape and was in the preintegration stage. With the creation of character towns, consumer parks have made full use of the preferential policies such as tax rebates and rent exemptions provided by local governments to attract businesses that focus on consumer experience, such as recreation, tourism, and leisure. Entrepreneurial investment in rural character commodities has gradually begun to destroy the original rural living conditions and stimulate tourism demand. Venture investment, consumption, and commoditized heritage interacted and entered the stage of primary industry integration. Commercial investment continues to increase, local heritage seekers and the real experienter of local heritage increase, and scenic spots are initially formed. The government is the leading developer at this stage. At this time, the similarity of local resources is still large. The support of local residents in the park has further attracted external tourists, private commercial investment has become more active, forming a “sightseeing + commercial shopping” model, and consumer products have gradually diversified. The public development policy continued and entered the advanced commercialization stage of commercial equilibrium [49]. At this stage, the natural landscape was deeply restored, and the post-production heritage landscape was formed. Local residents gradually realized the negative impact of commercialization on the original lifestyle. As the character towns’ landscape functions and consumer products diversified, investment companies began to replace government departments to play a leading role in the town. The stage of primary destruction begins with the retreat of the original residents in the park, and the town is shifted from government-led to enterprise-led. According to the demand for product quality and the space environment by the newly entered high-end consumer groups, the original scenic spot is driven by the market to form innovation and space innovation. Character towns have begun to change the consumption atmosphere of traditional parks, formed a stable tourism atmosphere through governance innovation, and entered the stage of advanced destruction. Subsequently, the policy changes, special development policies, and actions of consumer-oriented parks replaced public development policies and actions and started a new round of consumer product innovation. The entire park tourism industry structure was upgraded, and single consumption replaced experience consumption. The consumption structure was upgraded and entered the advanced destruction stage. The park has been completed, transformed, and upgraded [50, 51].

The character towns with the health industry as a major industry also contain the towns with nonmedical health services mainly focusing on healthcare. It is a kind of consumer-oriented town in parks. Most of them are located

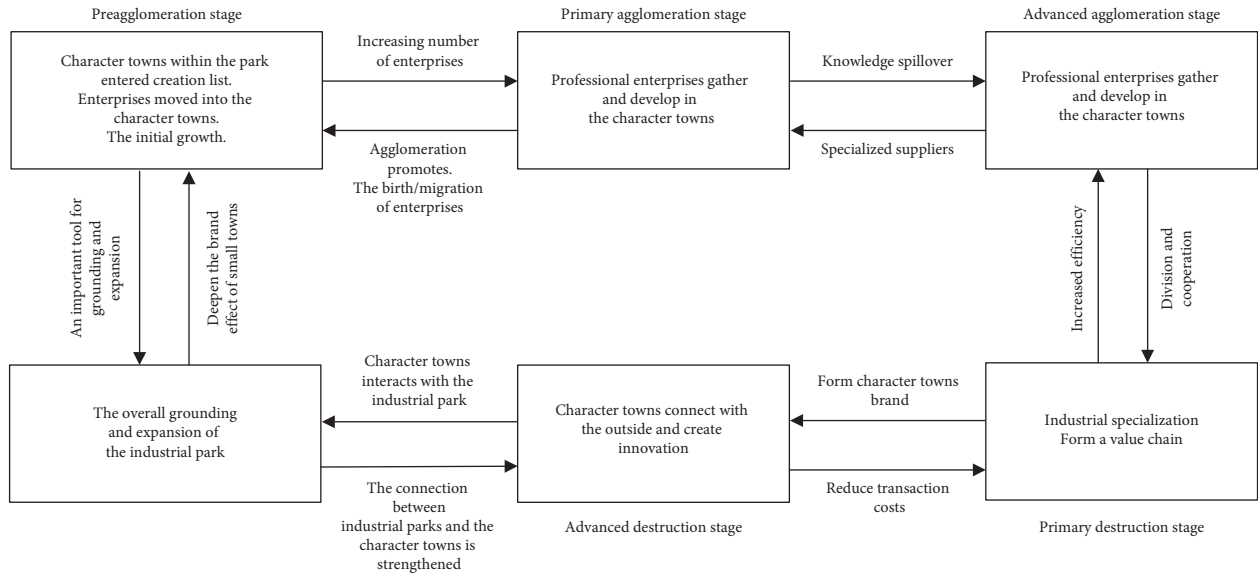


FIGURE 3: The creative destruction path of product-oriented park under the concept of character towns.

in places with beautiful environment and clean air. Health and leisure places such as nursing homes and health parks are built in these character towns. These character towns often hold sports events such as aviation and running, relying on the advantages of local mountains and water resources. For citizens living in cities, who are crouching in a tiny cubicle in office all day, these character towns provide a new type of consumption for health, a new place for family holiday leisure time, a new place for physical exercise, and a place for civilian populations. The creative destruction process of consumer-oriented parks is illustrated in Figure 4.

4.3. The Creative Destruction Path of Trade-Oriented Parks. In the traditional trade stage, the market scope is limited. According to Adam Smith’s theory of labor division, the division of labor is restricted by the scope of the market. Where consumption power is insufficient, the degree of industrial specialization is inadequate, and the product supply is insufficient, which restricts the division of labor and reduces the exchange capacity. At this time, the economic, transportation, and technological levels of the park are insufficient. The hinterland of the logistics node is limited to the surrounding area. The density of the logistics network is low, and the spatial structure is simple. The emergence of the Internet has broken through the old market. The reconstruction of character towns based on the Internet has diversified customer needs. The emergence of the sharing economy has accelerated the change of consumer behavior, forming a model of common use and common payment by consumers. The transition of the market foundation has shifted from the buyer’s personal confrontation with the seller’s group to the confrontation between the buyer’s group and the seller’s group. The relative positions of the supply and demand sides have changed. The transition of the market foundation has promoted the transformation of traditional postal transportation methods

to new logistics methods, the existing logistics nodes expanded [52], and the formation of local logistics centers entered the initial stage of open Internet. As a new form of trade, cross-border e-commerce has transformed the industry of trade-oriented parks from manufacturing demand-oriented to service demand-oriented. With the emergence of new trading channels such as O2O, both supply and demand parties can interact without physical channels. Traditional distribution channels are gradually withdrawn. The innovative combination of products and services in the town has crossed the bottom of the “smile curve” to form “hardware + software + services” mode. The Internet uses big data to mine consumers’ derivative needs in the process of using products, and the industrial foundation realizes the transition of intelligence and digitalization. The character towns have built a comprehensive logistics center and adopted an integrated supply chain management model [53, 54]. As an industrial policy, the preferential policies of character towns will further expand the industrial base. The hinterland of the logistics node broke through the local area and expanded to the outside area. The logistics node has developed from a single, time-based service to a diversified and full-range service and entered the “Internet +” advanced opening stage [55]. “Internet+” combines offline physical trade and online virtual trade to understand customers’ needs and emotions in specific situations through virtual scenes [56]. Online virtual scenes and offline physical products collaborate with each other, breaking the traditional distribution model and product presentation model. The Internet scene is materialized in life so that customers have psychological belonging needs for products and form a stable customer group. Market segmentation is based on a stable customer base, shaping the fixed relationship between products and customers. A comprehensive model of “scene + product” where manufacturers and consumers create value together is formed [57]. The logistics nodes of the old model are gradually shrinking and are being replaced by new

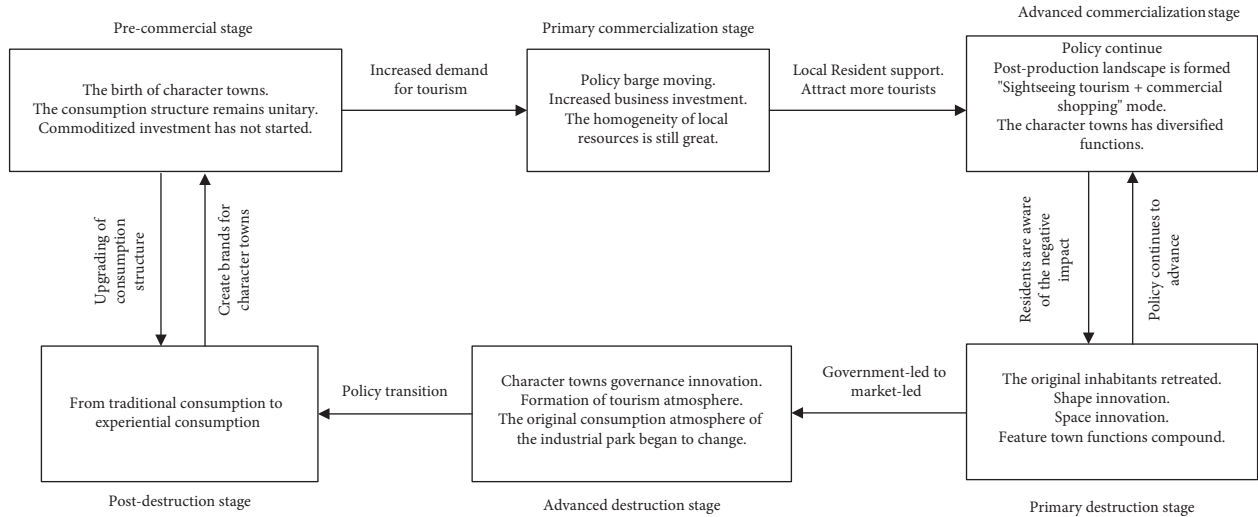


FIGURE 4: The creative destruction process of consumer-oriented park under the concept of character towns.

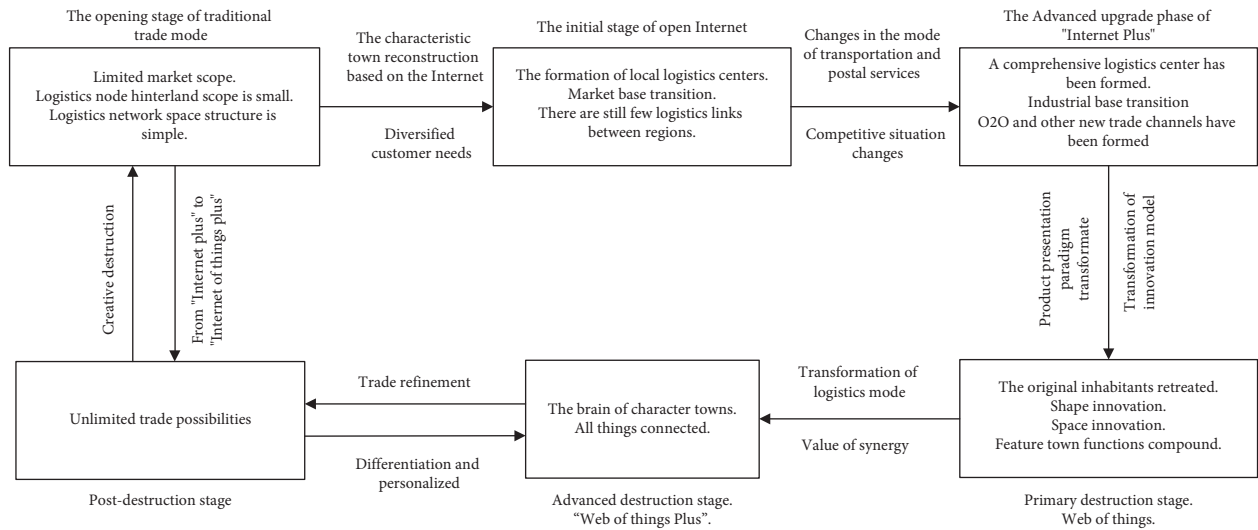


FIGURE 5: The creative destruction process of trade-oriented park under the concept of character towns.

TABLE 1: Summary of creative destruction path of the industrial park under the concept of character towns.

| Type of industrial park | Transformation and upgrading path | Critical node of the path | Primary destruction stage | Advanced destruction stage | Self-organizing evolution/replication stage |
|-------------------------|-------------------------------------|--|---|---|---|
| Product-oriented park | Replacement of industrial structure | Formation of industrial specialization | The division of labor and cooperation among enterprises | Connect to the outside of the character towns | Form a new industry industrial specialization |
| Consumer-oriented park | Upgrading of consumption structure | Upgrade of commercial model | Spatial form innovation | Change in consumption patterns | Experiential consumption park |
| Trade-oriented park | Trade control points changed | Replacement of logistics mode | Product plus scene | The brain of character towns | Unlimited trade possibilities |

logistics nodes. “Internet +” has realized the decentralization of industrial institutions and digitalization of enterprises, which has spawned the Internet of Things in social life, strengthened the connection between the industrial chain,

and promoted the transformation of warehousing, logistics, and other industries. At this time, the character towns became the control point of the Internet of Things in the trade park and entered the stage of primary destruction to the

park. With the transformation of the logistics model, character towns have become the “town brain” of the trade-oriented park. With the “Internet of Things +,” the “town brain” can connect people, finances, machines, and things in the park at any time and any place. In the park, all kinds of materials and information are connected to each other through the character towns and enter the advanced destruction stage. The differentiated and personalized service enters the postdestruction stage, which makes unlimited trade possible, and the trade-oriented park realizes transformation and upgrading.

The success of the transformation and upgrading of the trade-oriented parks will expand the originally limited market range. With the refined development of trade, it will further activate the “town brain” and enable the refined development of the Internet of Things dominated by it. Use the “town brain” to control living facilities, embed life and ecological elements in the trade-oriented park, and change the trade mode and the lifestyle of it. The creative destruction process of trade-oriented parks is illustrated in Figure 5.

Table 1 summarizes the creative destruction path of different industrial parks under the concept of character towns, and it shows the critical path node of different parks and three steps of destruction stages for each kind of park.

5. Discussion

According to the findings of this paper, different types of parks have different industrial transformation processes. The main path of production parks is industrial structure replacement, and the key node lies in the formation of industrial specialization. The main path of the transformation of consumer parks is the upgrading of consumption structure, and the key node is the upgrading of commercial mode. The key to the transformation of trade parks lies in the change of trade control points and the replacement of logistics mode. Different types of economic activities have different innovation ways, and the transmission path of these innovation ways, namely, the path of creative destruction, is also different, requiring different costs and the support of big data technology and methods.

6. Conclusion

Characteristic towns provide new possibilities for solving the bottlenecks in the transformation and upgrading of traditional industrial parks with the concepts of morphological innovation, spatial innovation, institutional innovation, and governance innovation. The transformation of the park promoted in the platform of a character town is essentially a process of industrial creative destruction and spatial self-organization under the guidance of regional policies. The transformation and upgrading of different types of traditional parks are inseparable from the replacement of the dominant industrial structure and the change of the form and function. However, due to the complexity of the industrial cluster evolution process, the transformation and

upgrade paths and key nodes of different industrial types of parks are different. The parks face different major contradictions and require a different solution. Based on the classification of industrial parks, under the guidance of creative destruction and self-organization theory, this paper has initially formed a theoretical framework for the transformation and upgrading of traditional parks and discussed the transformation path design of product-oriented parks, consumer-oriented parks, and trade-oriented parks with character towns. This paper only discusses the possible path of transformation and upgrading of the park under the concept of character towns from a theoretical level.

The outcomes of this paper provide a theoretical framework for the study of industrial transformation and upgrading of development zones from the new perspective of character towns and provide a new perspective for the study of urban citizens’ health platform. It is also the practice for Schumpeter’s “creative destruction” theory from China. It is worth further studying the following topics: how about the spillover effect of the character towns in the park, whether it has followed the laws of theory in practice, what innovative measures did the successful character towns have at each stage of the park transformation, what creative practices can be copied and promoted, and what is the problem of the failed character towns. In the next step of our work, the story of this paper can be verified by the actual cases of characteristic towns. In the future, the actual data can also be used to verify the occurrence process of creative destruction in each type of park by using the dimensionality reduction techniques of big data to obtain the data principal component supporting the further research.

Data Availability

The data used to support the findings of this study are from the official website of the Ministry of Commerce, PRC, <http://www.mofcom.gov.cn/xglj/kaifaqu.shtml>, the official website of the Ministry of Science and Technology, PRC, <http://www.most.gov.cn/>, the General Administration of Customs of the People’s Republic of China’s official website, <http://www.customs.gov.cn/>, the official website of Ministry of Culture and Tourism, PRC, <https://www.mct.gov.cn/>, and the official website of the Ministry of Industry and Information Technology, PRC, <http://www.miit.gov.cn/>.

Conflicts of Interest

The authors declare that there are no conflicts of interest with respect to the research, authorship, and/or publication of this paper.

Acknowledgments

The authors acknowledge the Center for Urban Development and Land Policy, Peking University-Lincoln Institute, for this thesis scholarship.




References

- [1] W. Li and Y. Wang, "Risk prevention and response in economic transition," *Development Research Center of the State Council: Three Major Risks of China's Industrial Transformation and Upgrading*, China Economic Report, vol. 3, 2018.
- [2] C. Zhang, "Research on the Transformation of "Industrial Parks to Character Towns" under New Urbanization -- A Case Study of Ningbo CRRC Industrial Base in Zhejiang Province, Rational Planning for Sustainable Development," in *Proceedings of the 2017 Annual Meeting of Chinese Urban Planning (19 Small Town Planning)*, Dongguan, China, 2017.
- [3] Y. Zhao and Z. Wenxia, *Cluster or Accumulation: Reflections on the Construction of Local Industrial Parks*, China's Industrial Economy, no. 1, Shanghai, China, 2008.
- [4] B. Ma, "Character towns: a grand strategy for economic transformation and upgrading in Zhejiang province," *Zhejiang Social Sciences*, vol. 3, 2016.
- [5] J. Lan, "Accelerating the transformation and upgrading of industrial character towns," *Zhejiang Economy*, vol. 19, 2015.
- [6] R. Zhang, "Eight Characteristics of the Fourth-generation Industrial Park," *Urban Development*, vol. 6, 2019.
- [7] W. Li and H. Chen, "Connotation analysis and planning suggestions for the integration of industry and city," *Urban Planning Journal*, vol. S1, 2012.
- [8] S. Sheng, "Zhang weiming: character towns: a form of industrial spatial organization," *Zhejiang Social Sciences*, vol. 3, 2016.
- [9] J. Weng, "High-quality Promotion of Character Town Construction," *Zhejiang Economy*, vol. 8, 2016.
- [10] W. Zhang and Y. Ma, "Innovation of Governance Mechanism in Character Towns under the Guidance of Social Governance," *Journal of Zhejiang Provincial Party School*, vol. 5, 2018.
- [11] A. Marshall, "Principles of economics: an introductory volume," *Social Science Electronic Publishing*, vol. 67, no. 1742, p. 457, 1920.
- [12] L. Qi, "Research on Investment and Financing Mode Selection of Industrial Park Infrastructure," Master's Thesis, Lanzhou University, Lanzhou, China, 2016.
- [13] Y. Tian and G. Qiao, "A study on the construction of industrial Parks in small towns," *Economic Geography*, vol. 2, 2001.
- [14] X. Zhang, Li Cai, and B. Ge, "Basic function and interaction of high-tech industrial development zone," *Science Research*, vol. 2, 1998.
- [15] J. Wu and J. Huang, "An analysis of polysemous site planning in creative industrial parks," *The Planner*, vol. 6, 2008.
- [16] Y. Wang, G. Li, and Q. Zhao, "A study on the spatial transformation of suzhou industrial park: from the perspective of the evolution from a "production-oriented society" to a "consumption-oriented society," *The Planner*, vol. 2, 2015.
- [17] Y. Dong, "A Study on the Development Strategy of Logistics Service in Yangshan Bonded Port Area," Master's thesis, Shanghai Jiaotong University, Shanghai, China, 2008.
- [18] Z. Feng, C. Gu, and Y. Zong, "Functional Orientation of Urban Development Zones under the Revitalization Strategy of Old Industrial Bases: A Case Study of Huludao Economic and Technological Development Zone," *Human Geography*, vol. 5, 2005.
- [19] J.-C. Wang, P.-F. Li, and P. Chen, "Industrial clusters in China from the perspective of geographical shift of manufacturing activities," *Regional research and development*, vol. 5, 2007.
- [20] M. Zhu, "Problems and Strategies facing the development of High-tech industrial parks in China," *Science and Technology Management Research*, vol. 10, 2008.
- [21] J. Wang, "Observation and reflection on the phenomenon of industrial parks in China," *The Planner*, vol. 9, 2011.
- [22] F. Zhangxian, S. Wang, and Y. Zhang, "Functional Transformation and Structural Optimization of Development Zones under the Background of Central City Polarization," *Urban Development Research*, vol. 1, 2010.
- [23] S. Zhang, "Development Status and Existing Problems of China's Cultural and Creative Industrial Parks," *Journal of Hohai University*, vol. 2, 2011.
- [24] H. Bu, W. Min, and N. Lin, "Research on service quality of tourist scenic spots based on servqual model -- A case study of gulangyu island," *Agricultural Resources and Zoning in China*, vol. 39, no. 9, 2018.
- [25] Y. Li and G. Chen, "Reflections on the brand building of leisure agriculture in changzhutan area," *Jiangsu Business Theory*, no. 11, 2011.
- [26] C. Xie and Z. Zhanyi, "problems and Solutions of Logistics Park construction in China," *Science and Technology Information*, vol. 24, 2008.
- [27] T. Veblen, "Why is economics not an evolutionary science?" *Quarterly Journal of Economics*, vol. 12, no. 4, pp. 373–397, 1898.
- [28] J. Schumpeter, *Capitalism, Socialism, and Democracy*, The Commercial Press, Beijing, China, 1979.
- [29] J. Schumpeter, *Theories of Economic Development*, Commercial Press, Beijing, China, 1991.
- [30] R. R. Nelson and S. G. Winter, *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA, USA, 1982.
- [31] M. Dai and Y. J. Yin, "dyschel: innovation theory:1912-2012 -- commemorating the 100th anniversary of schumpeter's first edition of economic development theory," *Economic Dynamics*, vol. 4, 2012.
- [32] J. Howland, *Hidden Order: Adaptive Complexity*, Shanghai Century Publishing Group, Shanghai, China, 2011.
- [33] X. Shen and T. Wu, *Philosophy of Self-Organization: A New View of Nature and Science*, Party School press, Beijing, China, 1993.
- [34] K. Doffey, *Evolutionary Economics: Guidelines and Scope*, Higher Education Press, Beijing, China, 2004.
- [35] S. A. Kauffman, "Final theory in biology. (Book reviews: the origins of order. Self-organization and selection in evolution.)," *Science*, vol. 260, no. 5113, pp. 1531–1533, 1993.
- [36] H. Shen, *Self-Organization Theory of Economic System: Modern Science and Economics Methodology*, China Social Sciences Press, Beijing, China, 1991.
- [37] U. Witt, "Evolutionary Economics: Some Principles," *Evolution in Markets and Institutions*, Physica-Verlag HD, Freiburg, Germany, 1993.
- [38] J. Yi, W. Meng, and X. Yang, "A Phased Study on the Evolution of Regional Innovation Networks," *Scientific research Management*, vol. 5, 2005.
- [39] J. Yi, "Research on the Influence of Regional Innovation Network and Generic Technology R&d on Industrial Innovation Ability," doctoral dissertation, Chongqing University, Chongqing, China, 2007.
- [40] G. B. Richardson, "The organisation of industry," *Economic Journal*, Royal Economic Society Royal Economic Society, vol. 82, no. 327, pp. 883–896, 1972.
- [41] J. Wang, "innovation of Industrial Cluster," *China industry and information Technology*, vol. 8, 2019.

- [42] G. B. Richardson, "The organization of industry," *Economic Journal*, vol. 82, no. 327, pp. 883–896, 1972.
- [43] B. Qiu, "The breadth and depth of character towns," *Architectural Design Management*, vol. 3, no. 5, 2017.
- [44] J. Zhou, "A Comparative Study on the Dynamic Mechanism and Path of Industrial Cluster Transformation and Upgrading in China's Coastal Developed Regions," doctoral dissertation, Jinan University, Jinan, China, 2016.
- [45] S. Wang and J. Li, "Character Towns Boost Industrial Cluster Transformation and Upgrading in Zhejiang Province -- A Case Study of Tonglu Pen Industry," *China Economic and Trade Guide*, no. 11, 2017.
- [46] C. J. A. Mitchell, "Entrepreneurialism, commodification and creative destruction: a model of post-modern community development," *Journal of Rural Studies*, vol. 14, no. 3, pp. 273–286, 1998.
- [47] C. J. A. Mitchell and S. B. D. Waal, "Revisiting the model of creative destruction: st. Jacobs, Ontario, a decade later," *Journal of Rural Studies*, vol. 25, no. 1, pp. 0–167, 2009.
- [48] J. Zhong, "Discussion on the Integrated Development of Character Towns and Regional Tourism," *Open Herald*, vol. 2, 2017.
- [49] L. Jiang and Q. Su, "Creative destruction and transformation of local identity in Zhouzhuang ancient town," *Geographical Journal*, vol. 8, 2013.
- [50] M. Lin and J. Bao, "Tourism commercialization in China's historic villages and towns -- application test of creative destruction model," *Tourismus Journal*, vol. 4, 2015.
- [51] J. Zhang, "A Study on the Commercialization of Traditional Village Space from the Perspective of Creative Destruction," *Comparative Analysis of Likeng Village*, Jiangxi Province, Southern Architecture, vol. 4, no. 1, Wangkou Village and Jiangwan Village in Wuyuan, 2017.
- [52] D. Zhang, *A Study on the Evolution Mechanism and Layout Optimization of Logistics Park*, Ph.D. Dissertation, Central South University, Changsha, China, 2006.
- [53] D. K. Fleming and Y. Hayuth, "Spatial characteristics of transportation hubs: centrality and intermediacy," *Journal of Transport Geography*, vol. 2, no. 1, pp. 3–18, 1994.
- [54] K. O'Connor, "Airport development in southeast asia," *Journal of Transport Geography*, vol. 3, no. 4, pp. 269–279, 1995.
- [55] H. Li, "A Hub-And-Spoke Model of Anhui Logistics Regional System," Master thesis, Anhui Normal University, Wuhu, China, 2004.
- [56] H. Lu, "Evolution Mechanism and Planning of Regional Logistics Hubs," Ph.D. Dissertation, Beijing Jiaotong University, Beijing, China, 2015.
- [57] Z. Zhao, "Internet +" Cross-border Management: A Perspective of Creative Destruction," *China's Industrial Economy*, vol. 10, 2015.

Research Article

Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis

V. Durga Prasad Jasti ¹, Abu Sarwar Zamani ², K. Arumugam ³, Mohd Naved ⁴,
Harikumar Pallathadka ⁵, F. Sammy ⁶, Abhishek Raghuvanshi ⁷,
and Karthikeyan Kaliyaperumal ⁸

¹Department of Computer Science and Engineering, VR Siddhartha Engineering College, Vijayawada, India

²Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

³Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

⁴Amity International Business School (AIBS), Amity University, Noida, UP, India

⁵Manipur International University, Imphal, Manipur, India

⁶Department of Information Technology, Dambi Dollo University, Dambi Dollo, Ethiopia

⁷Mahakal Institute of Technology, Ujjain, India

⁸IT @ IoT - HH campus, Ambo University, Ambo, Ethiopia

Correspondence should be addressed to Abhishek Raghuvanshi; abhishek14482@gmail.com and Karthikeyan Kaliyaperumal; karthikeyan@ambou.edu.et

Received 18 January 2022; Revised 9 February 2022; Accepted 10 February 2022; Published 9 March 2022

Academic Editor: G. Thippa Reddy

Copyright © 2022 V. Durga Prasad Jasti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Breast cancer is the most lethal type of cancer for all women worldwide. At the moment, there are no effective techniques for preventing or curing breast cancer, as the source of the disease is unclear. Early diagnosis is a highly successful means of detecting and managing breast cancer, and early identification may result in a greater likelihood of complete recovery. Mammography is the most effective method of detecting breast cancer early. Additionally, this instrument enables the detection of additional illnesses and may provide information about the nature of cancer, such as benign, malignant, or normal. This article discusses an evolutionary approach for classifying and detecting breast cancer that is based on machine learning and image processing. This model combines image preprocessing, feature extraction, feature selection, and machine learning techniques to aid in the classification and identification of skin diseases. To enhance the image's quality, a geometric mean filter is used. AlexNet is used for extracting features. Feature selection is performed using the relief algorithm. For disease categorization and detection, the model makes use of the machine learning techniques such as least square support vector machine, KNN, random forest, and Naïve Bayes. The experimental investigation makes use of MIAS data collection. This proposed technology is advantageous for accurately identifying breast cancer disease using image analysis.

1. Introduction

Any region of the body might be affected by cancerous cell development. Normal cells become crowded out as the cancerous growth spreads throughout the body, making it difficult for the body to operate properly [1]. Cancer is not a single disease but rather a collection of diseases. There are many different types of cancerous growths. Malignant

growth can occur in any internal organ, as well as in the blood cells, and it is not limited to one area. When it comes to the development and spread of malignant growths, there is a distinct difference between them.

Tumors and lumps form as a result of cancer growth. Some anomalies, however, may not be harmful. To determine whether a tumor or lump is cancerous, a little sample is removed by the doctor and examined under a microscope. If

it is not cancerous, it is referred to as a benign tumor. Malignancies other than tumors, such as leukemia (a blood illness), can occur in platelets or other cells of the body [2].

The aberrant development of cells is the beginning of breast malignant growth. Recurring knots or x-beams of these cells are typical features of tumors. Tumors can be harmful (diseases) if the cells within them grow into new, more potent substances or spread throughout the body. Various parts of the breast can get infected with breast cancer. The majority of breast cancers begin in the milk ducts before spreading to the areola. Some have their origins in the organs responsible for producing milk. Breast cancer can occur in a variety of ways, and some are more common than others. There are a few breast cancers that originate in distinct tissues. Even though there are many types of breast cancers, the condition results in a protrusion of the breasts. Mammogram screening can detect a wide range of breast illnesses. This aids in resolving issues before they become a problem [3].

An X-ray image of the breast is captured by mammography. Computerized mammography has eliminated the need for repeated mammograms in breast screening methods. Computer-based PC programs that warn radiologists to optimum variations in mammography and allow integrated film misleading are potential points of interest for DM [4].

Malignant breast and lung development, such as colon polyps, can be categorized using the computer-aided design framework. However, in the event that the master human audience is unavailable, these modalities can still be utilized. A compiled assessment of patient drawings was used by computer-aided design frameworks to show radiologist territories that seemed like anomalies. We should keep in mind that CAD can run differently depending on the settings, so any necessary adjustments are made to get the most accurate results [5].

Most of the CAD architecture is designed to help radiologists improve their accuracy and proficiency. They can be drawn by some highlights and overlook an injury; this can happen when they are trying to isolate prescription infections. Because of the wide range of opinions among mammograms, it is not uncommon for one of them to split an equal case in order to arrive at a different conclusion. Restorative networks must be supportive of outcomes [6], but unlike radiologists, CAD frameworks process images faster without reducing accuracy.

This article describes a machine learning and image processing-based evolutionary approach for detecting and classifying breast cancer. This model uses image pre-processing, image enhancement, segmentation, and machine learning algorithms to categorize and detect breast cancer.

2. Literature Survey

It is important to keep track of cancer statistics since it requires long-term planning, possible learning, and constant observation of every cancer patient [7]. The study's goal is to show how data mining processes can be used to improve the

statistical analysis results from cancer log data. The data were gathered from the Greece Cancer Log between the years 1998 and 2004. Data mining techniques and algorithms were used to process and train the data prior to examination. Using data mining techniques, the authors have developed a method for preparing and examining existing cancer data, which they hope this method may prove useful to other researchers in the future.

On the basis of the quantitative examination of bilateral mammographic picture element distinctions within the series of negative full-field digital mammography pictures, Tan et al. [8] developed and tested a novel computational model for forecasting near future breast cancer risk. A collection of 335 women's digital mammograms from four separate time periods comprised the historical dataset. With the help of leave-one-instance-out-based cross validation, three support vector machine risk replicas have been constructed and weathered. The results show a mammographic feature distinction-based risk model and a rising style of the near-period risk for the mammogram-identified breast cancer.

This study [9] uses evolutionary optimization techniques to extract two well-known datasets under machine learning by applying four different optimization methods. The Iris and Breast Cancer datasets were used to examine the suggested optimization strategies. The neural network is used with four optimization processes, such as the dragonfly, grey wolf, whale, and multiversity optimization, in this article's classification issue. In order to arrive at a precise conclusion, a number of control metrics were taken into consideration. Gray wolf and multiversity offer exact results over the other two methods in terms of convergence, runtime, and classification rate, according to the proportionate study.

Radiologists are increasingly using and requesting deep learning since it aids them in developing a precise diagnosis and enhances the accuracy of their predictions, according to Kaur et al. [10]. The Mini-MIAS dataset, which contains 322 images, is used in this article to demonstrate a new feature mining technique that uses K-mean clustering in order to select speed-up resilient features. In terms of a deep neural network and a multiclass support vector machine, a new segment is added to the categorization level that contributes 70% to training and 30% to testing. An autonomous decision-making method is based on K-mean clustering, and a multiclass support vector machine achieves higher precision rates than doing so manually.

A systematic review conducted [11] examines the progress made in computer-aided breast cancer diagnosis from the study's inception. Using a wide range of technical databases as a reference, the systematic review was able to be used for a wide range of papers in the field. Nevertheless, the scope of this article was limited to academic and scholarly publications, with no consideration given to commercial considerations. The results of this survey provide an overview of the current state of computer-aided diagnosis systems in relation to the picture modalities that are being used and the classifiers based on machine learning that are being used.

According to Mohant y, S.S and Mohant y, P.K. [12], breast cancer is the second most common form of cancer

worldwide, after lung cancer. A total of approximately 1.9 billion persons were overweight in 2016, with over 650 million of those individuals being obese. Therefore, it is clear that obesity and breast cancer risks are strongly linked. When it comes to oestrogen production through body fats, the author of this review explains that it is a peripheral area for oestrogen biosynthesis and oestrogen disclosure affecting body fat circulation.

Breast cancer detection accuracy and reduced diagnostic variation are critical, according to Wang et al. [13]. WAUCE (weighted area under the receiver operating characteristic curve ensemble) is a new model introduced by the authors, and its performance is compared with that of previous models utilizing datasets from Wisconsin diagnostic centers. When compared to existing ensemble models, the suggested approach achieves greater accuracy and a significantly lower variance in the detection of breast cancer. As an improved and more dependable alternative, the authors suggest that this methodology be used to diagnose other disorders.

According to Yang et al. [14], the deep learning-based classification of breast tissues from histology images has low accuracy because of the lack of training data and a lack of knowledge about structural and textual data that can span many layers. The ensemble of the multiscale convolutional neural network (EMCN) approach presented in this paper is used to categorize haematoxylin-eosin-stained breast microscope images into four categories, namely, benign lesion, normal tissue, invasive, and in-situ carcinoma. Prior to training the pretrained models, such as ResNet-152, DenseNet-161, and ResNet-101, each image is translated to many scales. The collected training bits are then used and upgraded during each scale. The EMS-net approach has better accuracy than the other three algorithms tested.

According to Xu et al. [15], the manual segmentation of ultrasound images takes a significant amount of time, making the automatic segmentation of ultrasound images essential. It has been suggested that images of breast ultrasound be divided into four categories: mass, skin, fat, and fibroglandular tissue by using convolution neural networks, as the authors have done. It appears from the quantitative measures and the Jaccard similarity index that the suggested strategy outperforms the alternatives by an impressive 80 percent. To help with the medical analysis of breast cancer and improve imaging for various types of medical ultrasound, the proposed technique may provide the segmentations needed.

The work [16] focused on a variety of ensemble approaches extensively employed in the field of bioinformatics for performing prediction tasks. Ensemble classification techniques for breast tumors are examined in terms of nine features, including publication domains, medical activities and research categories agreed upon various ensembles recommended, the sole methodologies used to build the ensembles and the validation structure adopted to examine these ensembles, the tools used to construct the ensembles, and optimizing. IEEE Explore, Scopus, ACM, and PubMed databases each had a total of 193 items published after the year 2000. Among the six medical jobs available, the

diagnosis remedial job appears to be the most frequently explored one, followed by the experimental-focused empirical form and evaluation-based research procedures. The use of ensemble approaches in the treatment of breast cancer is thoroughly examined in this article. For this reason, specialists in breast cancer research have provided suggestions in the form of a summary of findings.

Within the datasets, Kakti et al. [17] demonstrated greater precision in diagnosing breast cancer scenarios. Based on supervised learning in the hunt and decision tree algorithms, the MMDBM (mixed mode database miner) algorithm has been suggested. Based on empirical learning and comparison analysis, the proposed technique's output is more precise. Adding other datasets and attributes, as proposed by the author, could yield even better results.

For breast cancer detection, Ting et al. [18] recommended an algorithm called convolutional neural network improvement for breast cancer classification. It uses the convolutional neural network to improve the classification of breast cancer lacerations to help the professionals in the diagnosis of the disease. Classifying medical imaging as benign, malignant, or healthy patients is possible using the CCNI-BCC algorithm.

According to Chaudhury et al. [19], early detection and classification of breast cancer can help patients get the treatment they need. Using the notion of transfer learning, the authors have proposed a new deep learning framework for the diagnosis and categorization of breast cancer using breast cytology images. In contrast to current learning paradigms, transfer learning aims to use the information gained from one problem to solve a similar problem in the future. An attribute-mining structure is proposed that uses CNN architectures such as Google Net, VGGNet (visual geometry group network), and residual networks that have been previously trained and fed into a completely linked layer for the classification of malignant and benign cells with an average pooling classification method. Similar research work has been carried out in the area of breast cancer diagnosis [20–22]. Authors have developed a network model to classify medical-related data. Security protocols are also developed to secure medical-related wireless sensor data.

3. Methodology and Result Analysis

3.1. Methodology. The methodology consists of the following major phases. Figure 1 represents a block diagram of the proposed model.

- (i) Image preprocessing using geometric mean filter
- (ii) Feature extraction using AlexNet
- (iii) Feature selection using relief algorithm
- (iv) Classification using LS-SVM and other algorithms

Image preprocessing is important for the correct classification of disease images. Mammogram images contain various types of noise. These noises are removed using image filtering techniques. A geometric mean filter is used to remove noise from the input images [23, 24].

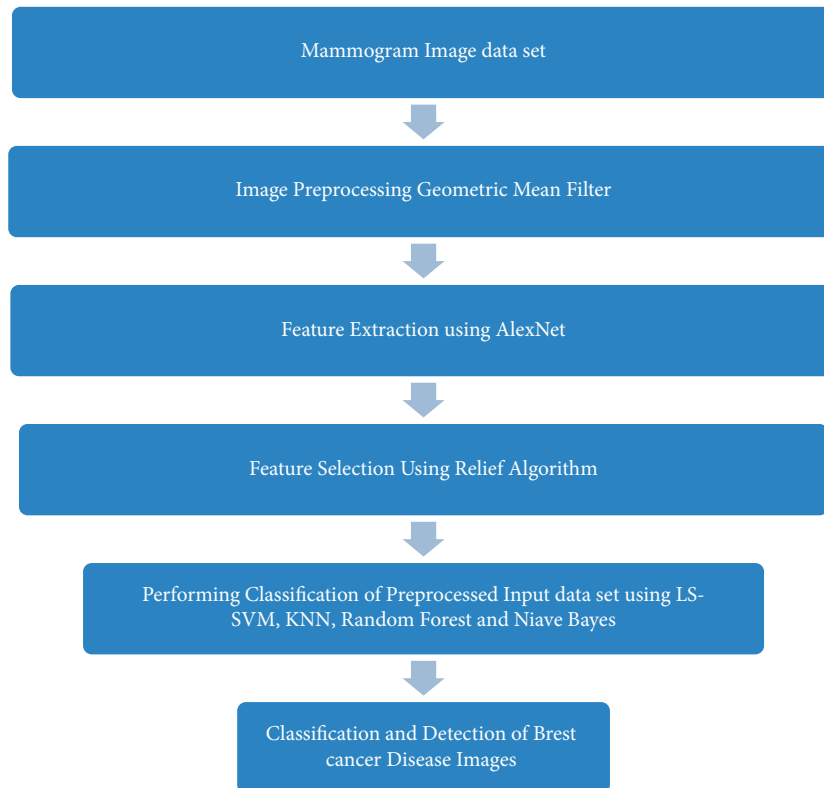


FIGURE 1: Computational technique for mammogram image classification and detection of breast cancer.

In order to extract features, the AlexNet uses a deep learning approach. An AlexNet CNN fully connected layer is utilized to extract features from the fused picture. There are 22 layers of feature extractor in the AlexNet CNN, all based on transfer learning, plus a fully connected (FC) layer with $1 \times 1 \times 64$ dimensions.

Inspired by instance-based learning techniques, Kira and Rendell developed the initial relief algorithm. Using relief's feature selection filtering process, each feature is given a proxy statistic that may be used to determine its "quality" or "relevance" to the target idea. Because it was designed for binary classification issues alone, the original relief method had no way to deal with missing data [16].

There are two types of supervised learning methods: SVM and LS-SVM. These help with classification and regression problems in a machine learning way. Both the SVM and the LS-SVM act as nonprobabilistic binary linear classifiers. They build a hyperplane, which is a line that separates the two classes. LS-SVM is an addition to SVM that is used to solve linear equations and also to find a training model for classification. SVM is used to solve quadratic equations, whereas LS-SVM is used to solve linear equations. LS-SVM classifier costs less than the SVM classifier. In contrast to SVM, LS-SVM is much easier to use because it only needs to solve a set of linear equations to work out how it works. In LS-SVM, there are only a few parameters that need to be set. There are other multivariate classifiers, such as the NB classifier and the NN classifier, but LS-SVM is better at dealing with linear and nonlinear multivariate

classification than these other classifiers. Radial basis function (RBF), linear, polynomial, quadratic, and MLP kernels are some of the kernels used in LS-SVM classifiers [25].

One of the simplest classifiers is the KNN classifier. It is called a "lazy" algorithm. A nonparametric algorithm is one that does not make any assumptions about how the data are spread out. The KNN classifier does not do this. The KNN algorithm is used to figure out what the unknown pattern looks like based on its closest neighbor. To classify the images, the nearest neighbor classifier is used. There are two parts to a KNN classifier. The first part is to figure out how far the unknown image is from each image used in the training phase. The second part is to figure out which training images are the most likely to be testing images. The Euclidean distance is used to classify the objects, and it is used to measure the distance between them. Euclidean distance is the most common way to figure out how far two points are from each other. It is the square root of the sum of the distance between the two points [26].

NB classification is a common way to do supervised learning. It is based on the Bayesian theorem with the assumption that each set of features is separate from the other. NB classification is also known as a person who wants to learn. I think it is very easy to build, and it is also very simple to understand. There is a very fast way to get the predicted class of the test data. It also works with a lot of data. This method of classifying looks at how each attribute and the class are related to each other for each individual case and comes up with a conditional probability for the relationships

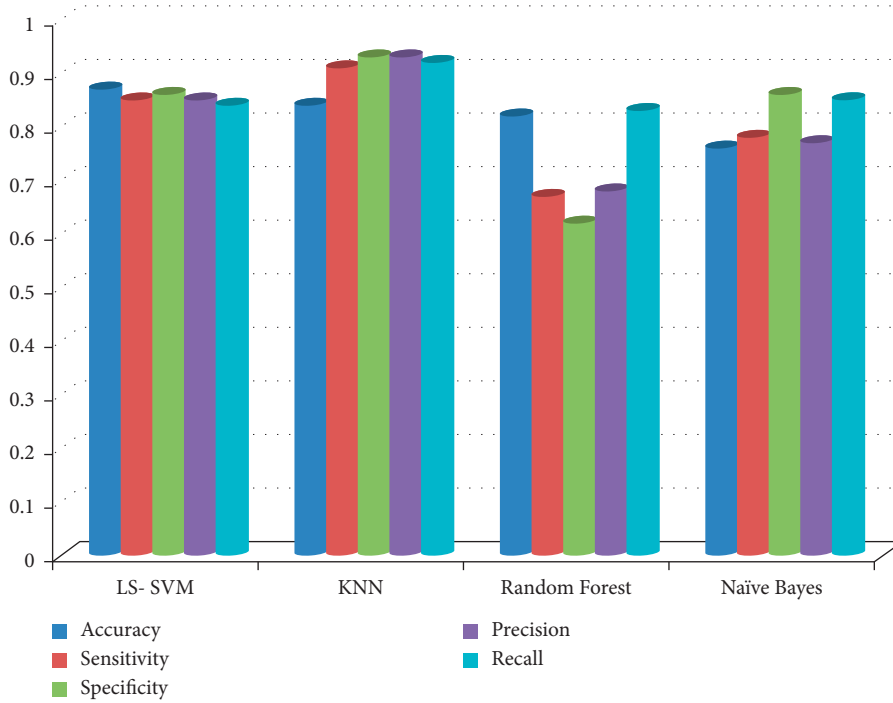


FIGURE 2: Result comparison of classifiers for breast disease detection with feature selection.

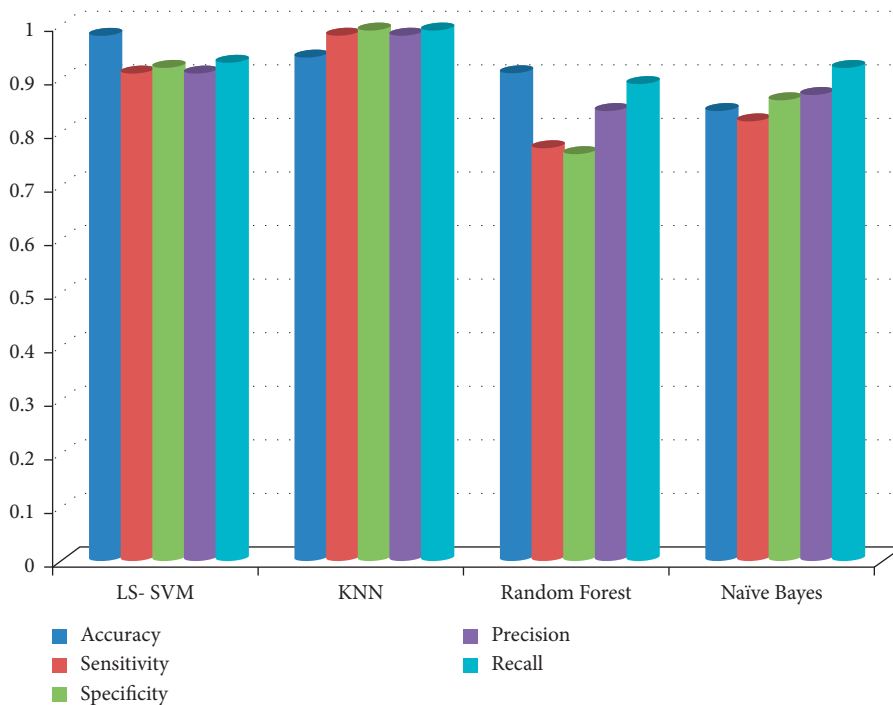


FIGURE 3: Result comparison of classifiers for breast disease detection with feature selection.

between attribute values and the class. It is used to figure out the chances of each class during training by counting how many times the training dataset shows up at different points in time.

The random forest tree (RFT) classifier is a group classification method. It is the same as the nearest neighbor

classifier method. FT makes more trees because it picks variables at random. A classifier learns by looking at a random set of data features to separate tree nodes. The RFT classifier is based on the concept of bagging, which means that each successive tree is made from a bootstrap sample of the data items. The majority vote is used to classify the data items.

3.2. Result Analysis. There are a lot of databases for breast cancer mammograms, and they are used to look at them. MIAS and DDSM are the two databases that are used the most often out of all the databases. It is used in this project to look at the MIAS database. In the MIAS database, there are 322 images of the right and left breasts from mammograms. Among 161 patients, 51 images were found to be malignant, 64 images were found to be benign, and 207 images were found to be normal. Most of the time, MIAS database images include background information, the pectoral muscle, and different types of noises. This picture has a lot of noise in its background. In order to get a better and more accurate analysis and interpretation of breast images, it is important to get rid of all the noise [27].

Three parameters accuracy, sensitivity, and specificity are used in this study to compare the performance of different algorithms.

$$\begin{aligned} \text{Accuracy} &= \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \\ \text{Sensitivity} &= \frac{\text{TP}}{(\text{TP} + \text{FN})}, \\ \text{Specificity} &= \frac{\text{TN}}{(\text{TN} + \text{FP})}, \\ \text{Precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})}, \\ \text{Recall} &= \frac{(\text{TP})}{(\text{TP} + \text{FN})}, \end{aligned} \quad (1)$$

where TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative

The accuracy, sensitivity, and specificity of LS-SVM, KNN, random forest, and Naïve Bayes classifiers for breast cancer disease detection are shown in Figures 2 and 3. Accuracy of LS-SVM is better than the rest of the classifiers. Sensitivity and specificity of the KNN algorithm are better than the rest of the classifiers.

4. Conclusion

Breast cancer is the most lethal type of cancer for women worldwide, affecting one out of every eight women. Because the etiology of breast cancer is still unclear, there are presently no effective treatments for preventing or treating the disease. Early detection and management of breast cancer are extremely successful techniques for diagnosing and managing the illness, and early detection may result in a higher chance of complete recovery. Mammography, the most effective method for detecting breast cancer, can be used to identify it early. Secondary advantages include the capacity to detect additional illnesses and the giving of information on the types of cancer, such as whether it is benign, malignant, or noncancerous. For the first time, an evolutionary approach for categorizing and identifying breast cancer using machine learning and image processing has been established. This model may be used to aid in

the classification and identification of skin problems by utilizing image preprocessing, feature extraction, feature selection, and machine learning methodologies. The geometric mean filter is used to improve the overall picture quality. To extract features from the data, AlexNet is employed. The characteristics to be utilized are chosen using the relief algorithm. To identify and diagnose various illnesses, the model employs machine learning algorithms such as the least square support vector machine, KNN, random forest, and Naïve Bayes. An MIAS data gathering system is used in the experimental inquiry. The proposed technique has the advantage of precisely detecting breast cancer sickness using image analysis, which is a considerable benefit.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] N. Howlader, A. M. Noone, M. Krapcho et al., *SEER Cancer Statistics Review*, Natl. Cancer Inst, Bethesda, MD, 2015.
- [3] T. C. Lewis, V. J. Pizzitola, M. E. Giurescu et al., "Contrast-enhanced digital mammography: a single-institution experience of the first 208 cases," *Breast Journal*, vol. 23, no. 1, pp. 67–76, 2017.
- [4] A. A. Tabl, A. Alkhateeb, W. ElMaraghy, L. Rueda, and A. Ngom, "A machine learning approach for identifying gene biomarkers guiding the treatment of breast cancer," *Frontiers in Genetics*, vol. 10, p. 256, 2019.
- [5] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.
- [6] M. Abdar, M. Zomorodi-Moghadam, X. Zhou et al., "A New Nested Ensemble Technique for Automated Diagnosis of Breast Cancer," *Pattern Recognit. Lett.* vol. 132, 2018.
- [7] I. Varlamis, I. Apostolakis, D. Sifaki-Pistolla, N. Dey, V. Georgoulas, and C. Lionis, "Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data in a regional cancer registry: the case of the island of Crete, Greece," *Computer Methods and Programs in Biomedicine*, vol. 145, pp. 73–83, 2017.
- [8] M. Tan, B. Zheng, J. K. Leader, and D. Gur, "Association between changes in mammographic image features and risk for near-term breast cancer development," *IEEE Transactions on Medical Imaging*, vol. 35, no. 7, pp. 1719–1728, 2016.
- [9] A. M. Hemeida, S. Alkhalaf, A. Mady, E. A. Mahmoud, M. E. Hussein, and A. M. B. Eldin, "Implementation of nature-inspired optimization algorithms in some data mining tasks," *Ain Shams Engineering Journal*, vol. 11, no. (2), 2019.

- [10] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification," *In-formatics in Medicine Unlocked*, vol. 16, Article ID 100151, 2019.
- [11] J. Wang, C. Xia, A. Sharma, G. S. Gaba, and M. Shabaz, "Chest CT findings and differential diagnosis of mycoplasma pneumoniae pneumonia and mycoplasma pneumoniae combined with streptococcal pneumonia in children," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–10, Article ID 8085530, 2021.
- [12] S. S. Mohant y and P. K. Mohant y, "Obesity as Potential Breast Cancer Risk Factor for Postmenopausal Women." *Genes & Diseases*, vol. 8, 2019.
- [13] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, no. 2, pp. 687–699, 2018.
- [14] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, "EMS-net: ensemble of Multiscale convolutional neural networks for classification of breast cancer histology images," *Neuro-computing*, vol. 366, pp. 46–53, 2019.
- [15] Y. Xu, Y. Wang, J. Yuan, Q. Cheng, X. Wang, and P. L. Carson, "Medical breast ultrasound image segmentation by machine learning," *Ultrasonics*, vol. 91, pp. 1–9, 2019.
- [16] T. Thakur, I. Batra, M. Luthra et al., "Gene expression-assisted cancer prediction techniques," in *Journal of Healthcare Engineering*, D. Zaitsev, Ed., vol. 2021, Article ID 4242646, 9 pages, 2021.
- [17] A. Kakti, S. Kumar, N. K. John, V. V. Ratna, S. Afzal, and A. Gupta, "Impact of patients approach towards healthcare costs on their perception towards health: an empirical study," *Tobacco Regulatory Science*, vol. 7, no. 6–1, pp. 7380–7390, 2021.
- [18] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications*, vol. 120, pp. 103–115, 2019.
- [19] S. Chaudhury, N. Shelke, K. Sau, B. Prasanalakshmi, and M. Shabaz, "A novel approach to classifying breast cancer histopathology biopsy images using bilateral knowledge distillation and label smoothing regularization," in *Computational and Mathematical Methods in Medicine*, D. Koundal, Ed., vol. 2021, Hindawi Limited, Article ID 4019358, 11 pages, Hindawi Limited, 2021.
- [20] S. Abbas, Z. Jalil, A. R. Javed et al., "BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm," *Peerj Computer Science*, vol. 7, 2021.
- [21] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and PUF-based Lightweight Authentication Protocol for Wireless Medical Sensor Networks," *IEEE Internet Of Things Journal*, vol. 1-1, 2021.
- [22] R. Thippa, S. Bhattacharya, P. K. R. Maddikunta, and S. Hakak, "Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset," *Multimed Tools Appl*, Springer, Berlin, Germany, 2020.
- [23] S. Suman, C. K. Loo, K. S. Yap, K. W. Wong, T. Beng, and . Huang, "Image enhancement using geometric mean filter and gamma correction for WCE images," in *Neural Information Processing. ICONIP 2014* vol. 8836, Cham, Switzerland, Springer, 2014.
- [24] M. Canayaz and M. H. COVIDNet, "MH-COVIDNet: diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images," *Biomedical Signal Processing and Control*, vol. 64, Article ID 102257, 2021.
- [25] Y. Ma, X. Liang, G. Sheng, J. T. Kwok, M. Wang, and G. Li, "Noniterative sparse LS-SVM based on globally representative point selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 788–798, 2021.
- [26] B. Sun and H. Chen, "A survey of k nearest neighbor algorithms for solving the class imbalanced problem," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5520990, 12 pages, 2021.
- [27] peipa.essex, "The mini-MIAS database of mammograms," 2012, <http://peipa.essex.ac.uk/info/mias.html>.

Research Article

An Efficient Blockchain Based Data Access with Modified Hierarchical Attribute Access Structure with CP-ABE Using ECC Scheme for Patient Health Record

F. Sammy ¹ and S. Maria Celestin Vigila²

¹Department of Information Technology, Dambi Dollo University, Dembi Dolo, Welega, Ethiopia

²Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu, India

Correspondence should be addressed to F. Sammy; sammy@dadu.edu.et

Received 21 January 2022; Revised 2 February 2022; Accepted 12 February 2022; Published 8 March 2022

Academic Editor: G. Thippa Reddy

Copyright © 2022 F. Sammy and S. Maria Celestin Vigila. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Secure patient health record (PHR) information exchange via cloud computing is a considerable security risk to user privacy. The fundamental reason of this issue is cloud computing's reliance on trustworthy third parties to share data across it. To exchange data securely, many conventional cryptographic algorithms employ various keying approaches. However, relying on a trusted third party compromises the privacy of consumers' data. To offer secure communication without the involvement of a third party, a distributed blockchain based (DBC) ciphertext-policy attribute-based encryption (CP-ABE) approach is employed in this study. Because of bilinear pairing and simple scalar multiplication factors, the proposed CP-ABE system is entirely dependent on elliptic curve cryptography to reduce complexity. Furthermore, the data requester provides dynamic attributes, and a user-centric access policy is created, allowing multiple authorities to manage the attributes and provide data access. Data confidentiality, data authentication, user authentication, and tamper-proof data are all guaranteed by the suggested method. The DBC-CP-ABE method is used to provide user-centric access policies and effective key management.

1. Introduction

The Internet of Things (IoT) is a new technology that allows items to communicate with one another across wireless networks. IoT devices are resource-constrained and have challenges with data processing, data storage, and energy consumption. Cloud computing provides a centralized solution to these resource restricted procedures to overcome these limits. The collected data is stored and processed in the cloud, but the cloud can be a trustless environment with major security issues such as single point failure, data tampering, lack of user privacy due to a lack of data access control, Denial of Service (DoS), Man-in-the-Middle attack (MiTM), and password phishing. As a result of permitted data/device access, the cloud environment is prone to security breaches, compromising users' privacy. Many public key secret writing techniques provide a fine-grained access

control strategy while also protecting the privacy of users. Among other public key encryption methods, CP-ABE scheme offers one-to-many access control which allows data to be shared across multiple users. But the implementation overhead incurs due to operations with bilinear pairing. It consumes more resources with high computational cost. To overcome this issue, less complex and less resource consumption scalar computation with elliptic curve cryptography (ECC) is used in this work. This reduces the computational requirement by two-three times that of bilinear pairing. This work focuses on building a security system with blockchain where hierarchical access control policy is achieved by combining CP-ABE and ECC. The experiment analysis shows that our scheme outperforms the compared work in cryptographic operations. The major focus of this work is concentrated to achieve the following criteria:

- (1) Adoption of straightforward scalar multiplication with ECC and CP-ABE approach reduces procedure overhead caused by bilinear pairing methodology
- (2) The proposed method ensures use of multiple authorities to manage attributes and shares multiple data attributes of single data user
- (3) To specify the access policy scheme with increased security, the Linear Secret Sharing Scheme (LSSS) is used
- (4) Attribute revocation for a data user is achieved with the help of RSA key pair in communication between the data user and Attribute Authority (AA)

The following are the last sections: Section 2 contains material from the research study that is relevant to the current effort, and Section 3 contains information on the proposed study's contribution. The preliminaries utilized on this project are explained in Section 4. The architecture of the blockchain based hierarchical access control scheme with CP-ABE using ECC is briefly described in Section 5. The modified hierarchical attribute access structure (MHAAS) with CP-ABE employing ECC is explained in Section 6. The integration of HACS-CP-ABE-ECC with blockchain is explained in Section 7. Section 8 summarizes the HACS-CP-ABE-ECC with blockchain security analysis, whereas Section 9 describes the performance evaluation conducted in this study.

2. Related Works

Cloud computing offers computation of massive data and data sharing in a promising way [1]. Data are encrypted and shared in cloud computing environment either with symmetric key encryption or public key standards [2–4]. This method has drawback in achieving security [2] and drawback in flexible access control [3] and shows poor performance [4]. To deal with these drawbacks, attribute-based encryption (ABE) is proposed. There are two types of attribute-based encryption: KP-ABE and CP-ABE. Bethencourt et al. [5] were the first to suggest CP-ABE. ABE scheme with bilinear pairing showing less efficiency was proposed [6]. ABE is further refined with CP-ABE involving hierarchical attributes as proposed by [7] to address key management problem [8]. A multiauthority-ABE with dynamic policy attributes is proposed, although the CP-ABE method demonstrates little improvement [9]. An access policy based on the DBDH scheme is proposed. All the CP-ABE methods described above use bilinear mapping using large sized keys. To lessen the complexity of CP-ABE, the decryption method is split into degrees: predecryption and final decryption degrees in [10]. But this method does not ensure forward security. This has been improved in other work [11] where encryption and decryption are outsourced and validated but lack improvement in encryption and decryption process. The work is also extended in [12] by redistributing the encoding and decoding system to fog nodes; however, they are easily attacked. Another decryption outsourcing work proposed in [13] resists against selective ciphertext.

Although the work in [14] provides outsourcing of encryption and decryption process, it uses bilinear pairing that remains as hurdle to performance improvement to CP-ABE.

CP-ABE does not ensure less storage overhead and good cost-effective solution as it depends on the use of bilinear maps. A bilinear map produces secret keys of larger values and ciphertext with linear associated attribute. And also it uses exponentiation factors for doing encryption and decryption process which relies on linear attributes defined in the access policy [15–17]. The problem of requiring a large key size necessitates the usage of elliptic curve cryptography (ECC) with a smaller key size. This paves the path for CP-ABE to define an access structure utilizing ECC [18–20]. Lightweight devices such as the CP-ABE with constant key size using ECC have been developed, but they are not appropriate for complex access structures [21, 22]. Another lightweight work using KP-ABE without bilinear pairing is proposed but suffers from poor scalability and lack of decryption outsourcing [23]. The overall computation overhead due to bilinear pairing is overcome with ECC [24]. Constant key size with CP-ABE using ABE addressing collision attack problem is proposed in [25]. Alternative to bilinear pairing with ECC to address secured data share is proposed in [26]. All the abovementioned work defines the access policy based on the set of attributes.

Bethencourt created the first tree-based access control structure in order to implement AND, OR, and OF strategies [27]; however, it is insecure. Many studies focus on improving access control strategies; however, the time it takes to encrypt and decrypt data grows as the number of attributes increases. The research was furthered by Lewko and Waters, who proposed a technique to convert tree access control to an LSSS and Waters enhanced CP-ABE with a matrix format [28]. With d -parallel BDHE assumption, this gives security. Many studies have refined the use of CP-ABE with flat access control [29–33], constant ciphertext [34], accountability and authorities with attribute revocation [35–38], and improvement in security through accountability and authorities. However, none of these structures support hierarchical file relationships.

Hierarchical CP-ABE based on LSSS matrix structure was also studied. By considering hierarchical heads sharing secret keys with users, these approaches lessen the burden of a single head [39, 40]. In this paper, we design a hierarchical based access relation for sharing multiple files [41] in a distributed blockchain context using LSSS. To address the privacy and security concerns, [45] present a unique pairing-free certificateless method that builds a novel reliable and efficient lightweight certificateless signature (CLS) scheme using a state-of-the-art blockchain technique and smart contract. Paper [46] addresses a lightweight and reliable authentication protocol for wireless medical sensor networks (WMSN), which is composed of cutting-edge blockchain technology and physically unclonable functions (PUF), to address physical layer security and the over-centralized server problem in WMSN. The elliptic curve digital signature algorithm (ECDSA), which is one of the essential building blocks of blockchain, is proposed in [47] as an efficient and large-scale batch verification technique with

group testing technology. Using edge computing and blockchain approaches, [48] introduces search efficiency, reliability requirements, and a resource allocation scheme to properly handle IoT devices. The study [49] demonstrates how to use erasure coding to overcome data integrity issues in IoT devices.

3. Our Contribution

We suggested a blockchain based hierarchical access scheme that uses CP-ABE with ECC in this paper. A hierarchical access hierarchy is defined here, with the user attribute satisfying partially or entirely alone allowing partial or complete access to the data. A root authority (RA) checks and joins all of the domain attribute authorities (AA) in the blockchain. For each AA, RA produces a public key and a master key. It also sends hierarchical access scheme to all AA. RA sends the public key to AA while keeping the master key hidden. AA takes an attribute from the users and generates an address, an RSA key pair, and a private key for that attribute. Based on this, AA distributes the attribute's address, RSA key pair, and private key to the user who satisfies the access structure to decrypt the data. The AA keeps track of the RSA key pair in order to revoke the user's attribute. To reduce computing complexity, the pre-decryption is outsourced to AA, and AA's trust is kept thanks to the presence of blockchain. The suggested method ensures that data is shared with several authorities and that different attributes of the user's identification are shared.

4. Preliminaries

4.1. Elliptic Curve Cryptography. ECC is a discrete logarithm problem-based public key cryptography (ECDLP). The elliptic curve E is defined by $FG(P)$, a finite field, and is written as $y^2 = x^3 + ax + b \pmod{p}$ and $4a^3 + 27b^2 \neq 0$. Calculate a point on the curve $Q = KG$, where G is the prime order r generator group over the polynomial time k . The plain texts are transferred to the elliptic curve's point Q . The ECC procedure is broken down into three phases.

(a) Key generation:

- (1) Both the data server and the data client have agreed to use the same elliptic curve $y^2 = x^3 + ax + b \pmod{p}$ and G
- (2) The data server generates a random number, $Sa \in Z_p$, as the private key, and $Pa = SaG$, as the public key
- (3) Data clients generate a private key using a random number $Sb \in Z_p$ and a public key using $Pb = SbG$

(b) Encryption:

The data server encrypts the message with Q by selecting a random number $K \in Z_p$, then computes the cypher text $C1 = KG$ and $C2 = Q + K Pb$, and sends both $C1$ and $C2$ to the data clients

(c) Decryption:

Data clients use $C2 - SbC1 = Q + kPb - SbKbG = Q$ to decrypt the message. The message is obtained by mapping to the curve's point Q .

4.2. Hierarchical Access Control Strategies. As demonstrated in Figures 1 and 2, a hierarchical access control technique allows numerous access structures to be combined into a single structure T .

4.3. Linear Secret Sharing Scheme. Beimel proposed the Linear Secret Sharing Scheme [33]. When all parties make up a share on vector Z_p , a Secret Sharing Scheme is defined across linear Z_p for various parties. Matrix M was created to generate shares for all parties. Consider the M matrix, which has p rows and q columns. Consider a row of a matrix M_i where $i = (1, 2, \dots, p)$ meets the criterion $1, 2, \dots, p) \rightarrow d$, and given a column vector $\bar{O} = (s, u_2, \dots, u_n)$ with the secret key $s \in Z_p$ and $u_2, \dots, u_n \in Z_p$ picked at random. M is made up of m shares of s , each of which is dependent. The share $m_i = (M \bar{O})_i$ belongs to a specific political party.

Consider an LSSS Π with T as the access tree structure and $S \in T$. This denotes an arbitrary permitted set, $L \subset \{1 \dots p\}$, and $L = \{i: m_i \in S\}$. $s = \sum_{i \in L} \omega_i m_i$ and m_i are arbitrary secret s specified by ω_i which is discovered in the matrix M in polynomial time. There is a vector $\omega = (1, 0, \dots, 0) = -1$ and $\omega \cdot M_i = 0$ for the unlawful set of rows $i \in L$.

When a j th secret of a nonleaf node is recovered from a set of n secrets, the set of attributes $\{\omega_{i \in Z_p}\}$ can be discovered in polynomial time by satisfying $\sum_{i \in L} \omega_{i,j} M_i^t = \epsilon_j$, where j denotes a row vector of length n with the j th element equal to 1 and all other elements equal to 0. As a result, secret share $s_j = \sum_{i \in L} \omega_{i,j} m_i$. The marking method defined by [34] is used to create the LSSS matrix. It translates a Boolean formula-defined access tree to the LSSS matrix technique. In hierarchical access control, this LSSS marking mechanism is employed. According to Figure 3, if the user characteristics only partially satisfy the access structure policy, just a portion of the information is decrypted.

5. Architecture of Blockchain Based Hierarchical Access Control Scheme with CP-ABE Using ECC (BHACS-CP-ABE-ECC)

Certificate authority (CA), attribute authority for personal, health, and insurance domain, cloud service provider, data owner (DO), data clients (DC), and edge nodes for pre-decryption process are all part of the proposed blockchain linked architecture. Figure 3 depicts the proposed scheme framework. The following is a description of each participant's functionality:

- (1) Root authority (RA): the main role RA is to provide identity of the communicating nodes by considering the security parameter (K) of the node and generates public parameter (PP). To generate a public key and master key for an attribute, this public parameter is submitted to the appropriate attribute authority.

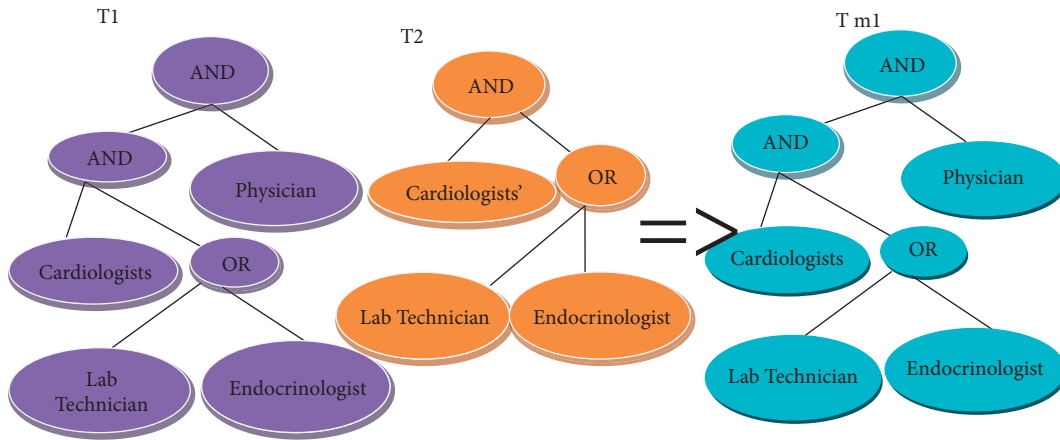


FIGURE 1: Part of integrated hierarchical access control structure.

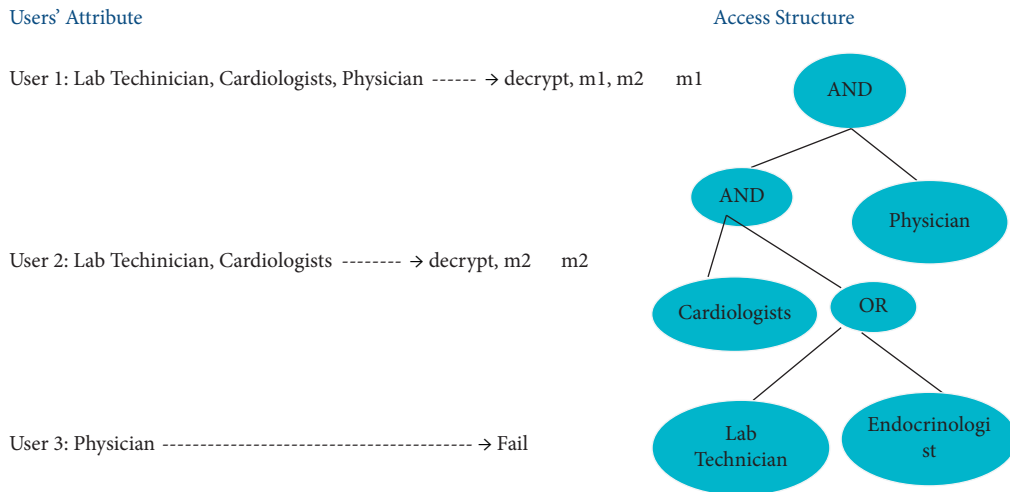


FIGURE 2: Part of integrated hierarchical access control process.

- (2) Attribute authority (AA): attribute authority of the domain extracts the attributes of their respective data clients. Attribute authority further generates public key and master key for that attribute.
- (3) Cloud storage: cloud storage serves to store encrypted data (CT) sent by the data owners.
- (4) Data owners (DO): the data is encrypted before being uploaded to the cloud server by the data owners. It creates ciphertext CT using plaintext B, the matching public key PK, and the access policy given by the LSSS matrix structure (M,m).
- (5) Data clients (DC): data clients are responsible for performing decryption on CT. Deciphering is done in two stages. First the local server near the DC serves as edge nodes and does partial encryption by inputting CT and SK. Finally the DC decrypt the partial decrypted CT to plaintext by considering CT' and DSK.

6. Modified Hierarchical Attribute Access Structure (MHAAS) with CP-ABE Using ECC

The following section explains the process carried out using hierarchical access policy structure (Schemes 1–5).

7. Integration of HACS-CP-ABE-ECC with Blockchain

The hierarchical access control scheme employing with ciphertext ABE using ECC is integrated with blockchain and its operation is explained below.

The operation of this method is explained as six principal components as initialization phase, authority creation, user creation, ciphertext data upload, creation and issuance of attributes, and revoke attribute. This process includes the reception of only the secret key of the attribute for a particular address in its wallet alone is specified in the process.

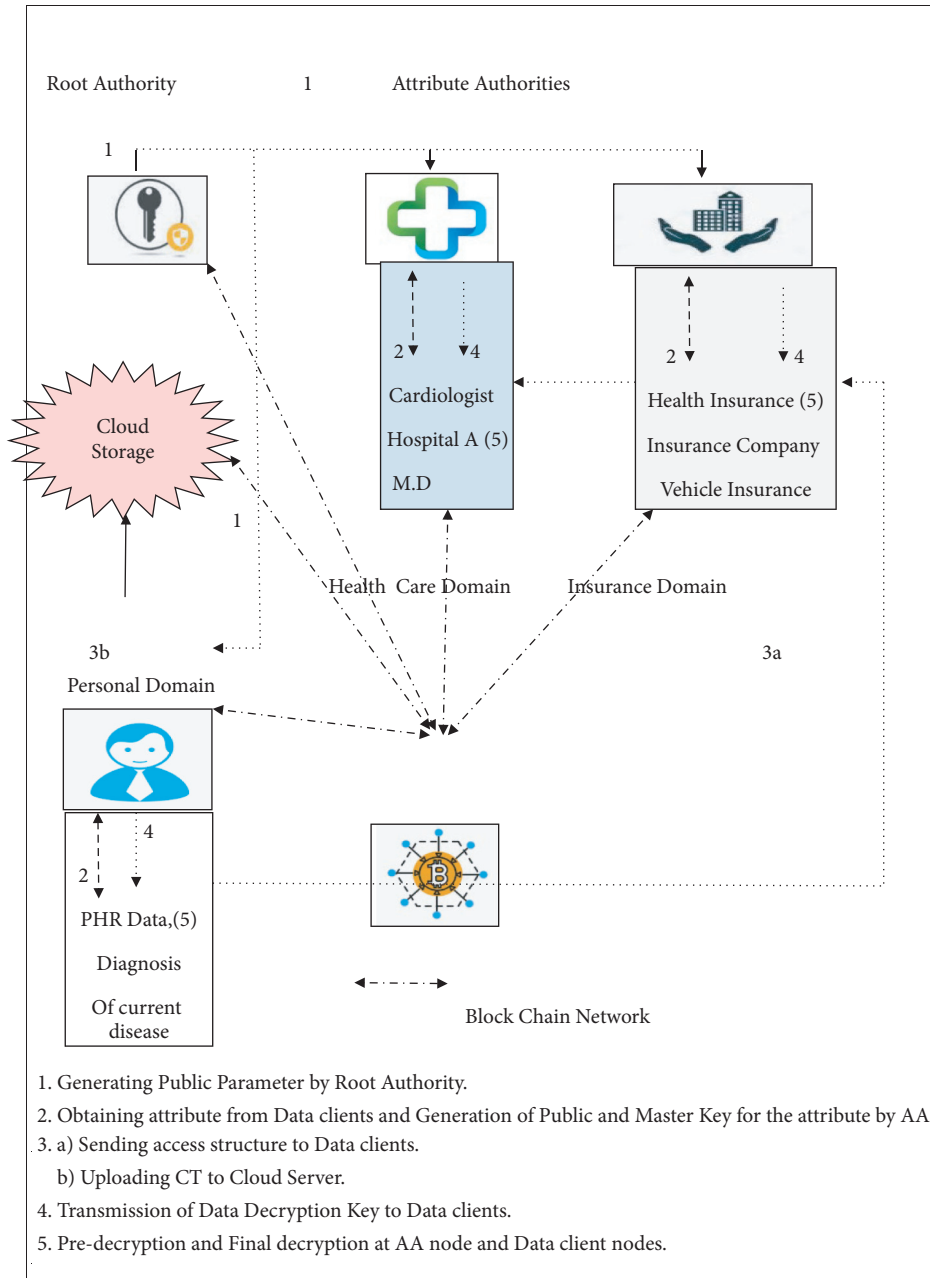


FIGURE 3: Architecture of proposed blockchain based patient centric data access with CP-ABE using ECC.

Input: Security parameter K .

Output: Public parameter PP .

1. Select G Generator of cyclic subgroups with prime order r on $E, GF(q)$ of order r , Elliptic Curve E defined over $GF(q)$.
2. Select $H: 0, 1^* Z_r$ as the hash function, and map the elements of Z_r to the users' GIDs.
3. Define the global attributes $A = \{a_1, a_2, \dots, a_n\}$. Each attribute is defined in the form of LSSS Matrix M_i , representing hierarchical access structure. These attributes are maintained by multiple authorities who generate the necessary key corresponding to the attributes.
4. PP is calculated from the set of $\{GF(q), E, G, h, A\}$.
5. PP is sent to Attribute Authority for generating public and Master key for an attribute.

SCHEME 1: System initialization at root authority.

Input: Public parameter PP .

Output: Public Key PK and Master Key MSK for attribute i .

1. Select two random numbers $\alpha_i, \beta_i \in Z_r$
 2. Generate Master Key = $\{\alpha_i, \beta_i, \forall_i\}$.
 3. Generate Public Key = $\{G\alpha_i, G\beta_i, \forall_i\}$.
-

SCHEME 2: Generation of public and master key by attribute authority.

Input: Plaintext set $\{B_j, j \in (p, q)\}$, PK , LSSS Matrix Structure (M, m)

Output: Cipher Text CT .

1. The data owners encrypt the plaintext message B using symmetric key c_r and generate the cipher text data $CT_d = E_{c_r}(B)$, using the symmetric key encryption algorithm $E(B)$.
2. Calculates the hash value on the cipher text data $H_{CT} = H(CT_d)G$. This ensures data integrity.
3. Data Owners defines LSSS structure (M, m) and sends to the data clients.
4. The encryption algorithm is divided into two stages

- a. Calculate $C_0 = c_r + sG$ where $s \in Z_p$.
 - b. Select two random vectors $\vec{v} = (s, v_2, v_3, \dots, v_m) \in Z_p$ and $\vec{u} = (0, u_2, u_3, \dots, u_m) \in Z_p$ and calculate $\lambda_x = M_{x, \vec{v}}$, $\omega_x = M_{x, \vec{u}}$, $C_{1,x} = \lambda_x G + \gamma_x y_{m(x)} G$, $C_{2,x} = \gamma_x G$, $C_{3,x} = \omega_x G + \gamma_x k_{m(x)} G$ if $m(x) \in$ normal attributes where γ_x random number \in and $x \in [1, p]$ and M_x is the row x of M .
-

5. Finally calculate $CT = \{(M, m), C_0, \{C_{1,x}, C_{2,x}, C_{3,x}\}, CT_d, H_{CT}\}$. Data Owners uploads CT to the Cloud Server Storage.
-

SCHEME 3: Encryption by data owners.

Input: PP, GID, MSK

Output: Generation of Private Key for the Data clients .

1. Generates Key $USR_{i, GID} = \gamma_i + H(GID)k_i$ for the user with GID of i th attribute.
 2. Attribute Authority generates a conversion key $USR_{EN, GID} = USR_{i, GID}, i \in s_i, GID$. It Sends this key to partial decrypting node and Data client node.
 3. Data client obtains the private key by calculating $USR_{i, GID} = \gamma_i + H(GID)k_i + z$ where z is a random number $\in Z_r$.
-

SCHEME 4: Key generation for data decryption by attribute authority.

The encryption, decryption, and retrieval of public key of the attribute are omitted here.

- (a) Initialization phase: this phase includes initialization of blockchain and setting of hierarchical access based policy scheme to all attribute authorities in different domains and provides permission chain through RA.

The RA performs mining of the genesis block. For all attribute authority domains, RA is in charge of producing public key PK and master key MSK . Everyone has access to the public key, while the master key is kept hidden in order to generate the private key for data decryption.

Input: Cipher Text CT ,

Output: Plain Text $\{B_p, j \in (p, q)\}$

Pre-Decryption Stage:

1. For attribute m_x the decryption is obtained as $D_x = C_{1,x} \cdot SK_{m(x), GID} \cdot C_{2,x} \cdot H(GID) \cdot C_{3,x}$. At the end of this stage $CT' = \{C_0, CT_d, H_{CT}, T_1, T_2\}$ is sent to the final decryption stage where $T_1 = \sum_{x \in X} C_x \cdot D_x = sG - z \sum_{x \in X} C_x \cdot \gamma_x \cdot G$ and $T_2 = \sum_{x \in X} C_x \cdot C_{2,x} = \sum_{x \in X} C_x \cdot \gamma_x \cdot G$

Final Decryption Stage:

2. Data Clients calculates $H'_{CT} = H(E_{c_r}(B)) \cdot G$ by using $C_r' = C_0 \cdot T_1 + zT_2$. If $H_{CT} = H'_{CT}$, the plaintext is valid.
-

SCHEME 5: Decryption by data clients.

- (b) Authority creation: RA uses MSK of the Attribute Authority Domains and generates private key SK_{AU} for each AA domain. For this, RA generates new address $\{p, RSA_{SK}, RSA_{PK}\}$ for each AA domain and transmits public key and RSA key pair to all AA domains. RSA key pair is used to transmit the symmetric encryption key securely to all the requester. RA provides “transmit”, “receive,” and permission rights to all AA domain and joins the blockchain system.
- (c) User creation: now the respective AA adds the individual users and obtains the attributes from the users in the domain. AA generates the address u , RSA key pair, and private key of the individual users SK_{USER} . AA transmits the address u and RSA key pair to the individual users and keeps the SK_{USER} with AA.
- (d) Ciphertext data upload: data owners create ciphertext of the data and uploads the hashed ciphertext to the cloud storage.
- (e) Attribute creation and distribution: data owners specify the LSSS policy for all data requesters via AA. AA holds the LSSS policy attributes and generates the data decryption private key for the each data requester. Data clients who satisfy the partial or full policy structure are granted access to the data; otherwise, access is refused.
- (f) Revoke attribute: since AA holds the attributes of the data clients, it can also withdraw the attributes as it holds the RSA key pair of the attribute for the particular data client. It revokes all the attributes of the data client.

Functions of RA:

- (1) Blockchain creation
- (2) Permits all AA to join the blockchain
- (3) Grants connect, send, and receive permission to all AAs

7.1.2. Attribute Authority

- (1) User creation
- (2) Obtains attributes from the data requester
- (3) Creates address and RSA key pair and private key for a particular attribute of the data requester
- (4) Sends LSSS access policy to all the data requester
- (5) Pre-Decryption of CT
- (6) Revoke attributes for a particular data requester

7.1.3. Data Owner

- (1) Define and send LSSS access policy to all AA
- (2) Data is encrypted and sent to cloud storage using the symmetric encryption technique
- (3) Shares the symmetric key securely using RSA key pair to the entire AA

7.1.4. Data Clients

- (1) Sends the attribute list to the AA of that domain
- (2) Performs final decryption of the data requested

8. Security Analysis of BHACS-CP-ABE-ECC

The following section examines the proposed scheme’s security. Under the assumption of DDH, the security model is considered as being secure.

The proposed method supports multiauthority and multiattribute from single data user. For each attribute, attribute authority generates a set of {address, RSA key pair, and private key}. The address, master key, and RSA key pair are preserved by the AA to secure the system against adversary attacks and to perform attribute revocation. Instead

7.1. Functions of the Blockchain Components

7.1.1. Root Authority

BHACS-CP-ABE-ECC:

- (1) Initialization phase-(PP \rightarrow PK,MSK)
- (2) Authority Creation(PK,MSK,P) \rightarrow SK_AA1,SK_AA2. . . SK_AAn

of bilinear pairing, the computing phase uses basic scalar multiplication, which makes the procedure more efficient with the compared models. Also, the decryption process is done at two stages: one at the AA end and the other at the data requester which makes the decryption process lighter at the data requester end.

8.1. Security Assumption under Decisional Diffie–Hellman.

The proposed model considers Decisional Diffie–Hellman (d-DDH) Assumption and described as follows: the challenger chooses F , a cyclic group with prime order s , and G , a cyclic group F generator, while y and k are chosen at random from Z_s . Despite being given a tuple of (G, yG, kG) , the adversary finds it difficult to validate y, k, G in polynomial time from random element $X \in F$. The algorithm A overcomes the DDH problem with a constant factor ρ which is obtained from $|\text{Fs}[A(G, yG, kG, Z = ykG) = 0] - \text{Fs}[A(G, yG, kG, Z = X) = 0]| \geq \rho$.

8.2. Security under Chosen Ciphertext Attack. The communication between adversary δ and the challenger ζ is given below. The summons is given an access structure (T, m) by the opponent. The initialization algorithm is run by challenger. The system's public parameter is used to compute the public and secret keys, with the public key being sent to the opponent. Stage 1: the adversary queries the secret keys of the attribute from the challenger. The challenger records the attributes provided by the adversary in the list and stores it with the corresponding adversary address in the attribute list.

Challenge phase: here the challenger picks out two identical-length messages $(B_0, B_1) \in Z_p$. Then the challenger selects $\beta \in \{0, 1\}$ and forwards B_β under matrix (T^*, m) to challenger δ .

Stage 2: the adversary inquires about the secret key with same input queries in Stage 1. Guess: the guess that the adversary creates is equal to $1/2$, the probability of guessing β_0 of β . The game is defined as $|\text{Fs}[\beta_0 = \beta] - 1/2|$. Thus our model is secure against selective ciphertext attack.

8.3. Data Security. The adversary is unable to obtain to decrypt the ciphertext as its attribute must satisfy the access policy defined in the matrix structure corresponding to a row of Tm . For unauthorized set of rows L , there exists a vector $\omega. (1, 0, \dots, 0) = -1$ and $\omega.M_i = 0$ for $i \in L$, where $L = \{i: m_i \in S\}$.s and ω is the polynomial time in matrix M . The adversary cannot calculate the first element of vector ω . Thus, the proposed scheme ensures the data security.

8.4. Forward Security. The attribute authority revokes the attribute of the user with the users address and RSA key pair. The user/address of the revoked attribute cannot decrypt the data again as AA has deleted/blocked the address of the corresponding attribute from the attribute list. Hence the proposed method ensures forward security.

8.5. Collision Resistance. The proposed method ensures resistance to collision attack. The hierarchical access structure policy has been defined in the system. There is a chance of colliding with same access policy generated by multiple users. In this system, a unique address is generated for each set of attributes defined by different users. Hence address of $\text{User}_A \neq \text{address of User}_B$. This uniqueness provides collision-free system.

8.6. Data Integrity. The owner of the data computes ciphertext with symmetric encryption key algorithm E to encode the plaintext message B using symmetric key c_r and calculates the ciphertext data $CT_d = E_{c_r}(B)$. Then the hash value on the ciphertext data $H_{CT} = H(CT_d)G$ is calculated. This ensures data integrity. The final cryptic message uploaded to the cloud is $CT = \{(M, m), C_0, \{C_{1,x}, C_{2,x}, C_{3,x}\}, CT_d, H_{CT}\}$. During decryption, the hash value is used to ensure data integrity at the data requester end.

8.7. Unauthorized Communication. In each level, the data is decrypted using the user address and an RSA key pair. For each set of attributes defined by distinct users, a unique address is generated in this system. As a result, user A 's address is different from user B 's address and the unauthorized users cannot be entirely in the system. This prevents unauthorized hierarchical communication between nodes.

9. Performance Evaluations

This section briefs the performance evaluation in terms of used system properties, communication overhead, and computation overhead with the previous works and the proposed work.

The subsequent section describes Table 1 that provides the comparative study of the proposed approach with other research works carried out.

From Table 1, it is found that all the works carried out rely on ECC scheme rather than bilinear pairing and [24, 44] use LSSS based access structure where [21, 42, 43] use AND gate access structure whose performance is lesser than the previous LSSS approach. The proposed work differs from all the above by adopting hierarchical access structure with LSSS. Each attribute set corresponds to a row of a matrix in LSSS structure. Thus, our scheme supports multiauthority and multiple data access method.

Table 2 describes the computation cost encountered in current research study and compared study. The entire scheme employs common ECC with 160 bit by $|G|$. It seems that our scheme considers a single row in matrix structure for performing cryptographic operation which is more efficient than the compared schemes.

It is observed that, from [21, 43], the computation overhead depends on the difference between the number of attributes defined in the access policy and the total attributes in the system. Schemes in [24, 44] depend on different attribute set to perform cryptographic process. The scheme in [42] uses KP-ABE and our scheme uses CP-ABE and has

TABLE 1: Performance comparison of the proposed study with other research studies.

| Research study | Pairing free | Access structure | Multiauthority | Decryption outsourcing |
|----------------------|--------------|-------------------|----------------|------------------------|
| [21] | Yes | AND gate | No | No |
| [24] | Yes | LSSS | No | No |
| [42] | Yes | AND gate | No | No |
| [43] | Yes | AND gate | No | Yes |
| [44] | Yes | LSSS | Yes | Yes |
| Proposed work | Yes | Hierarchical LSSS | Yes | Yes |

TABLE 2: Computation cost of the compared algorithms.

| Scheme | Encryption | Predecryption | Local decryption |
|----------------------|-------------------------|---------------|-------------------------|
| [21] | $(Nt - \Lambda + 2)g$ | — | $(Nt - \Lambda + 3)g$ |
| [24] | $(3Ns + 1)g$ | — | $(Dt + 1)g$ |
| [42] | $(Ns + 1)g$ | — | $(Dt + 1)g$ |
| [43] | $(Nt - \Lambda + 1)g$ | — | $(Nt - \Lambda + 2)g$ |
| [44] | $(4Ns + 1)g$ | $(Dt + 1)g$ | $(1)g$ |
| Proposed work | $(Ns + 1)g$ | $(Dt + 1)g$ | $(1)g$ |

Ns: total number of rows in the matrix Λ , Nt: system attributes, Dt: number of rows in access matrix Λ , Nt: total number of attributes in the system, Dt: number of attributes fulfilling the access policy, $|\Lambda|$: access policy attributes, and g: ECC scalar factor for multiplication.

TABLE 3: Communication overhead of the compared algorithms.

| Scheme | Private key (bits) | Public key (bits) | Ciphertext (bits) |
|----------------------|--------------------|-------------------|---------------------------|
| [21] | $2 * g $ | $(3Nt + 1) g $ | $(Nt - \Lambda + 3) g $ |
| [24] | $(\Lambda) g $ | $(2Nt + 2) g $ | $(2Ns + 1) g $ |
| [42] | $(Ns + 1) g $ | $(2Nt + 2) g $ | $(2Ns + 2) g $ |
| [43] | $1 * g $ | $(Nt + 1) g $ | $(Nt - \Lambda + 2) g $ |
| [44] | $(\Lambda) g $ | $(2Nt + 2) g $ | $(3Ns + 1) g $ |
| Proposed work | $(\Lambda) g $ | $(Nt + 2) g $ | $(Ns + 1) g $ |

Ns: total number of rows in the matrix Λ , Nt: system attributes, Dt: number of rows in access matrix Λ , Nt: total number of attributes in the system, Dt: number of attributes fulfilling the access policy, $|\Lambda|$: access policy attributes, and g: ECC scalar factor for multiplication.

efficient encryption process as the encryption process depends on set of attributes alone. But there is an overhead in decryption process in the scheme in [42] as the end node has to do total decryption process. But our scheme offloads the predecryption process to AA and final decryption at the end nodes. The proposed work excels other compared works in computation and communication cost.

Table 3 discusses the communication cost observed in our work and other compared works. Compared with other works, our scheme and the scheme in [44] have increased communication overhead as predecryption process is carried at the AA in our scheme and at edge nodes in the scheme in [44]. Also revocation process is carried at AA in our scheme without affecting other components in the system. Hence the communication overhead at the end data requester is minimum compared with others.

To excel our scheme, we implemented our proposed work in Ubuntu platform. It is deployed with Python and charm library to implement the ECC using simple scalar multiplication. The scheme is implemented with 512-bit ECC where 160 bits serves as ECC group order. In Figures 4 and 5, we show that our scheme excels the work done in

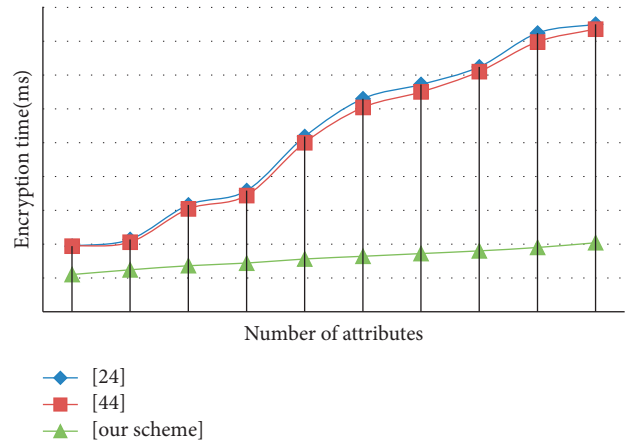


FIGURE 4: Comparison of encryption time (ms).

[24, 44] in executing cryptographic operations. Our scheme excels [24, 44] in executing encryption method and decryption outsourcing is performed in the proposed work and [44]; hence it has same decryption time.

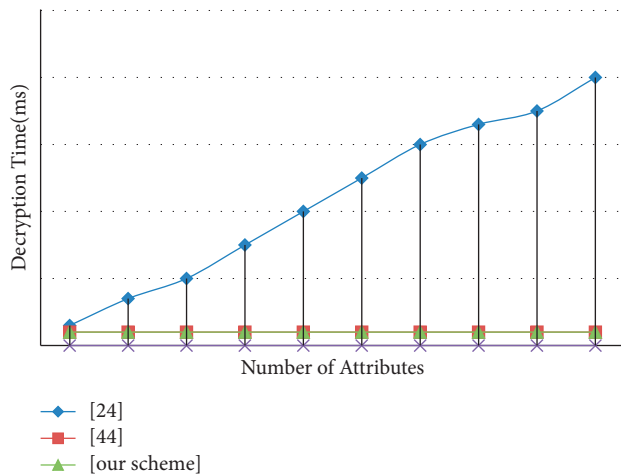


FIGURE 5: Comparison of decryption time (ms).

10. Conclusion

The proposed work uses CP-ABE using ECC in blockchain network with hierarchical access structure. The scheme considers multiple authorities and multiple data access by defining the attribute set. The attribute set is represented as row in LSSS matrix structure. For each attribute set, a unique address is generated along with RSA key pair. This pair is helpful in revoking the attribute, thereby providing security from unauthorized users. Further, the security mechanism of the proposed work is defined under d-DDH assumption. From the experimental analysis, it is found that our scheme shows better outcome than the compared work. The constraint of the current work affects the ciphertext length with increase in number of attributes. Hence, in future work an efficient CP-ABE will focus on alleviating this problem.

Data Availability

The data shall be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. F. Ransome, *Cloud Computing: Implementation, Management, and Security*, CRC Press, Boca Raton, Florida, 2017.
- [2] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," *ACM Transactions on Information and System Security*, vol. 9, no. 1, pp. 1–30, 2006.
- [3] M. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K. Fu, "March). Plutus: scalable secure file sharing on untrusted storage," *Fast Company*, vol. 3, pp. 29–42, 2003.
- [4] S. D. C. Di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati, "Over-encryption: management of access control evolution on outsourced data," in *Proceedings of the 33rd International Conference on Very Large Data Bases*, pp. 123–134, Vienna, Austria, 2007, September.
- [5] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *Proceedings of the 2007 IEEE Symposium on Security and Privacy (SP'07)*, pp. 321–334, IEEE, Berkeley, CA, USA, 2007, May.
- [6] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption," in *Advances in Cryptology - EUROCRYPT 2010*, pp. 62–91, Springer, Berlin, Heidelberg, 2010.
- [7] H. Deng, Q. Wu, B. Qin et al., "Ciphertext-policy hierarchical attribute-based encryption with short ciphertexts," *Information Sciences*, vol. 275, pp. 370–384, 2014.
- [8] X. Yan, H. Ni, Y. Liu, and D. Han, "Privacy-preserving multi-authority attribute-based encryption with dynamic policy updating in PHR," *Computer Science and Information Systems*, vol. 16, no. 3, pp. 831–847, 2019.
- [9] L. Aceto, I. Damgaard, L. A. Goldberg, M. M. Halldorsson, A. Ingolfsson, and I. Walukiewicz, *Automata, languages and programming: 35th international colloquium, ICALP 2008 Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part II*, Vol. 5126, Springer, Berlin, Heidelberg, 2008.
- [10] M. Green, S. Hohenberger, and B. Waters, "Outsourcing the decryption of abe ciphertexts," in *Proceedings of the USENIX security symposium*, vol. 2011, no. 3, San Diego, CA, USA, 2011, August.
- [11] W.-M. Li, X.-L. Li, Q.-Y. Wen, S. Zhang, and H. Zhang, "Flexible CP-ABE based access control on encrypted data for mobile users in hybrid cloud system," *Journal of Computer Science and Technology*, vol. 32, no. 5, pp. 974–990, 2017.
- [12] P. Zhang, Z. Chen, J. K. Liu, K. Liang, and H. Liu, "An efficient access control scheme with outsourcing capability and attribute update for fog computing," *Future Generation Computer Systems*, vol. 78, pp. 753–762, 2018.
- [13] C. Zuo, J. Shao, G. Wei, M. Xie, and M. Ji, "CCA-secure ABE with outsourced decryption for fog computing," *Future Generation Computer Systems*, vol. 78, pp. 730–738, 2018.
- [14] H. Zhong, Y. Zhou, Q. Zhang, Y. Xu, and J. Cui, "An efficient and outsourcing-supported attribute-based access control scheme for edge-enabled smart healthcare," *Future Generation Computer Systems*, vol. 115, pp. 486–496, 2021.
- [15] C. Chen, Z. Zhang, and D. Feng, "Efficient ciphertext policy attribute-based encryption with constant-size ciphertext and constant computation-cost," in *International Conference on Provable Security*, pp. 84–101, Springer, Berlin, Heidelberg, 2011.
- [16] Z. Zhou, D. Huang, and Z. Wang, "Efficient privacy-preserving ciphertext-policy attribute based-encryption and broadcast encryption," *IEEE Transactions on Computers*, vol. 64, no. 1, pp. 126–138, 2013.
- [17] F. Fuchun Guo, Y. Yi Mu, W. Susilo, D. S. Wong, and V. Varadharajan, "CP-ABE with constant-size keys for lightweight devices," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 763–771, 2014.
- [18] P. S. L. M. Barreto, H. Y. Kim, B. Lynn, and M. Scott, "Efficient algorithms for pairing-based cryptosystems," in *Advances in Cryptology - CRYPTO 2002*, pp. 354–369, Springer, Berlin, Heidelberg, 2002.
- [19] K. Lauter, "The advantages of elliptic curve cryptography for wireless security," *IEEE Wireless communications*, vol. 11, no. 1, pp. 62–67, 2004.
- [20] M. Zheng, Y. Xiang, and H. Zhou, "A strong provably secure IBE scheme without bilinear map," *Journal of Computer and System Sciences*, vol. 81, no. 1, pp. 125–131, 2015.

- [21] V. Odelu and A. K. Das, "Design of a new CP-ABE with constant-size secret keys for lightweight devices using elliptic curve cryptography," *Security and Communication Networks*, vol. 9, no. 17, pp. 4048–4059, 2016.
- [22] V. Odelu, A. K. Das, M. Khurram Khan, K.-K. R. Choo, and M. Jo, "Expressive CP-ABE scheme for mobile devices in IoT satisfying constant-size keys and ciphertexts," *IEEE Access*, vol. 5, pp. 3273–3283, 2017.
- [23] K. Sowjanya, M. Dasgupta, S. Ray, and M. S. Obaidat, "An efficient elliptic curve cryptography-based without pairing KPABE for Internet of Things," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2154–2163, 2019.
- [24] S. Ding, C. Li, and H. Li, "A novel efficient pairing-free CP-ABE based on elliptic curve cryptography for IoT," *IEEE Access*, vol. 6, pp. 27336–27345, 2018.
- [25] N. Raj and A. R. Pais, "CP-ABE scheme satisfying constant-size keys based on ECC," *ICETE*, vol. 2, pp. 535–540, 2020.
- [26] Y. Tian, T. Shao, and Z. Li, "An efficient scheme of cloud data assured deletion," *Mobile Networks and Applications*, vol. 26, pp. 1–12, 2020.
- [27] L. Cheung and C. Newport, "Provably secure ciphertext policy ABE," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, pp. 456–465, Alexandria, Virginia, USA, 2007, October.
- [28] B. Waters, "Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization," in *Public Key Cryptography - PKC 2011*, pp. 53–70, Springer, Berlin, Heidelberg, 2011.
- [29] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," in *Computational Intelligence and Neuroscience*, A. A. Abd El-Latif, Ed., vol. 2021pp. 1–11, 2021.
- [30] H. He, J. Zhang, J. Gu, Y. Hu, and F. Xu, "A fine-grained and lightweight data access control scheme for WSN-integrated cloud computing," *Cluster Computing*, vol. 20, no. 2, pp. 1457–1472, 2017.
- [31] P. Ratta, A. Kaur, S. Sharma, M. Shabaz, and G. Dhiman, "Application of blockchain and Internet of Things in healthcare and medical sector: applications, challenges, and future perspectives," in *Journal of Food Quality*, R. Khan, Ed., vol. 2021pp. 1–20, 2021.
- [32] J. Li, W. Yao, Y. Zhang, H. Qian, and J. Han, "Flexible and fine-grained attribute-based data storage in cloud computing," *IEEE Transactions on Services Computing*, vol. 10, no. 5, pp. 785–796, 2016.
- [33] J. Li, Y. Zhang, X. Chen, and Y. Xiang, "Secure attribute-based data sharing for resource-limited users in cloud computing," *Computers & Security*, vol. 72, pp. 1–12, 2018.
- [34] W. Teng, G. Yang, Y. Xiang, T. Zhang, and D. Wang, "Attribute-based access control with constant-size ciphertext in cloud computing," *IEEE Transactions on Cloud Computing*, vol. 5, no. 4, pp. 617–627, 2015.
- [35] F. Ajaz, M. Naseem, S. Sharma, M. Shabaz, and G. Dhiman, "COVID-19: challenges and its technological solutions using IoT," in *Current Medical Imaging Formerly: Current Medical Imaging Reviews* vol. 17, , 2021.
- [36] T. Naruse, M. Mohri, and Y. Shiraishi, "Provably secure attribute-based encryption with attribute revocation and grant function using proxy re-encryption and attribute key for updating," *Human-centric Computing and Information Sciences*, vol. 5, no. 1, pp. 1–13, 2015.
- [37] J. Li, W. Yao, J. Han, Y. Zhang, and J. Shen, "User collusion avoidance CP-ABE with efficient attribute revocation for cloud storage," *IEEE Systems Journal*, vol. 12, no. 2, pp. 1767–1777, 2017.
- [38] L. Xue, Y. Yu, Y. Li, M. H. Au, X. Du, and B. Yang, "Efficient attribute-based encryption with attribute revocation for assured data deletion," *Information Sciences*, vol. 479, pp. 640–650, 2019.
- [39] Y. Wang, F. Li, J. Xiong, B. Niu, and F. Shan, "Achieving lightweight and secure access control in multi-authority cloud," in *IEEE Trustcom/BigDataSE/ISPA* vol. 1, , pp. 459–466, IEEE, 2015.
- [40] M. Chase, "Multi-authority attribute based encryption," in *Theory of Cryptography Conference*, pp. 515–534, Springer, Berlin, Heidelberg, 2007.
- [41] S. Wang, J. Zhou, J. K. Liu, J. Yu, J. Chen, and W. Xie, "An efficient file hierarchy attribute-based encryption scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1265–1277, 2016.
- [42] X. Yao, Z. Chen, and Y. Tian, "A lightweight attribute-based encryption scheme for the Internet of Things," *Future Generation Computer Systems*, vol. 49, pp. 104–112, 2015.
- [43] A. K. Junejo and N. Komninos, "A lightweight Attribute-based security scheme for fog-enabled cyber physical systems," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 2145829, 18 pages, 2020.
- [44] R. Cheng, K. Wu, Y. Su, W. Li, W. Cui, and J. Tong, "An efficient ECC-based CP-ABE scheme for power IoT," *Processes*, vol. 9, no. 7, p. 1176, 2021.
- [45] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for IIoT devices," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, 2021.
- [46] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, vol. 14, no. 8, 2021.
- [47] H. Xiong, C. Jin, M. Alazab et al., "On the design of blockchain-based ECDSA with fault-tolerant batch verification protocol for blockchain-enabled IoMT," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [48] L. Zhang, Y. Zou, W. Wang, Z. Jin, Y. Su, and H. Chen, "Resource allocation and trust computing for blockchain-enabled edge computing system," *Computers & Security*, vol. 105, Article ID 102249, 2021.
- [49] D. Liu, Y. Zhang, W. Wang, K. Dev, and S. A. Khawaja, "Flexible data integrity checking with original data recovery in IoT-enabled maritime transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2021.

Research Article

Cognitive-Behavioral Therapy (CBT) Is Applied in Post-Traumatic Stress Disorder (PTSD) of Chinese Shidu Parents Who Lost Their Only Child

Guilin Yu ¹, Hongfeng Liu,^{1,2} Chiang-Hanisko Lenny,³ Daijun Chen,¹ Yin Yu,¹ and Chanyuan Sun¹

¹Hubei Province Key Laboratory of Occupational Hazard Identification and Control, Wuhan University of Science and Technology, Wuhan, China

²Suizhou Vocational and Technical College, Suizhou 441300, China

³Christine E. Lynn College of Nursing, Florida Atlantic University, Boca Raton, FL 33431, USA

Correspondence should be addressed to Guilin Yu; yuguilin46@163.com

Received 20 January 2022; Revised 7 February 2022; Accepted 10 February 2022; Published 8 March 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Guilin Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. The objective is to help Chinese Shidu parents who have lost their only child to relieve post-traumatic stress disorder. **Methods.** A qualitative phenomenology study using the hermeneutical phenomenological method was employed in a major metropolitan city in China. Participants were 46 parents who had lost their only child and suffered from post-traumatic stress disorder. Three major themes are: (1) to conduct semistructured in-depth interviews with 46 Shidu parents; (2) to develop and implement a psychological intervention program combining with group intervention and individual counseling based on cognitive-behavioral therapy; (3) to assess the effect of psychological intervention through PTSD Checklist for DSM-5 (PCL-5). **Results.** Those Shidu parents who lost their only child got great relief from PTSD. The Shidu parents got great relief and changed a lot after the intervention. They became active to participate in group counseling and willing to help others because they experienced the value and fun in the process. They rebuilt their attachment and looked forward to their future life. **Conclusion.** Cognitive-behavioral therapy can alleviate and even cure post-traumatic stress disorder in Shidu parents who have lost their only child.

1. Introduction

The Shidu parents refer to the parents who have lost their only child. The death of the only child is a disaster for any one-child family. Since then, those who lost their only child had no spiritual sustenance, lost their life goals and pursuits. In the first few months, years, or even decades, they felt intense painful experiences and extreme behavioral reactions. As a result, they became depressed, autistic, sensitive, and vulnerable, and even retaliated against themselves or society because of their irrational cognition [1]. According to the research done by Liu [2] and other researchers, the PTSD rate of Shidu parents ranged from 69.5% to 80%, and they generally avoided society.

The main symptoms of post-traumatic stress disorder (PTSD) [3] include repeated flashbacks of painful experiences and painful memories, avoidance of people and things related to the loss of original injury, increased alertness, and borderline paranoid personality. Shidu parents with PTSD cannot extricate themselves from grief, guilt, anger, and struggle all day long, which eventually pose a serious threat and heavy burden to themselves and society. A review of literature in Chinese and foreign languages in recent years reveals that there have been many crossregional studies on PTSD with large samples and large data, focusing on the correlation of symptoms, factors affecting the quality of life, and the moderating effect of social support, etc., but there are very few studies on psychological intervention. The

researchers collected through the early stage of the data and found that cognitive-behavioral therapy was the standard therapy for post-traumatic stress disorder. [4] Cognitive-behavioral therapy (CBT) is a structured, short-term, cognitive-oriented psychotherapy developed by A. T. Beck in the 1960s. It is mainly aimed at psychological disorders such as depression, anxiety, post-traumatic stress disorder, and psychological problems caused by unreasonable cognition. A. T. Beck once said that maladaptive behaviors and emotions were both from maladaptive cognition. Cognitive-behavioral therapy can change the maladaptive emotions and behaviors of Shidu patients by changing their wrong cognition of themselves, people, or things and finally achieve the purpose of curing post-traumatic stress disorder. Considering the particularity of Shidu parents, this study adopts a qualitative research method and uses cognitive-behavioral therapy to conduct a psychological intervention for post-traumatic stress disorder of those Shidu parents, so as to provide the scientific basis for the government and social organizations to carry out relevant psychological work.

2. Experimental Details

2.1. Object. This study was carried out with the support of a community health and family planning department, community grid members, social volunteers, and psychological experts in Wuhan. In the family activity center of a community in Wuhan, we held health lectures and reading meetings every Thursday morning to build a close relationship with the Shidu parents. We finally identified 46 participants suffering from post-traumatic stress disorder (PTSD) from 190 Shidu parents who lost their only child through group discussion and questionnaire survey (Table 1). The participants were in a good mental state and have good expression ability, and they all signed informed consent.

3. Methods

3.1. PTSD Checklist for DSM-5 (PCL-5) [5]. By using the scale to conduct a questionnaire survey, 33 points were used as the positive demarcation score to screen out the patients with post-traumatic stress disorder; the scale was used to evaluate the effect of psychological intervention in the process of psychological intervention. The PTSD score decreased, indicating that the intervention was effective. When the PTSD score decreased to less than 33 points, the patients with post-traumatic stress disorder were cured.

3.2. Data Collection and Analysis—Interpretative Phenomenological Analysis. Upon receiving the University IRB approval, the data were collected by trained researchers through semistructured in-depth interviews. The data collected were tape-recorded with the participants' consent, and notes were taken to enable the precision of statements. Each face-to-face, tape-recorded interview lasted 2 to 3 hours. And the information was obtained with PTSD Checklist for DSM-5 (PCL-5) as the topic.

3.3. Implementation of Cognitive-Behavioral Therapy. A psychological intervention program combining individual counseling and group intervention based on cognitive-behavioral therapy was developed and gradually implemented. During the intervention period, PTSD scores were evaluated every 3 months, and the program was modified according to the evaluation results before implementation. The frequency of treatment was once a week in the first month, followed by once every two weeks for a total period of a year.

3.4. Case Counseling—Cognitive-Behavioral Therapy (CBT) Helps Shidu Parents to Accept Themselves. First, cognitive-behavioral therapy was applied to help Shidu parents to accept reality and to restore the self. Through professional psychological counseling, Shidu parents were encouraged, and they clearly realized that the child has passed away. Instead of denying reality, they should strive to live a strong and valuable life. The application of cognitive-behavioral therapy helped Shidu parents get rid of the pain of losing their only child and pay more attention to the personal value realization.

Second, cognitive-behavioral therapy (CBT) was applied to help Shidu parents to correct irrational cognition and rebuild the cognitive structure. (1) Explain the significance and the implementation steps of cognitive-behavioral therapy. (2) Find out the irrational cognition of the Shidu parents by talking with them. (3) Guide Shidu parents to discover their irrational cognition, help them to realize that irrational cognition will lead to emotional and behavioral obstacles, which will eventually affect their normal life, inspiring the determination to correct irrational cognition. (4) By using the techniques of arguing with irrational cognition, imagining, and relaxation training, we help Shidu parents to correct their irrational cognition, establish correct and rational cognition, and improve their cognitive, emotional, and behavioral self-management abilities. (5) Summarize the examples of good changes in Shidu parents' lives after the intervention, and we further strengthen the significance of establishing rational cognition.

Third, the application of cognitive-behavioral therapy helped Shidu parents learn to self-heal: relax the body, release, and transfer bad emotions. Shidu parents were guided to master relaxation training methods such as breathing, muscles, and imagination so that they could adjust themselves at any time and control their own physiological and psychological changes independently.

3.5. Group Intervention—Cognitive-Behavioral Therapy (CBT) to Help the Shidu Parents Rebuild Their Attachment Relationship. First, Shidu parents often identify with other parents who lost their only child as Tong Ming Ren, which means "those who share the same destiny" [6]. Shidu parents believe that only in groups they can better face the pressures and challenges of life. In the group, Shidu parents can get psychological comfort, support each other, and stay together for warmth [7]. The researchers encouraged Shidu parents to

TABLE 1: Data of research subjects.

| Characteristics | | N | Percentage (%) |
|--------------------------------------|----------------------------|----|----------------|
| Gender | Male | 21 | 45.7 |
| | Female | 25 | 54.3 |
| Age | 50–60 | 12 | 26.1 |
| | 60–70 | 19 | 41.3 |
| | ≥70 | 15 | 32.6 |
| Marital status | Couple | 29 | 63.04 |
| | Single | 17 | 36.96 |
| Education | Middle school or under | 26 | 56.5 |
| | Senior high school | 11 | 23.9 |
| | Undergraduate or above | 9 | 19.6 |
| Annual household income (yuan/month) | ≤2000 | 25 | 54.3 |
| | 2000–5000 | 14 | 30.4 |
| | ≥5000 | 7 | 15.3 |
| Residence status | Solitude | 15 | 32.6 |
| | Live with family/relatives | 31 | 67.4 |
| Self-care ability | Fully care | 30 | 65.2 |
| | Barely provide for oneself | 15 | 32.6 |
| | Cannot provide for oneself | 1 | 2.2 |
| Time since the child's death | 6 months~1 year | 2 | 4.3 |
| | 1 year~5 years | 18 | 39.1 |
| | 5 years~ | 26 | 56.6 |
| Cause of death | Accident | 26 | 56.5 |
| | Disease | 19 | 41.3 |
| | Missing | 1 | 2.2 |

get familiar with each other and participate in group activities as soon as possible through health lectures, readings, and games.

Second, spring and autumn outings, farmhouse entertainment, festival celebrations, and other activities were organized to help those Shidu parents gradually get out of the group and contact the society, enhance their interpersonal communication skills, feel the joy of life, and thus overcome their social avoidance behaviors and mentality.

Third, the Shidu parents are guided to participate in the activities of love, such as accompanying, playing with the children in the orphanage, holding Thanksgiving education lectures in the young labor camps, establishing “One Helps One” psychological intervention to link with university students, and motivate Shidu parents’ emotion as fathers and mothers as well as the responsibility to society. By feeling love and being loved at the same time, they could repair the heart ruins and rebuild benign self.

Fourth, expert lectures on the meaning of life should be carried out to encourage those Shidu parents to discuss their ideal of life and gradually guide them to transfer their attachment to other more positive and lasting aspects, such as their own health, interests, and hobbies, and helping others.

Reassessment of PTSD scores is carried out.

4. Results

After psychological intervention through cognitive-behavioral therapy, the psychological state of the Shidu parents has undergone a great change.

Shidu parents accept the fact that their child has left and keep connected with their child in a positive way so that their grief can be greatly relieved.

We applied cognitive-behavioral therapy for case studies, such as subject P1 kept saying, “My son, I miss him, I do not know what to do, I’m so broken. . .” Only by helping P1 unload her psychological burden could she start a new life. The researchers performed the true story of the child’s death in the form of “psychodrama”, allowing the Shidu parents to naturally vent their emotions, role-playing their children, and guiding them to say the words they have been hiding in their mind, including guilt and missing, to help them formally bid farewell to their children. Finally, subject P1 went to the child’s grave to say goodbye and said to the child, “Son, you go. . . ., you go to study, you go to serve the country, you go busy with your own things, I will take care of myself, I also want to live my own life”. Then, she burned all the child’s relics. Subject P1 said: “Since I said goodbye to my baby, I have felt a lot lighter and I do not have to be sad all the

time. Now I sing when I have free time. I love singing.” “I still miss my son. When I miss him, I sing songs of longing until tears run out of my eyes. After singing, I feel very happy.”

Shidu parents should establish correct and positive cognitive concepts, learn to self-heal, be able to control their own emotions and behaviors, and accept the present self-more.

Irrational cognition leads to negative emotions and destructive behaviors, which will become a major source of stress over time. For example, subject P4 said before: “The death of children is caused by myself, and I am a disaster, so I should live an unhappy life”. Subject P5 said: “I always feel sick. . . .”; subject P16 thought: “People nowadays are snobbish and deliberately look down on me. . . .”. Through cognitive-behavioral therapy, the Shidu parents could correct the irrational cognition of stress and reconstruct the positive and adaptive cognitive structure. Under the guidance of correct cognition, they were more positive and optimistic in mood, and their behaviors were more conducive to their own development. In the process of psychological intervention, self-healing methods were taught to the research subjects to help those Shidu parents control bad emotions and relax their body and mind. After the intervention, subject P4 said: “I rarely blame myself, even if occasionally remorse, I will quickly transfer attention. People’s life may be long or short, but only a few decades of time. My poor son left before me, but I also must go one day. There is really nothing to blame and care about. I want to live a good life, which is what my son wants me to do.” Subject P5 said: “When I’m not feeling well, I just take a deep breath or go outside to relax and feel comfortable.” Subject P16 mentioned: “Don’t care too much about other people’s eyes, my husband is good to me on the line, I’ve been working hard for most of my life, and I will love myself for the rest of my life. I will praise and hug myself every day.”

Participate actively in group activities, enjoy helping others, and experience value and fun from these activities.

By the early group intervention of cognitive behavior therapy, the relationship between the researchers and participants became increasingly close. The participants have also mentioned that participating in the outdoor activities made them relaxed and joyful and helping others manifested their own value, and Shidu parents were then no longer obsessed with the pain of losing their only child and were willing to put the limited energy into pleasure and value. Subject P18 said: “I’m older and I’ve got rheumatoid arthritis, but I feel a lot better and my pain get relieved when I come out and gather with people”; subject P22 said: “People of my generation have suffered hardships and are dedicated. We are willing to offer help to those in need.” Subject P25 mentioned: “I am a member of the Communist Party of China, I was eager to work overtime in the unit in the past time. Now that I am old, I need to spend more time taking care of myself, but if I am needed, I am willing to help others as long as my health is available.” Subject P31 established a connection with a student in a university in Wuhan who was mentally reduced and physically limited after brain tumor surgery. Subject P31 always encouraged the students and their

parents. Subject P31 said: “This child, I like at first sight, He is about the same age as my son, but he is also very poor, I want to encourage him, to help him fight the disease.”

Reconstruct attachment relationship [8] and be full of longing for future life.

The Shidu parents have the right connection with their children in the early stage, and they can face themselves and others positively and experience the fun and value from these things. After that, they hope to establish a more lasting and positive attachment relationship. Through expert lectures and group intervention discussions, the participants can find their personal meaning of life. Subject P35 said: “I had the idea of writing a book when I was young, and now I have the time to do it.” Subject P37 said: “I love dancing, I would like to take you to the square dance, I hope to have the opportunity to participate in the square dance competition.” Subject P39 said: “I want to travel while I am still healthy, I want to visit the motherland’s beautiful mountains and rivers, visit the good scenery at home and abroad.” Subject P40: “My father is 100 years old, my aunt is 97 years old, and I’m 71 years old now, I’m trying to live to be 108 years old.” Subject P43 said: “I like playing online games, watching TV, traveling and shopping. There are many interesting things waiting for me.” Subject P46 said: “I will keep doing Tai chi every day, waiting for my grandson to get married, and then I will have great grandchildren.”

Post-traumatic Stress Disorder was alleviated in the Shidu parents (Table 2).

With the deepening intervention of cognitive-behavioral therapy, the scores of post-traumatic stress disorder of Shidu parents who lost their only child gradually decreased, and some of them had dropped to less than 33 points, indicating that the post-traumatic stress disorder of the Shidu parents had been cured. There are still two Shidu parents whose scores were over 33 points, but compared to the original score, the decline is obvious.

5. Discussion

5.1. Necessity and Importance of Psychological Intervention for the Psychological Trauma of the Shidu Parents. This study found that although the loss of the original trauma occurred for several years or more, some of Shidu parents still suffered from PTSD. In this way, these Shidu parents should be carried out psychological intervention after the loss of their only child. However, all participants reported no prior psychological intervention from any organization or individual, and all of the participants reported that they would get over their grief sooner and faster if someone had been involved with them, especially in the immediate aftermath of their child’s death. At present, the psychological intervention on the psychological trauma of the Shidu parents has not been caused widespread attention, and there is a big flaw and insufficiency. It needs to call on the government, social organizations, and professionals in the psychological field to focus on the mental health of Shidu parents, using scientific and systematic psychological therapy providing early psychological intervention for the Shidu parents. These

TABLE 2: PTSD scores of subjects before and after cognitive-behavioral therapy intervention (unit: points).

| Subjects | Initial scores | 1 year after the intervention | <i>t</i> | <i>p</i> |
|--------------------------------|----------------|-------------------------------|----------|----------|
| PTSD scores | 43.11 ± 13.42 | 14.15 ± 5.25 | 13.783 | 0.000** |
| Intrusive symptoms | 11.13 ± 4.22 | 4.15 ± 1.83 | 9.441 | 0.000** |
| Withdrawal symptoms | 3.93 ± 1.61 | 1.89 ± 1.45 | 6.511 | 0.000** |
| Cognition and negative emotion | 16 ± 5.39 | 3.15 ± 2.37 | 14.669 | 0.000** |
| Increased alertness | 12.04 ± 3.93 | 4.96 ± 2.98 | 11.010 | 0.000** |

* $p < 0.05$, ** $p < 0.01$.

measures can help Shidu parents to relieve the pain of original injury, early resumption of ourselves, and return to society. The purpose of this study is to explore the scientific and effective psychological intervention methods for post-traumatic stress disorder of the Shidu parents who have lost their only child.

5.2. Cognitive-Behavioral Therapy can Alleviate Post-traumatic Stress Disorder in Shidu Parents Who Have Lost Their Only Child. After 1 year of psychological intervention with cognitive-behavioral therapy, all of the participants have accepted the reality, rebuilt their attachment relationships, and actively faced life, indicating that the post-traumatic stress disorder of the Shidu parents has been significantly relieved or even cured in some cases. Two of them still had obvious symptoms of post-traumatic stress disorder. The possible factors were as follows: first, both of them were female, with relatively sensitive and fragile character, especially subject P27 suffered from an anxiety disorder when she was young. Second, they are very close to their children and regard them as the only ideal in their life. Third, they place high hopes on their children's path to success, but their children die shortly after becoming successful, and it is a huge blow. Although they did not get rid of PTSD, their previous major symptoms such as depression and withdrawal were significantly alleviated. Currently, only mild guilt and insomnia symptoms remain, which should be alleviated with further intervention. Most studies confirmed that the cognitive-behavioral therapy intervention on depression [9], sorrow [10], anxiety [11], psychological problems such as post-traumatic stress disorder [12, 13] was effective. Moreover, it has many advantages, such as low requirements on therapists, convenient and easy implementation, and extensive coverage, so the cognitive-behavioral therapy on Shidu parents with post-traumatic stress disorder has extensive application prospects in China.

5.3. Limitations and Prospects of Cognitive-Behavioral Therapy in the Process of Psychological Intervention. Cognitive-behavioral therapy has developed into a very systematic and mature therapy, which can directly solve existing psychological problems and correct dysfunctional thoughts and behaviors, and has been successfully applied to completely different psychiatric patients and populations. [14]. The participants of this study were persuaded to voluntarily accept psychological intervention, and their psychological status was relatively stable,

and their mentality was more positive than that of the ordinary Shidu parents. The research process of cognitive-behavioral therapy for the post-traumatic stress disorder of the Shidu parents was very smooth, and the post-traumatic stress disorder of the participants was relieved to a certain extent. However, we should not assume that this type of psychotherapy is appropriate for all Shidu parents with PTSD. In China, there are many more severe cases of post-traumatic stress disorder among Shidu parents, and they cannot get over the grief of losing their only child, lose hope for life, and refuse all help [8]. It is difficult for researchers to intervene in the Shidu parents with such strong resistance to psychotherapy. There are also those Shidu parents with extreme thoughts and paranoid terror. Because they complain about injustice, hate others, or even revenge against society, they already have serious personality disorders. In this case, researchers could try a combination of CBT therapy, drugs, and electric shocks [15, 16]. In conclusion, CBT therapy has many advantages, such as low requirements for therapists, convenient implementation, and wide range of involvement. Therefore, CBT therapy has a wide application prospect in the intervention of post-traumatic stress disorder in parents who lost their only child in China.

6. Conclusion

To sum up, cognitive-behavioral therapy can alleviate the post-traumatic stress disorder of Shidu parents, and the combination of group intervention and individual counseling is very practical and effective, which is worth popularizing.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank all those who have helped in the course of writing this paper. This work was supported by the National Social Science Foundation of China (grant no. 17BSH119).

References

- [1] B. Liu, "Formation of borderline personality disorder in shidu patients and Transference focus therapy," *Journal of Henan Institute of Education (Philosophy and Social Science Edition)*, vol. 39, no. 04, pp. 89–94, 2020.
- [2] B. Liu, "Research on post-traumatic stress disorder and its rescue mechanism in shidu loss group," *Journal of Xinyang Normal University (Natural Science Edition)*, vol. 40, no. 05, p. 2, 2020.
- [3] Eli Buzohre, Y. Zhou, Y. Liang et al., "A profile analysis of post-traumatic stress disorder and depressive symptoms among Chinese Shidu parents," *European Journal of Psychotraumatology*, vol. 11, no. 1, p. 1766770, 2020.
- [4] Z. Zhu, F. Li, G. Zhou et al., "Research progress on mental health and psychological support of families of the shidu parents," *Chinese Journal of Health Psychology*, vol. 27, no. 06, pp. 954–957, 2019.
- [5] M. Deng, "New advances in clinical research of post-traumatic stress disorder (DSM-5 new standard)," *Chinese Journal of Health Psychology*, vol. 24, no. 05, pp. 641–650, 2016.
- [6] Y. Zheng and T. R. Lawson, "Identity reconstruction as shiduers: narratives from Chinese older adults who lost their only child," *International Journal of Social Welfare*, vol. 24, no. 4, pp. 399–406, 2014.
- [7] S. Fang, "Spiritual community and Double introversion: a Study on the Construction of self-organization of the Elderly who lost their only child in China -- Based on the Comparative Analysis of the self-organization of four elderly who lost their only child," *Theory Monthly*, no. 05, pp. 174–181, 2018.
- [8] D. Zalaznik, M. Weiss, and J. D. Huppert, "Improvement in adult anxious and avoidant attachment during cognitive behavioral therapy for panic disorder," *Psychotherapy Research*, pp. 1–17, 2019.
- [9] Z. Huang, Q. Han, Lu Yian et al., "Effects of cognitive behavioral therapy on insomnia and depression in patients with insomnia and depression," *Nursing Research*, vol. 35, no. 01, pp. 80–85, 2021.
- [10] R. A. Bryant, L. Kenny, J. Amy et al., "Predictors of treatment response for cognitive behaviour therapy for prolonged grief disorder," *European Journal of Psychotraumatology*, vol. 8, no. sup6, 2019.
- [11] C. Yuan, "Application of cognitive behavioral therapy in test anxiety intervention," *Journal of Primary and Secondary School Mental Health Education*, vol. 10, pp. 53–56, 2021.
- [12] H. Martine and A. L. Méli ssande, "Latent class analysis of post-traumatic stress symptoms and complex PTSD in child victims of sexual abuse and their response to Trauma-Focused Cognitive Behavioural Therapy," *European Journal of Psychotraumatology*, vol. 11, no. 1, 2020.
- [13] S. Åkerblom, S. Perrin, M. Rivano Fischer, and M. McC. Lance, "Treatment outcomes in group-based cognitive behavioural therapy for chronic pain: an examination of PTSD symptoms," *European Journal of Pain*, vol. 24, no. 4, 2020.
- [14] K. Lee Raby and M. Dozier, "Attachment across the lifespan: insights from adoptive families," *Current Opinion in Psychology*, vol. 25, pp. 81–85, 2019.
- [15] H. Qi, S. Jufeng, and Ma Jun, "Cognitive behavioral therapy combined with repeated transcranial magnetic stimulation improves sleep and anxiety and depression in patients with stroke," *Neural Injury and Functional Reconstruction*, vol. 16, no. 11, pp. 669–671, 2021.
- [16] H. Haixiao, L. Cuilv, and Z. Lijun, "Application of handwork combined with Network cognitive behavioral therapy in rehabilitation of patients with mental disorders and hypertension," *Guangdong medical journal*, vol. 41, no. 22, pp. 2343–2347, 2020.

Research Article

Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal

M. M. Venkata Chalapathi ¹, M. Rudra Kumar ², Neeraj Sharma,¹ and S. Shitharth ³

¹School of Engineering, Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore, Bhopal, India

²Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India

³Department of Computer Science and Engineering, Kebri Dehar University, Kebri Dehar 001, Ethiopia

Correspondence should be addressed to S. Shitharth; shitharths@kdu.edu.et

Received 13 January 2022; Revised 2 February 2022; Accepted 7 February 2022; Published 28 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 M. M. Venkata Chalapathi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the recent past, handling the high dimensionality demonstrated in the auditory features of speech signals has been a primary focus for machine learning (ML-)based emotion recognition. The incorporation of high-dimensional characteristics in training datasets in the learning phase of ML models influences contemporary approaches to emotion prediction with significant false alerting. The curse of the excessive dimensionality of the training corpus is addressed in the majority of contemporary models. Modern models, on the other hand, place a greater emphasis on merging many classifiers, which can only increase emotion recognition accuracy even when the training corpus contains high-dimensional data points. “Ensemble Learning by High-Dimensional Acoustic Features (EL-HDAF)” is an innovative ensemble model that leverages the diversity assessment of feature values spanned over diversified classes to recommend the best features. Furthermore, the proposed technique employs a one-of-a-kind clustering process to limit the impact of high-dimensional feature values. The experimental inquiry evaluates and compares emotion forecasting using spoken audio data to current methods that use machine learning for emotion recognition. Fourfold cross-validation is used for performance analysis with the standard data corpus.

1. Introduction

Emotions have a profound influence on the physical and psychological well-being in humans. How well patients convey their emotions and how well their therapists recognize and respond to them determine improvement in therapeutic settings. [1] Therapists must deal with enormous volumes of data over a lengthy period of time, which is difficult because they must see numerous patients throughout that time. A platform that can give meaningful speech-based emotion identification insights, for example, might be useful in therapy sessions. EmoViz allows users to take voice samples and use a neural network to determine emotional feelings (such as joyful, sad, angry, surprised, or neutral). Emotional information may be

obtained through the examination of spoken audio signals without the need of intrusive technology such as facial recognition or internal signal-based physiological sensor data. Users may view how their emotions have evolved over time and how they have grouped audio and texts based on their emotions using the application EmoViz. [2] Emotion is important in everyday interpersonal connections and is seen as a necessary skill for human communication. [2] It facilitates communication by expressing emotions and responding to individuals being communicated with. Many cognitive and affective computing tasks, such as rational decision-making, perception, and learning, benefit from emotional input. As intelligent systems grow more ubiquitous, emotion identification is becoming increasingly crucial. [3].

Computer games, banking, call centers, video monitoring, and psychiatric diagnosis are just a few examples of real-world applications for emotion detection systems. Other practical applications for emotion detection systems include online learning, business applications, clinical investigations, and entertainment [4, 5]. Voice signals incorporate emotions when it comes to the creation of intelligent systems known as “emotion recognition from speech.” Because of a host of intrinsic socio-economic benefits, speech signals are a great source for emotional computing. Because of their inexpensive cost, they are more appealing for speech emotion recognition research than other physiological signals such as electroencephalograms, electrooculograms, and electrocardiograms [6].

Despite modest development, the accuracy of this approach in identifying fear is lower than for other emotions [7, 8]. When Semwal and colleagues [8] integrated fundamental frequency, ZCR (zero-crossing rate), MFCC, and energy, they were able to identify fear with a 77 percent accuracy. Sun et al. [9] revealed that a deep learning neural network model identified bottleneck information with an accuracy of 62.50 percent in detecting fear.

1.1. Motivation. A number of processes are utilized by machine learning techniques to obtain a collection of speech features that may be used to properly categorize emotions. To build appropriate emotion recognition from a speech system, a suitable collection of characteristics must be chosen from which to train the selected learning algorithm. Emotion recognition algorithms mainly rely on features extracted from spoken audio signals [3, 10]; however, identifying an appropriate feature set is challenging [11]. Speech emotion recognition is challenging for a variety of reasons, including an imperfect description of an emotion and the blurring of the boundaries between distinct emotions. Emotion identification from speech is being improved by introducing new aspects, as demonstrated in [12], with an accuracy of 91.75 percent on an acting corpus when employing PLP characteristics. This accuracy is rather low when compared to the 95.20 percent accuracy attained for the synthesis of acoustic characteristics focusing on MFCC and pitch for recognizing speech emotion. Some studies have sought to agglutinate numerous auditory characteristics to increase the accuracy and precision of speech emotion identification [7, 8].

1.2. Problem Statement. “Ensemble learning” refers to the process of combining various learning models with the goal of producing a better learner as a result. Such algorithms are used in a variety of fields, including medical investigations [13] and dialect prediction [14]. Bagging [15] and boosting [16] are two of the most common ensemble approaches. In terms of accuracy, ensembles of core estimation methods have been shown to outperform single hypotheses [17]. Quinlan [18] tested boosting and

bagging ensemble models on a variety of datasets and found them to be remarkably effective. Bagging, as the name implies, aims to train several estimators on random subsets of the dataset. If the training samples are drawn with replacement, they are referred to as “bootstrap samples.” Ensemble methods were also used to analyse audio data. Schuller et al. [19] investigated ensemble learning methods for recognizing speaker emotion through speech and found an increase in the accuracy of movie content. Morrison et al. [20] combined several classifiers for emotion recognition tasks in call center practices using an unweighted vote method. However, the majority of the contributions indeed are limited to opt the classification decision delivered by the majority of classifiers used in the ensemble of diversified classifiers. The crux of high-dimensional features remains the same. Hence, rather than the ensemble of diversified classifiers, the focus shall be on handling the high dimensionality of the features.

1.3. Organisation of the Manuscript. This paper’s structure includes an introduction to the previously stated ensemble learning by high-dimensional acoustic features for emotion recognition from speech audio signals. In Section 2, we look at related work and numerous models for emotion recognition from speech audio signals. Section 3 covers the methods and materials connected to the suggested model. In Section 4, experimental research is conducted, and the proposed model is compared to other modern models. The conclusion of this article is explained in Section 5, followed by references.

2. Related Work

There have been a few studies on support vector machine ensemble learning [21]. Hu et al. [22] used such an ensemble to solve the problem of rotating machinery failure detection. However, studies of this nature are few and far between.

Bhavan et al. [23] used a bagged ensemble approach on the Emo-DB and achieved a prediction accuracy of 92.45 percent. Shegokar and Sircar [24] proposed a CWT with prosodic elements for recognizing emotion in speech audio signals. Using PCA feature transformation and SVM with quadratic kernel as a classification approach, they achieved an overall accuracy of 60.1 percent on the RAVDESS database. The EMD (empirical mode decomposition) method, which uses the reconstructed signal’s optimal features, was used to analyse reconstructed speech signals. On the Spanish database, they were able to achieve an average classification accuracy of 91.16 percent using the RNN technique.

As stated in the introduction, there are numerous reasons why emotion identification remains a major challenge. There is a disconnect between acoustic qualities and human emotions, as well as a theoretical framework for linking voice characteristics to a speaker’s emotional state [10, 24–26]. Because of these underlying difficulties, there is

disagreement in the research about which elements are better for recognizing emotion recognition. [10, 26]. When several different types of auditory characteristics are combined, researchers have shown promising results in speech emotion identification [10, 26–29]. They have, however, struggled to find a way to combine the various elements in a way that is both effective and efficient. The study [3, 10, 27] emphasises the importance of identifying appropriate features in order to improve the stability of speech emotion recognition systems. Researchers frequently use specialist software to simplify the extraction, selection, and unification of speech features across multiple sources. Diverse learning algorithms for speaker emotion recognition have been demonstrated to be learned and verified using specific features extracted from public databases.

Multiple neural networks are fused together to achieve the goal of increasing the recognition efficiency from multiple perspectives. When a trained model is applied to an unprepared platform, gradient disappearance and overfitting can easily occur. The ability to generalise is crucial in speech emotion recognition. Ensemble learning has a number of advantages, including the ability to generalise and parallelism. The accuracy and reliability of each individual expert are crucial in ensemble learning [30, 31].

The use of ensemble learning and traditional machine learning approaches in speech emotion recognition has recently increased [32]. Weighted sum fusion was used by Mao et al. [33] to demonstrate that separating complex language features from emotional aspects in speech improves the recognition rate. Liu et al. used a variety of emotional feature subsets to train subclassifiers, which were then used to create a decision-making layer fusion, resulting in improved recognition results. Existing ensemble learning relies heavily on expert credibility allocation, which is a significant flaw in the system. In contrast, the data root for the initial decision is speech features, and acquisition methods are limited, resulting in slight variations across samples and inaccurate grouping information [34, 35].

On this basis, ensemble learning models can be used to make more stable decisions by combining multiple models. On the other hand, each expert’s credibility is updated online based on their accuracy rate. Both generalisation and recognition of speech emotions have improved [36].

The most recent attempt to conduct ensemble learning by fusing together diverse categorization strategies was HAF [37], which combined various classification algorithms. Despite the model’s superior performance, the high dimensionality of the training corpus remains a problem. This contribution depicts an ensemble learning model for clustering the speech audio signals of the dataset used as input to the training phase to mitigate the negative impact of the high-dimensional features. The suggested method uses the distribution diversity of feature values spanned over different records of divergent emotions to determine the best aspects. The proposed model is motivated by the previously described model titled “Speech Emotion Recognition Using Supervised Bayes Learning on Digital Features (SBL-DF)” [38]. The SBL-DF, on the other hand, does not address the negative impact of high-dimensional features.

3. Methods and Materials

This section explores the materials and methods used in the proposed model that enables to predict emotions in speech audio signals. The materials and methods explored in this section are centric to handle the adverse impact of high-dimensional features towards emotion prediction, feature extraction, feature optimization, and ensemble classification using the adaptive boosting technique as represented in Figure 1. The detailed description of these materials and methods is explored in following sections.

3.1. Dimensionality Reduction. The Fuzzy C-Means [39] clustering technique has been employed to handle the high dimensionality of the training corpus that leads to low sensitivity and specificity, which causes intolerable false-alarming.

The FC-Means method divides the input data $\{r_i \mid \exists r_i \in tC \wedge 1 \leq i \leq |tC|\}$ into clusters, with each cluster retaining a group of records with a substantial association. Concerning this:

Take the records randomly as centroids and perform fuzzy clustering using Fuzzy C-Means, such that one or more records would be in more than one cluster.

Find the optimal centroids of the resultant clusters and perform the clustering of records recurrently till there is no change in the centroids.

The records that may settle in more than one cluster can be scaled for their relationship by measuring their distance from the centroids of the corresponding clusters having those records.

The algorithm works by distributing membership to each record, resulting in each cluster centroid being proportional to the related format of the distance between each record and the corresponding centroid. The closer the data is to the cluster’s centroid, the closer their membership is to a specific core of the cluster. Following the membership iteration, the cluster’s centroid shall be revised using the following formulas:

$$\bigvee_{j=1}^{|C|} \left\{ \mu_{ij} = \left[\sum_{k=1}^{|tC|} \left(\frac{|c_j \cap r_{ik}|}{|c_j|} \right)^{(2/f_i-1)} \right]^{-1} \right\}, \quad (1)$$

$$\bigvee_{j=1}^{|C|} \left\{ c_j = \left[\frac{\left(\sum_{i=1}^{|tC|} (\mu_{ij})^{f_i} * |r_i| \right)}{\left(\sum_{i=1}^{|tC|} (\mu_{ij})^{f_i} \right)} \right] \right\}. \quad (2)$$

The number of records representing the record is indicated by the notation $|tC|$. The notation c_j reflects the record having aspect with the highest support towards the j^{th} cluster, while the notation $f_i \in [1, \infty]$ reveals index fuzziness. Centroids are indicated by the set $C = \{c_1, c_2, \dots, c_{|C|}\}$. The notation μ_{ij} denotes the Euclidian distance of the i^{th} record of the record $\{r_i \mid \exists r_i \in tC \wedge 1 \leq i \leq |tC|\}$ towards the current centroid c_j of the j^{th} cluster. The depiction represents the Euclidean distance between the j^{th} cluster centroid and the records of



FIGURE 1: Data flow diagram of the model.

record $\{r_i \exists r_i \in tC \wedge 1 \leq i \leq |tC|\}$. This Fuzzy C-Means main algorithm's purpose is fading:

$$J(U, V) = \sum_{i=1}^{|tC|} \sum_{j=1}^{|C|} \left\{ (\mu_{ij})^{f_i} \left\| \frac{c_j \cap r_i}{|c_j|} \right\|^2 \exists i \leq |tC| \wedge j \leq |C| \right\}. \quad (3)$$

$|c_j \cap r_i|/|c_j|$ // is the Euclidean distance of the j^{th} cluster centroid c_j as well as the i^{th} record $\{r_i \exists r_i \in tC \wedge 1 \leq i \leq |tC|\}$.

The steps involved in Fuzzy C-Means clustering are as follows:

(i) The set $tC = \{r_1, r_2, \dots, r_i, r_{i+1}, \dots, r_{|tC|-1}, r_{|tC|}\}$ represents a set of records such that each record $\{r_i \exists r_i \in tC \wedge 1 \leq i \leq |tC|\}$ represents the record, whereas the notation $C = \{c_1, c_2, \dots, c_{|C|}\}$ indicates set of centroids of all clusters.

- (1) The cluster centroid c_j of the j^{th} cluster has been selected randomly.
- (2) The fuzzy membership μ_{ij} has been computed by utilizing

$$\mu_{ij} = \frac{1}{\sum_{m=1}^{|tC|} \left(|c_j \cap r_{im}|/|c_j| \right)^{(2/f_i-1)}}. \quad (4)$$

- (3) Here, the fuzzy centroid v_j has been measured by utilizing

$$c_j = \left\{ \frac{\left(\sum_{i=1}^{|tC|} (\mu_{ij})^{f_i} * |r_i| \right)}{\left(\sum_{i=1}^{|tC|} (\mu_{ij})^{f_i} \right)} \right\}. \quad (5)$$

- (4) The afore two steps (2&3) are recurrent till the condition $\beta > \|U(m+1) - U(m)\|$ is true or the value of the notation j is minimal.

The notation m in this case reflects the iteration's progress. Criterion termination is indicated by the use of the notation β that ranges between 0 and 1. The notation $U = |C| * (\mu_{ij}) * |tC|$ illustrates a fuzzy membership matrix. Finally, the depiction J denotes the fitness estimation process.

Let the number of fuzzy clusters that have been generated be of the size $|fC|$ of the set $fC = \{f_{c_1}, f_{c_2}, \dots, f_{c_{|fC|}}\}$, which contains fuzzy clusters in the chronological order.

3.2. Optimal Feature Selection. For each set D_j of the records representing j^{th} the label, find the optimal features (x-coordinates of the given speech audio signal) compared to the counterpart set $\{D_k \exists k \neq j\}$. For each set D_j , a feature (x-coordinate) x_i is optimal if the values projected to the i^{th} set's feature D_j are having distribution diversity with the values

projected for the same feature x_i in other sets $\{D_k \exists k \neq j\}$. For each feature x_i of the set D_j , the process shall estimate the diversity weight towards each of the other sets $\{D_k \exists j \neq k\}$, which is the absolute difference between the maximum similarity one and the probable similarity observed ($0 \leq p\text{-value} \leq 1$). The mathematical model of identifying optimal features from each pair of sets is portrayed in the following description:

```

forall_{i=1}^{|X|} {x_i \exists x_i \in X} Begin // for all the feature attributes
  forall_{j=1}^{(n-1)} {[x_i^j \exists [x_i^k] \in D_j, j \neq k]} // Begin
    dw_{x_i \Rightarrow D_j} = 1 // the overall diversity of the feature x_i
    concerning the set D_j (label)
    forall_{k=1}^{(n)} {[x_i^k \exists [x_i^j] \in D_k, j \neq k]} // Begin
      p_{ks} = KS - test ([x_i^j], [x_i^k]) // performing the
      fusion of diversity estimation of the feature x_i between
      the sets D_j, D_k
      d(x_i)_{D_j \Leftrightarrow D_k} = d\tau // the diversity d(x_i)_{D_j \Leftrightarrow D_k} of
      the feature x_i between sets D_j \Leftrightarrow D_k has initialized to
      distance threshold d\tau
      if (p_{ks} < p\tau) Begin // the probable similarity
      value (p_{ks}) observed for the feature x_i between the sets
      D_j, D_k has found to be greater than the given prob-
      ability threshold p\tau
        d(x_i)_{D_j \Leftrightarrow D_k} = p_{ks} // the diversity d(x_i)_{D_j \Leftrightarrow D_k}
        of the feature x_i between sets D_j \Leftrightarrow D_k has been dis-
        covered from the ks-test
        End // of the condition
        dw_{x_i \Rightarrow D_j} = dw_{x_i \Rightarrow D_j} \otimes d(x_i)_{D_j \Leftrightarrow D_k}
      End // of the iterations
      if (dw_{x_i \Rightarrow D_j} \geq d\tau) Begin // if the diversity weight
      dw_{x_i \Rightarrow D_j} of the feature x_i towards the set D_j (label) is
      greater than or equal to the given diversity threshold
      d\tau,
        fD_j \leftarrow x_i // then consider the feature x_i is opti-
        mal to the set D_j and move that to the optimal features
        set fD_j
        End
      End // of the iterations
    End // of iterations
  // Preprocess the datasets of diversified labels//
  forall_{j=1}^{|C|} forall_{i=1}^{|X|} {x_i \exists x_i \in X \wedge x_i \notin fD_j} Begin // for each feature
  x_i that is selected as an optimal feature of the set D_j of
  the label j,
    {D_j} [x_i] // discarding the feature x_i and values
    projected to the corresponding feature from the set D_j
  End
  
```

3.3. The Classifier

3.3.1. Classification Procedure. This section describes the classifier employed in this proposal, as well as the training stage model and the classification procedure's objective function.

The proposed classifier was built using adaptive boosting. The classifier was designed to combine a large number of Boolean classifiers, also known as weak classifiers, that have been built using decision trees. Each weak classifier was built using the best features taken from a series of quantitative steps. These weak classifiers categorize the provided test data based on whether the condition is true or false. Another bad classifier might label the negatives as bipartite, which includes both false positives and false negatives. This procedure has been repeated until the overall weak classifier determines that the task has been finished. Furthermore, the outcomes obtained, all weak classifiers, in general, are combined into the rating scale and provide the final result.

In this article's projected model, each weak classifier was employed to highlight the coherently ideal features gathered during the quantitative seed phase towards binary classification. The classification technique was also repeated for each risk management implementation using a weak classifier; the corpus component that could not be accurately identified was the focus of the next classifier iteration, known as "boosting." Furthermore, weight classification revealed an inferior classifier, which is employed on each iteration. Completing weak classifiers iteratively results in accurately categorised records from all of these weak classifiers. Each weak classifier, according to the projected method, recommends a certain n-gram for classification accuracy. Furthermore, the classification results of weak classifiers would be justified in order to discover the polarity of the given record. When compared to other binary classification challenges, the Adaboost classifier has been demonstrated to be a feasible approach for optimising DT output (decision trees). It has the potential to be widely employed to improve the performance of various machine learning approaches. The label prediction approach for an unlabelled record consists of the steps listed below:

- (i) Extract the values of all considered features from the unlabelled records
- (ii) The adaptive boosting classification strategy recommended in this proposal shall be used to predict the germination quality of seed samples as:
- (iii) Discover the standard measures of the fitness coefficients of the features towards all weak classifiers
- (iv) Consider the values of the features in the given input record; the considered features are optimal in regard to one or more weak classifiers
- (v) Prepare the normal distribution for each optimal feature that uses the input value of the feature as a standard measure
- (vi) Find the fitness confidences of the input record towards all optimal features of the corresponding weak classifier
- (vii) Compare the standard measures of the fitness coefficients discovered during the training phase and fitness confidences of the respective features to predict the label
- (viii) There shall be a label assigned to each input record after completing this prediction phase

4. Experimental Study

This section focuses on the proposed model's practical implementation in comparison to some of the latest methods discussed in the literature. This section describes the dataset in detail, the changed programme's requirements, and the system conditions that are critical for performance study. Python [40] is used to execute the model, while PyCharm [41] is used to write the code.

4.1. The Data. For the experimental analysis of the proposed model, the dataset RAVDESS [42] was used, which is a corpus of speech audio signals reflecting a variety of emotions. 247 people who were typical of untrained adult researchers assessed the emotional relevance, expressiveness, and authenticity of the RAVDESS dataset. A total of 72 volunteers have also been made available for the dataset's cross-validation. It has been reported that emotional relevance, reliability, and cross-reliability are all higher. 6204 speech audio signal records were used in the experiment, each of which was labelled with the emotions identified in the corresponding speech audio signal. The following are the counts of records representing different emotions: anger, disgust, fear, joy, neutral, surprise, and sad, where the records labelled as anger, fear, joy, and sad each counted at 1128, disgust counted at 576, neutral counted at 564, and surprise counted at 552. Overall, the 200 words spoken by 200 different people in 200 different emotional contexts represent a wide range of emotions.

4.2. Data Processing. The speech audio signals of the dataset are transmuted into the digital format [43] such that each speech signal transformed to a set of y -coordinates representing the corresponding x -coordinates. It is viewed as a two-dimensional matrix of digital representations of each speech audio signal. A total of seven datasets in the CSV format, each representing one of the emotion labels, are generated after data processing.

4.3. Performance Analysis. This approach has been evaluated for performance using metrics from the confusion matrices of all other contemporary models, including those that use "hybrid acoustic features (HAF)" [37] and "Supervised Bayes Learning of Digital Features (SBL-DF)" [38]. It has to divide the records of each label into two sets to perform cross-validation. The suggested EL-HDAF and contemporary models HAF and SBL-DF have undergone fourfold cross-validation to demonstrate the superiority of EL-HDAF over the existing HAF and SBL-DF models. Table 1

The overall number of records taken for the experimental study is 6204. The records used for training are 4653, and the overall records used for testing are 1551.

In order to evaluate the multilabel cross-validation adopted in the performance analysis, the metrics including precision (positive predictive value) and sensitivity should be used. Some other metrics for analysis that are not deemed significant include the weighted sensitivity,

TABLE 1: The mean and deviation of the assessment metric values depicted from multifold cross-validation.

| Average of 10-fold result and deviations | | | |
|--|-------------------------|-------------------------|-------------------------|
| Metrics | EL-HDAF | HAF | SBL-DF |
| Precision | 0.944679 ± 0.032751 | 0.894171 ± 0.058389 | 0.880646 ± 0.065648 |
| Sensitivity recall | 0.954865 ± 0.012566 | 0.908286 ± 0.010922 | 0.896128 ± 0.017381 |
| Specificity | 0.954897 ± 0.005458 | 0.907739 ± 0.002584 | 0.893572 ± 0.010829 |
| F-score | 0.951408 ± 0.018341 | 0.899709 ± 0.031556 | 0.886927 ± 0.034786 |
| Decision accuracy | 0.948429 ± 0.030393 | 0.894171 ± 0.058389 | 0.880646 ± 0.065648 |

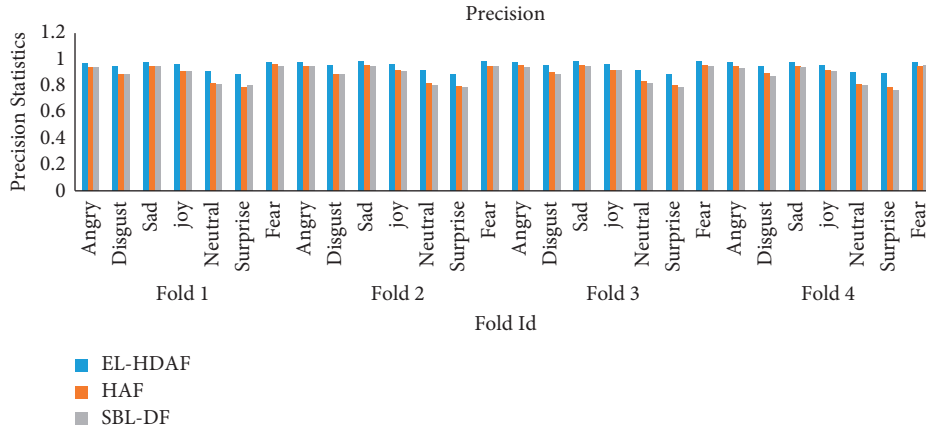


FIGURE 2: Fourfold cross-validation determined the positive prediction rate (precision).

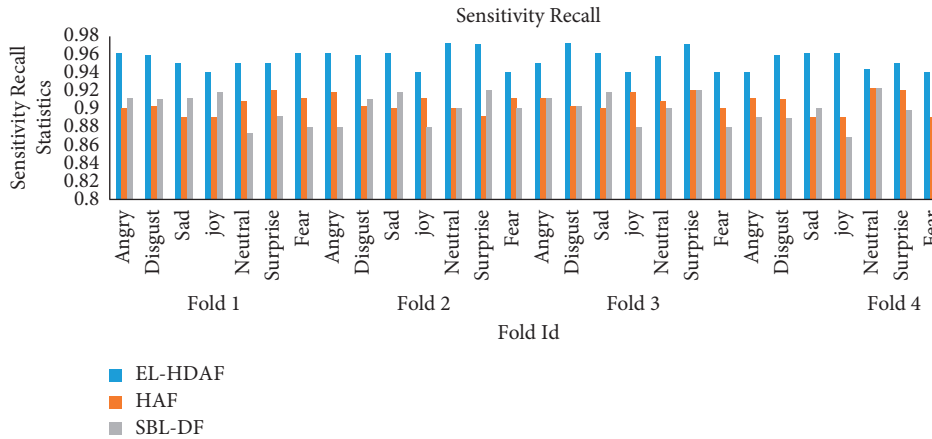


FIGURE 3: Prediction of emotions sensitivity (recall) as determined by cross-validation, fourfold.

weighted measures of F-score, and precision metrics. The breadth of the solution's implementation and its effectiveness can be determined at the micro-level study of the associated assessment metrics.

When compared to the HAF and SBL-DF approaches, the recommended EL-HDAF strategy shows a more consistent rate of accuracy for all emotions, according to the statistical data shown in Figure 2.

Figure 3 shows that the EL-HDAF has similar performance advantages of emotion prediction sensitivity (recall) compared to contemporaneous models HAF and SBL-DF.

The F-measure and distinct labels are used to display the graphs in Figure 4, with the F-measure representing the harmonic-mean of the precision and sensitivity. The EL-HDAF surpassed the other frameworks, HAF and SBL-DF,

that were used for comparison, according to the statistical statistics as in graphical representation.

Figure 5 specifies some factors of which one of the critical measures, the ratio defined for true negative amongst the cumulative set of actual negatives, is considered. The graphical representation of the performance refers to the conditions that refer to the fact that the proposed model is EL-HDAF and is performing superior in comparison to other key models HAF and SBL-DF reviewed for the corpus of requirement specifications. The comparison of the two models is shown in the form of graph with the help of fourfold labels as angry, disgust, fear, glad, neutral, sad, and surprised. Thus, it has been concluded that the performance of the proposed model in terms of specificity is better in all the labels while compared to the contemporary models.

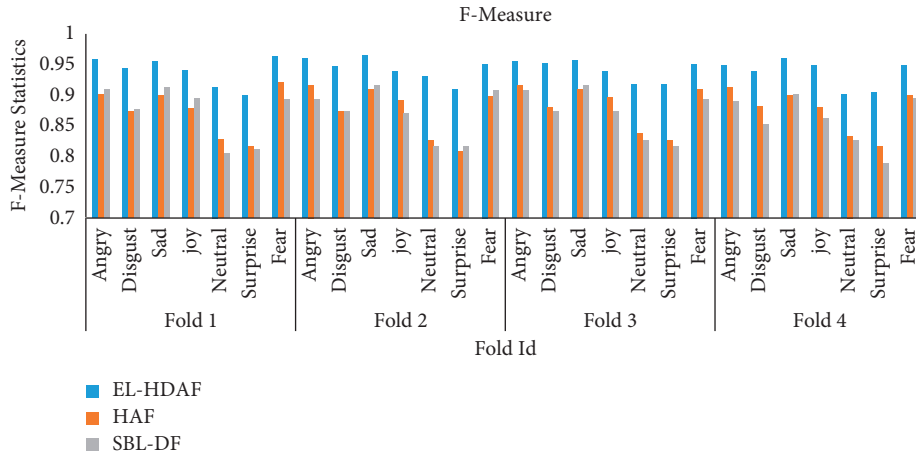


FIGURE 4: Fourfold cross-validation of EL-HDAF, HAF, and SBL-DF contributed a mean.

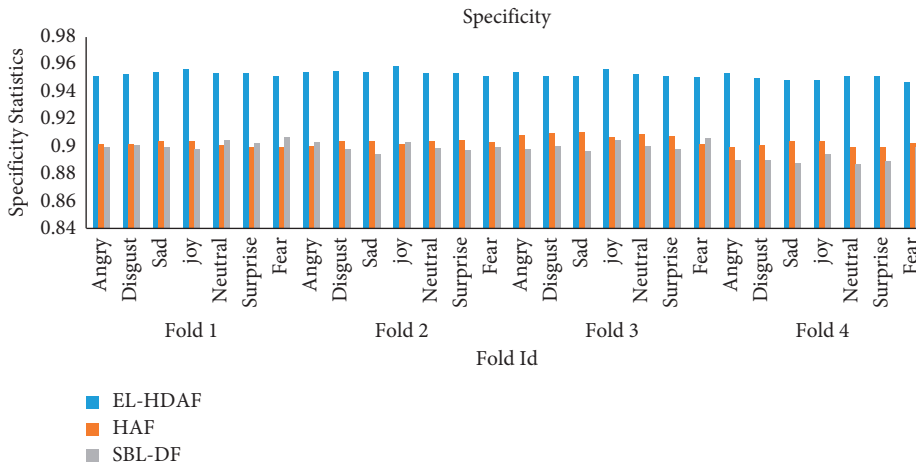


FIGURE 5: Specificity observed for the proposed EL-HDAF and contemporary models HAF and SBL-DF in terms of metric specificity over fourfolds.

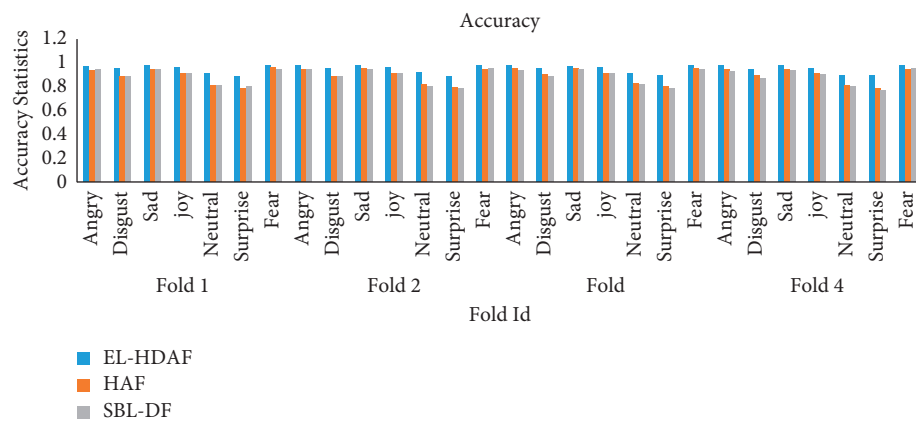


FIGURE 6: Accuracy of EL-HDAF, HAF, and SBL-DF over fourfolds.

The accuracy metric has been used for measuring the performance of EL-HDAF, HAF, and SBL-DF over the fourfolds as exhibited in Figure 6. The comparison of the three models is shown in the form of graph with the help of fourfold labels as angry, disgust, fear, glad, neutral, sad and surprised.

Therefore, it has been concluded that the performance of the proposed model in terms of accuracy is better in all the labels compared to other contemporary models.

Weighted measures of accuracy, recall, and F-score are all essential metrics in determining the strength of the

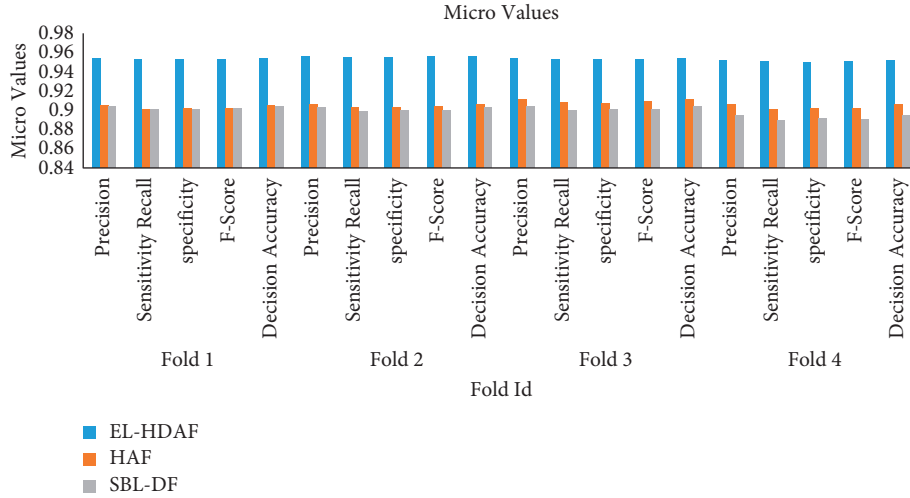


FIGURE 7: Micromeasures of precision, sensitivity (recall), f-measure, as well as accuracy.

multilabel classification performance because they assist to understand the classifier’s performance overall. The metric values represent the classifier’s ability to scale its performance based on the precision, sensitivity, and evaluation accuracy factors that include the harmonic-mean of the precision and sensitivity. The micromeasures of the corresponding metrics precision, sensitivity, accuracy, and f-score are also critical to assess the performance of multilabel classification.

For EL-HDAF, HAF, and SBL-DF, the weighted measures of the corresponding metrics observed for each emotion prediction are the essential inputs to determine the micromeasures of the corresponding metrics. The micromeasures of the corresponding cross-validation metrics are represented in Figure 7. The fourfold cross-validation process and the resultant micromeasures of precision, sensitivity, f-score, and class prediction accuracy indicate that the model EL-HDAF outperforms the models SBL-DF and HAF.

5. Conclusion

In recent years, predicting emotional states from acoustic features of spoken audio signals has been a prominent objective in the field of speech audio signal processing. Machine learning models with a high feature dimension are used to recognize empathy from audio data. To reduce the effect of high-dimensional data on the proposed model during training, the feature values of various classes were analysed for diversity, and a novel clustering approach was devised. It is also worth mentioning that the adaptive boosting classification technique is intended to learn from the various clusters in the training corpus. Ensemble Learning by High-Dimensional Acoustic Features (EL-HDAF) is a projected model that has been evaluated against two existing models, HAF and SBL-DF, using the benchmark dataset RAVDESS using fourfold cross-validation. In performance analysis, the cross-validation metrics and accompanying micromeasures were

investigated. The results of the suggested and current measurements demonstrate that EL-emotion HDAF detection beats the existing methods HAF and SBL-DF with the fewest false alarms and the highest decision accuracy. In the future, the acoustic features of the speech stream can be adjusted utilizing evolutionary computing methodologies to increase the performance of ensemble learning models in predicting emotion. The contribution would motivate future research towards emotion detection through acoustic features of speech signals, where an evolutionary technique has an optimal scope in feature optimization.

Abbreviations

| | |
|-----------------------------------|--|
| ML: | Machine learning |
| EL-HDAF: | Ensemble learning by high-dimensional acoustic features |
| ZCR: | Zero-crossing rate |
| EMD: | Empirical mode decomposition |
| HAF: | Hybrid acoustic features |
| SBL-DF: | Speech emotion recognition using supervised Bayes learning on digital features |
| c_j : | Cluster centroid |
| μ_{ij} : | Euclidean distance |
| v_j : | Fuzzy centroid |
| $ fC $: | Fuzzy clusters |
| D_j : | Diversity |
| x_i : | Feature |
| (p_{ks}) : | Probable similarity value |
| $p\tau$: | Probability threshold |
| $d\omega_{x_i \Rightarrow D_j}$: | Diversity weight |
| DT: | Decision trees. |

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request (shitharths@kdu.edu.et).

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Authors' Contributions

M M Venkata Chalapathi conceptualised the study, curated the data, performed a formal analysis, devised the methodology, contributed to the software, and wrote the original draft; M. Rudra Kumar supervised the study, wrote and reviewed the content, edited the article, and helped with project administration and visualization; Neeraj Sharma supervised the study, wrote and reviewed the software, validated the content, and wrote the original draft and was responsible for devising the methodology; S. Shitharth wrote, reviewed, and edited the article, helped acquire funding, and contributed to the visualization and formal analysis, and also software development.

References

- [1] D. Fabiano and S. Canavan, "Emotion recognition using fused physiological signals," in *Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 42–48, IEEE, Cambridge, United Kingdom, September 2019.
- [2] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, and A. Košir, "Audio-visual emotion fusion (AVEF): a deep efficient weighted approach," *Information Fusion*, vol. 46, pp. 184–192, 2019.
- [3] K. Sarker and K. R. Alam, "Emotion recognition from human speech: emphasizing on relevant feature selection and majority voting technique," in *Proceedings of the 3rd International Conference on Informatics, Electronics & Vision (ICIEV)*, pp. 89–95, Dhaka, Bangladesh, May 2014.
- [4] R. Subhashini and P. R. Niveditha, "Analyzing and detecting employee's emotion for amelioration of organizations," *Procedia Computer Science*, vol. 48, pp. 530–536, 2015.
- [5] A. Rychalski and S. Hudson, "Asymmetric effects of customer emotions on satisfaction and loyalty in a utilitarian service context," *Journal of Business Research*, vol. 71, pp. 84–91, 2017.
- [6] M. Papakostas, E. Spyrou, T. Giannakopoulos et al., "Deep visual attributes vs. hand-crafted audio features on multi-domain speech emotion recognition," *Computation*, vol. 5, no. 2, p. 26, 2017.
- [7] A. Khan and U. K. Roy, "Emotion recognition using prosodie and spectral features of speech and Naïve Bayes Classifier," in *Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSP-NET)*, pp. 1017–1021, IEEE, Chennai, India, March 2017.
- [8] N. Semwal, A. Kumar, and S. Narayanan, "Automatic speech emotion detection system using multi-domain acoustic feature selection and classification models," in *Proceedings of the 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pp. 1–6, IEEE, New Delhi, India, February 2017.
- [9] L. Sun, B. Zou, S. Fu, J. Chen, and F. Wang, "Speech emotion recognition based on DNN-decision tree SVM model," *Speech Communication*, vol. 115, pp. 29–37, 2019.
- [10] I. Luengo, E. Navas, and I. Hernaez, "Feature analysis and evaluation for automatic emotion identification in speech," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 490–501, 2010.
- [11] S. Basu, J. Chakraborty, A. Bag, and M. Aftabuddin, "A review on emotion recognition using speech," in *Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 109–114, IEEE, Coimbatore, India, March 2017.
- [12] H. K. Palo, M. Chandra, and M. N. Mohanty, "Emotion recognition using MLP and GMM for Oriya language," *International Journal of Computational Vision and Robotics*, vol. 7, no. 4, pp. 426–442, 2017.
- [13] A. Ozcift and A. Gulten, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 443–451, 2011.
- [14] X. Sun, "Pitch accent prediction using ensemble machine learning," in *Proceedings of the Seventh International Conference on Spoken Language Processing*, Denver, CL, USA, September 2002.
- [15] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [16] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.
- [17] H. J. V. Veen, *Le Nguyen the Dat Armando Segnini*, Kaggle Ensembling Guide, 2015.
- [18] J. R. Quinlan, "Bagging, boosting, and C4. 5," in *Proceedings of the the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pp. 725–730, Portland, Oregon, August 1996.
- [19] B. Schuller, S. Reiter, R. Muller, M. Al-Hames, M. Lang, and G. Rigoll, "Speaker independent speech emotion recognition by ensemble classification," in *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo*, pp. 864–867, IEEE, Amsterdam, Netherlands, July 2005.
- [20] D. Morrison, R. Wang, and L. C. De Silva, "Ensemble methods for spoken emotion recognition in call-centres," *Speech Communication*, vol. 49, no. 2, pp. 98–112, 2007.
- [21] H. C. Kim, S. Pang, H. M. Je, D. Kim, and S. Y. Bang, "Support vector machine ensemble with bagging," *Pattern Recognition with Support Vector Machines*, Springer, in *Proceedings of the International Workshop on Pattern Recognition with Support Vector Machines*, pp. 397–408, August 2002.
- [22] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 688–705, 2007.
- [23] A. Bhavan, P. Chauhan, R. R. Hitkul, and R. R. Shah, "Bagged support vector machines for emotion recognition from speech," *Knowledge-Based Systems*, vol. 184, Article ID 104886, 2019.
- [24] P. Shegokar and P. Sircar, "Continuous wavelet transform based speech emotion recognition," in *Proceedings of the 2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pp. 1–8, IEEE, Surfers Paradise, Gold Coast, Australia, December 2016.
- [25] S. Parthasarathy and I. Tashev, "Convolutional neural network techniques for speech emotion recognition," in *Proceedings of the 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 121–125, IEEE, Hitotsubashi Hall in Tokyo, Japan, September 2018.
- [26] W. Jiang, Z. Wang, J. S. Jin, X. Han, and C. Li, "Speech emotion recognition with heterogeneous feature unification of deep neural network," *Sensors*, vol. 19, no. 12, p. 2730, 2019.
- [27] Z. T. Liu, M. Wu, W. H. Cao, J. W. Mao, J.-P. Xu, and G. Z. Tan, "Speech emotion recognition based on feature

- selection and extreme learning machine decision tree,” *Neurocomputing*, vol. 273, pp. 271–280, 2018.
- [28] H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking: studies on acted and spontaneous speech,” *Computer Speech & Language*, vol. 29, no. 1, pp. 186–202, 2015.
- [29] J. B. Alonso, J. Cabrera, M. Medina, and C. M. Travieso, “New approach in quantification of emotional intensity from the speech signal: emotional temperature,” *Expert Systems with Applications*, vol. 42, no. 24, pp. 9554–9564, 2015.
- [30] W. Zehra, A. R. Javed, Z. Jalil, H. U. Khan, and T. R. Gadekallu, “Cross corpus multi-lingual speech emotion recognition using ensemble learning,” *Complex & Intelligent Systems*, vol. 7, pp. 1–10, 2021.
- [31] O. Obulesu, K. Suresh, D. Gaurav et al., “Adaptive diagnosis of lung cancer by deep learning classification using wilcoxon gain and generator,” *Journal of Healthcare Engineering*, vol. 2021, Article ID 5912051, 13 pages, 2021.
- [32] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, “An ensemble machine learning approach through effective feature extraction to classify fake news,” *Future Generation Computer Systems*, vol. 117, pp. 47–58, 2021.
- [33] Q. Mao, X. Zhao, and Y. Zhan, “Extraction and analysis for non-personalized emotion features of speech,” *Advances in Information Sciences and Service Sciences*, vol. 3, no. 10, pp. 225–263, 2011.
- [34] S. Shitharth, B. M. Gouse, R. Kadiyala, and B. Vidhyacharan, *Prediction of COVID-19 Wide Spread in India Using Time Series Forecasting Techniques*, Springer, Berlin, Germany, 2021.
- [35] G. T. Reddy, S. Bhattacharya, S. S. Ramakrishnan et al., “An ensemble based machine learning model for diabetic retinopathy classification,” in *Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*, pp. 1–6, IEEE, Vellore Institute of Technology, Vellore, India, February 2020.
- [36] A. K. Cherian, E. Poovammal, N. S. Philip, K. Ramana, S. Singh, and I. H. Ra, “Deep learning based filtering algorithm for noise removal in underwater images,” *Water*, vol. 13, no. 19, p. 2742, 2021.
- [37] K. Zvarevashe and O. Olugbara, “Ensemble learning of hybrid acoustic features for speech emotion recognition,” *Algorithms*, vol. 13, no. 3, p. 70, 2020.
- [38] M. V. Chalapathi, “Speech emotion recognition using supervised Bayes learning on digital features of multi-label data corpus,” *Design Engineering*, pp. 1065–1078, 2021.
- [39] M. Shasidhar, V. S. Raja, and B. V. Kumar, “MRI brain image segmentation using modified fuzzy c-means clustering algorithm,” in *Proceedings of the 2011 International Conference on Communication Systems and Network Technologies*, pp. 473–478, IEEE, Katra, India, June, 2011.
- [40] “Python. (n.d.),”.
- [41] “pycharm. (n.d.),”.
- [42] S. R. Livingstone and F. A. Russo, “The ryerson audio-visual database of emotional speech and song (RAVD ESS): a dynamic, multimodal set of facial and vocal expressions in north American English,” *PLoS One*, vol. 13, no. 5, Article ID e0196391, 2018.
- [43] “wav-to-csv. (n.d.),”.

Research Article

Exploration of Environmental Protection-Oriented Ecoenvironmental Performance Audit System

Jie Wang ^{1,2}, Xiaomei Wang,² and Na Li²

¹School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

²School of Economics and Management, Chuzhou University, Chuzhou 239000, China

Correspondence should be addressed to Jie Wang; wangjie_nust@163.com

Received 13 January 2022; Revised 28 January 2022; Accepted 3 February 2022; Published 25 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Jie Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the era of the COVID-19 (SARS-CoV-2) virus, it has become imperative to audit the environment surrounding us to prioritize public health and the healthcare system. This study aims to accelerate the reform of the economic development modes and the construction of a resource-friendly society. The environmental protection-oriented ecoenvironmental performance audit system is studied here. Based on the current situation of regional ecological pollution in the Yangtze River Economic Belt, the situation and existing problems are analyzed for the ecoenvironment in the Yangtze River Economic Belt from the perspectives of biodiversity, water ecoenvironment, wetland ecoenvironment, and forest ecoenvironment. Then, the situation is deeply analyzed for the ecoenvironmental performance audit project in the Yangtze River Economic Belt. Afterward, we explore the basic characteristics of the environmental audit and summarize the implementation path of resource and environmental audit. The results show that there are some problems in the control and utilization of funds for environmental protection and the development and protection of resources in the cities along the Yangtze River Economic Belt. Then we analyze the problems in the ecoenvironmental performance audit of the Yangtze River Economic Belt and give some suggestions for improving the ecoenvironmental performance audit work. The objective of this paper is to improve the application effect of ecoenvironmental performance audit and provide references for future environmental protection work which will have great impacts on public health and the development of healthcare services.

1. Introduction

Recent decades have witnessed the frequent occurrence of global extreme climate, major natural disasters, and global pandemics which seriously affect socioeconomic development, the standard of living and production, and public health. In China, the northeast is prone to drought, while landslides and mudslides are common in the southwest. The resource and environmental problems such as the recent COVID-19 virus bring damage to residents, thereby hindering sustainable socioeconomic development. In this context, ecoenvironmental issues have become the focus of social concern.

The Yangtze River is the largest in China, and the basin area covers about 2.05 million square kilometers, occupying 21% of the total land area of China [1, 2]. Meanwhile, the

total GDP and population of the Yangtze River Economic Belt exceed 40% of those in China. The topographic features of the Yangtze River Economic Belt are widely distributed and the ecosystem is diversified. However, due to the barbaric exploitation and loose government supervision in the past, the ecoenvironment system of the Yangtze River Economic Belt has become very vulnerable, with excessive depletion of natural resources and serious soil erosion. Many factors, such as industrial pollution, biodiversity damage, overutilization of mineral resources, over fishery, deforestation, grain for green projects, and uncontrolled groundwater extraction seriously affect the stability of ecological resources in the Yangtze River. Although some progress has been made in biodiversity conservation in the Yangtze River Economic Belt, the habitat and reproduction of some animals and plants have been irreversibly destroyed due to

human activities which are not conducive to environmental protection. Environmental pollution in the Yangtze River can have an extremely adverse effect on public health and the sustainable economic development of provinces along the Yangtze River Economic Belt. So in this paper, we study the ecoenvironmental performance audit system of the Yangtze River Economic Belt to improve the application effect of environmental audit and therefore provide references for future environmental protection work.

Nowadays, the awareness of ecoenvironmental protection is deepening in China, and the Chinese government has been vigorously promoting environmental protection policies. Meanwhile, advanced data processing and analysis technologies are maturing, and hardware facilities are being improved. Relevant data show that, to protect the ecoenvironment, the state investment in environmental pollution control has increased from 338.7 billion RMB in 2007 to 953.9 billion RMB in 2020 [3]. Besides, many policies are issued to conserve resources and protect the environment, including the construction of an ecoenvironment monitoring network and the construction of the overall framework of big data of ecoenvironment. These policies have enhanced environmental protection and pollution control and promoted the development of the environmental protection industry. In this context, the environmental performance audit has brought great opportunities and challenges. Since the 1980s and 1990s, the Chinese government has carried out audits of special funds for environmental protection. Through decades of rapid development, the environmental audit projects in China have been significantly developed. The successful implementation of large-scale environmental audit projects also provides scientific and effective empirical data for the development of environmental audits in China. However, there are still many problems and deficiencies in the existing environmental performance audit. Based on the case of regional ecological pollution in the Yangtze River Economic Belt, we have analyzed the current situation and existing problems of the ecoenvironment in the Yangtze River Economic Belt were from the aspects of biodiversity, water ecoenvironment, wetland ecoenvironment, and forest ecoenvironment. Then, we have carefully studied the ecoenvironmental protection audit in the Yangtze River Economic Belt which specifically includes the background, the implementation, and the problems of ecoenvironmental performance audit in the Yangtze River Economic Belt. Finally, we put forward some optimization suggestions for improving the audit of the Yangtze River Economic Belt.

The main contribution of our study lies in enriching the research literature of environmental protection, analyzing the key problems in the ecoenvironmental performance audit in China, and putting forward some suggestions for a better environmental protection-oriented ecoenvironmental performance audit system. Nowadays, the ecoenvironmental performance audit is a new audit field in China, and there is no mature empirical model or system. Although some ecoenvironmental performance audits have been carried out in China, most of them are for a certain type of resource, and there is no comprehensive audit for multiple resources.

Therefore, in this paper, the environmental protection-oriented ecoenvironmental performance audit system is studied, theory and practice are combined, the key problems in the current ecoenvironmental performance audit system in China are focused on, and finally, optimization suggestions are proposed.

The remainder of this paper is organized as follows: Section 2 discusses the research method and related literature on environmental audit, Section 3 presents the ecological environment in the Yangtze River Economic Belt, Section 4 analyzes the ecoenvironmental performance audit in the Yangtze River Economic Belt, and we summarize and conclude in Section 5.

2. Research Method and Related Literature on Environmental Audit

2.1. Research Method. In this paper, we adopt the methods of literature review, case analysis, and inductive analysis to explore the concept, organization, implementation, technical methods, and application of environmental audit in the context of big data. First, the existing problems are analyzed in the ecoenvironment governance of the Yangtze River Economic Belt through available environment data from relevant yearbooks and literature. Second, the ecoenvironmental protection audit project of the Yangtze River Economic Belt is chosen for case analysis. Third, we analyze the ecoenvironment governance based on the regional environmental monitoring data of the Yangtze River Economic Belt. Finally, we discuss the advantages and disadvantages of environmental audit projects in the Yangtze River Economic Belt in the context of big data, as well as the optimization countermeasures of ecological environmental protection audit in the Yangtze River Economic Belt.

2.2. Related Literature on Environmental Audit. The concept of traditional audit is conservative and the audit methods are backward, which is not conducive to finding audit clues, carrying out relevant work, and implementing follow-up audit rectification. Therefore, the environmental audit should be renovated to adapt to the current situation through theoretical and practical innovation. There is a lot of the latest research on environmental audit. For example, Jiang and Tan (2021) examined the causal relationship between the national environmental audit and regional energy efficiency from the perspective of institutional environment and difference in development level [4]. They found that the implementation of national environmental audit could prominently promote regional energy efficiency, and the external governance effect of national environmental audit on regional energy efficiency is more significant in developed areas and areas with better institutional environments. Through the exploration of the internal mechanism of the national environmental audit affecting regional energy efficiency improvement, the research also found that the national environmental audit can improve regional energy efficiency by promoting technological progress. Nazarova et al. (2020) demonstrated the necessity of an environmental

auditing system, and they argued that, due to the complexity of the audit objectives and the contradictory and multilevel relationships between its main elements, it is complicated in conducting an environmental audit in the forestry sector. The research emphasized the importance of green economy audit which may be a suitable auditing support mode for forestry enterprises [5]. Marwa et al. (2020) discussed the relationship between environmental disclosure quality and environmental audit. The results showed that the timely disclosure of environmental information was affected by the environmental audit committee, CSR (Corporate Social Responsibility) committee, the auditors, earnings management behavior, firm size, and the industry, while no evidence has been found that there is a statistically significant relationship between CSR committee and the voluntary disclosure of environmental information [6]. Silva et al. (2018) aimed to assess the potential and application of UAVs (Unmanned Aerial Vehicles) in environmental audit and proposed a method for data acquisition and identification of environmental impacts [7]. In conclusion, much literature has attached the importance to environmental audit especially in the era of the COVID-19 (SARS-CoV-2) virus, but unfortunately, so far, few studies have directly explored the environmental protection-oriented ecoenvironmental performance audit in the Yangtze River Economic Belt. Given the importance of the ecological environment in the Yangtze River Basin, we investigate the problem in the ecoenvironmental performance audit of the Yangtze River Economic Belt and give some suggestions for improving audit quality.

3. Analysis of the Ecological Environment in the Yangtze River Economic Belt

3.1. Overview of the Ecological Situation in the Yangtze River Economic Belt

3.1.1. The Current Situation of Biodiversity in the Yangtze River Economic Belt. In this paper, we chose the Yangtze River Economic Belt as the experimental base. The Yangtze River Economic Belt is a densely populated area with rich biodiversity, and human production activities can have an important impact on the quantity and living conditions of organisms in the area. The daily activities of the surrounding residents will disturb the ecological factors, such as the local soil, atmosphere, and water. Meanwhile, the development of tourism may destruct the habitat and reproduction of plants and animals, resulting in ecosystem imbalance and squeezing the living space of plants and animals [8, 9]. With the improvement of law enforcement and nature reserves management system, nature reserves throughout China are less subject to human interference, which provides an effective way for the restoration of natural vegetation and the ecosystem. Besides, the implementation of many policies, such as ecological migration and grain for green projects, has effectively protected the habitats of many wild animals and plants. Although some achievements have been made in biodiversity protection, human activities are still interfering with the survival of wild animals and plants. In terms of water pollution, there are about 400,000 chemical

enterprises along the Yangtze River Economic Belt, and nearly one-third of the heavily polluted enterprises are distributed within 5 kilometers of the water source, which has seriously destroyed the water environment in the Yangtze River region. Water pollution has also caused serious ecological problems in the Yangtze River, such as the disappearance of the original biological population, a sharp decline in the number of wild animals and plants, and the destruction of biodiversity [10].

3.1.2. Ecological Situation of Water Environment in the Yangtze River Economic Belt. The Yangtze River has a total length of 6,300 kilometers with eight first-class tributaries and many important inner lakes, so water resources in the Yangtze River Basin are quite abundant. However, due to the lack of adequate protection, the water ecological environment in the Yangtze River Economic Belt is not optimistic. The main problems of water ecology are manifested in the following aspects. (1) With the construction of dams, hydropower stations, and other water conservancy facilities, the main Yangtze River and many tributaries are no longer directly connected with the upstream, midstream, or downstream, and many rivers are blocked. (2) Due to the demand for water storage and flood control, the natural fluctuation law in the Yangtze River Basin has been affected, and because of the changes in water temperature, water quantity, and hydrological law in the Yangtze River Basin, the dry season of the Dongting Lake and Poyang Lake in the Yangtze River Economic Belt has advanced, which brings great challenges to the protection of the lake ecology [11]. (3) With the acceleration of urbanization, there are many sand mining sites in the Yangtze River Economic Belt and the speed of sediment exploitation has exceeded the limit of natural recovery in the Yangtze River Basin, which has made the water quality safety and the living space of aquatic animals seriously threatened. Xin et al. (2019) studied a monitoring system to collect real-time information for monitoring water environment information [12].

The water system of the Yangtze River Basin is shown in Figure 1. The Yangtze River, the longest river in China and the third-longest river in the world, originates from Tanggula Mountain in Qinghai Province; it has 49 tributaries, including the Yalong River, Minjiang River, Jialing River, and some other important tributaries shown in Figure 1. Finally, the Yangtze River empties into the East China Sea at Chongming Island in Shanghai.

The comparison of runoff, sediment, and total phosphorus flux in the hydrological control station of the mainstream of the Yangtze River from 2019 to 2020 is shown in Figure 2.

3.1.3. Ecological Situation of Wetlands in the Yangtze River Economic Belt. The wetland resources are abundant in the Yangtze River Basin. The upper reaches of the Yangtze River are dominated by forest wetlands and alpine wetlands, while the middle and lower reaches of the Yangtze River are dominated by freshwater lake wetlands and beach wetlands. The wetland area of the Yangtze River Basin has changed

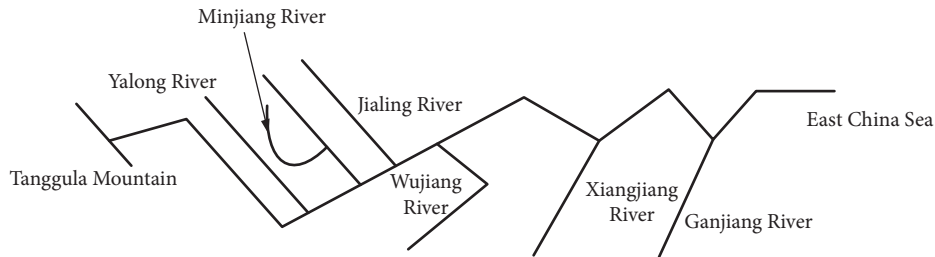


FIGURE 1: Water system of the Yangtze River Basin.

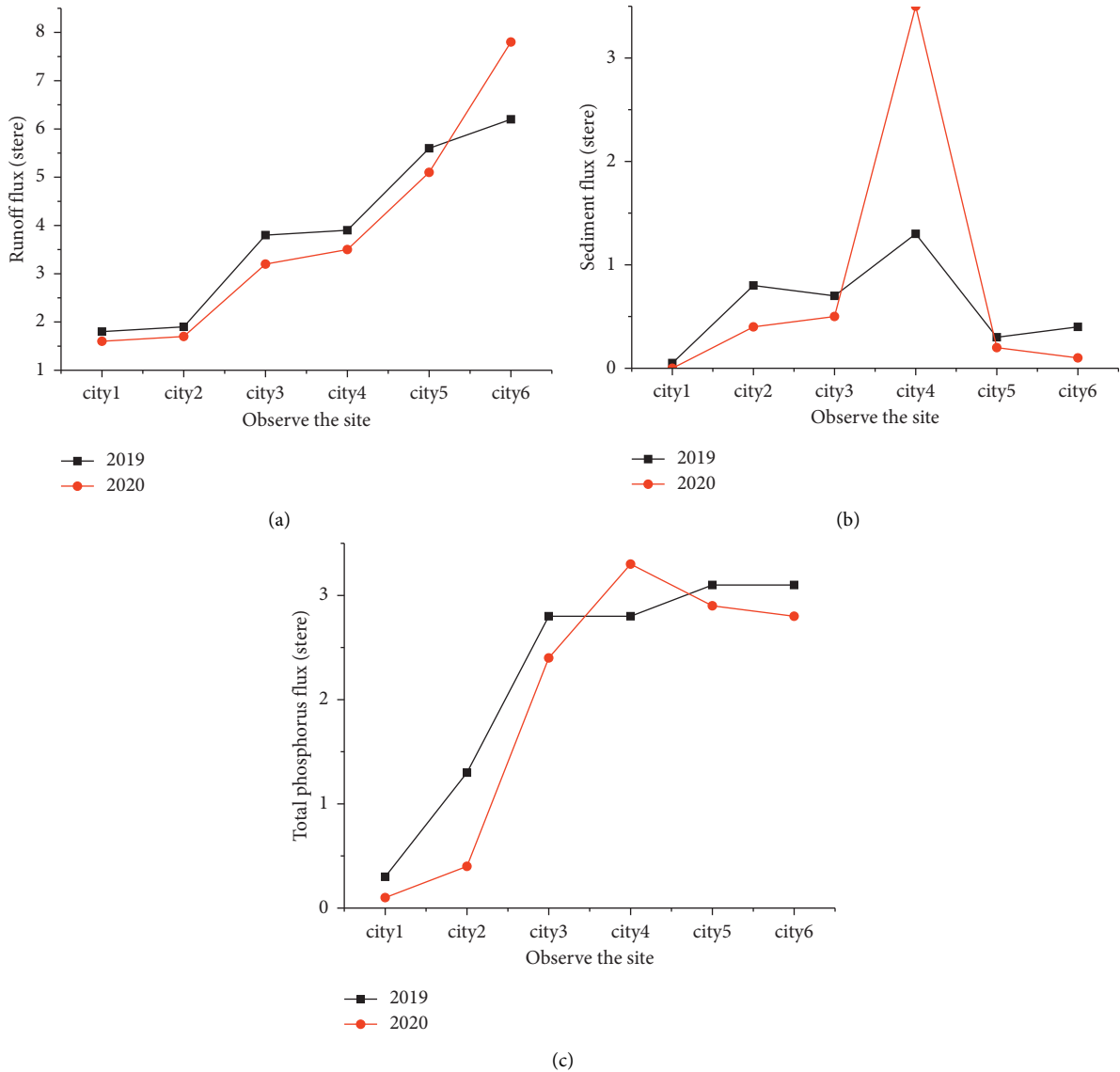


FIGURE 2: Runoff, sediment, and total phosphorus fluxes of main urban hydrological stations in the Yangtze River mainstream. (a) Runoff. (b) Sediment. (c) Total phosphorus flux.

greatly. By the year 2000, the lakes in the Yangtze River Basin have decreased significantly, while aquaculture farms have increased significantly, resulting in serious water pollution and eutrophication of the water body. After the year 2000, with the improvement of the environmental protection awareness of the public and the strengthening of national

environmental protection governance, the number of lakes has significantly increased, and the wetland ecology has been developing well. LV et al. (2019) studied the space distribution features of BIM (Building Information Modelling) geospatial big data and correspondingly proposed the data store and management model for BIM geospatial big data [13].

3.1.4. Ecological Situation of Forest in the Yangtze River Economic Belt. In forest ecology, forestry is crucial to the ecological construction of the Yangtze River Economic Belt and plays an important role in improving the ecosystem. Forest ecology can restore wetland ecosystems, improve the desert ecosystem, build and protect forest ecosystems, and maintain biodiversity. With the construction of key ecological projects, such as desertification control, grain for green projects, construction of protective forest systems in the Yangtze River Basin, and natural forest protection, the forest ecology of the Yangtze River Economic Belt has been significantly improved. The forest coverage rate of the Yangtze River Basin has exceeded 30%, and the intensity of soil and water loss has decreased by 42% [14]. Nevertheless, the Yangtze River Economic Belt still faces problems such as uneven distribution of forest resources, insufficient forest volume, and insufficient forest resources per capita. The forest ecoenvironment in the Yangtze River Economic Belt needs to be further improved [15].

Water is essential to human survival, so are land, sunlight, and minerals. Thus, water, land, sunlight, and minerals constitute the natural resource environment. Natural resources are defined as processes that can produce economic benefits under specific conditions. Ecoenvironment refers to the sum of various natural forces that are closely related to human beings and affect human production and life. The composition of natural resources is shown in Figure 3.

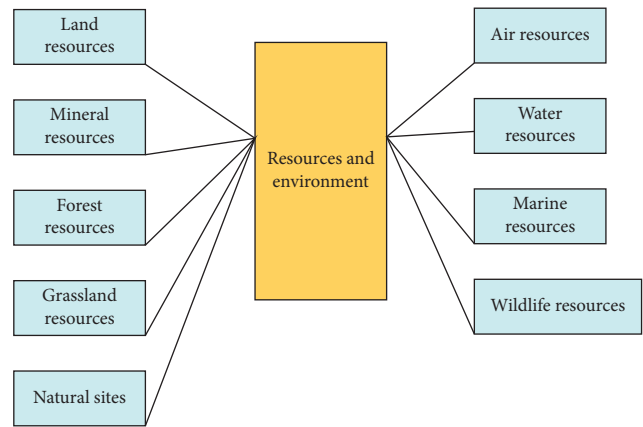


FIGURE 3: Schematic diagram of natural resources composition.

3.2. Analysis of the Yangtze River Wetland Ecosystem and Existing Problems. The evolution law of the Yangtze River is unique. Different types of wetland ecosystem complexes are evolved through the hydrological process, biogeochemical process, and ecological process. The wetland area of the Yangtze River Basin is 115,400 hectares, including the marsh-wet meadow complex ecosystem and the alpine Canyon River wetland in the upper reaches of the Yangtze River, the river-flood-wetland-Lake Wetland complex ecosystem in the middle reaches of the Yangtze River, and the delta-coastal wetland complex ecosystem in estuaries. As is known to all, the Yangtze River wetland ecosystem has very important ecological, social, and economic benefits. The annual runoff of the Yangtze River is 980 billion cubic meters, accounting for more than 35% of the total water resources in China. The Yangtze River ensures the livelihood and ecological water resources for the provinces along the Yangtze River basin and supplies water for other provinces, such as Henan, Hebei, and Tianjin. The Yangtze River Basin is located in the subtropical monsoon climate zone with two distinct climates in the rainy season and the dry season. The Yangtze River system and the Yangtze River wetland ecosystem can store floods and reduce the frequency of floods and droughts, and they are important barriers to the ecological security of the basin [16, 17]. To sum up, the Yangtze River wetland ecosystem is the core capital of the Yangtze River Economic Belt strategy. Figure 4 illustrates the changing trend of ecosystem service value in the Yangtze River Economic Belt, and Figure 5 reveals the changing

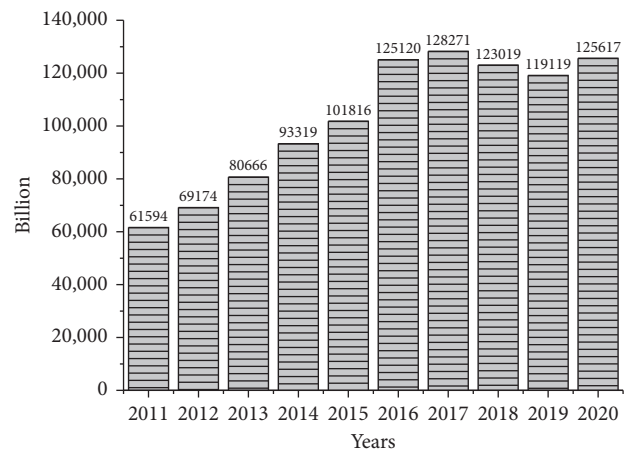


FIGURE 4: Change trend of ecosystem service value in Yangtze River Economic Belt.

trend of ecosystem service value of administrative units in the Yangtze River Economic Belt.

There are three deficiencies in wetland ecological management in the Yangtze River Economic Belt. The first one is multimanagement and piecemeal management, in which the overall interests often give way to local interests, and the interests of the vulnerable groups are deprived by the strong groups. The governance of the Yangtze River Basin still lacks a powerful comprehensive management organization. The second is that wetland protection lacks legislative authority. Currently, China's land law defines a wetland as unused land. Therefore, the wetland has become the victim of the land policy of balance between occupation and compensation, from which a few people get exorbitant profits while violating the interests of the mass [18]. The third is the weak awareness of wetland protection. Most people have not realized the severity of wetland ecosystem degradation in the Yangtze River, and they lack an accurate understanding of the evolution status and future trend of the Yangtze River wetland, which leads to unfriendly ecological protection environment [19].

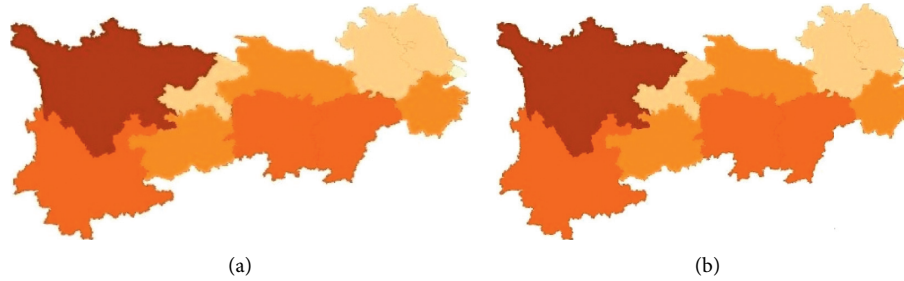


FIGURE 5: Changes of ecosystem service value of administrative units in Yangtze River Economic Belt: (a) 2019 and (b) 2020.

3.3. Analysis of the Yangtze River Water Environment Ecosystem and Existing Problems. The Yangtze River basin is a strategic water source in China with its vast territory, diverse internal water environment, and rich biological resources. The Yangtze River bears huge pressure from regional development and plays a key role in ecological security. There are about 300 fish species in the Yangtze River Basin, of which about 170 species are unique to the Yangtze River Basin. Among them, rare birds, such as Siberian Crane and Chinese Merganser, are also widely distributed in the Yangtze River Basin. Additionally, the water environment and ecosystem of the Yangtze River Basin provide shelters for the East Asia-Australia migratory birds. According to the evaluation results of relevant agencies, the Yangtze River Basin occupies 5 of the top 10 in the ranking of 1,031 important wetlands for the complete migration route of migratory birds. However, the current ecological imbalance of the Yangtze River Basin is becoming more and more serious. It is necessary to deeply analyze the outstanding problems in the Yangtze River Basin and take corresponding measures to solve them.

According to relevant research, the problems existing in the ecological governance of the Yangtze River Economic Belt can be summarized into the following three aspects. The first is the improper use of environmental protection funds, which is manifested in the unbalanced use of funds, unreasonable use of funds, or ineffective use of funds. By 2020, a total of 1.3 billion RMB of funds from the comprehensive management of the eight provinces and municipalities along the Yangtze River was deposited into the local financial department, and about 900 million RMB has been deposited into the project authority and related construction units. Additionally, the government and other relevant departments in the eight provinces and municipalities along the Yangtze River illegally used 200 million RMB of ecoenvironmental protection funds to compensate for administrative expenses or project expenditures.

The second problem is the insufficient protection of the ecological environment, which is manifested in the unauthorized provision of water resources, and the lack of oversight on illegal network sales. Data analysis indicates that more than 500 enterprises from different provinces and municipalities along the Yangtze River have sold water resources without permission. More than 50 enterprises have exceeded the water consumption limit. Meanwhile, 600 coastline projects that illegally occupy land from several provinces have not been demolished. Thus, the protection of

ecological diversity in many provinces and municipalities has not been implemented at the grassroots level. To obtain a harmonious development with nature, the intensity of further punishment should be increased and real-time systems should be employed.

Third, there are material weaknesses in pollution control. Specifically, the sewage disposed by the sewage treatment plants does not meet the corresponding standards, and the waste disposal in landfill plants does not being carried out effectively. By December 2020, the sewage disposed by many sewage treatment plants in several provinces in the Yangtze River Basin has failed to meet the national Class A discharge standards. The sewage treatment capacity is limited due to the damage to the irrigation and drainage network. In 2020, five provinces and municipalities directly discharged 221 million tons of sewage into the Yangtze River. A total of 45 enterprises in 7 provinces have failed to dispose of the hazardous waste properly, 10 garbage treatment plants in 3 provinces are overloaded, and 5 health care enterprises in 4 provinces failed to dispose of the medical waste according to the regulations.

Figure 6 illustrates the distribution of some chemical parks in the Yangtze River Economic Belt.

4. Case Analysis of Ecoenvironmental Performance Audit in the Yangtze River Economic Belt

4.1. The Background of Ecoenvironmental Performance Audit in the Yangtze River Economic Belt. The Yangtze River is the longest in China. The Yangtze River Economic Belt represents the economy zone near the Yangtze River. The Yangtze River Economic Belt covers 11 provinces and municipalities including Guizhou, Yunnan, Sichuan, Chongqing, Hunan, Hubei, Jiangxi, Anhui, Zhejiang, Jiangsu, and Shanghai and covers an area of about 2.05 million square kilometers, with a GDP and population exceeding 40% of those in China. After decades of development since the reform and opening up, the Yangtze River Economic Belt has become one of the most developed regions in China with a high level of economic development, dense populations, and a complete industrial chain. However, there are many urgent problems with the development of agriculture, industry, and commerce in the Yangtze River Economic Belt, which are mainly the conflicts of interest caused by the inadequate/an imbalance of economic development between regions [20]. The

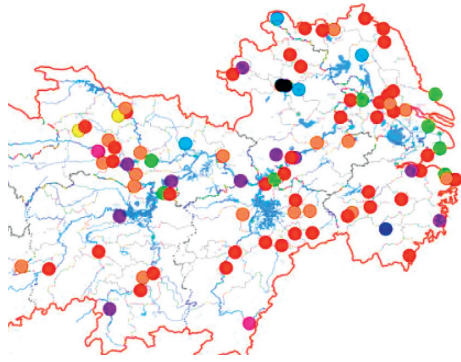


FIGURE 6: Distribution of some chemical parks in the Yangtze River Economic Belt.

political division of the Yangtze River Economic Belt is shown in Figure 7. The basic situation of each province and municipality in 2020 is shown in Figure 8.

In December 2020, the National Audit Office has organized relevant auditors to audit the implementation of policies for ecoenvironmental protection and the use of funds in cities along the Yangtze River Economic Belt and has achieved good audit results. Here, we cite audit results to analyze and summarize audit experience, to provide a reference for the audit work in other regions. The related concepts of ecoenvironmental performance audit are shown in Figure 9. Environmental performance audit is the combination of environmental audit and performance audit. Environmental audit is the process of supervising, authenticating, and evaluating the environment of an organization. Performance audit refers to the audit of the economy, efficiency, and effect of the economic activities of the auditee. Meanwhile, environmental performance audit is an activity in which the audit institution (i.e., the national audit institutions, internal audit institutions, or accounting firms) supervises and evaluates the environmental management system of the auditees as well as the environmental problems and responsibilities arising from economic activities [21].

4.2. Audit Project Implementation

4.2.1. Preaudit Investigation Stage. In the preaudit investigation stage, the auditors should determine the scope, objectives, and objects of the audit investigation [22]. The audit objectives can be divided into overall objectives and specific objectives. The overall objective of the ecoenvironmental performance audit is to judge whether the local governments of the provinces along the Yangtze River Economic Belt have implemented the ecological and environmental protection policies by the planning outline. The specific objectives can be separated into two aspects. The first one is to evaluate whether the special funds are used in accordance with regulations and whether there is any surplus or misappropriation. The second one is about resource exploitation and the prevention and control of pollution, which is to evaluate whether resources are overexploited or illegally extracted.



FIGURE 7: Political division of the Yangtze River Economic Belt.

4.2.2. Audit Method. In the audit process, auditors use the inspection method, inquiry method, letter method, and comprehensive analysis method to obtain reasonable audit evidence to ensure the credibility of audit results. When evaluating the use of special funds, auditors shall inquire the user of funds and the project manager, verify the use of special funds, and check the cash accounts and approval documents of special funds through inspection. Besides, auditors verify the integrity and existence of funds through the letter of confirmation.

4.3. Audit Results. The results of the ecoenvironmental performance audit can be divided into the following three aspects: (1) high-pressure management and use of funds related to ecoenvironmental protection: there are two kinds of problems in this aspect. The first one is that financial departments retain special funds, and the second one is that financial departments do not implement financial support funds. Specifically, by the end of 2020, a total of 1.3 billion RMB was earmarked for water pollution control in eight cities along the Yangtze Economic Belt, while 200 pollution treatment projects in 10 cities along the Yangtze River Economic Belt have yet to start. (2) Resource development and ecological protection: there are two types of problems in this aspect. The first is that the project density is greater than the bearing standard, and the second is the illegal water intake. By the end of 2020, the provinces and municipalities in the Yangtze River Economic Belt have completed the construction of 250 million small hydropower stations, and the minimum distance between small hydropower stations is only 100 meters, so the construction intensity of small hydropower stations is quite high. In the 10 provinces and municipalities in the Yangtze River Economic Belt, more than 400 enterprises take water without licenses, and more than 100 enterprises take water in excess. The newly built high polluting projects in many provinces and municipalities have not yet gone through the EIA (Environmental Impact Assessment) approval procedures. (3) Pollution control: in terms of pollution control, there are two problems. Firstly, the governance policies are not being carried out effectively, and secondly, the projects are not started in time or do not meet the expected requirements. Specifically, the water quality of the national key lakes which have been under long-term treatment is still very poor, and the centralized sewage treatment equipment has not been established in many development zones.

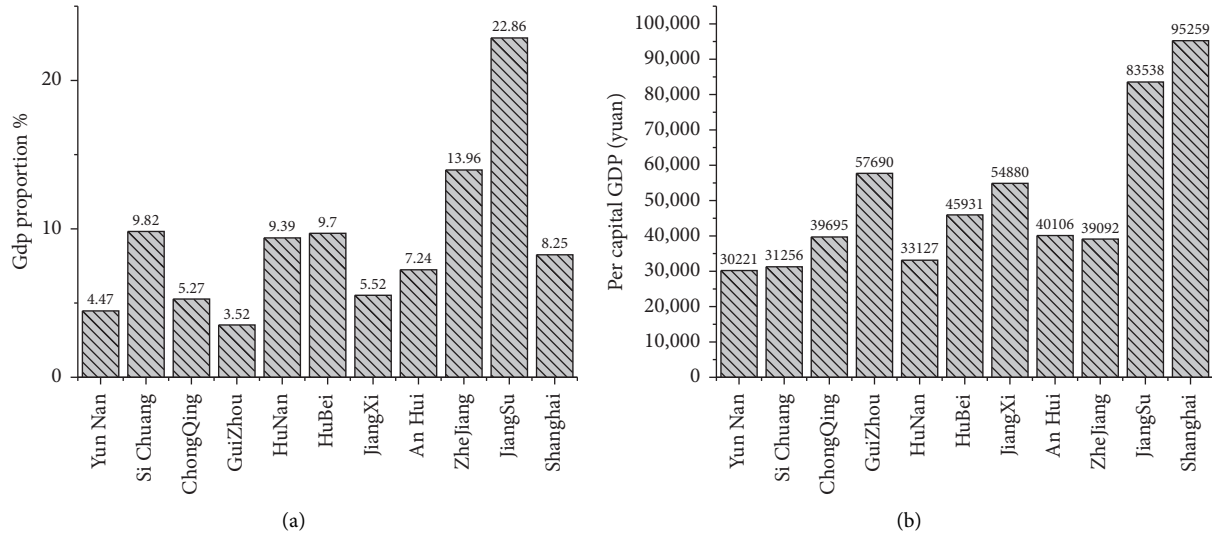


FIGURE 8: The basic situation of each province and municipality in 2020. (a) GDP proportion and (b) GDP per capita.

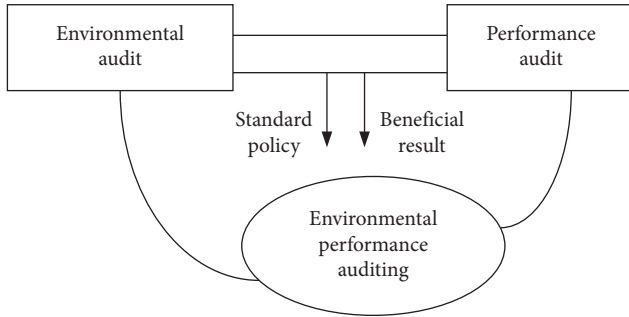


FIGURE 9: The concept map of ecoenvironmental performance audit.

4.4. Analysis on the Problems of the Ecoenvironmental Performance Audit in the Yangtze River Economic Belt. According to the actual situation in China, environmental performance audit has not formed a relatively well-developed audit mode such as financial audit, and it is still in the initial and exploratory stage. At this stage, there are many problems in the ecoenvironmental performance audit of the Yangtze River Economic Belt which can be summarized as follows. First, the proprietorship of water resources in the River Basin is not clear, and the root of the confusion lies in the lack of clear provisions in the water law. The inconsistency of administrative regulations and departmental rules leads to the lack of evaluation criteria in the ecoenvironmental performance audit. Second, the overlapping of government administrative powers leads to the inefficiency of the ecoenvironmental performance audit of the Yangtze River Economic Belt. Nowadays, the management of the Yangtze River Economic Belt mainly adopts a single management mode, which leads to the serious overlapping of functions. As for water resources management, due to the overlapping of functions, all departments try to benefit from the affairs with greater interests, while all departments are unwilling to manage the affairs without interests, which set up so many obstacles to protect the environment of the

Yangtze River Basin and also cause the inefficiency of the audit. Third, the integrity and long-term nature of the environmental pollution problem increased the difficulty of ecoenvironmental performance audit. The problem of water environmental pollution in the Yangtze River Basin is caused by the long-term activities of human beings and cannot be formed and spread in a short time. It is hard to tackle water environmental pollution in a short period due to its complex causes. In the process of water pollution control, different stakeholders have different goals and make different decisions, which makes the environmental standards inconsistent. In this case, it is difficult to solve the pollution problem only by the ecoenvironmental performance audit.

4.5. Optimization Suggestions for Improving the Audit of the Yangtze River Economic Belt. The quantity of natural resources in the Yangtze River Economic Belt is far less than the demand of humans, and the development of administrative regions of the Yangtze River Economic Belt is uneven, provinces in the lower reaches of the Yangtze River Economic Belt have developed better than those in the upper reaches. Due to geographical factors, the provinces in the lower reaches of the Yangtze River Economic Belt can achieve ecoenvironmental protection results in the upper reaches of the Yangtze River Economic Belt at a relatively low cost. With the development of transportation, the possibility of cooperation between cities in different regions of the Yangtze River Economic Belt is increasing. The cooperation between provinces or cities is conducive to the reallocation of resources and the overall progress and development of the basin [23]. The purpose of the ecoenvironmental performance audit is to further promote the improvement of national governance objectives and governance capacity. Therefore, auditors should focus on the development planning of the Yangtze River Basin. During the audit, auditors should strictly review the rationality of

laws and regulations for the protection of the Yangtze River Basin and evaluate the medium and long-term environmental strategic planning of the Yangtze River Basin, thereby promoting the overall development of the economy, environment, and ecological construction in the Yangtze River Basin. Besides, the auditors should learn to use 3S technology to conduct the ecoenvironmental performance audit. The so-called 3S technology, in fact, is remote sensing (RS), geographic information system (GIS), and global satellite positioning system (GPS) collectively. In the ecoenvironmental performance audit of the Yangtze River Economic Belt, auditors can use 3S technology to analyze the area and distribution of the natural resources and make a comparative analysis with the previous year, to evaluate and judge the development and utilization of natural resource assets. The application of 3S technology can effectively improve the accuracy of audit results.

5. Conclusions

As the government attaches more and more importance to environmental protection, government audit plays an increasingly important role in environmental supervision. Global pandemic situations triggered more and more audit practices to study the ecoenvironmental audit, to improve the ecoenvironmental audit system, and to improve the healthcare system. In this paper, the Yangtze River Economic Belt in the ecoenvironmental protection audit is specifically analyzed using the literature review, case analysis, and inductive analysis from many aspects, such as the audit concept, organization and implementation, technical methods, and result application. Meanwhile, the problems existing in the ecological environmental management of the Yangtze River Economic Belt are analyzed through environmental data, such as the relevant yearbook and literature. The results find that the current audit evaluation system is not perfect, and the reliability of audit results is not high. Consequently, the optimization suggestions are put forward for improving the audit work of the Yangtze River Economic Belt. Our research concludes that auditors should focus on the national environmental protection policies, strictly implement environmental performance audit procedures, and learn to use advanced technologies to conduct the ecoenvironmental performance audit to improve the quality of ecoenvironmental performance audit. However, there are still some deficiencies in this paper. During the audit analysis, we only conducted the environmental performance audit on water resources, while audit on other resources has not been performed. In the future work, with the development of environmental audit technology, audit on other environmental pollution problems such as carbon emission will be further improved.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work was supported by the 2017 general project of Humanities and Social Sciences Research in Universities in Anhui Province (project number: SK2017B05); the 2017 major project of Humanities and Social Sciences Research in Universities in Anhui Province (project number: SK2017A0417); the 2019 project of Chuzhou University (project number: RJZ01); the 2021 project of Scientific Research and Practice Innovation for Postgraduate in Jiangsu Province (project number: KYCX21_0391).

References

- [1] Y. Pan, Z. Liu, and Y. Wang, "Research on pricing of renewable water resources in supply chain environment," *IOP Conference Series: Materials Science and Engineering*, vol. 612, no. 5, Article ID 052023, 2019.
- [2] M. N. Chisola, M. v. der Laan, and K. L. Bristow, "A landscape hydrology approach to inform sustainable water resource management under a changing environment. A case study for the Kaleya River Catchment, Zambia," *Journal of Hydrology: Regional Studies*, vol. 32, no. 1, Article ID 100762, 2020.
- [3] L. Meng, W. Wang, T. Li, C. Liao, L. Zhao, and Y. Chen, "Evaluation of the effects of shear stress on crucian carps passing through turbines," *IOP Conference Series: Earth and Environmental Science*, vol. 774, no. 1, Article ID 012147, 2021.
- [4] Q. Jiang and Q. Tan, "National environmental audit and improvement of regional energy efficiency from the perspective of institution and development differences," *Energy*, vol. 217, Article ID 119337, 2021.
- [5] K. Nazarova, V. Hotsuliak, V. Miniailo, M. Nezhyva, and V. Mysiuk, "Accounting, analysis and environmental audit as an imperative of the development of green economy in the state's economic security system," *E3S Web of Conferences*, vol. 166, Article ID 13003, 2020.
- [6] M. Marwa, B. Salhi, and A. Jarboui, "Environmental audit and environmental disclosure quality," *Scientific Annals of Economics and Business*, vol. 67, no. 1, pp. 93–115, 2020.
- [7] C. A. Silva, C. R. Duarte, J. A. B. Sabadia, and M. V. S. Souto, "Drone in the environmental audit: potentialities and applications," *Anuário do Instituto de Geociências - UFRJ*, vol. 41, no. 3, pp. 207–215, 2018.
- [8] F. Morante-Carballo, N. Montalván-Burbano, P. Carrión-Mero, and K. Jácome-Francis, "Worldwide research analysis on natural zeolites as environmental remediation materials," *Sustainability*, vol. 13, no. 11, p. 6378, 2021.
- [9] K. Wu and M. Tian, "Research on environment education strategy water resource utilization based on dynamics," *Journal of Coastal Research*, vol. 115, no. sp1, p. 498, 2020.
- [10] M. Polemio and K. Walraevens, "Recent research results on groundwater resources and saltwater intrusion in a changing environment," *Water*, vol. 11, no. 6, p. 1118, 2019.
- [11] V. Mikhailov, N. Kudrevatykh, and T. Tyuleneva, "The research of environmental-and-economic risks of the coal mining enterprise impact on water resources," *E3S Web of Conferences*, vol. 134, no. 2, Article ID 01019, 2019.
- [12] W. Xin, T. Can, W. Wei, and L. Ji, "Change detection of water resources via remote sensing: an L-V-nstc approach," *Applied Sciences*, vol. 9, no. 6, pp. 1223–1235, 2019.
- [13] Z. Lv, X. Li, and H. Lv, "BIM big data storage in WebVRGIS," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 99, p. 1, 2019.

- [14] L. Kong, H. Zheng, E. Rao, Y. Xiao, Z. Ouyang, and C. Li, "Evaluating indirect and direct effects of eco-restoration policy on soil conservation service in Yangtze River Basin," *The Science of the Total Environment*, vol. 631-632, pp. 887–894, 2018.
- [15] F. Zhang, Z. Zhang, R. Kong et al., "Changes in forest net primary productivity in the Yangtze River basin and its relationship with climate change and human activities," *Remote Sensing*, vol. 11, no. 12, p. 1451, 2019.
- [16] S. Lu, X. Tang, X. Guan, F. Qin, X. Liu, and D. Zhang, "The assessment of forest ecological security and its determining indicators: a case study of the Yangtze River Economic Belt in China," *Journal of Environmental Management*, vol. 258, Article ID 110048, 2020.
- [17] D. Zhang, X. Wang, L. Qu et al., "Land use/cover predictions incorporating ecological security for the Yangtze River Delta region, China," *Ecological Indicators*, vol. 119, Article ID 106841, 2020.
- [18] M. Silver, M. Barosky, and C. Aviles, "Water audit expands reuse opportunities at the brewery," *World Water: Water Reuse & Desalination*, vol. 8, no. 4, pp. 21–23, 2017.
- [19] X. Li, X. Yu, L. Jiang, W. Li, Y. Liu, and X. Hou, "How important are the wetlands in the middle-lower Yangtze River region: an ecosystem service valuation approach," *Ecosystem Services*, vol. 10, pp. 54–60, 2014.
- [20] A. V. Glushchenko, I. F. Gorlov, N. V. Filipov, D. A. Mosolova, E. P. Kucherova, and N. I. Mosolova, "Internal ecological audit of environmental facilities of agricultural enterprises," *IOP Conference Series: Earth and Environmental Science*, vol. 677, no. 3, Article ID 032047, 2021.
- [21] R. Huang and Y. Li, "Undesirable input–output two-phase DEA model in an environmental performance audit," *Mathematical and Computer Modelling*, vol. 58, no. 5-6, pp. 971–979, 2013.
- [22] S. Aslam, R. U. Rehman, and M. Asad, "Linking environmental management practices to environmental performance: the interactive role of environmental audit," *Pakistan Journal of Commerce and Social Sciences (PJCSS)*, vol. 14, no. 1, pp. 99–119, 2020.
- [23] S. Fu, H. Zhuo, H. Song, J. Wang, and L. Ren, "Examination of a coupling coordination relationship between urbanization and the eco-environment: a case study in Qingdao, China," *Environmental Science and Pollution Research*, vol. 27, no. 19, pp. 23981–23993, 2020.

Research Article

Somewhat Homomorphic Encryption: Ring Learning with Error Algorithm for Faster Encryption of IoT Sensor Signal-Based Edge Devices

V. Subramaniaswamy,¹ V. Jagadeeswari,¹ V. Indragandhi,² Rutvij H. Jhaveri ,³
V. Vijayakumar,⁴ Ketan Kotecha ,⁵ and Logesh Ravi⁶

¹School of Computing, SASTRA Deemed University, Thanjavur, India

²School of Electrical Engineering, Vellore Institute of Technology, Vellore, India

³School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

⁴School of Computer Science and Engineering, University of New South Wales, Sydney, Australia

⁵Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India

⁶Department of Computer Science and Engineering,

Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

Correspondence should be addressed to Rutvij H. Jhaveri; rutvij.jhaveri@sot.pdpu.ac.in

Received 16 December 2021; Revised 10 January 2022; Accepted 17 January 2022; Published 24 February 2022

Academic Editor: Mamoun Alazab

Copyright © 2022 V. Subramaniaswamy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, Homomorphic Encryption (HE) has shown the possibility of securely running a computation arbitrarily without performing the data decryption. Many authors have shown Somewhat Homomorphic Encryption (SHE) or Fully Homomorphic Encryption (FHE) schemes implemented practically on both the addition and multiplication operations for SHE. The recent methods for implementing the FHE methods completely depend on arbitrarily reducing the time taken to perform the encrypted multiplication operation to increase the computation power required by SHE methods. This paper aims to accelerate the encryption primitives in an integer-based SHE based on the duration between each data transmission from the sensor and data packaging method. If the number of sensors increases exponentially in an edge device environment, the signals have to be encrypted faster in a packed mode in the edge environment and transferred to the cloud without a loss in data. The presented SHE method reduces the time taken for encryption based on the input number from the sensor and invariably increases the performance of the edge device. This advantage also helps the deploying healthcare application obtain end-to-end privacy in transmitting sensitive patient data.

1. Introduction

The ecosystem required for the Internet of Things (IoT) is growing exponentially because of the large-scale availability of low-cost sensors, actuators, microprocessors, and high-speed Internet infrastructure. These devices can be integrated seamlessly for gathering data from the required environment, depending on the application. The collected data have to be processed and monitored continuously for effective implementation. The healthcare industry is one of the largest revenue-generating sectors in India with a market

share of 133 billion USD by 2022 [1]. The continuous monitoring of hospitalized patients and a timely prediction of complications that arise from disease leads to early recovery and reduces patient hospital stay. Apart from increased revenue, the reduced hospital stay of the patient will translate into less strain in the current healthcare infrastructure as well as saving the patient lives through the early detection of diseases.

With the established Internet framework, multiple sensor devices integrated with individual patients transmit their vital information through the network in the Edge

Computing (EC) [2–4] environment before processing critical information in cloud computing. In this current scenario, the transmitted data from the device will be transferred insecurely through wifi networks and then to the centrally located servers. There is a possibility of altering the patient details and vital signs through hacking. They are transmitted through the nonsecured integrated IoT devices to cloud computing for further processing. In this scenario, the patient will be misdiagnosed, which also leads to further complications. This type of problem can be solved by adding a layer of security before transmitting the patient’s vital signs into the cloud server for further processing and timely prediction. The time is taken to encrypt the patient data, send the encrypted data, process the encrypted data, and transmit urgent messages in case of emergency to the healthcare provider as early as possible. Another scenario is the number of patients increased by 100-fold from the average number of available beds at any particular time also leads to network congestion or bottleneck in transmitting the data through IoT devices. This paper explains a new method using Fully Homomorphic Encryption with reduced encryption time before transmitting data when compared to the MORE or PORE method. Also, this new method helps in the elimination of bottleneck problems. Even with the increase in data, the time required for processing in the Edge Computing environment is less before sending the data to the cloud environment on a priority basis. The computation method increased the performance in the edge environment by carrying a data prediction in Edge Computing and storing and analyzing data in the cloud [5–8].

This paper focuses on the security issues required while transmitting data from edge devices into a cloud server. We specifically concentrate on preserving the privacy of patients’ information and enhancing the security of transmitted data by encrypting them with a modified lightweight algorithm based on Somewhat Homomorphic Encryption-Ring Learning with Error. This presents a new way to securely encrypt IoT sensor signal value based on the frequency of transmitting the signal to the edge environment. Depending on the encryption function selected, the edge device will require less computing space and time to encrypt and transmit the data to the cloud server. Table 1 presents the list of abbreviations used in this paper to help readers to have a better understanding.

2. Motivation

The healthcare system delivers a quality of service to the people whenever and wherever they need to increase the quality of life and decrease the mortality rate. Healthcare in public life is mainly due to the advanced need to stay healthy and contribute to the economy for a long period. The health and lifestyle do not match at all period.

Chronic diseases are noncommunicable diseases that increase due to changes in the lifestyle of people. For instance, dramatic changes in food quality, mainly an increase in fast food, excess weight gain, work stress, alcohol, and so on, are the main reasons for chronic diseases. As per the World Health Organization (WHO) report, it increases the

TABLE 1: List of abbreviations used.

| | |
|------|---|
| SHE | Somewhat Homomorphic Encryption |
| FHE | Fully Homomorphic Encryption |
| HE | Homomorphic Encryption |
| IoT | Internet of Things |
| USD | United States Dollar |
| EC | Edge Computing |
| WHO | World Health Organization |
| HBP | High Blood Pressure |
| MORE | Matrix Operation and Randomization Encryption |
| TTP | Trusted Third Party |
| EHR | Electronic Health Record |
| ECG | Electrocardiogram |
| SHA | Secure Hash Algorithm |
| RLwE | Ring Learning with Errors |

mortality rate. Hypertension [9] is one of the chronic conditions in which blood pressure increases in the arteries also called High Blood Pressure (HBP). HBP does not usually cause symptoms, but it leads to a major risk influence in heart failure, kidney failure, heart attack, artery disease, and so on. The systolic and diastolic are the two measurements expressed to measure blood pressure. The systolic pressures are within the range of 100–130 mmHg and for diastolic 60–80 mmHg. The BP is at or above 130/80, or 140/90 mmHg is diagnosed to be hypertension. It is further classified into primary and secondary hypertension. The HBP is classified into gestational hypertension, preexisting hypertension, and preeclampsia during pregnancy. It is essential to monitor pregnant women continuously because hypertension causes globally 16% of maternal death approximately.

The secure send and storing of personal data without being compromised play a major role in protecting the privacy of the patient [10–12]. The patient’s health records or diagnosis stored over the cloud-based server need to be accessed by any physician at any time without being compromised or transferred securely.

In general, Figure 1 depicts the IoT-based fog computing healthcare monitoring system with security consisting of three steps. First, the patient’s health data, especially blood pressure data, are collected by placing IoT smart matches. The collected data are sent to fog computing through wireless networks such as Bluetooth, Wi-Fi, and WiMAX. Second, fog is an extension of the cloud for the analysis of data where we use mobile in the place of fog to perform the security operation. The Homomorphic Encryption technique is carried out in fog computing. Third, the secured data are stored in the cloud subsystem.

If a patient is getting treatment from multiple hospitals, the patient’s records are converted into Electronic Health Record (EHR) to be used anywhere at any time for future reference. The way of converting patients’ medical records into electronic format has been implemented throughout the world with the sole purpose of available patient records accessed throughout the world at any point in time. The converted medical records were stored in the local server maintained by the institution or in the cloud for easy access anywhere in the world, with the availability of new sensors

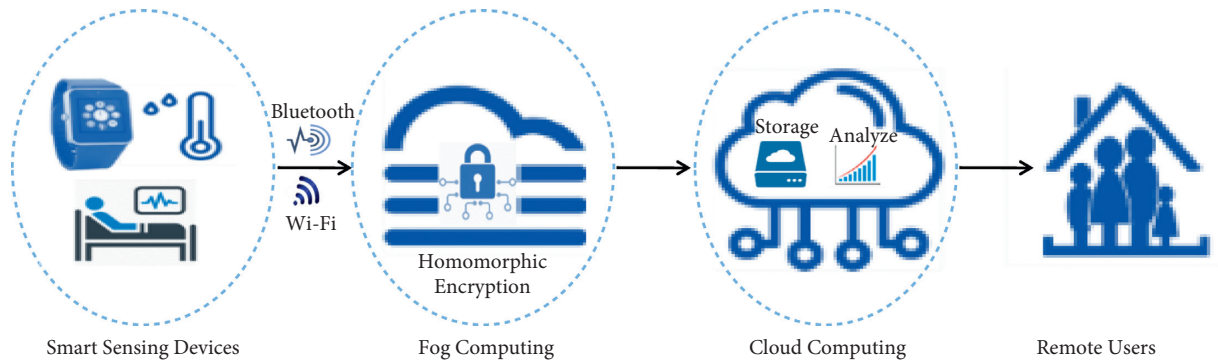


FIGURE 1: A conceptual framework of IoT-based fog computing healthcare system.

which can be integrated with the patient to monitor their health in real-time and store it in a server or cloud-based system. In case of emergency, integration of the entire tool alerts the medical crew; accurate and timely treatment at the hospital was carried out seamlessly in a matter of minutes. The systematic collection of that information is static and dynamic. The static contains the patient's personal information such as name, age, and gender that need not be changed, while dynamic information contains the streaming information generated by sensors which are changed from time to time.

In the combination of the Internet of Things, big data analytics, and cloud computing, the patient is continuously monitored, and disease is predicted using ensemble methods. Then the proposed method is to securely store patient data and analyze them without requiring personal information. The next step is to process the stored data and diagnose based on disease symptoms. The final step is to alert the required medical personnel without compromising patient privacy. This paper only concentrates on security and privacy for the IoT healthcare data and patient information stored in EHR.

The remainder of the paper is organized as follows: Section 2 describes the relevant works as a literature survey, and Section 3 introduces Fully Homomorphic Encryption. Later Section 4 describes the experimental setup and discusses in detail the results obtained. Finally, Section 5 concludes with a summary of the work and provides future work directions.

3. Literature Survey

In order to reduce the delays inferred by cloud computing, Fan et al. [13] proposed fog-cloud computing to increase efficiency. The authors also addressed the security and privacy challenges using ciphertext policy-attribute-based encryption. Hariss et al. [14] designed a MORE (Matrix Operation and Randomization Encryption) approach Homomorphic Encryption to provide privacy for real-world applications. Sanchez-Guerrero et al. [15] proposed a novel adaptive extended Merkle tree structure to provide privacy to extremely sensitive information stored in EHR. However, all these schemes were applied to already existing stored data, not streaming data.

Pham et al. [7] proposed a smart home healthcare service for the elders who are staying alone using various IoT sensors as a real-time application in a cloud environment. The authors also developed a robot assistant using the gradient boosting decision tree algorithm to find the activities of the body. A cloud-based system is proposed [16] to control and monitor the H1N1 virus. Doctors upload the patients' treatment information in the Amazon EC2 service for future reference in this system. To diagnose Chikungunya, Sood and Mahajan [17] designed a fog-assisted cloud-based system in which users are continuously monitored to collect information stored in the cloud and analyzed in Edge Computing. With the advancement of IoT, Rani et al. [18] also diagnosed Chikungunya and stored patients' information in the cloud. Mobile environment jointly merged with cloud computing to monitor ECG pattern of the patients in mobile designed by Zhang et al. [19]. The system has the details of patients such as name, identification number, gender, age, medical record, and ECG report. However, all these systems fail to meet security measures to prevent user's sensitive health-related data from unauthorized access.

Using the condition-based methodology, Verma and Sood [20] developed a cloud-centric diagnostic system to predict the possible disease in users using medical devices and sensors [21]. The user's personal information and their diagnosed diseases were stored and analyzed in the cloud. Here, Trusted Third Party (TTP) provides security to the users' information, but it can be hacked easily with the public key. Hossain and Muhammed [22] proposed a framework for monitoring Electrocardiogram (ECG) of disabled and older people. The monitoring information is gathered at the cloud platform to be accessed everywhere at the cost needed. Watermarking and signal enhancement techniques are used to secure the data on the client side.

In case medical treatment is done for the patient in a remote network-constrained environment, the medical data have to be transferred securely, and only relevant information can be accessed by authorized users such as healthcare providers. The secure encryption and transmission of data without any loss is also a vital part of security. The data from the sensors will be transmitted to the local edge device; the device further encrypts the data using a public key. The batch process of encryption is done using a new algorithm to encrypt to form the ciphertext whose final

size is larger than the original text size. If a hacker accesses the secure data, the data can only be viewed as a random number and is not possible to decrypt without a secure private key.

The FHE encrypted data transmitted through the edge device will perform the error estimation of the transmitted data. The final data analytics in the cloud server can be performed without any decryption for processing the final output. If the processed data breaches a certain threshold set by the healthcare provider, an immediate alert message will be delivered to the corresponding physician for further diagnosis.

4. Fully Homomorphic Encryption (FHE)

In the FHE method, the security key generated for encryption consists of both private key and public key for secure encryption. The general steps in the process involve four steps such as generation of the encrypted key (symmetric or asymmetric), encryption of data into ciphertext, decryption of the ciphertext using a private key, and evaluation of the transmitted data. In the symmetric encryption method, the general public key will be used for both encryption and decryption of data. The asymmetric encryption method is done using a general public key, while the decryption is done using a secure private key. The final text will be processed in an encrypted ciphertext without decrypting the original message sent by the sensors. In order to evaluate the robustness of the encryption method, the final processed ciphertext will be decrypted using the private key, and the plain final processed results will be displayed.

Fully Homomorphic Encryption technique encrypts the patient's input information to produce the ciphertext without knowing any information about the plaintext, which matches the operation performed on plaintext. Homomorphic Encryption is said to be FHE when it satisfies both properties of addition and multiplication. The basic properties are as follows:

For addition,

$$[\text{Enc}_k(a) + \text{Enc}_k(b)] \bmod X = [\text{Enc}_k((a + b) \bmod X)] \cdot \bmod(X). \quad (1)$$

For multiplication,

$$[\text{Enc}_k(a) * \text{Enc}_k(b)] \bmod X = [\text{Enc}_k((a * b) \bmod X)] \cdot \bmod(X). \quad (2)$$

The mobile IoT device where the encryption occurs should protect the patient's real identity from public view, as well as in case of an emergency; authorized users can effectively trace the patient using the mobile edge device. The identity can also be morphed by adding a randomly generated pseudonym to the user, but this will further increase the computational cost for each random generation of names in resource-constrained IoT encrypted terminals.

4.1. Concept and Proposed Model. The major drawback of the FHE is the size of the encrypted ciphertext and the minimum storage space required in the cloud server, which is directly

related to the performance of the analysis. For an FHE to be fully realizable in the real application with high security, the encrypted file size should be reduced, and the time taken for encryption should be appreciably reduced with the increase in the number of connected sensors. FHE scheme implemented in a low-level language and using a parallel processor reduces the edge domain's encryption time. These are packed to a certain extent and implemented using Somewhat Homomorphic Encryption. The edge device also performs the threshold comparison with the encrypted data without knowing the raw transmitted data, and the result will be a simple yes or no. The basic point in privacy is preserving encrypted data by evaluating some ideal properties such as accuracy, no reversibility, diversity, revocability, randomness, and performance.

The Edge Computing devices, if implemented properly, will have less latency and stringent quality of service requirements. Figure 2 presents the architecture of polynomial Homomorphic Encryption in the healthcare system. The main idea is to bring the core cloud computing to mobile devices, connected sensor devices, and actuators. Most of the edge devices will be on the move. Therefore, the transmitted signal should not get retransmitted or duplicated in another edge node. This will further lead to memory constraints in the cloud computing storage.

The proposed model will strip the sensor data into a simple number with comma-separated phrases and fill them into a linear algebraic expression. Each polynomial will be added with a simple operator, and the idea is to shorten the sent polynomial into a manageable packet size. The original data will be added to the final text for further processing.

The network connection between the sensor terminals and the edge device is where most of the attacks can take place. The transmitted data from the sensor can be attacked by Eavesdropping, Sybil attack, sleep deprivation attack, and a man in the middle attack. The data are transmitted between the sensor and edge devices using SHA message verification techniques. The medical sensor signals have to be routed through a trusted secure network channel. Another way to prevent the attack is to add a new layer of security to the transmitted signal in the edge device before transmitting through the cloud network. One such type of protection requires less time to encrypt the continuous signal transmitted by the sensor into an encrypted signal using a public key provided by the user within a few milliseconds. Then, the encrypted signal is transferred to the cloud for further processing. The advantage of Fully Homomorphic Encryption is that the data need not be decrypted for processing; only if the signal value exceeds a certain threshold, will the authorized user receive the alert text message, including the physician in charge. The data will be decrypted by the authorized user using a private key provided by the medical institution. The physician has access only if it is decrypted by the user.

5. Preliminaries

In this section, we review our associated cryptographic schemes and mathematical concepts.

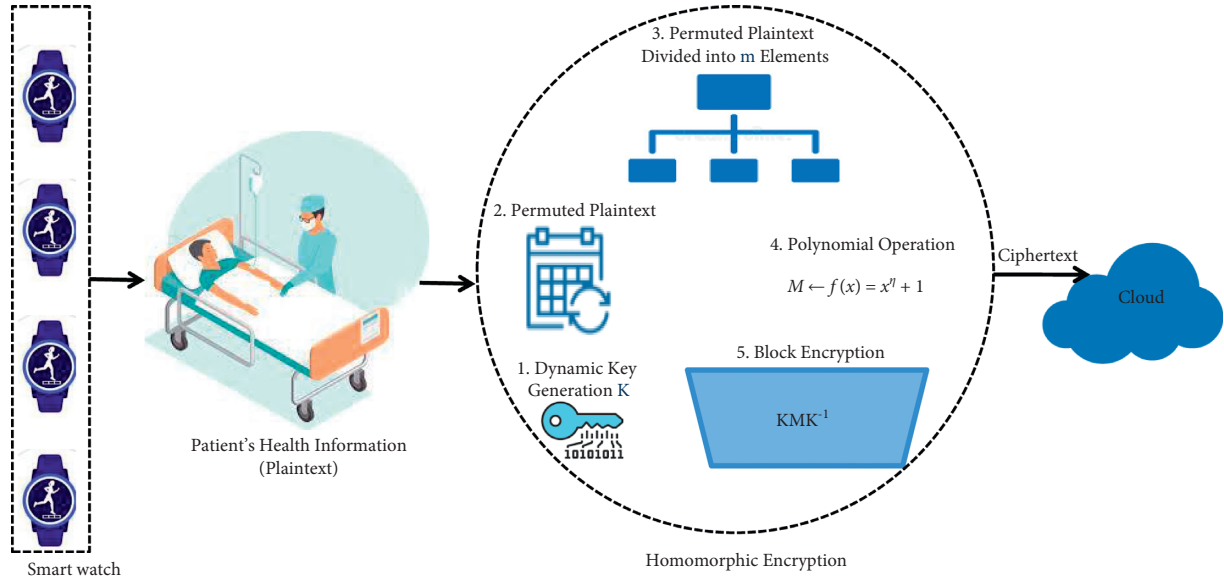


FIGURE 2: Architecture of polynomial Homomorphic Encryption in the healthcare system.

5.1. *Homomorphism.* Assume that G^* is a cyclic multiplicative group and G and G' are two cyclic additive groups of a prime order. Let kernel $K(f)$ be a kind of function from $G \rightarrow G'$ and a and b belong to G . φ is mapping from $G/K \rightarrow G'$, which satisfies the following properties:

- (i) φ is well-defined: if $f(a)$ is equal to $f(b)$, then $\varphi(ka) = \varphi(kb)$, $ka = kb$ if and only if $ab^{-1} \in k$. Then, $f(a * b^{-1}) = e'$.
- (ii) φ is one-one function: let $\varphi(ka) = \varphi(kb)$.

$$\begin{aligned}
 f(a) \Delta [f(b)]^{-1} &= e^{-1}, \\
 f(a) \Delta f(b^{-1}) &= e^{-1}, \\
 f(a * b^{-1}) &= e^{-1}, \\
 ab^{-1} &\in k, \\
 ka &= kb.
 \end{aligned} \tag{3}$$

- (iii) φ is onto function: for every a which belongs to G and $x \in G'$, $f(a) = x$; then, $\varphi(ka) = x$.
- (iv) φ is homomorphism: let $ka, kb \in G/k$; then $\varphi(ka \otimes kb) = \varphi(k(a * b))$.

5.2. *Secret Sharing Scheme.* A secret sharing scheme divides the secret k into x shares with a $y+1$ share that is unable to reconstruct. Based on Lagrange's theorem, the order of any subgroup H of a finite group G divides the order of G .

5.2.1. *Secret Construction Phase*

- Step 1: assign a polynomial function $f(z)$ with degree s , in which all coefficients belong to Z .
- Step 2: compute the share $x_i = f(i) \pmod{N}$ for $i = 1, 2, 3, \dots, n$.

Step 3: user publishes a list of n shares like (x_1, x_2, \dots, x_n) and each x_i is circulated among patients.

5.2.2. *Secret Reconstruction Phase*

- Step 1: any $y+1$ share $(x_{n_1}, x_{n_2}, \dots, x_{n_{y+1}})$ can be able to reconstruct the secret k .
- Step 2: Compute $x = f(0) = \sum_{i=n_1}^{n_{y+1}} x_i (\prod_{j=n_1}^{n_{y+1}} y_j / y_j - y_i \pmod{N})$.

The encryption method can compute the average, standard deviation, and regression coefficient of the prediction without performing any decryption algorithm. The method of Ring Learning with Errors (RLWE) from Peikert and Regev [23] has some assumptions described in the following. The ring $(r_i) = Z_p(x)/f(x)$; the ring has n degree of polynomials with coefficient in Z_p ; the assumption is that all the polynomial numbers of samples are of the same form with a random number and error number. The ciphertext is equivalent to the noise present in the error distribution instead of a uniform number. The polynomial $f(x)$ chosen for the operation is $x^n + 1$, with n having a power of two. The ring elements after multiplication will have lesser than L^2 norm multiplication of the component, which is basic polynomial multiplication, and the addition is componentwise coefficient addition. The error distribution is of the Gaussian form D with $\epsilon > 0$ based on the probability density function. The multiplication function performed on the ring Z_p will only increase the size of the ciphertext by a small amount and reduced time. The time, size of the key, and ciphertext are computed using MATLAB software before implementing in the edge devices environment.

5.3. *Somewhat Homomorphic Encryption Algorithm.* SHE = Enc (keygen, encryption, addition, multiplication, and decryption).

Step 1. Create the asymmetric encryption keygen (public and private key).

The parameter for key generation is based on a polynomial function $f(x) = x^n + 1$, where n has a power of two. The modulus s is a prime number that is related by modulus $(2n)$. Then, s variable defines the ring parameter $ri = \text{mod}(Z_p(x)/f(x))$ described earlier. The Gaussian functions have a discrete error value given by $\varepsilon = (D(Z_p), \varepsilon 1)$ with the standard deviation. The discrete Gaussian $D > 0$; then only the multiplication method will work. They have to be chosen based on the hidden security parameter K . The ring element will define the secret key (pv.K) with a uniform random element. An additional error term will be added to the key. The public key (pub.K) contains the less random element and an error term to be available for encryption by the edge device. $\text{pub.K} = \text{mod}(\text{number} * \text{key} + \text{mod}(t) * \varepsilon)$ and $\text{pv.K} = \text{mod}(\text{number} * \varepsilon - \text{pub.K})$. The private key is provided with the authorized person, whereas the public key for encryption is available on the edge device for encryption protocol.

Step 2. Encrypt the plain data into a secure ciphertext using a public key and private key.

The message or continuous number will be encrypted based on the degree of the polynomial (n) with a continuous or a particular text message. In general, the computation of the normal text (norm) or number into an encrypted ciphertext contains all the elements belonging to the general variable.

$$C_p = a_0 * x + t * x + \text{norm}(a_1) + \text{mod}(t) * x. \quad (4)$$

Step 3. Packet the additional data into a smaller ciphertext for easy transmission.

The packets from the edge device will be split into easy modules after encryption based on $\text{mod}(t)$ value. If the degree of the polynomial is larger, then the encryption time is also more. Depending on the encryption device, the public key will be selected based on the available RAM in the edge device for easy encryption and transmission time. The complexity increases with the increase in the number of devices connected to a single edge device.

Step 4. Process the encrypted text for finding the critical threshold for the patient transmitted data.

The decryption of the whole normal text or number can be performed with the appropriate error value added to the public key, which is also a part of a private key generation. The decryption method with less error is achieved by the RLwE method. In general, the decrypted text which will be computed for $\text{mod}(t)$ is given by

$$\text{decrypt text} = \sum_0^{\delta} C p_i P v.K_i + \text{number} \in Z_x. \quad (5)$$

The number in the decryption key will be based on the preprogrammed packaged size of the ciphertext encrypted in the edge device.

Step 5. If the threshold for the sensor is breached, alert the user and transfer encrypted data for cloud transfer ahead of the data packet. Additional functional methods are designed in the edge device, such that after the encryption of the number, the device will perform a simple logical operation. If the encrypted number of the ciphertexts exceeds a certain threshold value, an alarm text will be delivered to the patient and the attending physician as an added advantage.

Step 6. Process the sent data in the cloud without decryption, and if a critical condition is breached, alert the medical team immediately. The data pack sends to the server in a sequential time because of the large volume of the data transmission to the cloud server. The timing of the device will be adjusted based on the encryption timing required for sending the data. The low-cost sensor module will not have a facility to store the large volume of transmitting signal. Therefore, the storage, encryption, and timely delivery have to be taken care of by the edge device attached to the sensor module. Even if the network availability is not there, the device will store a certain value for a certain period and upload them when the connection is established with the cloud server.

The advantage of the current method is that the transmitted data will be sent in packets and reassembled in the cloud storage device for further processing.

6. Implementation

The biological signal-like ECG device was mimicked using a random signal generator in the MATLAB environment. The signal from the generator is stored in a series of column values for every 30 seconds before sending it to an encryption algorithm. Similarly, in the edge environment, the sensor data will be processed in a batch mode every 30 seconds before the encryption algorithm processes the incoming message. Each column value will be encrypted based on the number size sent from the sensor. For example, suppose the temperature and pressure sensor are attached to the patient. In that case, a single bit temperature and pressure value will be taken at every 2-minute interval, whereas the ECG-like signal will be transmitted for every second from the patient. The encryption text size will vary at the Edge Computing device depending on the monitoring sensor. Therefore, the size of the number sent for encryption is given by toy (2), medium (128), or large numbers (1024). The function for encryption is defined in MATLAB for easy software implementation, and later it can be converted into a basic C code for edge device implementation. The current simulation for Somewhat Fully Homomorphic Encryption with RLwE was implemented for simple cases with three different scenarios in Intel Core 2 Duo machine at 1.6 GHz with 16 GB RAM, which were performed to analyze the encryption time and size in the edge environment before transmitting to a cloud server. The size of the encryption along with the public key also plays a vital role in the performance of the edge device to process the incoming values continuously without any delay.

7. Results and Discussion

The algorithm helps to solve the time taken and the computation necessary for the edge device to handle many sensors from each patient in a shorter period of time to encrypt the message without affecting the performance of the system. The code should safeguard from a certain type of attack against the vulnerabilities posed by the FHE code. One way is to implement the secure wireless connection from the sensor module to the edge device module with SHA1 hashes to verify the integrity of the received signal from the wireless module. The current paper explains the time required to perform the encryption in edge device based on the numbers and text received from the various sensor modules.

8. Encryption Time of Text and Numbers

The encrypted text and values at various levels are based on the received text size and the time available at the edge node to perform the required encryption operation. In the case of numbers, the sum and product of the two can be directly encrypted from the available space without any delay. If the text message is sent from the sensor, then the text has to be converted into a number first before performing the encryption operation. Such conversion will save the time required for encryption at the edge interface. The encryption performed using a public key at the edge device will have to separate the encryption timing based on the received value as well as the integrity of the incoming number or text. The encryption of a smaller number will require less time to process; the encryption based on n number of prime numbers chosen indicates that the $\text{mod}(t)$ should be greater than the length of the received signal. The initial parameters s and n also need to be smaller for resource constraint edge device or based on the time interval at which the signal is transferred from the sensor (e.g., temperature sensor and pressure sensor) attached to the patient. The physiological signals like ECG have to be monitored continuously for any variation in the patient's heart rate, which is critical for monitoring patient health. The larger signal will have to be packed in a certain order for easy processing at the edge environment. The addition operation of the encryption performed much faster, even for the larger size of the number. The multiplication of the prime number polynomial in the encryption takes a larger time to convert and encrypt. The idea here is to choose the multiplication factor based on the size of the text and the degree of a polynomial the edge device will select based on the incoming signal. This will be a trade-off in the edge device performance versus the selection of appropriate encryption for each received signal from the sensor module.

9. Implementing Numbers Packaging in Edge Device

Here, we propose converting the incoming signal into a separate column for easy conversion of signals into a useful number. First, the ciphertext encrypted with a smaller text

will be placed separately with a sensor number in the cloud server, and the larger number will be encrypted into multiple files for easy transfer. For example, the encrypted size of the ciphertext based on the incoming package will have a separate number selected during the private key generation. The selected number will add a single value at the appropriate interval for performing the decryption mechanism. The ciphertext sends from the edge device will have a unique number that will match the private key based on the size of the ciphertext received.

10. Decryption and Comparison of Encrypted Text

The ciphertext sent from the edge device after performing the specific encryption depends on the degree of polynomial selected for encryption based on the size of the packed number. In a Fully Homomorphic Encryption system, the received signals can be operated without performing any decryption algorithm. For example, the average value of the temperature sensor from the encrypted message can be calculated in the cloud server without performing any decryption of the real data. If the threshold is breached in the server, an alert message would be delivered from the cloud platform for a list of authorized users. If the decoding of data has to be performed, it will be performed only with the patient consent that possesses the authorized key. The decryption of the encoded message was based on the size of the packed number sent and the duration of the value received by the edge platform. The personal information and sensitive patient details will be saved using bitwise encryption rather than packed encryption as described previously. The sensitive data requires more encryption power than the packed encryption performed for fast encryption of the sensor readings.

To avoid the issue of large bit size key, the prime number was chosen for the public key to determine the size and the time required for the encryption of the number and the text is also converted into a random number before performing an encryption mechanism. The two ring elements were used for both the public key and the ciphertext encryption, and the whole method was performed in a single degree reduction during the multiplication operation of the encryption step. The developed method was implemented in MATLAB based on the message space modulus t and power of the prime number factor n dimension; the ciphertext degree for each type of encryption method is shown in Table 2. The time required for encryption using a public key and a private key in general is based on the text size chosen based on packet size (tiny, medium, and large). The public encryption key required for performing individual addition and multiplication based on the general public key is evaluated to assess the efficiency of the multiplication encryption mechanism. From the calculated result, it is evident that the addition requires less time than the multiplication of prime numbers. The calculation also shows the maximum time required for performing simple encryption if the complexity of the ciphertext size increases. The encrypted size of the public key size increases with the degree of the

TABLE 2: Encryption operation time and size for various message spaces.

| Message space (mod(t)) | n (ciphertext degree) | Encryption operation time | | | | Encrypted size | |
|----------------------------|-------------------------|---------------------------|------------|---------|----------|----------------|---------------|
| | | Pub. key ms | Pv. key ms | Add. ms | Mult. ms | Pub. key KB | Ciphertext KB |
| 2 | 512 (1) | 25 | 57 | <1 | <1 | 12 | 24 |
| | 4096 (4) | 210 | 520 | 2 | 150 | 386 | 781 |
| | 8192 (10) | 425 | 1410 | 7 | 860 | 2256 | 4526 |
| 128 | 1024 (1) | 55 | 112 | <1 | <1 | 29 | 60 |
| | 4096 (4) | 225 | 610 | 4 | 160 | 490 | 975 |
| | 8192 (10) | 448 | 1654 | 7 | 1140 | 5534 | 11052 |
| 1024 | 1024 (1) | 52 | 108 | <1 | <1 | 35 | 69 |
| | 4096 (4) | 225 | 620 | 4 | 214 | 535 | 1091 |
| | 8192 (10) | 455 | 1825 | 9 | 554 | 6010 | 12105 |

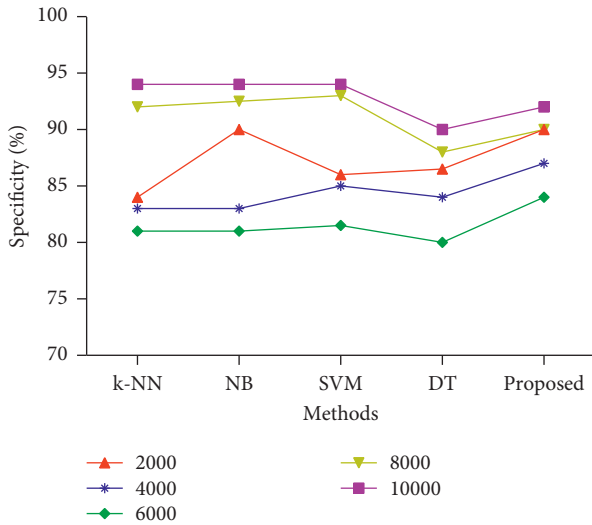


FIGURE 3: Analysis of specificity.

ciphertext to perform the operation. The algorithm helps to solve the time taken and the computation necessary for the edge device to handle in a shorter period of time to encrypt the message to the server without affecting the performance of the system. The code should safeguard from a certain type of attack against the vulnerabilities posed by the FHE code. The code implemented by the SHE based on the new algorithm is lightweight more suited for the encryption at the edge device itself for preserving the privacy of the patient.

For Somewhat Homomorphic Encryption (SHE), the RLWE method using the parameters given in the algorithm is implemented in MATLAB. The time and size are taken for key encryption along with addition and multiplication operation performed during encryption operation.

11. Experimental Results

The experiment has been accomplished to evaluate the proposed model using different instances with four different models such as k-Nearest Neighbor (k-NN), Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT).

The specificity analyzed between the proposed model and four different models is presented in Figure 3 for clear

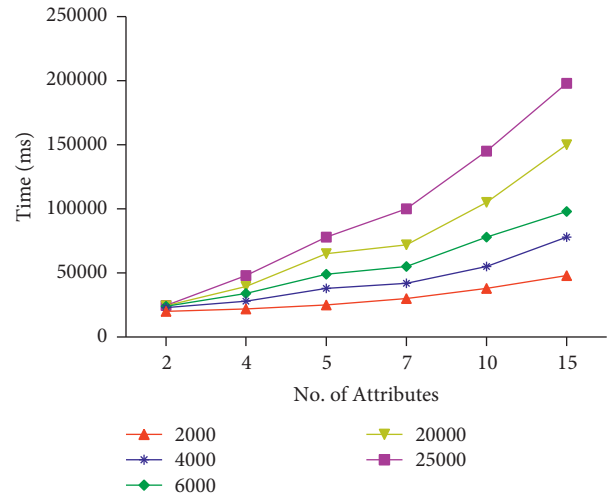


FIGURE 4: Time taken to encrypt attributes.

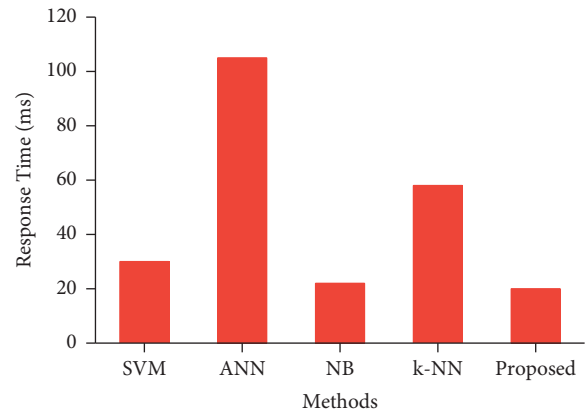


FIGURE 5: Analysis of response time.

understanding. From the figure, it is noted that the proposed method outperforms the existing baselines.

It is observed from Figure 4 that the total time taken to encrypt 2000 records is less than others. Although, it increases gradually when there are more numbers of records in the input dataset.

It is noted from Figure 5 that the response time taken by the proposed model is less when compared with other existing baseline models.

12. Conclusions

In the current paper, we proposed a modified LWRE method based on Somewhat Homomorphic Encryption, which can be practically applied to a medical-oriented sensor device that requires confidentially protecting patient information. This method helps to solve the problem by systematically monitoring patient health in real-time without divulging sensitive data. The proposed work provides a secure key that the patient can only authorize to decrypt the original raw data from the sensor attached to the patient. The encryption performed in the MATLAB software helps to emulate a real device scenario and helps to fine-tune the operation required for the particular type of sensor and the number of packaging methods required for efficient operation of the edge device time. The encryption scheme can be programmed based on the number of sensors attached to the particular edge environment by estimating the number of times a device transmits data to the edge device. In the future, the proposed approach will be enhanced to meet the requirements of time-sensitive applications. Also, the proposed SHE shall be considered for the Federated Learning applications for ensuring better security.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the Science and Engineering Research Board (SERB), Department of Science and Technology, India, for financial support through the Mathematical Research Impact Centric Support (MATRICS) scheme (MTR/2019/000542). The authors also acknowledge SASTRA Deemed University, Thanjavur, for extending infrastructural support to carry out this research work.

References

- [1] Indian Healthcare Industry Analysis, "Indian Healthcare Industry Analysis," October-2018, <https://www.ibef.org/industry/healthcare-presentation>.
- [2] C. S. Nandyala and H.-K. Kim, "From cloud to fog and IoT-based real-time U-healthcare monitoring for smart homes and hospitals," *International Journal of Smart Home*, vol. 10, no. 2, pp. 187–196, 2016.
- [3] A. M. Rahmani, T. N. Gia, B. Negash et al., "Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: a fog computing approach," *Future Generation Computer Systems*, vol. 78, pp. 641–658, 2018.
- [4] P. Verma and S. K. Sood, "Fog assisted-IoT enabled patient health monitoring in smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1789–1796, 2018.
- [5] M. S. Hossain, M. A. Rahman, and G. Muhammad, "Cyber-physical cloud-oriented multi-sensory smart home framework for elderly people: an energy efficiency perspective," *Journal of Parallel and Distributed Computing*, vol. 103, pp. 11–21, 2017.
- [6] S. K. Sood and I. Mahajan, "Fog-cloud based cyber-physical system for distinguishing, detecting and preventing mosquito borne diseases," *Future Generation Computer Systems*, vol. 88, pp. 764–775, 2018.
- [7] M. Pham, Y. Mengistu, H. Do, and W. Sheng, "Delivering home healthcare through a cloud-based smart home environment (CoSHE)," *Future Generation Computer Systems*, vol. 81, pp. 129–140, Apr. 2018.
- [8] V. Subramaniaswamy, R. Logesh, M. Abejith, S. Umasankar, and A. Umamakeswari, "Sentiment analysis of tweets for estimating criticality and security of events," *Journal of Organizational and End User Computing*, vol. 29, no. 4, pp. 51–71, 2017.
- [9] D. Li, H. W. Park, E. Batbaatar et al., "Application of a mobile chronic disease health-care system for hypertension based on big data platforms," *Journal of Sensors*, vol. 2018, pp. 1–13, 2018.
- [10] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for IIoT devices," *IEEE Transactions on Industrial Informatics*, 2021.
- [11] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and T. Reddy, "CANintelliIDS: Detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU," *IEEE Transactions on Network Science and Engineering*, vol. 8, 2021.
- [12] R. H. Jhaveri, S. V. Ramani, G. Srivastava, T. R. Gadekallu, and V. Aggarwal, "Fault-resilience for bandwidth management in industrial software-defined networks," *IEEE Transactions on Network Science and Engineering*, IEEE, vol. 8, no. 4, pp. 3129–3139, 2021.
- [13] K. Fan, J. Wang, X. Wang, H. Li, and Y. Yang, "A secure and verifiable outsourced access control scheme in fog-cloud computing," *Sensors*, vol. 17, no. 7, p. 1695, 2017.
- [14] K. Hariss, H. Noura, and A. E. Samhat, "Fully Enhanced Homomorphic Encryption algorithm of MORE approach for real world applications," *Journal of Information Security and Applications*, vol. 34, pp. 233–242, 2017.
- [15] R. Sanchez-Guerrero, F. A. Mendoza, D. Diaz-Sanchez, P. A. Cabarcos, and A. M. Lopez, "Collaborative eHealth meets security: privacy-enhancing patient profile management," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1741–1749, 2017.
- [16] R. Sandhu, H. K. Gill, and S. K. Sood, "Smart monitoring and controlling of pandemic influenza A (H1N1) using social network analysis and cloud computing," *Journal of Computational Science*, vol. 12, pp. 11–22, 2016.
- [17] S. K. Sood and I. Mahajan, "Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus," *Computers in Industry*, vol. 91, pp. 33–44, 2017.
- [18] S. Rani, S. H. Ahmed, and S. C. Shah, "Smart health: a novel paradigm to control the chikungunya virus," *IEEE Internet of Things Journal*, vol. 4662, p. 1, 2018.
- [19] X.-S. Zhang, F.-Y. Leu, C.-W. Yang, and L.-S. Lai, "Healthcare-based on cloud Electrocardiogram system: a medical center experience in middle taiwan," *Journal of Medical Systems*, vol. 42, no. 3, p. 39, 2018.

- [20] P. Verma and S. K. Sood, "Cloud-centric IoT based disease diagnosis healthcare framework," *Journal of Parallel and Distributed Computing*, vol. 116, pp. 27–38, 2018.
- [21] H. Mostafaei and M. S. Obaidat, "Learning automaton-based self-protection algorithm for wireless sensor networks," *IET Networks*, vol. 7, no. 5, pp. 353–361, 2018.
- [22] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial internet of things (IIoT) - enabled framework for health monitoring," *Computer Networks*, vol. 101, pp. 192–202, 2016.
- [23] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," *Advances in Cryptology—EUROCRYPT 2010*, vol. 6110, pp. 1–23, 2010.

Research Article

Ecological Welfare of China's Forest Towns: Concept, Formation Mechanism, and Evaluation Index System

Chen Chen,^{1,2} Haitao Sun,² and Yingli Huang ¹

¹School of Economics and Management, Northeast Forestry University, Harbin, 150040, China

²College of Accounting and Finance, Heilongjiang Polytechnic, Harbin, 150010, China

Correspondence should be addressed to Yingli Huang; lqq2015010092@nefu.edu.cn

Received 15 January 2022; Revised 29 January 2022; Accepted 3 February 2022; Published 24 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Chen Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The economic subsystem, social subsystem, and natural ecosystem of forest towns constitute forest towns' complex ecosystem. Forest towns' complex system provides good ecological welfare for the residents' survival and development, especially the beautiful ecological environment has played a positive role in promoting the residents' healthcare. First of all, this paper expounds on the concept of forest towns' ecological welfare from the construction goals of forest towns. Secondly, it analyzes the formation mechanism of forest towns' ecological welfare based on the DPSIR conceptual model. Finally, taking forest towns in Heilongjiang Province as an example, the index system for ecological welfare evaluation is constructed, and the influence degree of the evaluation index is analyzed through using the analytic hierarchy process. The results show that employment and income are the most important indicators that affect the ecological welfare of forest towns in Heilongjiang Province. This paper provides a theoretical basis for evaluating the level of ecological welfare level, level of healthcare level, and planning development path of Chinese forest towns in the future.

1. Introduction

Forest town is a town that relies on forest resources to provide ecological products and ecological services such as forest sightseeing, leisure vacation, sports, and health as its main characteristics. And it is a type of characteristic town that integrates industry, culture, tourism, and community functions. Since 2015, Zhejiang Province has taken the lead in the construction of forest towns. As of December 2020, there have been more than 1,000 provincial forest towns, and national-level forest characteristic towns also announced 50 construction pilot projects in August 2018. Forest towns' construction has set off an upsurge in China. The construction process of the forest town is a process of human-earth interaction. The natural ecosystem has the provisioning function with food and water supply, the regulation function with maintaining air quality and regulating the climate, the supporting function with producing oxygen and forming soil, and the cultural function with providing aesthetic appreciation, entertainment, and

recreation for human beings [1]. It is an essential contribution to human survival and development and maintaining a good quality of life, including healthcare [2]. Human activities act on the natural ecosystem of the forest town, forming a "human and Earth circle" of deep penetration and interactive influence between nature, economy, and society. Forest towns develop a series of related industries such as accommodation, catering, and local specialties through tourism. It has brought into play the economic, ecological, and social benefits and maintained the healthcare of the residents of the forest town. However, due to the different development models of forest towns, the structure of the human-earth system is affected, and the ecological welfare of people is affected. In some forest towns, due to the limitations' technology and the management policies, the multiple functions of forest towns have not been fully utilized compared with the abundant forest resources, and even the forest ecological curse effect has appeared. Some forest towns adopt the extensive development model of "high investment, high consumption, and high emission."

Although they have attracted more tourists and achieved economic and social development, the forest ecological environment faces severe pressure. The deterioration of the environment results in people's increasing demand about the ecological welfare which cannot be met. At present, the research on forest towns is in its infancy stage. The qualitative research conducted by most scholars on forest towns mainly focuses on the concept, characteristics, construction significance, and development path of forest towns [3–5]. And some scholars also discussed the distribution, characteristics, and influencing factors of forest towns from a spatial perspective [6]. Although some scholars have evaluated characteristic towns from different perspectives, the evaluation of forest towns, especially the study of ecological welfare evaluation, has not yet been formed [7–10]. In the construction of a “people-centered” town, bringing more ecological welfare to people is the ultimate goal of town construction and development.

Therefore, this paper takes the forest town as the research objects, and first explains the ecological welfare concept of the forest town. Secondly, the formation mechanism of ecological welfare is explored through the DPSIR conceptual model. Finally, constructing the ecological welfare evaluation index system of forest towns as Heilongjiang Province an example, and using the analytic hierarchy process to analyze the influencing factors. The research of this paper fills the blank of ecological welfare evaluation of forest towns, improves the theoretical basis of forest towns' ecological welfare, and provides a scientific basis for exploring the development path of forest towns' ecological welfare.

2. The Concept of Ecological Welfare in Forest Towns

Welfare is a synonym of well-being, which refers to all the beautiful living conditions or living environment of human beings. However, due to the late beginning of research on ecological welfare, the research has shown significant complexity, so the concept has not yet been able to form a unified understanding [11]. Zhang believes that ecological welfare is the result of the extension of welfare connotation and the popularization of ecological movement, and the extension of the concept of harmonious development between man and nature and society [12]. Tang believes that ecological welfare is an important indicator of human welfare [13]. Huang et al. believe that “ecological welfare” reflects the effect of the ecological environment on the level of human welfare and defines it as all material and non-material, direct, and indirect benefits that a good ecological environment brings to people and the economy and society [14]. Zang et al. propose that ecological welfare is “the products and services that humans obtain or enjoy from the natural environment and are directly related to human well-being provided” [15]. Hu et al. considered that ecological welfare could divide into broad and narrow sense. In a narrow sense, ecological welfare refers to the direct or indirect benefits obtained from the natural ecosystem, and ecological welfare is an integral part of social welfare [16]. In a broad sense, ecological welfare is the comprehensive

human welfare produced by human beings based on natural capital and combining human and material capital investment [17]. The economic and social development of forest towns is based on the advantages of abundant natural resources. At the same time, the natural ecosystem will also be affected by the economy and society. Therefore, forest towns can be regarded as a complex ecosystem composed of economy, society, and natural environment. Through field investigations in small towns, China's goal of building forest towns is organic to integrate “production, life, ecology, and culture.” According to the goal of the construction, ecological welfare is not limited to the products and services provided by the natural ecosystem, it should be proposed from a broad perspective. Hence, the concept of ecological welfare is that forest towns' complex ecosystem provides the production function welfare, life function welfare, ecological function welfare, and cultural function welfare to the residents of the forest town through direct or indirect means (Figure 1).

3. The Formation Mechanism of the Ecological Welfare of Forest Towns

In 1993, the European Environment Agency (EEA) proposed the DPSIR model, which looks at the interaction between humans and environmental systems from the system analysis [18]. The DIPSIR model comprises five dimensions: driving force, pressure, state, influence, and response. Among them, the development of society, economy, and population as the driving force (D) is the fundamental reason for the generation and development of the system. The driving force acts on the ecological environment, putting pressure on the natural ecological environment (P), prompting changes in the state of the ecological environment (S), and then having various impacts on humans and the ecological environment (I). These influences prompt people to respond to changes in the environmental state (R). The response measures directly act on environmental pressure, state, and influence or act on the driving force constituted by the economy, society, and population. The ecological welfare of forest towns is also produced under the joint action of economy, society, and natural ecosystem. Hence, this paper adopts the DPSIR conceptual model to analyze the formation mechanism of ecological welfare.

In the forest town complex ecosystem, in order to promote economic and social development, increase the use of forest resources, and achieve ecological welfare, the driving force (D) is divided into three aspects: economic driving force, social driving force, and resource driving force. However, promoting urbanization in forest towns, the total population increases and the development and utilization of forest resources will inevitably consume ecological resources. And it has an impact on the ecological environment. So, the driving force causes pressure on the ecological environment (P). Driven by economic and social development, under the pressure of the destruction of the ecological environment, the affected ecological environment causes changes in the ecosystem in the forest town, disturbing the state of ecosystem services (S). Then, under the

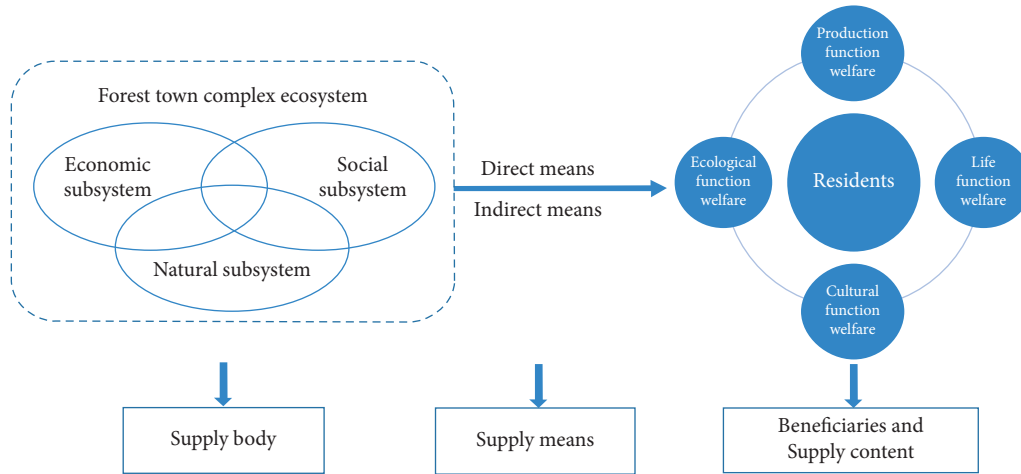


FIGURE 1: Ecological welfare of forest towns.

influence of the driving force (D), pressure (P), and state (S) of the forest town, the ecological welfare of the forest towns is formed, including eight aspects: income, medical care, employment, education, culture, social security, ecological environment, and infrastructure construction. Then, the town managers formulate policies and measures based on the current level of residents' ecological welfare and the states of the economic, social, and natural environment, which forms a response (R). The formulation of policies will promote the transition of economic and social development of production and lifestyle to green and low-carbon. By increasing financial investment in environmental and ecological governance, the pressure on the ecological environment will relieve, the condition of the ecosystem will improve and restore, and the level of ecosystem services will improve and then improve the ecological welfare of human beings. Meanwhile, response (R) will give feedback to ecological welfare (I) to verify whether the forest town development policy promotes the improvement of ecological welfare. The construction of the DPSIR model framework forms a multidimensional closed-loop circular structure of economy-society-nature-ecological welfare policy, which is helpful to analyze the formation process of ecological welfare (Figure 2).

4. Construction of Ecological Welfare Evaluation Index System of Forest Towns

4.1. Study Area-Forest Town in Heilongjiang Province. Heilongjiang Province is one of the largest forestry provinces in China. The total forestry management area of Heilongjiang is 31.75 million hectares, accounting for 2/3 of the province's land area, and the forest coverage rate reaches 47.3%. It is the most important state-owned forest area and the largest timber production base in China. There are more than 100 species of forest trees and more than 30 species of higher utilization value in Heilongjiang. In 2018, Heilongjiang occupies 8 of the 50 forest towns' construction pilot projects selected by the state. Abundant forest resources give Heilongjiang the advantage of building a forest town.

4.2. Research Method. The analytic hierarchy process (AHP) is a comprehensive evaluation method for system analysis and decision-making. It transforms human judgment into a comparison of the importance of several factors and more reasonably solves the shortcomings of decision maker's subjective judgments. This paper uses the analytic hierarchy process to calculate the weight of the influencing factors of the forest town's evaluation index so that we obtain the importance of the influencing factors. The calculation process is as follows:

- (1) Establish a judgment matrix. Put the 16 influencing factors of the forest towns' sustainable development into a matrix $M = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$, the elements in the matrix correspond to a_{ij} ($i, j = 1, 2, \dots, n$); a_{ij} represents the importance of i compared with index j , and $a_{ij} > 0$ and $a_{ij} \times a_{ji} = 1$, and it forms a positive and negative judgment matrix. Then, five experts in the group compare the indicators in the matrix pair by pair, score the importance of the two indicators, and fill in the average score in the matrix.
- (2) Calculate the weight of the judgment matrix and check the consistency. Though using SPSSAU software, calculating the CI, CR, and weight values of the judgment matrix, conduct a consistency test. When $CR < 0.1$, the judgment matrix is considered to have satisfactory consistency, the weight of each indicator pass test.

4.3. Selection of Evaluation Index. The forest town is a functional community that integrates "suitable for production, livability, and travel," meeting people's survival and development needs by the forest town's "production, life, ecology, and culture." For the residents of forest towns, stable economic income, suitable jobs, a beautiful living environment, a healthy body, sound social security, perfect infrastructure construction, and a strong cultural

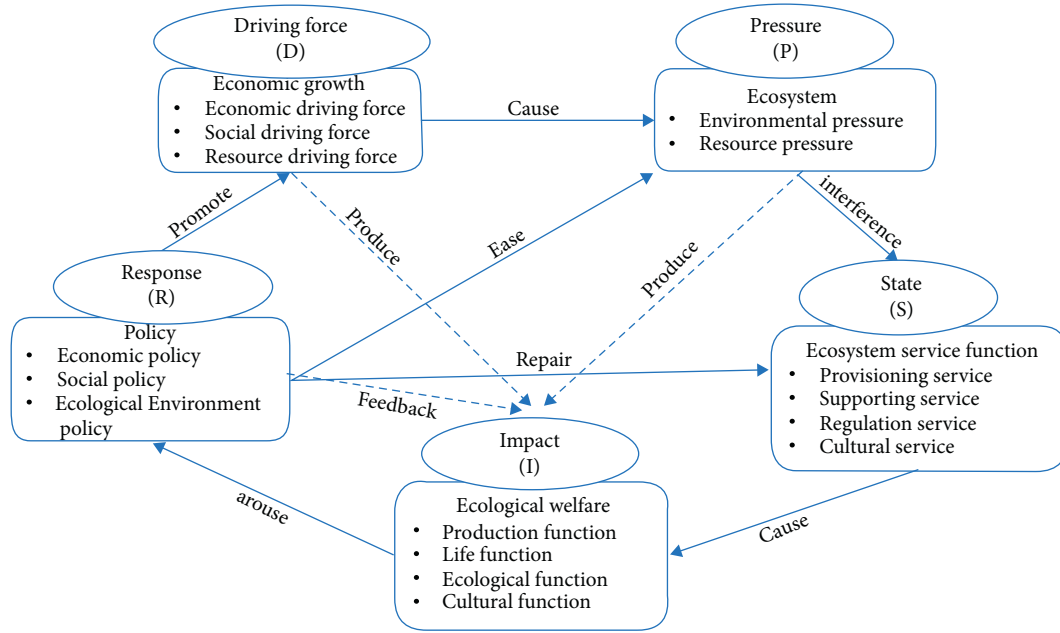


FIGURE 2: The formation mechanism of the ecological welfare based on the DPSIR conceptual model.

TABLE 1: Evaluation index system of forest town.

| Target level | Criterion level | Index level | Weights |
|-------------------------|-----------------------|---|---------|
| Production function | Income level | Per capita GDP of forest town residents | 11.55% |
| | Consumption level | Per capita consumption level of local residents | 7.27% |
| Life function | Medical service | Number of doctors per 10,000 people | 7.70% |
| | | Average years of education | 7.56% |
| | Education level | Basic medical participation rate | 5.87% |
| | | Minimum living security rate | 3.98% |
| | Social security | Employment rate of residents in forest towns | 14.82% |
| | | Road area per capita in the town | 2.65% |
| | Employment level | Per capita housing area of forest town residents | 3.62% |
| Infrastructure services | | Internet penetration rate | 3.85% |
| Ecological function | Ecosystem environment | Green coverage rate in built-up area | 3.55% |
| | | Forest cover rate | 8.15% |
| | | Good air quality rate | 8.15% |
| | | Average negative oxygen ion concentration | 3.39% |
| Cultural function | Cultural service | Number of heritages with forest culture and spiritual value | 3.96% |
| | | Number of A-level scenic spots in forest town | 3.96% |

atmosphere are their most cherished functions. Therefore, combined with the construction goals of forest towns, this paper will evaluate the ecological welfare of forest towns from the four aspects of the production function, life function, ecological function, and cultural function. By learning the welfare evaluation indicators from Hu et al. [16], Xiao Liming (2021) [19], Guo Lihua (2017) [20], Gao Bofa (2020) [21], and Ding Linlin (2017) [22], we selected the general indicators of the ecological welfare evaluation of forest towns. While choosing the characteristic indicators of the ecological welfare by referring An huben (2015) [23], Qin Yan et al. (2010) [24], Zhu Lin et al. [25], and Huang et al. [14]. The ecological welfare index system of forest towns with eight specifically functional activities are constructed, including income, consumption, medical care, education, social security, employment, infrastructure services, and cultural services (Table 1).

5. Analysis of Evaluation Results

It can be seen from Table 2 that the CR value of the judgment matrix is 0.039, which is less than 0.1, so the judgment matrix has a satisfactory consistency result. It can be concluded that, among the 16 indicators from the weight values in Table 1, the top four most important factors affecting the forest towns' ecological welfare are employment rate of residents, per capita GDP of forest town residents, forest cover rate, and good air quality rate. It shows that the production function is the most important and fundamental function for forest towns' residents. The employment welfare and income welfare brought by the production function are the foundation of survival for the residents of the forest towns. Besides that, an excellent ecological environment is also an essential factor influencing the ecological welfare of forest towns. The development of the town relies on abundant

TABLE 2: The consistency test results.

| Largest characteristic root | CI value | RI value | CR value | The consistency test |
|-----------------------------|----------|----------|----------|----------------------|
| 16.94 | 0.063 | 1.594 | 0.039 | Pass |

forest resources. The higher forest coverage rate is not only a guarantee for production but also an ecological barrier for residents' lives. However, the weight ratios of infrastructure services and cultural function welfare evaluation indicators in life functions are relatively small. This is because the residents of the forest towns primarily engage in forest tourism, forest product planting and breeding, or forest management and protection, and their overall income level is not high. Therefore, compared with living facilities and cultural atmosphere, they are more concerned about the economic benefits brought by forest resources.

6. Conclusion

Good ecological welfare can meet not only the increasing material and cultural needs of the people but also the ultimate goal of the development of forest towns. This paper puts forward the concept of forest town's ecological welfare based on the construction goal of forest town. Then, it explores the formation mechanism of ecological welfare through the DPSIR conceptual model. In addition, taking Heilongjiang Forest Town as an example, we construct four aspects of forest town ecological welfare evaluation index system including the production function welfare, life function welfare, ecological function welfare, and cultural function welfare. At last, through using the analytic hierarchy process to evaluate the importance of the indicators, the research of this paper provides a theoretical basis for the analysis of the ecological welfare development level of forest towns in the future. However, the collection of ecological welfare data of forest towns in this paper should be strengthened. Therefore, the data collection, the research on the improvement path of forest towns' ecological welfare and the interaction between ecological welfare and health-care still need to be followed up.

Data Availability

The raw data supporting the conclusions of this article are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work was supported by "The D-type project of the Fundamental Research Funds for the Central Universities" of "The model, PATH, and experimental demonstration of natural resources capitalization taking the forest ecological bank as an example" (2572020DZ09) and "New sustainable development based on the perspective of green finance to

release resource advantages Mechanism Research" (2572020DY06).





References

- [1] Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: A Framework for Assessment*, Island Press, Washington DC, 2003.
- [2] S. Li, "How to scientifically measure the contribution of nature to mankind-A social-ecosystem analysis framework based on ecosystem services and its application," *Academic Frontiers*, vol. 11, pp. 28–35, 2020.
- [3] S. Zhang, "Some thoughts on the development of forest towns," *China Economic Report*, vol. 3, pp. 95–98, 2017.
- [4] Y. Ge, C. Cui, and Z. Chang, "Policy interpretation and path return to the healthy development of forest town-Taking Wugong forest town in Xianyang City as an example," *Journal of Beijing Forestry University*, vol. 3, pp. 49–54, 2020.
- [5] J. Zhang, Y. Zeng, Y. Zhang, and B. Cheng, "Literature review and analysis of domestic forest town research," *Forestry Economy*, vol. 12, pp. 37–42+68, 2019.
- [6] Y. Ge and J. Zhang, "Spatial distribution characteristics and influencing factors of forest characteristic towns in Zhejiang Province," *Journal of Zhejiang A&F University*, vol. 2, pp. 374–381, 2020.
- [7] B. Zhu, "Influencing factors and paths of the coordinated development of "big towns and small towns"-An empirical study based on the construction of characteristic towns in Zhejiang," *Academic Forum*, vol. 1, pp. 116–121, 2018.
- [8] C. Wang and S. Jia, "Characteristic index system and evaluation of Chinese characteristic towns," *Nanjing Social Sciences*, vol. 2, pp. 79–86+92, 2019.
- [9] J. Tan and Y. Pan, "Performance evaluation of ecological welfare in characteristic towns-taking Chongqing as an example," *Technology Management Research*, vol. 24, pp. 40–46, 2019.
- [10] T. Jiang, "Construction of the evaluation system for the tourism function of characteristic towns in Zhejiang Province," *China's Agricultural Resources and Regionalization*, vol. 6, pp. 227–232, 2019.
- [11] Li Yan, S. Li, Y. Gao, and Y. Wang, "A framework for the classification of ecosystem services connecting multi-level human well-being," *Acta Geographica Sinica*, vol. 8, pp. 1038–1047, 2013.
- [12] J. Zhang, "The rise of the concept of ecological welfare and the transformation of medical security model," *Ecological Economy*, vol. 1, pp. 90–92+116, 2009.
- [13] J. Tang, "Three issues of ecological welfare-based on the perspective of ecological values," *Journal of Nanchang University*, vol. 4, pp. 12–20, 2020.
- [14] Y. Huang, Z. Qiao, W. Zou, H. Qin, and X. Zhang, "Research on the evaluation index system of the ecological welfare supply system of national forest parks based on ISM," *Acta Ecologica Sinica*, vol. 10, pp. 4090–4098, 2021.
- [15] Z. Zheng, D. Zheng, C. Sun, and X. Zou, "A multi-scale empirical test of the ecological curse effect in mainland China," *Geographical Research*, vol. 5, pp. 851–863, 2016.

- [16] M.-j. Hu, Z.-j. Li, Z.-s. Ding, N.-x. Zhou, D.-l. Qin, and C. Zhang, "Urban ecological well-being intensity and driving mode based on three-dimensional well-being: Taking the Yangtze Delta as an example," *Journal of Natural Resources*, vol. 36, no. 2, pp. 327–341, 2021.
- [17] Z. Zheng and X. Zou, "Test of convergence characteristics of water resource intensity in mainland China: empirical evidence based on inter-provincial panel data," *Journal of Natural Resources*, vol. 6, pp. 920–935, 2016.
- [18] European Environment Agency, *Halting the Loss of Biodiversity by 2010: Proposal for a First Set of Indicators to Monitor Progress in Europe*, EEA Copenhagen, Denmark, 2005.
- [19] L. Xiao and Q. Xiao, "Analysis of the differentiation and spatial convergence of urban ecological welfare performance pattern in the Yellow River Basin," *Soft Science*, vol. 2, pp. 46–53, 2021.
- [20] L. Guo and F. Wang, "Evaluation of changes in the welfare of ecological migration in pastoral areas from the perspective of feasible capabilities: taking Inner Mongolia and Qinghai as examples," *Heilongjiang Ethnic Series*, vol. 2, pp. 44–51, 2017.
- [21] B. Gao, Li Cong, S. Li, and X. Han, "Research on the welfare status and influencing factors of relocated farmers in ecologically fragile areas," *Resources and Environment in Arid Regions*, vol. 8, pp. 88–95, 2020.
- [22] L. Ding, Q. Wu, and Y. Li, "Research on the welfare changes of land-lost farmers under the background of new urbanization," *China Population, Resources and Environment*, vol. 3, pp. 163–169, 2017.
- [23] H. An, Y. Fan, and B. Yang, "Research on sustainable development evaluation of forestry resource-based cities based on DPSIR model-taking Yichun as an example," *Science and Technology Management Research*, vol. 5, pp. 74–78, 2015.
- [24] Y. Qin, S. Shen, and F. Wu, "The calculation method and application of the cultural function value of forest ecosystems: taking Zhangjiajie Forest Park as an example," *Journal of Central South University of Forestry and Technology*, vol. 4, pp. 26–30, 2010.
- [25] L. Zhu, Li Lan, Z. Li, B. Fan, and D. Zhang, "Status and analysis of foreign forest cultural value evaluation indicators," *World Forestry Research*, vol. 5, pp. 92–96, 2015.

Research Article

Healthcare Security Incident Response Strategy - A Proactive Incident Response (IR) Procedure

Ying He ¹, Leandros Maglaras ², Aliyu Aliyu ², and Cunjin Luo ^{3,4}

¹School of Computer Science, University of Nottingham, Nottingham, UK

²School of Computer Science and Informatics, De Montfort University, Leicester, UK

³School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

⁴Key Lab of Medical Electrophysiology, Ministry of Education, Institute of Cardiovascular Research, Southwest Medical University, Luzhou, China

Correspondence should be addressed to Cunjin Luo; cunjin.luo@essex.ac.uk

Received 4 December 2021; Revised 16 January 2022; Accepted 18 January 2022; Published 23 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Ying He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The healthcare information system (HIS) has become a victim of cyberattacks. Traditional ways to handle cyber incidents in healthcare organizations follow a predefined incident response (IR) procedure. However, this procedure is usually reactive, missing the opportunities to foresee danger on the horizon. Cyber threat intelligence (CTI) contains information on emerging attacks and should be ideally utilized to inform the IR procedure. However, current research shows that the IR has not been effectively informed by CTI, especially in healthcare organizations. This paper fills in this gap by proposing a proactive IR response procedure based on the National Institute of Standards and Technology (NIST) IR methodology. This paper then presents the NHS WannaCry case study to demonstrate the use of the proposed IR methodology. We collate cyber security advisories from different CTI sources such as US/UK CERT to protect interconnected systems and devices from Ransomware attacks. This research provides novel insights into the IR in healthcare through embedding CTI advisories into IR processes and concludes that our proposed IR procedure can be used to counteract WannaCry Ransomware using CTI advisories. It has the significance of transforming the way of IR from reactive to proactive using the CTI in healthcare.

1. Introduction

Cyber security attacks such as Ransomware [1] have caused major incidents to the Critical National Infrastructure (CNI) within various industries, especially in healthcare [2]. Existing work shows that a staggering 34% of ransomware attacks are targeted at healthcare organizations [3]. Recent research shows that over the past five years time, thousands of healthcare-related data breaches have been reported, affecting more than 154 million health records in total [4]. A typical example is the WannaCry malware, one of the most historic ransomware attacks, that had targeted the UK National Health Service (NHS) causing 19,000 appointments to be canceled, costing the NHS £20 million between 12 May 2020 and 19 May 2020 and £72 million in the subsequent cleanup and upgrades to its IT systems [5].

Healthcare organizations are a favored target as it has many critical systems within their medical infrastructure [6-11]. Once a healthcare organization is infected with ransomware, healthcare services will fail to operate as expected [3] and human lives will be jeopardized. It is imperative to defend against the threats such as ransomware, especially in healthcare.

Traditional ways to handle adverse events in healthcare organizations follow a predefined incident response (IR) procedure, which includes preparation, detection and analysis, containment, eradication, and recovery, and post-incident activities [12-14]. What is not so readily considered is that countermeasures themselves can have unintended consequences, whether in crime prevention, physical security [15], or cybersecurity [16]. What is even less often considered is the fact that countermeasures can actually

cause harm, whether to the infrastructure or to some or all of its users. This harm can be as minor as the disruption and additional security burden of using a system to the negative impact on entire groups of users, forcing them to leave the system/service or placing them in a position where they are more physically or psychologically impacted [17].

However, most organizations are still adopting a reactive method [18], which obscures their capability to foresee potential security attacks in the future. Emerging threats need to be handled with a proactive approach [19]. CTI is the provision of evidence-based knowledge about impeding threats aiming to support organizations' security defense at strategic, operational, and tactical levels. A proactive approach relies on the CTI as a consultative practice, built with people processes, and technology to achieve continuous improvement of cyber security.

To address this challenge, this research proposed a proactive IR procedure through embedding the CTI that contains information on emerging attacks, root causes, affected assets, the course of actions. This allows the organizations to be well aware of the emerging attacks, get prepared and respond to those attacks proactively. This paper then presents the NHS WannaCry Ransomware case study to demonstrate the use of the proactive IR procedure. This paper makes the following contributions,

1. Reviews the current IR literature and practices and identifies the gap of the research in IR informed by CTI;
2. Proposes a proactive IR methodology that is informed by CTI, through embedding CTI into the traditional IR procedure;
3. Presents the NHS WannaCry case study to demonstrate the use of the proposed proactive IR procedure.

The remainder of this paper is structured as follows. Section 2 presents related work of cyber threat intelligence, security incident response, and Ransomware. Section 3 introduces the proactive incident response (IR) methodology. Section 4 applied the proactive IR methodology to analyse the NHS WannaCry Ransomware case study. Section 5 concludes the research and outlines future work.

2. RELATED WORK

This session introduces related work in cyber threat intelligence (CTI), security incident response (IR), and Ransomware. TABLE 1 provides a summary of related work.

2.1. Cyber Threat Intelligence (CTI). CTI contains information such as attack vectors, attack actors, victims, courses of actions, affected organisations and is presented in the form of CTI feeds using different standardised languages (e.g. MAEC, STIX, TAXII, CYBOX) [20]. The security communities have established CTI platforms (e.g. UK/US Cert, Microsoft, MISP, and MITRE) to facilitate threat exchange [21]. Proactive defense should be ideally informed by CTI [22]. The detective and preventive capabilities needed to resolve attacks have been improved by CTI as it has provided

advisories and security recommendations during security operations. The knowledge base of threat information and the way in which data is represented concludes the successfulness of the CTI within the cyber domain. This purpose is served by the use of taxonomies [23], CTI sharing standards [20], and ontologies [24] in security defence.

CTI is classified into four different types, namely Strategic, Operational, Tactical, and Technical threat intelligence [21]. Strategic Threat Intelligence [25] can help the decision-makers understand current and identify further risks related to the aims of an organisation. It is consumed by the board level of decision-makers is often short and concise, focusing on business impact and risk. Operational Threat Intelligence [26] provides information on the details of the incoming attack, the identity, and capability of the attack actor, and also the probability of the attack. This information is consumed by the security managers and the incident response team lead. Tactical Threat Intelligence provides details on the threat actors, their tools, and methodologies, which is also known as the Tactics, Techniques, and Procedures (TTPs) [27]. It is consumed by architects, internet administrators, security analysts, etc. Technical Threat Intelligence involves the technical details of an attacker's capabilities for example their tools, Command and Control (C2) channels, and infrastructure. It is usually consumed by staff at the first line of defense i.e. SOC analysts.

2.2. Security Incident Response (IR). Security incident response procedure includes preparation, detection and analysis, containment, eradication and recovery, and post-incident activities [12,13,14,28]. The Preparation stage establishes the incident response capability that will enable the organization to be ready to respond when an adverse event occurs. Main activities include the preparation of the communication routes and facilities, hardware, software, network diagrams, security plans, and predefined mitigation strategies. In the detection phase, organizations use precursors and indicators (e.g. information collected from log files, intrusion detection systems, and antivirus software) to detect incidents. Accurately detecting and assessing possible incidents have proven to be difficult, especially when determining the type, extent, and magnitude of it due to the high volumes of potential incidents. Their intrusion detection systems [29] can receive thousands if not millions of alerts per day. The majority of incidents require containment before performing the eradication and recovery as it may reduce the resources used and the damage caused to the business processes. Various containment strategies are available and predefined. After major incidents have occurred, organizations should hold a "lessons learned" meeting with all parties involved [14,30]. The meeting will help the organization when improving its security measures and the incident response handling itself.

2.3. Ransomware. Ransomware [31] is a variation of malicious software that once installed will encrypt files on a machine. The attacker will then demand a ransom to which the victim will have to pay to get their files decrypted back to

TABLE 1: Proactive Incident Response (IR) informed by Cyber Threat Intelligence (CTI) in the context of counteracting ransomware

| | Author(s) | Description |
|------------|----------------------------|---|
| CTI | Barnum [20] | Standard description of CTI using structured threat information expression |
| | Tounsi and Rais [21] | A survey on technical threat intelligence and its CTI sharing platforms |
| | He et al. [22] | Proactive cyber defence strategy through feeding CTI into IR processes |
| | Burger et al. [23] | Taxonomy model for cyber threat intelligence information exchange technologies |
| | Qamar et al. [24] | Data-driven analytics for CTI through mapping CTI feeds to Web Ontology Language (OWL) ontologies |
| | Dog et al. [25] | Strategic cyber threat intelligence sharing and a case study on IDS logs |
| | Li et al. [26] | Operational threat intelligence and a comparative analysis of CTI |
| | Maymí et al. [27] | Tactical threat intelligence (tactics, techniques, and procedures) |
| IR | Cichonski et al. [12] | NIST IR model: computer security incident handling guide |
| | Souppaya and Scarfone [13] | NIST malware incident prevention and handling |
| | Ahmad et al. [14] | A case study on information systems and security incident response processes |
| | Moreno et al. [28] | IR processes enhanced by blockchain technologies |
| | Grispos et al. [30] | IR processes (follow-up stage) improved by Agile methodology |
| Ransomware | Field [5] | NHS WannaCry ransomware incident investigation and response |
| | Brewer [34] | Ransomware IR detection, prevention, and cure |
| | Hassan [32] | Ransomware IR definition and its variants |
| | Kyurkchiev et al. [33] | CryptoLocker ransomware analysis and investigation |

their original form. It can deny legitimate users access to systems or data. It blocks access to the systems or data and threatens to release the victim's data unless a ransom is paid. Such attacks typically use a Trojan as their attacking method. The Trojan can be disguised as a legitimate file waiting to be downloaded and opened by the users. There are various variants of ransomware such as Petya, Locky, and Samas [32].

Two types of ransomware making headlines all across the world in recent months are called CryptoLocker [33] and CoinVault. Both types of ransomware operate, in the same way, as they infect a computer as soon as an unsuspecting user clicks an unknown link or opens up an attachment sent via email. The high profile Ransomware is the UK NHS WannaCry [5]. Several hospitals and GP surgeries were forced to shut down their entire IT systems over the weekend, after ransom notes from hackers appeared on computer screens, threatening to delete all of their files within seven days unless a ransom of \$300 in bitcoin currency was paid. It can be distributed through spam emails and fake ads, which trick users into downloading the virus onto their computer. It then sets about creating encrypted copies of files on the victim's computer, and deleting the originals, leaving the victim with only the encrypted copies, which cannot be accessed without a decryption key.

3. Proactive IR Methodology

In this section, we propose the proactive IR methodology by mapping the National Institute of Standards and Technology (NIST) IR methodology [12,13] to the extracted CTI advisories from different sources including US CERT [35–37]. and industrial best practices. Through embedding CTI into the IR lifecycle, organisations can benefit from an informed IR with the CTI advisories from different sources. organisations should be able to take this information and map it to their own IR processes to enhance their networks, systems and applications security against potential attacks. Fig. 1 provides an overview of

our proposed proactive IR procedure with actions in different stages of IR. The IR procedure starts from Planning & Preparation, finishing in the stage of Post-incident Activity. We have listed the actions required to be taken in each of the IR stages. The actions include both the ones from NIST and newly added ones derived from CTI advisories. The rest of this section detailed the actions in different stages.

3.1. Planning & Preparation. The Preparation stage establishes the security plan and incident response capability that enables the organisation to be ready to respond to an incident. It helps to prevent incidents by ensuring sufficient security for systems, networks, and applications. The initial preparation phase involves the creation of an incident response team and acquiring the necessary tools and resources, as well as implementing a set of controls on their assets, some of which will be based on the risk assessment results in an attempt to limit the number of incidents that have already been identified. This stage identifies the facilities needed throughout the life cycle. An incident reporting mechanism facility should be implemented, allowing individuals or teams to declare a potential incident to a wider view of people or a person of higher authority. This can be done via phone numbers, email addresses, online forms, security management systems, etc. Issue tracking system facilities can be implemented, containing information about the case owner, case status update as well as for the uses of report generating and learning purposes. This stage should prepare software and hardware to analyse and mitigate incidents. For example, a clean image of the operating system, fresh application installations, digital forensic software, additional workstations, servers, and networking equipment. Organisations should also consider implementing systems purely for backups and to analyse incidents in a controlled environment. Incident analysis resources should also be incorporated in the preparation stage such as risk assessment.



FIGURE. 1: Proactive Incident Response (IR) informed by Cyber Threat Intelligence (CTI)

Network diagrams and a list of critical assets should also be identified. This will allow for the addition of specific security controls based on high-end organisational-assets. System and staff training is also required. The incident response team can play a key role in the risk assessment and training process.

CTI can strengthen the Preparation capacity by providing additional information on emerging threats that the organizations are facing. Such CTI can be obtained from different sources, such as CTI advisories provided by US/UK

CERT, security incident reports, and CTI sharing platforms. The CTI advisories especially those within the same industry or business domain can help the organizations prepare tailored IR plans and capacity to counteract emerging threats and proactively react to the incidents.

3.2. Detection. The detection of an incident can be via a variety of forms, each with a varied level of detail and fidelity. One form of incident detection uses automated capabilities,

for example, network/host-based intrusion detection systems [38,39], antivirus software, and log analyzers. Another uses manual needs, such as users reporting problems. Some incidents have blatant signs that can be easily detected, for example, a defaced website, yet others are almost impossible to detect. An organization will typically receive high volumes of potential incidents. Their intrusion detection sensor alerts can receive thousands if not millions of alerts per day. Precursors and indicators are two categories that can show signs of an incident. A precursor is “a sign that an incident may occur in the future”. Whereas, an indicator is “a sign that an incident may have occurred or may be occurring now”. These precursors and indicators can be obtained from antivirus software, file integrity checking software, and intrusion detection prevention systems. Log files can also be used for detection purposes and they can be obtained from the operating system, services and application logs, and network device logs.

CTI sources can provide an exhaustive list of indicators. These indicators are usually publicly available via the CTI advisories provided by US/UK CERT. These indicators can also be downloaded from CTI sharing platforms such as MISP [40]. Those can then be used to update the signatures of the IDPS, allowing the IDPS to detect emerging threats proactively. Some CTI sharing platforms provide enabling functionalities to continuously feed updated indicators to the IDPS and other monitoring systems.

3.3. Analysis & Assessment. Once an incident has emerged, the incident response team should work quickly to analyze and validate it, making sure to follow pre-defined processes and documenting every step taken. When the incident becomes apparent to the team, performing rapid initial analysis should be performed to determine the incident scope. For example, which networks, systems, or applications are affected; how the incident is occurring (e.g., what tools or attack vectors are being used, what vulnerability is being exploited (5)); and who or what originated the incident. The initial information produced from this analysis enables the team to prioritize subsequent activities (e.g., containment strategy of the incident and further analysis regarding the effects of the incident).

When analyzing the incidents, the precursors or indicators may not be accurate. For example, intrusion detection systems regularly produce false positives – incorrect indicators. This shows the difficulties that incident response teams face, especially when every indicator has to be examined to determine whether it is legitimate. Additionally, thousands of indicators are identified every day, making it an extremely daunting task to identify the real security indicators. Even an indicator has shown to be accurate, it still does not necessarily conclude that an incident has occurred. For example, the modification of a critical file may not be a security incident but a human error. Some incidents do not have clear symptoms, for example, one-character modification within a file name.

The incident response team may be unable to fully determine the cause and nature of an incident. This could have

major effects on the organization, for example not having enough information to make decisions on whether to contain or eradicate the incident. In this scenario, organizations should seek assistance from CTI sources such as US-CERT, and CTI sharing platforms to determine the full scope and cause of the incident. There are also criteria such as YARA rules available from CTI sources to identify and verify an incident.

3.4. Containment & Eradication. Incident containment can reduce the resources used and the damage caused. The majority of incidents require containment. Organisations should implement containment techniques to provide additional time to develop a tailored remediation strategy. For example, Sandboxing is a containment technique that allows the organisation to monitor the activity and gather more evidence. The decision-making process is a crucial part of containment for example whether to shut down a system, disconnect it from a network and disable functions. Pre-determined strategies and procedures make containing the incident easier. Within these strategies, the organisation should define acceptable risks. The criteria should be documented clearly to facilitate containment decision-making. Criteria for determining the appropriate strategy include potential damage to and theft of resources; the need for evidence preservation; service availability (e.g. network connectivity, services provided to external parties); time and resources needed to implement the strategy; effectiveness of the strategy (e.g. partial or full containment); and duration of solution (e.g. incident related components to be removed urgently, within 5 hours (temporary) or permanent solution).

Information about the attacking host and the incident-related components could be from the CTI. Once the attacker’s IP address and the incident have been identified, a CTI search could lead to more information about the attack such as attack vectors and threat actors of similar attacks. The use of CTI sources, for example, the national vulnerability databases (NVD) is key to identifying the attacker host. CTI communities have collected and consolidated related incidents from numerous organizations into a database. This shared information can be presented in several ways, such as real-time blacklists and trackers.

Eradication deletes the incident-related components and disables all user accounts that were infected, as well as mitigates all the identified vulnerabilities that were exploited. During the eradication process, identifying all the victim hosts within the organisation is important so they can be remedied. Eradication can also be performed during the recovery stage. Organisations can check the CTI sources for possible eradication solutions.

3.5. Recovery. Within recovery, the administrator will restore systems back to their normal state and confirm that the systems are functioning normally; and if applicable, remediate vulnerabilities to prevent similar incidents. The recovering process may involve the use of clean versions disks when restoring, rebuilding the system from scratch,

replacing comprised files with clean versions, installing patches, changing passwords, tightening security parameters, high-level logging, and network monitoring. CTI sources can be consulted regarding the backup and recovery of data in different categories.

3.6. Post-Incident Activity. The post-incident activity aims to improve technology and learn lessons. After major incidents have occurred, organizations should hold a “lessons learned” meeting with all parties involved. The meeting will help the organization when improving its security measures as well as the incident response handling itself. The question that could be answered in the meeting includes, exactly what happened; how well did staff and management perform in dealing with the incident; were the documents procedures followed, were they adequate? what would the staff and management do differently next time if a similar incident occurs; what corrective actions can prevent similar incidents in the future; what indicators or precursors should be watched for in the future to detect similar incidents; what additional tools or resources are needed to detect, analyze, and mitigate future incidents?

CTI advisories from US/UK CERT provide recommendations (e.g. in the form of a business continuity plan) on what can be improved in order to prevent a similar incidents in the future. Such information is also available in the CTI course of action attributes in the format of CTI feeds shared by different CTI platforms.

4. Case Study

This section uses the NHS WannaCry Ransomware case study to demonstrate the use of the proactive IR procedure proposed in Section 3. In the study, we collate cyber security best practices and advisories to protect interconnected systems and devices from Ransomware. CTI can be used in numerous ways to help an organisation defend against the incidents [41]. However, there is limited research in applying CTI to counteract Ransomware. This section maps the relevant CTI information about Ransomware to our proposed IR procedure by following the NIST IR methodology [12,13]. The information provided in this case study covers the basics of Ransomware as well as cyber security best practices and general prevention techniques. We followed our proposed proactive IR procedure to examine how NHS has reacted to the WannaCry Ransomware attack and identify the opportunities where can be improved using the proactive procedure. This is achieved through mapping the proposed proactive IR procedure to the NHS WannaCry Ransomware Investigations Report [42] and Ransomware CTI advisories [35-37]. Fig. 2 presents the NHS WannaCry Ransomware attack IR procedure, including the actions evidenced by NHS, the missed actions, and the added proactive actions from CTI advisories. TABLE 2 elaborates the proactive actions in Figure 2. As we can see, the NHS has not fully addressed all actions in a traditional IR life cycle. The key missing items are, a clear security plan in the planning & preparation stage, the indicators, and precursors

in the detection stage, the lack of central directions on the mitigation strategy in the containment & eradication stage. NHS seems to have done well in the recovery and post-incident activity stages. NHS could have benefited from additional CTI sources to enable a proactive IR. The reports warning ransomware risks can improve NHS’s awareness in the planning & preparation stage, the indicators and precursors identified from CTI can be fed into their detection stage. By checking the impact score of ransomware incidents and the YARA rules from CTI sources can help identify and verify the incident, the CTI advisories (i.e. MS17-010 SMB) provide solutions for containment & eradication and recovery. CTI advisories on business continuity plans can have improved the IR capacity as part of the post-incident activity.

5. Discussion

The healthcare information system (HIS) has become a victim of the cyber attacks, such as the UK NHS WannaCry Ransomware attack. Cyber-attacks have been categorised as a Tier One Priority Risk in the UK National Security Strategy [19]. The UK government has established the National Cyber Security Centre (NCSC) to proactively mitigate cyber security risks. These initiatives recognise the importance of tracking and forecasting upcoming changes in the cyber landscape in order to proactively respond to potential cyber threats. This requires the organisation to be well aware of the threat landscape while responding to the emerging threats [24, 43, 44].

CTI can be used in numerous ways to help an organisation defend against the incidents, however, there is limited research found applying CTI into IR especially in healthcare. Our proposed proactive IR procedure contributes to the national strategy on cyber security through the research of proactive incident response informed by CTI. Proactive incident response can enhance the organisations’ capability in defending against attacks. Being aware of the threat landscape can reduce the uncertainty in making security decisions in IR processes [21].

Within the traditional IR, the preparation phase involves the creation and training of an incident response team, acquisition of the necessary tools and resources, the implementation of a set of controls on their assets based on the risk assessment. The detection and analysis stage collects and correlates the precursors, indicators, and log files to determine an incident and define the scope. The containment eradication and recovery phase include the incident response team attempting to mitigate the incident by containing it through defined strategies; eradicating by deleting the incident-related components from systems, networks, and applications then ideally recovering from it. The final post-incident activity includes a report issued by the organisation, detailing the cause of the incident as well as costs and steps that should be taken to prevent incidents in the future.

Throughout this phase, there tends to be a cycle of activities back to the detection and analyses, see figure 5.1. For example, to see if additional hosts are infected by malware



- Traditional IRactivities evidenced by NHS
- - - Traditional IRactivities not carried out by NHS but recommended
- Proactive IR activities added

FIGURE 2: Ransomware (WannaCry) Proactive IR informed by CTI

while eradicating the incident (5). Once the incidence is handled adequately,

There have been extensive CTI sources into the IR life cycle including preparation, detection, analysis, containment, eradication, and recovery of incidents. It will be important for an organisation to map their IR to CTI in order to get better coverage and ultimately provide a more robust security system. Teams within the organisation have the job of creating or implementing a vulnerability-free

system to protect their assets. There are numerous CTI sources that provide incidents information regarding assets, the exploited vulnerability (e.g. NVD [45]), and emerging threats (e.g. through US/UK CERT advisories, CTI sharing platforms) and security recommendations (e.g. security incident reports). Organisations should take advantage of this information to strengthen their IR capacity [46].

Our proposed proactive IR methodology addresses this challenge by embedding CTI into each stage of the

TABLE 2: Cyber Threat Intelligence (CTI) advisories for Ransomware (WannaCry) Incident Response (IR)

| WannaCry IR stages | CTI advisories |
|-----------------------------|---|
| Planning and preparation | NHS has taken inadequate actions against the alerts published in July 2016 warning that cyberattacks could jeopardise access to critical patient record systems. NHS will benefit from ransomware CTI advisories [35–37] on how to prevent such incidents; example solutions include rehearsing the IR plan before implementing it straight away. |
| Detection | The WannaCry incident report does not indicate whether NHS has used a monitoring system to identify the indicators. NHS can use the ransomware CTI advisories [35] to identify and feed indicators into the monitoring system through signature updates. Indicators include but are not limited to mssecsvc.exe, diskpart.exe, lhdfgui.exe, ransomware07_no_detection.exe, and WCry_WannaCry_ransomware.exe. |
| Analysis and assessment | NHS confirmed the WannaCry incident and identified the scope and impact. NHS can still benefit from CTI advisories [35–37] for the verification. The CTI advisories show that the impact can be “temporary or permanent loss of sensitive or proprietary information, disruption to regular operations, financial losses incurred to restore systems and files, and potential harm to an organisation’s reputation” [35]. |
| Containment and eradication | NHS lacked central direction and formalised process to respond to WannaCry incident. They failed to shut down/isolate the systems in time. Example solutions from CTI advisories are to apply MS17-010 SMB vulnerability dated March 14, 2017; enable spam filters to prevent phishing emails; and manage the use of privileged accounts. |
| Recovery | NHS worked with the IT suppliers to recover the system. CTI advisories [35–37] also provide a list of solutions to consider, e.g., backing up sensitive and important data regularly and testing the backups to ensure they work correctly upon use. |
| Lessons learned | NHS learned the lessons from this incident; they conducted causal analysis and took actions to improve the security controls and policies. CTI advisories [35–37] also provide some solutions like implementing a business continuity plan. |

traditional IR processes. It fills in the gap of CTI utilization in order to strengthen the IR capacity. The practitioners can benefit from informed IR decision-making using the CTI. This research also uses a case study to demonstrate the feasibility of the proactive IR methodology. The disadvantages are the lack of practical application in real practices and the lack of detailed explanation for each item in different IR stages. For example, cost estimation is embedded in the risk assessment item in the preparation stage but not displayed in the high-level proactive IR processes.

6. Challenges and Future Directions

The incident response includes the proper deployment of strategies, policies, and hardware and software security solutions in the organization [47]. The process of deciding which countermeasures and security policies will be applied against cyberattacks using CTI should take into account also the cost-perspective of the company. Even by applying a standard password policy that forces complex passwords, the company would increase the security budget due to extra costs induced from password creation and storage [48]. As stated in [49], correct modeling of the behavior of attackers and general users and proper calculation of the cost associated with the behavior of each entity could result in cost-efficient security policies.

Incident response of an organization could become more efficient by taking into account supply chain management, emerging technologies, privacy preservation techniques and even business analytics. The authors in [50] investigated the perspective of exploiting information analysis using BA inside Incident response plans in order to address the dynamic and uncertain cybersecurity threat environment. This

initial analysis could lead to the incorporation of BA into IR. Moreover, supply chain cybersecurity analysis can be used in order to calculate attack propagation and cascading effects [51]. This analysis can further improve IR and future work on this area is very promising especially if emerging technologies like blockchain or 5G are also taken into account [52–54]. Finally, secure arrangements for IoT healthcare privacy-preserving data collection must be taken into account when selecting the proper security solutions [55].

Regarding our proposed proactive IR procedure, our future work will focus on applying it in real practice in healthcare organisations. We will also consider integrating the proactive IR procedure with existing IR products used in Healthcare such as Security Information and Event Management (SIEM), Intrusion Detection Systems, Orchestration Automation and Response (SOAR) [18], and Security Operations Centre (SOC) in healthcare. This involves a careful mapping of CTI to each different component of the existing IR products. Future work will also consider expanding the current proactive IR processes by elaborating the items listed in each IR stage.

7. Conclusions

CTI contains knowledge of impending attacks, such as threat vectors, threat actors, victims profiles, courses of action, etc. and is shared via different CTI platforms such as UK/US Cert, Microsoft, MISP, and MITRE, with the intention to create a proactive line of cyber defense and should be ideally used to inform incident response, however, there is limited research in applying CTI into IR especially in healthcare. This paper addresses this gap by proposing a proactive IR procedure that is embedded with CTI. We examined

different stages of the IR procedure and identified the points where CTI can be fed into the IR processes. This paper then presented the NHS WannaCry Ransomware case study to demonstrate the use of the proposed proactive IR procedure. This research has significance for the IR practices within healthcare organizations. The practitioners can use the proposed proactive IR procedure to counteract Ransomware and other security-related adverse events in a systematic manner. Healthcare organizations can benefit from an informed proactive IR using CTI.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Kharaz, S. Arshad, C. Mulliner, W. Robertson, and E. Kirda, "Unveil: a large-scale, automated approach to detecting ransomware," in *Proceedings of the 25th USENIX Security Symposium USENIX Security 16*, pp. 757–772, Austin, TX., August 10–12, 2016.
- [2] M. K. Kagita, N. Thilakarathne, T. R. Gadekallu, P. K. R. Maddikunta, and S. Singh, "A review on cyber crimes on the internet of things," arXiv:2009.05708, 2020.
- [3] F. Donovan, "Healthcare industry takes brunt of ransomware attacks," [Online]. Available: <https://healthitsecurity.com/news/healthcare-industry-takes-brunt-of-ransomware-attacks>, 2019.
- [4] J. G. Ronquillo, J. Erik Winterholler, K. Cwikla, R. Szymanski, and C. Levy, "Health it, hacking, and cybersecurity: national trends in data breaches of protected health information," *JAMIA Open*, vol. 1, no. 1, pp. 15–19, 2018.
- [5] M. Field, *Wannacry cyber attack cost the nhs£ 92m as 19,000 appointments cancelled*, The Telegraph, 2018.
- [6] C. Luo, H. Soygazi, H. Janicke, and Y. He, "Security defense strategy for intelligent medical diagnosis systems (IMDS)," in *Proceedings of the 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 23–27 July 2019.
- [7] Y. He and C. Johnson, "Challenges of information security incident learning: an industrial case study in a Chinese healthcare organization," *Informatics for Health and Social Care*, vol. 42, no. 4, pp. 393–408, 2017.
- [8] Y. He and C. Johnson, "Improving the redistribution of the security lessons in healthcare: an evaluation of the generic security template," *International Journal of Medical Informatics*, vol. 84, no. 11, pp. 941–949, 2015.
- [9] M. Evans, Y. He, L. Maglaras, I. Yevseyeva, and H. Janicke, "Evaluating information security core human error causes (is-heck) technique in public sector and comparison with the private sector," *International Journal of Medical Informatics*, vol. 127, pp. 109–119, 2019.
- [10] M. Evans, Y. He, C. Luo et al., "Real-time information security incident management: a case study using the is-heck technique," *IEEE Access*, vol. 7, pp. 142147–142175, 2019.
- [11] M. Evans, Y. He, C. Luo, I. Yevseyeva, H. Janicke, and L. A. Maglaras, "Employee perspective on information security related human error in healthcare: proactive use of is-heck in questionnaire form," *IEEE Access*, vol. 7, pp. 102087–102101, 2019.
- [12] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, *Computer security incident handling guide*, NIST Special Publication, vol. 800, no. 61, pp. 1–147, 2012.
- [13] M. Souppaya and K. Scarfone, *Guide to malware incident prevention and handling for desktops and laptops*, NIST Special Publication, vol. 800, p. 83, 2013.
- [14] A. Ahmad, S. B. Maynard, and G. Shanks, "A case analysis of information systems and security incident responses," *International Journal of Information Management*, vol. 35, no. 6, pp. 717–723, 2015.
- [15] S. Dekker, *The field guide to understanding 'human error'*, CRC Press, 2017.
- [16] S. L. Pfleeger and R. K. Cunningham, "Why measuring security is hard," *IEEE Security & Privacy Magazine*, vol. 8, no. 4, pp. 46–54, 2010.
- [17] Y. T. Chua, S. Parkin, M. Edwards et al., "Identifying unintended harms of cybersecurity countermeasures," in *Proceedings of the 2019 APWG Symposium on Electronic Crime Research (eCrime)*, pp. 1–15, IEEE, Pittsburgh, PA, USA, 13–15 Nov. 2019.
- [18] C. Islam, M. A. Babar, and S. Nepal, "A multi-vocal review of security orchestration," *ACM Computing Surveys*, vol. 52, no. 2, p. 37, 2019.
- [19] P. Hammond, *National Cyber Security Strategy 2016 to 2021*, Her Majesty's Government, London, 2016.
- [20] S. Barnum, *Standardizing cyber threat intelligence information with the structured threat information expression (stix)*, Mitre Corporation, vol. 11, pp. 1–22, 2012.
- [21] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Computers & Security*, vol. 72, pp. 212–233, 2018.
- [22] Y. Ying He, L. A. Maglaras, H. Janicke, and K. Jones, "An industrial control systems incident response decision framework," in *Proceedings of the 2015 IEEE Conference on Communications and Network Security (CNS)*, pp. 761–762, IEEE, Florence, Italy, 28–30 Sept. 2015.
- [23] E. W. Burger, M. D. Goodman, P. Kampanakis, and K. A. Zhu, "Taxonomy model for cyber threat intelligence information exchange technologies," in *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, pp. 51–60, ACM.
- [24] S. Qamar, Z. Anwar, M. A. Rahman, E. Al-Shaer, and B.-T. Chu, "Data-driven analytics for cyber-threat intelligence and information sharing," *Computers & Security*, vol. 67, pp. 35–58, 2017.
- [25] S. E. Dog, A. Tweed, L. Rouse et al., "Strategic cyber threat intelligence sharing: a case study of ids logs," in *Proceedings of the 2016 25th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–6, IEEE, Waikoloa, HI, USA, 1–4 Aug. 2016.
- [26] V. G. Li, M. Dunn, P. Pearce, D. McCoy, G. M. Voelker, and S. Savage, "Reading the tea leaves: a comparative analysis of threat intelligence," in *Proceedings of the 28th Security Symposium (Security 19)*, pp. 851–867.
- [27] F. Maymí, R. Bixler, R. Jones, and S. Lathrop, "Towards a definition of cyberspace tactics, techniques and procedures," in *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)*, pp. 4674–4679, IEEE, Boston, MA, USA, 11–14 Dec. 2017.
- [28] J. Moreno, M. A. Serrano, E. B. Fernandez, and E. Fernández-Medina, "Improving incident response in big data ecosystems by using blockchain technologies," *Applied Sciences*, vol. 10, no. 2, p. 724, 2020.

- [29] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, 2016.
- [30] G. Grispos, W. B. Glisson, and T. Storer, "Enhancing security incident response follow-up efforts with lightweight agile retrospectives," *Digital Investigation*, vol. 22, pp. 62–73, 2017.
- [31] R. Brewer, "Ransomware attacks: detection, prevention and cure," *Network Security*, vol. 2016, no. 9, pp. 5–9, 2016.
- [32] N. A. Hassan, *Ransomware families Ransomware Revealed*, pp. 47–68, Springer, 2019.
- [33] N. Kyurkchiev, A. Iliev, A. Rahnev, and T. Terzieva, "A new analysis of cryptolocker ransomware and welchia worm propagation behavior. some applications. iii," *Communications in Applied Analysis*, vol. 23, no. 2, pp. 359–382, 2019.
- [34] R. Brewer, "Cyber threats: reducing the time to detection and response," *Network Security*, vol. 2015, no. 5, pp. 5–8, 2015.
- [35] U. S. CERT, "Indicators associated with WannaCry ransomware," [Online]. Available: <https://www.us-cert.gov/ncas/alerts/TA17-132A>, 2017.
- [36] Stop Ransomware, "Ransomware," [Online]. Available: <https://www.us-cert.gov/security-publications/Ransomware>, 2019.
- [37] Petya Ransomware, "Alert (TA17-181A). Petya ransomware," [Online]. Available: <https://www.us-cert.gov/ncas/alerts/TA17-181A>, 2017.
- [38] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *Journal of Information Security and Applications*, vol. 50, p. 102419, 2020.
- [39] A. Ahmim, M. A. Ferrag, L. Maglaras, M. Derdour, and H. Janicke, "A detailed analysis of using supervised machine learning for intrusion detection," in *Strategic Innovative Marketing and Tourism*, pp. 629–639, Springer, 2020.
- [40] C. Wagner, A. Dulaunoy, G. Wagener, and A. Iklody, "MISP: the design and implementation of a Collaborative threat intelligence sharing platform," in *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, pp. 49–56.
- [41] M. Conti, T. Dargahi, and A. Dehghantanha, *Cyber threat intelligence: challenges and opportunities*, Springer, 2018.
- [42] D. of Health, *Investigation: WannaCry cyber attack and the NHS*, [Online]. Available: <https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf>, 2018.
- [43] S. Samtani, K. Chinn, C. Larson, and H. Chen, "Azsecure hacker assets portal: cyber threat intelligence and malware analysis," in *Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pp. 19–24, IEEE, Tucson, AZ, USA, 28–30 Sept. 2016.
- [44] S. Samtani, R. Chinn, H. Chen, and J. F. Nunamaker Jr, "Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence," *Journal of Management Information Systems*, vol. 34, no. 4, pp. 1023–1053, 2017.
- [45] M. A. Williams, S. Dey, R. C. Barranco, S. M. Naim, M. S. Hossain, and M. Akbar, "Analyzing evolving trends of vulnerabilities in national vulnerability database," in *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*, pp. 3011–3020, IEEE, Seattle, WA, USA, 10–13 Dec. 2018.
- [46] B. Ndibanje, K. Kim, Y. Kang, H. Kim, T. Kim, and H. Lee, "Cross-method-based analysis and classification of malicious behavior by api calls extraction," *Applied Sciences*, vol. 9, no. 2, p. 239, 2019.
- [47] L. Maglaras, M. A. Ferrag, A. Derhab, M. Mukherjee, H. Janicke, and S. Rallis, "Threats, protection and attribution of cyber attacks on critical infrastructures," *arXiv:1901.03899*, 2019.
- [48] L. Maglaras, H. Janicke, and M. A. Ferrag, "The cost perspective of password security," in *Handbook of Research on Multimedia Cyber Security*, pp. 319–330, IGI Global, 2020.
- [49] S. Gong and C. Lee, "Cyber threat intelligence framework for incident response in an energy cloud platform," *Electronics*, vol. 10, no. 3, p. 239, 2021.
- [50] H. Naseer, S. B. Maynard, and K. C. Desouza, "Demystifying analytical information processing capability: the case of cybersecurity incident response," *Decision Support Systems*, vol. 143, p. 113476, 2021.
- [51] A. Yeboah-Ofori and S. Islam, "Cyber security threat modeling for supply chain organizational environments," *Future Internet*, vol. 11, no. 3, p. 63, 2019.
- [52] L. Maglaras and I. Kantzavelou, *Cybersecurity Issues in Emerging Technologies*, CRC Press, 2021.
- [53] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and puf-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, p. 1, 2021.
- [54] H. Xiong, C. Jin, M. Alazab et al., "On the design of blockchain-based ecdsa with fault-tolerant batch verification protocol for blockchain-enabled iomt," *IEEE Journal of Biomedical and Health Informatics*, p. 1, 2021.
- [55] J. Song, Z. Han, W. Wang, J. Chen, and Y. Liu, "A new secure arrangement for privacy-preserving data collection," *Computer Standards & Interfaces*, vol. 80, p. 103582, 2022.

Research Article

A Novel One-Shot Object Detection via Multifeature Auxiliary Information

Yu Song , Min Li , Weidong Du, Yao Gou , Zhaoqing Wu , and Yujie He

Xi'an Institute of High Technology, Xi'an, Shaanxi 710025, China

Correspondence should be addressed to Yu Song; huogongdaoren@126.com

Received 14 December 2021; Accepted 17 January 2022; Published 21 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Yu Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advantage of using only a limited number of samples, few-shot learning has been developed rapidly in recent years. It is mostly applied in the object classification or detection of a small number of samples which is typically less than ten. However, there is not much research related to few-shot detection, especially one-shot detection. In this paper, the multifeature information-assisted one-shot detection method is proposed to improve the accuracy of one-shot object detection. Specifically, two auxiliary modules are applied to the detection algorithm: Semantic Feature Module (SFM) and Detail Feature Module (DFM), which, respectively, extract semantic feature information and detailed feature information of samples in the support set. Then these two kinds of information are then calculated with the feature image extracted from the query image to obtain the corresponding auxiliary information that is used to complete one-shot detection. Thanks to the two auxiliary modules, which can retain more semantic and detailed information of samples in the support set, the proposed method can enhance the utilization rate of sample feature information and improve object detection accuracy by 2.97% compared to the benchmark method.

1. Introduction

Deep neural networks have been widely used in computer vision, such as posture recognition [1] and plant disease recognition [2], and object detection is the research hotspot in this field. Generally speaking, object detection algorithms can be divided into two categories according to different training strategies: one-stage object detection and two-stage object detection. The popular algorithms are the YOLO algorithms [3–5] and R-CNN algorithms [6–8], which dramatically improve the object detection effect and enhance detection efficiency. However, those algorithms rely on object annotation information, which cannot be easily obtained. Therefore, researchers gradually focus on few-shot detection.

Few-shot detection is derived from few-shot learning, a particular case of meta-learning. At present, the learning methods can be roughly broken down into four categories: measurement learning-based learning, meta-learning-based learning, data enhancement-based learning, and multimodal approaches-based learning, among which the meta-

learning-based learning method is the most popular. In the meta-training stage, by compositing several samples from different classes to take different meta-task, the model can learn the differences between examples of various categories and the similarities between samples of the same type. While in the meta-testing stage, the recognition task can be completed without retraining or only with a small amount of rapid training for a new category. However, few-shot learning is mainly used in few-shot classification rather than few-shot detection.

Few-shot object detection is used to complete detection for objects with very few samples in the dataset. The existing few-shot detection methods fall into three categories: fine-tuning, model structure-based learning, and metric-based learning. The few-shot detection training strategy generally contains two stages: meta-training stage and fine-tuning stage. In the meta-training, N categories were randomly chosen from the training set, each containing K samples to form the support set of the model, namely, a meta-task. Next, a small number of object samples were selected to fine-tune the model. The purpose was to train the model to detect

N classes of objects from $N \times K$ data and then generalize the knowledge to adapt to new classes. This task is called N -way K -shot. In few-shot learning, the K value is usually smaller than ten; when $K = 1$, it is called one-shot learning.

Existing metric-based few-shot detection mainly divides the dataset into the support set and the query set. It selects several image samples from the two sets to form the minimum training unit task (meta-task) and then trains the model through specific strategies. The detection algorithm first obtains the corresponding features of the images in the two sets, then measures the distance between the two features, and judges the object category according to the distance. According to the label's location information, the regression operation is performed to complete the object positioning. We note that multiple features with different scales will be generated when the convolutional neural network extracts features of the support images. However, as the current algorithm only conducts simple distance measurement, the utilization rate of object feature information is extremely low.

To solve this problem, this paper proposed a novel one-shot detection method on the basis of metric-based learning. The main contributions of this paper are as follows:

- (1) This novel method integrated the Semantic Feature Module (SFM) and the Detail Feature Module (DFM), which generated features about the support images of two sizes (7×7 , 3×3). These features were then operated with query images' feature and generated the corresponding multifeature auxiliary information (MFAI) of Semantic Feature Auxiliary Information and Detailed Feature Auxiliary Information.
- (2) Experimental results showed that both the SFM and the DFM could increase the accuracy of one-shot detection. A combination of the two modules could even increase the detection accuracy by 2.97% compared to the original algorithm.

2. Related Works

In recent years, research on few-shot learning has attracted a lot of interest, which can effectively solve the classification and detection task using only a few labeled samples. The recent related works of few-shot classification, few-shot object detection, and one-shot object detection are listed in Table 1.

In the general one-shot object detection method, the weight extracted from the image is mainly used to measure the object feature distance. However, the semantic feature information and detailed feature information of the support set object are not fully utilized. This paper introduced the SFM and the Object Detail Module based on one-shot object detection, inspired by literature [18]. By using more object features to train the deep neural network, the detection effect of our model was better.

3. Method

The semantic information and detailed information need to be generated separately. In theory, the support images can obtain feature images of different sizes through different

modules. In our method, the 7×7 feature can retain more object details, while the 3×3 feature contains more semantic information about the object. The 7×7 feature and feature of the query image were used for dot product operation. The 3×3 feature is convolved with the feature of the query image. The corresponding MFAI of Semantic Feature Information and Detail Feature Information was generated through the above two operations.

3.1. Training Strategy. As mentioned above, in the training stage, assume that the dataset is D and divided into D_{base} and D_{novel} . D_{base} represents an object image dataset which contains a large number of annotated images, of which category is C_{base} ; and D_{novel} represents a dataset containing a few of samples with annotations with category as C_{novel} . With $D_{\text{base}} \cap D_{\text{novel}} = \emptyset$, $C_{\text{base}} \cap C_{\text{novel}} = \emptyset$, so the ultimate goal of one-shot detection is to classify and locate the object of query image in D_{novel} . Similar to the literature [16], the whole training process is divided into two steps. Firstly, the data in C_{base} are used to train the model, so the model can learn the meta-features; then this trained model can have a good detection effect on the object of the base class. Lastly, the $D_{\text{base}} \cup D_{\text{novel}}$ dataset is utilized for model training and fine-tuning to adapt to the new category, then generalizing the knowledge learned in the first step to the new object category.

Based on few-shot learning, we innovatively utilized the semantic information and detailed information of the object to complete the one-shot detection. Specifically, we used the smallest training unit $T = \{(S_i, x_i)\}_{i=1}^{|T|}$ in the training stage. The data in T are all from the randomly sampled support set in D_{base} , S_i , and the query image, x_i . i represents the i -th task. S_i contains N categories, with K samples in each class. Through multitask training, we obtained an object model with a detection base class. Next, the fine-tuned model was continuously adapted with the $D_{\text{base}} \cup D_{\text{novel}}$ dataset, and the detection model $f_\theta(x|S)$ was fine-tuned to fit the new category, in which θ is the parameter that the model needs to learn. The final one-shot detection could be completed by f_θ tuned well.

3.2. Multifeature Information-Assisted Detection Method. Firstly, the feature extraction module was used to extract the 7×7 feature image of query image. Then the support images were input into the DFM, SFM, and Weighted Module (WM), respectively, to get the corresponding 7×7 , 3×3 , 1×1 feature maps, while the channels were consistent. Next, we applied dot product 7×7 feature of support images with 7×7 feature generated by query image, and finally generated 7×7 feature—Detail Feature Auxiliary Information (DFAI), then the 7×7 feature generated by query image was convolved with the 3×3 feature of support images as a filter, and finally generated 7×7 feature—Semantic Feature Auxiliary Information (SFAI). The average operation was carried out on the two kinds of auxiliary information. The next step was to convolve the processed averaged feature (7×7 feature) with the weight information (1×1 feature) generated by the support images; thus, the 7×7 MFAI to be detected was

TABLE 1: Recent related works.

| Category | Ref. | Methods |
|---------------------------|------|---|
| Few-shot classification | [9] | The Mahalanobis distance in a state-of-the-art few-shot learning approach (CNAPS [10]) is adopted to improve performance |
| | [11] | Presents a novel network to learn and preserve the feature manifold's topology formed by different classes |
| | [12] | Proposes the similarity ratio as an indicator for the generalization performance of a few-shot model |
| | [13] | Takes advantage of the earth mover's distance (EMD) to measure the distance between dense image representations which determines image relevance |
| | [14] | Merges three learning methods: visual feature learning, knowledge inferring, and classifier learning, into a unified framework |
| Few-shot object detection | [15] | Introduces the oPen sEt mEta LEaRning (PEELER), which randomly selects a set of novel classes, maximizes the posterior entropy over every sample, and utilizes the Mahalanobis distance as a new metric |
| | [16] | Improves the CentreNet detector for the few-shot learning and a class-specific code generator is modeled by meta-learning |
| | [17] | Uses Attention-RPN, multirelation detector and contrastive training strategy to detect novel objects |
| One-shot object detection | [18] | Proposes the model that uses labeled base categories and quickly improves to new categories, utilizing a meta-feature learner and a new upgraded module |
| | [19] | Develops coattention and coexcitation framework (CoAE) that contributes to several technical aspects |
| | [20] | Develops a new algorithm to guide the parameter posterior towards its true distribution to remedy the posterior fading problem that compromises the effectiveness of shared weights |

generated. Finally, the detection feature map was put into the detection network to generate the classification and location information. The process was described as shown in Figure 1.

To explain the function of each module in detail, we elaborated the function into two steps. The first step was to combine the Semantic Feature Auxiliary Module (SFAM) with the WM to output the feature to be detected; then combine the dot product information auxiliary module with the WM to output the feature to be detected. The former is illustrated in Figure 2. Firstly, query image and support images were input into both the SFM and the feature extraction module to extract their respective features. Then the feature extraction module outputs the $1024 \times 7 \times 7$ feature I_Q , the SFM outputs the $N \times 1024 \times 3 \times 3$ feature S_C , and the WM outputs the $N \times 1024 \times 1 \times 1$ feature S_W . After the convolution operation of S_C and I_Q , F_C was generated, as shown in

$$F_C = S_C * I_Q. \quad (1)$$

Then, S_W and F_C were convolved to get the $N \times 1024 \times 7 \times 7$ feature Y_C to be detected, as shown in

$$Y_C = S_W * F_C, \quad (2)$$

$$Y_C = S_W * (S_C * I_Q). \quad (3)$$

The DFM is illustrated in Figure 3. Similarly, the query image and support images were put into the feature extraction module and the DFM to extract their respective features. So, the feature extraction module outputs the $1024 \times 7 \times 7$ feature I_Q , the DFM outputs the $N \times 1024 \times 7 \times 7$ feature S_D , and the WM outputs the $N \times 1024 \times 1 \times 1$ feature S_W . After the dot production operation of S_D and I_Q , F_D was generated, as shown in

$$F_D = S_D \otimes I_Q. \quad (4)$$

Then, S_W and F_D were convolved to obtain the $N \times 1024 \times 7 \times 7$ feature Y_D to be detected, as shown in

$$Y_D = S_W * F_D, \quad (5)$$

$$Y_D = S_W * (S_D \otimes I_Q). \quad (6)$$

3.3. Loss Function. To handle the various objects which need to be detected in one-shot detection, the model in this paper adopted a softmax layer [18]. The predicted score on classification for the i -th class was represented by $\hat{c} = e^{c_i} / \sum_{j=1}^N e^{c_j}$. To get better model convergence, the cross-entropy loss over the calibrated scores \hat{c} was adopted, as shown in

$$L_c = - \sum_{i=1}^N O(\cdot, i) \log(\hat{c}_i), \quad (7)$$

where $O(\cdot, i)$ is an indicator function. When the current anchor box fits into class i , its value is 1. Otherwise, the value is 0. For the bounding box regression calculation and object determination method, we followed the same detection way as YOLOV3. After anchors with different aspect ratios were preset, coordinate classification of anchors would be processed through calculation and finally predicted the object. In this paper, corresponding loss functions were adopted, such as the Mean Squared Error (MSE) loss and the Binary Cross-Entropy (BCE) loss. The loss function of multifeature information-assisted one-shot detection proposed in our model is shown in

$$L_{\text{det}} = L_c + L_{\text{bbx}} + L_{\text{obj}}, \quad (8)$$

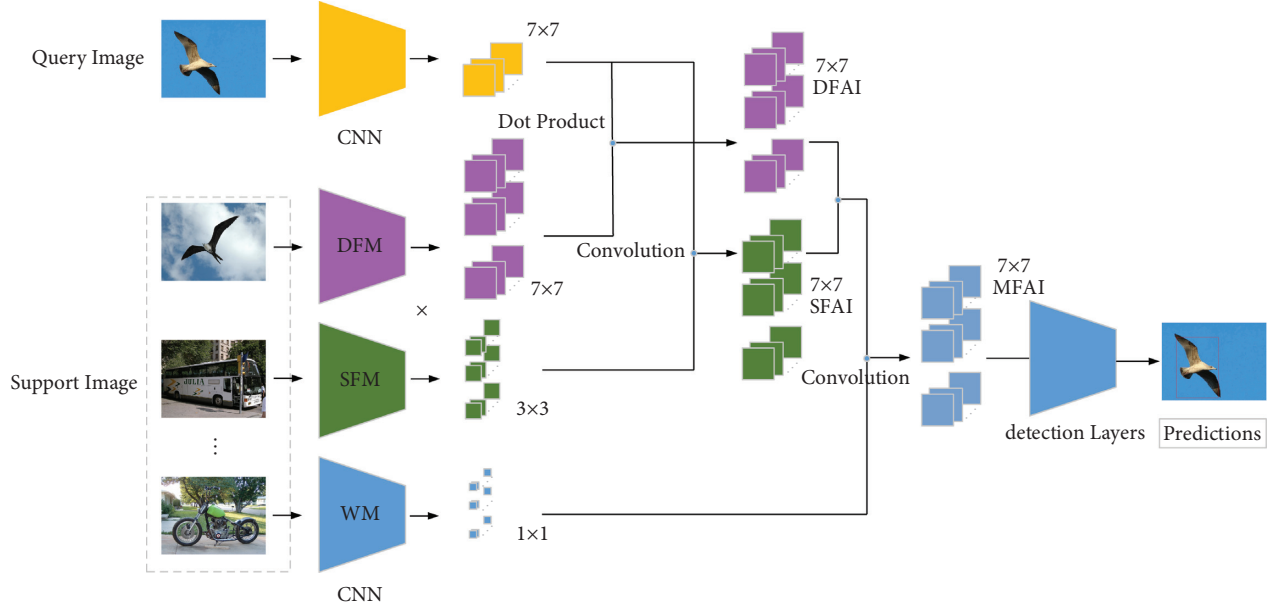


FIGURE 1: The structure of the one-shot detection assisted by multiple feature information.

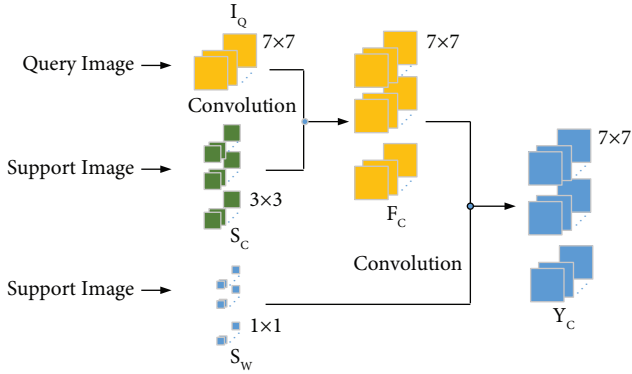


FIGURE 2: Schema of the semantic feature auxiliary module.

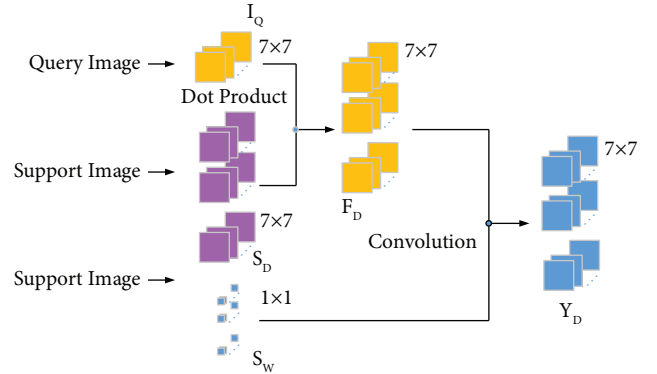


FIGURE 3: Schema of the detail feature auxiliary module.

where L_{bbox} is expressed as

$$L_{bbox} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (2 - w_i \times h_i) \left[(a_i - \hat{a}_i)^2 + (b_i - \hat{b}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right], \quad (9)$$

where S means the grid set in YOLO and S^2 represents 13×13 , 26×26 , and 52×52 . B stands for prediction box. I_{ij}^{obj} means the box at i, j , which is 1 if it is an object; otherwise, it is 0. a_i, b_i, w_i, h_i represent the central point coordinates and the width and height of the object, respectively. $\hat{a}_i, \hat{b}_i, \hat{w}_i, \hat{h}_i$ represent the predicted values of the center point coordinates, width, and height, respectively.

L_{obj} can be expressed as in

$$L_{obj} = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (c_i - \hat{c}_i)^2, \quad (10)$$

where I_{ij}^{noobj} means the box at i, j which is 1 if it is not an object; otherwise, it is 0. \hat{c}_i is the prediction confidence on the object for class i . \hat{c}_i is set to 1 if the object is the true value of a certain class; otherwise, it is 0.

4. Experiment

4.1. Experimental Environment. The running environment of the algorithm verification experiment is shown in Table 2.

4.2. Dataset. The datasets we adopted are VOC 2007 [21] and VOC 2012 [22], which are the widely used object detection benchmarks. Out of the 20 categories, we selected samples of five categories as novel datasets and the remaining 15 categories as the base datasets. The model training was divided into two stages: base training stage and one-shot fine-tuning stage. In the base training stage, the images of base samples were used to train the normal model in the supervised mode. And in the one-shot fine-tuning

TABLE 2: Runtime environment.

| Item | Parameter |
|----------------------------------|-----------------------------|
| | Titan Xp (12G video memory) |
| GPU | CUDA 10.0 cuDNN 7.0 |
| Operating system | Ubuntu 16.04 |
| Python version | 3.6 |
| Iterations | 60 |
| Learning rate | 0.001 |
| Momentum | 0.9 |
| Momentum attenuation coefficient | 0.00004 |

stage, the images of novel ones were used to ensure that each class of objects only had one annotated bounding box.

4.3. Experiment and Analysis. A large number of experiments have been done on these datasets. To illustrate the effectiveness of the different methods, several representative datasets were selected. In the test phase, we used five unseen categories of data in training: bird, bus, cow, motorbike, and sofa. Due to the space limitation, we only present the results data of bird and bus categories in Figures 4 and 5.

As can be seen from the above Figure 4, there are a total of 12 images arranged in four rows, with three object images in each row. All the object images used the same kind of detection algorithm. The first line to the fourth line, respectively, shows the detection results of the original algorithm [18], the detailed feature information auxiliary algorithm, the semantic feature information auxiliary algorithm, and the multifeature information auxiliary algorithm proposed in this paper. Different detection results of the same object are presented in each column for those four algorithms. For column (a), the object is conspicuous, so all four algorithms can correctly identify it. As the background of the object in column (b) is relatively complex, the original algorithm and the auxiliary algorithm of detailed feature information cannot detect the object well. In contrast, the algorithms in the third and fourth lines that integrate semantic information can detect the object more accurately. Since the objects in column (c) belong to multiobject detection in a complex background; the detection effect of the second row is not ideal. The algorithms in the first and third rows can completely detect conspicuous objects. While the fourth line algorithm can detect multiple objects, the second object detection is not complete because there is little difference between the background color and the object color.

Similarly, there are also 12 images in Figure 5, which are arranged in the same manner. The object image of each row uses the same kind of detection algorithm. The first row to the fourth row, respectively, represents the detection results of different buses by the four algorithms. For column (a), the object is prominent, so all the four algorithms can correctly identify it. In column (b), the background of images is relatively complex. Although the original algorithm and the

detailed feature assist algorithm can locate the object more accurately, there is still a misjudgment in the classification of extracted features. As a result, the bus is misclassified as a train. The algorithms in the third and fourth rows incorporate semantic information to detect objects more accurately. Since the object in column (c) is similar to a train, the four algorithms misjudge the result. That is why the accuracy of bus detection results is low.

4.4. Ablation Experiments. The accuracy improvement of our proposed method was due to the multifeature auxiliary detection mechanisms, that is, SFM and DFM. To illustrate the importance of these modules, we implemented ablation experiments by disabling different modules.

4.4.1. Semantic Feature Auxiliary Detection Algorithm. Figure 6 shows the flow chart of the semantic feature auxiliary detection algorithm. Two modules were constructed, namely, the SFAM and the WM. The semantic feature-assisted detection algorithm was compared with the original algorithm in the experiment. As can be seen in Table 3, semantic feature assistance achieves better performance.

4.4.2. Detail Feature Auxiliary Detection Algorithm. The detail feature auxiliary detection algorithm was implemented, as shown in Figure 7. Two modules were also constructed: the detail feature auxiliary module (DFAM) and the WM. In the experiment, this detail feature-assisted detection algorithm was also compared with the original algorithm. As indicated in Table 3, the performance of the detail feature-assisted detection algorithm is only better than that of the basic weight network.

4.4.3. Multifeature Auxiliary Detection Algorithm. In the multifeature auxiliary detection algorithm, both the above modules were introduced into the network and fused with the basic weighted network structure.

The detection results of the above three algorithms and the original benchmark algorithm are shown in Table 3.

Table 3 shows the comparison results of those algorithms proposed in this paper and the original algorithm, from which we can see the performance of the designed modules. On the left side of the table are different algorithms, while on the right side are the corresponding experimental results. The average precision (AP) of objects in the five classifications has also been calculated separately for the bird, bus, cow, motorbike, and sofa. The mAP represents the mean of AP for these five novel classes. The first row is the original algorithm, and the other three are related algorithms proposed in this paper. The second and third rows are the results of ablation experiments. The second algorithm adds the DFAM to the original algorithm, which is called the detail feature auxiliary detection algorithm. It can be observed that the AP for the bird, cow, motorbike, and sofa is higher than

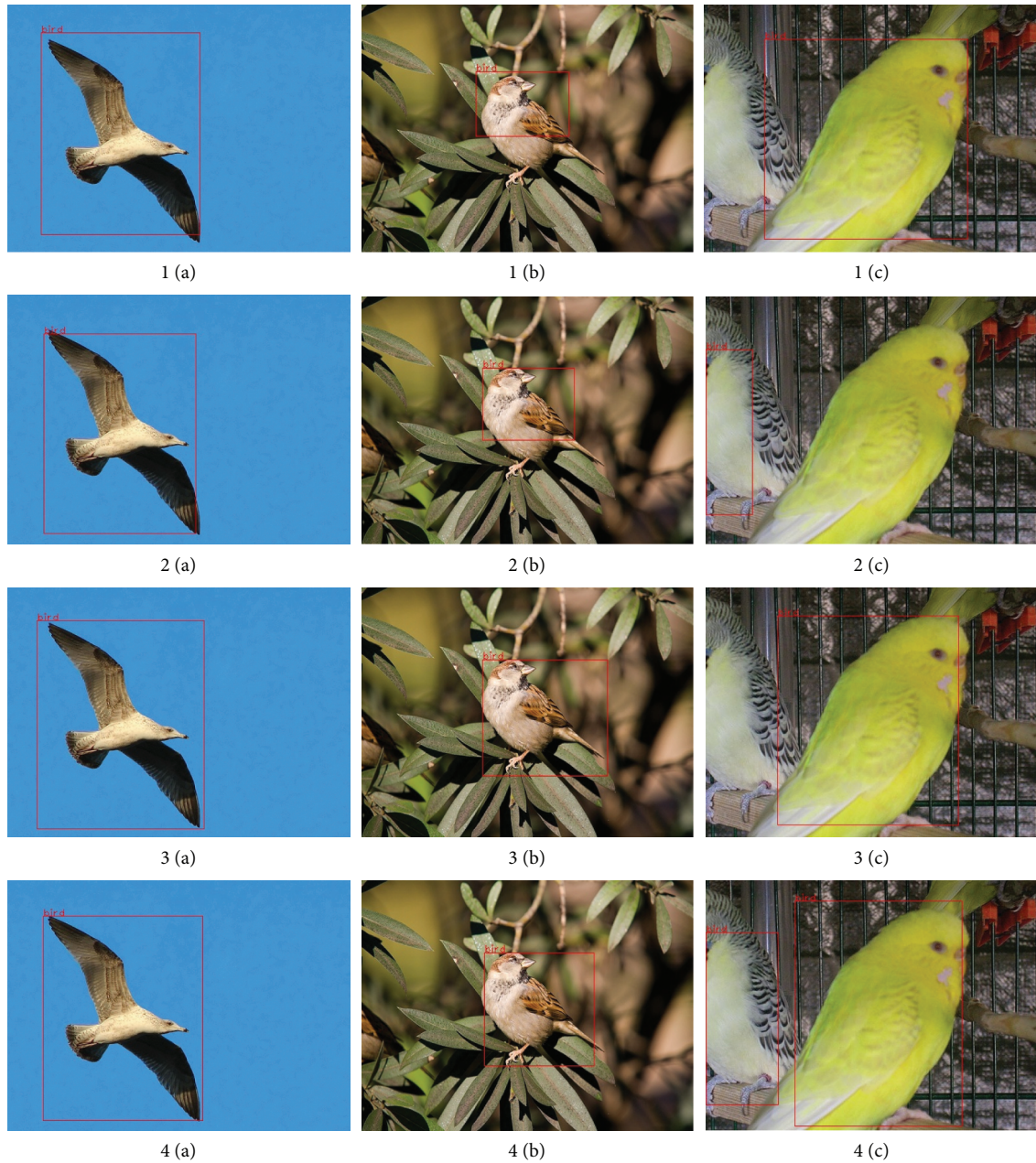


FIGURE 4: Bird object detection results.

that of the original algorithm, while it is not the case for the bus. Well, the performance is improved by less than 1%. The third row adds the SFAM to the original algorithm, which is called the semantic feature auxiliary detection algorithm. It can be observed that the AP of the algorithm is higher than that of the original algorithm, with an increase of 2.58% in AP and 1.75% in mAP for the bus. It can also be seen that the SFAM proposed in this paper does enhance the semantic information of objects and promotes classification accuracy and detection precision. The fourth row is the result of the

multifeature auxiliary detection algorithm proposed in this paper. Notably, the algorithm has enhanced the detailed information and semantic information, and the detection results of all those five novel objects are superior to the original algorithm. In particular, the AP of the bus has been improved by 5.06%, and the mAP has been improved by 2.97%. The detection results in the fourth row show that the detail feature and semantic feature, two auxiliary information introduced in the algorithm, successfully improve the AP and mAP of one-shot object detection.



FIGURE 5: Bus object detection results.

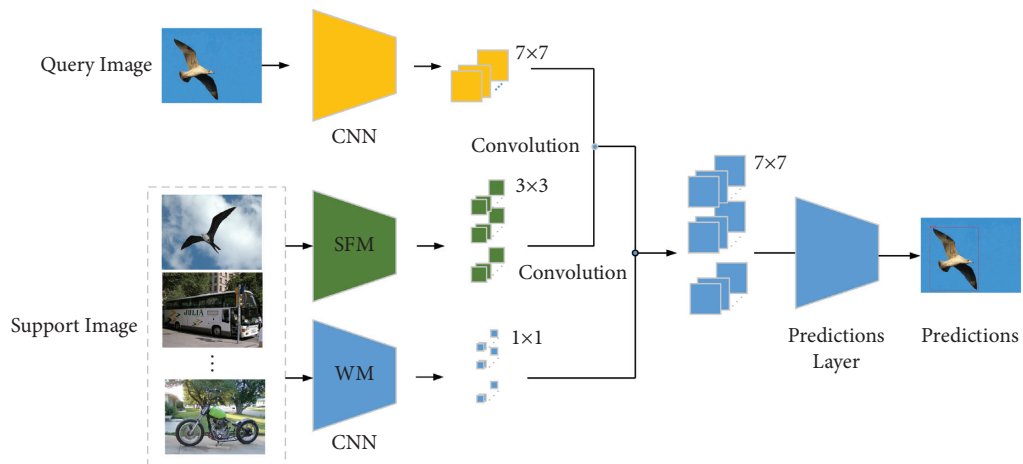


FIGURE 6: One-shot detection algorithm structure assisted by semantic information.

TABLE 3: Ablation experiment results.

| Methods | Novel | | | | | |
|--------------------------------------|--------------|-------------|--------------|--------------|--------------|--------------|
| | Bird | Bus | Cow | Mbike | Sofa | mAP |
| Benchmark algorithm [16] | 24.12 | 3.45 | 24.97 | 26.44 | 27.51 | 21.3 |
| Detail feature auxiliary detection | 24.52 | 3.27 | 26.58 | 27.03 | 28.14 | 21.91 |
| Semantic feature auxiliary detection | 25.37 | 6.03 | 26.32 | 28.09 | 29.44 | 23.05 |
| Multifeature auxiliary detection | 27.02 | 8.51 | 27.8 | 28.67 | 29.35 | 24.27 |

The bold values indicate that the performance of the proposed method is better than that of the Benchmark algorithm.

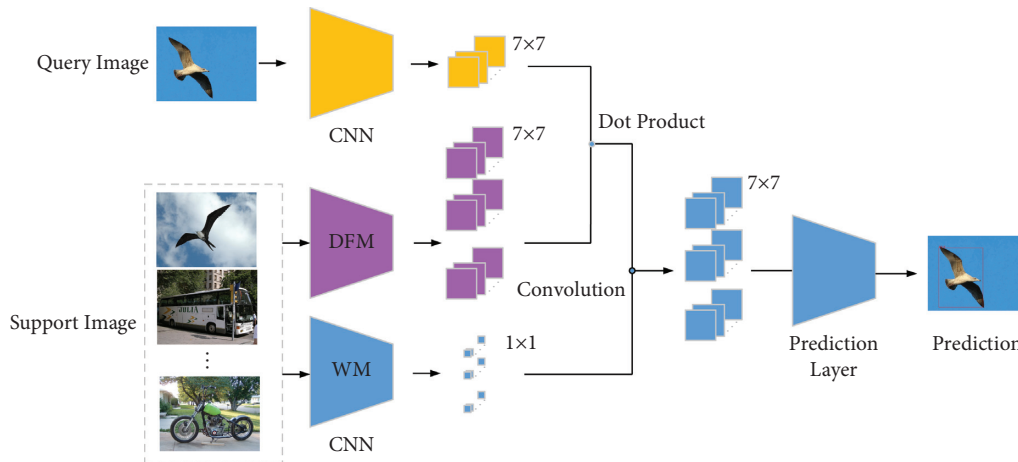


FIGURE 7: One-shot detection algorithm structure assisted by detailed information.

5. Conclusion

In this paper, a novel one-shot detection method based on multifeature auxiliary information was proposed. Compared to previous studies, this algorithm utilized two auxiliary mechanisms: the Semantic Feature Module and the Detail Feature Module, which significantly improved the detection effect of a single sample object. Experimental results on public datasets demonstrate that the new proposed algorithm has better one-shot detection performance than the original method. To further evaluate the advanced performance, an ablation experiment was conducted. Experiments showed that the two auxiliary modules play a positive role in the detection results. The combined detection accuracy of the two modules has been increased by 2.97% compared to the benchmark algorithm. In the future, we will apply this proposed method to other types of datasets, including infrared images and SAR images. Although this method is mainly for one-shot object detection, we also look forward to its application in few-shot object detection.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant no. 62006240.

References

- [1] T. R. Gadekallu, M. Alazab, R. Kaluri, and P. Reddy, "Hand gesture classification using a novel CNN-crow search algorithm," *Complex & Intelligent Systems*, vol. 7, no. 6, 2021.
- [2] T. R. Gadekallu, D. S. Rajput, M. Reddy et al., "A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU," *Journal of Real-Time Image Processing*, pp. 1–14, 2020.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, IEEE, Las Vegas, NV, USA, June 2016.
- [4] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, IEEE, Honolulu, HI, USA, July 2017.
- [5] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2021, <https://arxiv.org/pdf/1804.02767.pdf>.
- [6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, IEEE, Columbus, OH, USA, June 2014.

- [7] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, IEEE, Santiago, Chile, December 2015.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [9] P. Bateni, R. Goyal, V. Masrani, F. Wood, and L. Sigal, "Improved few-shot visual classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14493–14502, IEEE, Seattle, WA, USA, June 2020.
- [10] J. Requeima, J. Gordon, J. Bronskill, S. Nowozin, and R. E. Turner, "Fast and flexible multi-task classification using conditional neural adaptive processes," 2021, <https://arxiv.org/abs/1906.07697v2>.
- [11] X. Tao, X. Hong, X. Chang, S. Dong, X. Wei, and Y. Gong, "Few-shot class-incremental learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12183–12192, IEEE, Seattle, WA, USA, June 2020.
- [12] L. Zhou, P. Cui, X. Jia, S. Yang, and Q. Tian, "Learning to select base classes for few-shot classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4624–4633, IEEE, Seattle, WA, USA, June 2020.
- [13] C. Zhang, Y. Cai, G. Lin, and C. Shen, "DeepEMD: few-shot image classification with differentiable earth mover's distance and structured classifiers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12203–12213, IEEE, Seattle, WA, USA, June 2020.
- [14] Z. Peng, Z. Li, J. Zhang, Y. Li, G. Qu, and J. Tang, "Few-shot image recognition with knowledge transfer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 441–449, IEEE, Seoul, South Korea, November 2019.
- [15] B. Liu, H. Kang, H. Li, and G. Hua, "Few-shot open-set recognition using meta-learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8798–8807, IEEE, Seattle, WA, USA, June 2020.
- [16] J.-M. Perez-Rua, X. Zhu, H. Timothy, and T. Xiang, "Incremental few-shot object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13846–13855, IEEE, Seattle, WA, USA, June 2020.
- [17] Q. Fan, W. Zhuo, C.-K. Tang, and Y. Tai, "Few-shot object detection with attention-RPN and multi-relation detector," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4013–4022, IEEE, Seattle, WA, USA, June 2020.
- [18] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8420–8429, IEEE, Seoul, South Korea, November 2019.
- [19] T.-I. Hsieh, Y.-C. Lo, H.-T. Chen, and T. L. Liu, "One-shot object detection with co-attention and co-excitation," 2021, <https://arxiv.org/abs/1911.12529>.
- [20] X. Li, C. Lin, C. Li et al., "Improving one-shot NAS by suppressing the posterior fading improving one-shot NAS by suppressing the posterior fading," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13836–13845, IEEE, Seattle, WA, USA, June 2020.
- [21] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [22] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: a retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

Research Article

Factors Influencing Green Entrepreneurship of Returning Migrant Workers under the Dual-Carbon Background

Li Beiwei,¹ Yue Zhengliang ,^{1,2} and Liu Hongtao²

¹School of Management, Jilin University, Changchun City, Jilin Province, China

²Zhuhai College of Science and Technology, Zhuhai City, Guangdong Province, China

Correspondence should be addressed to Yue Zhengliang; yuezhengliang@zcst.edu.cn

Received 15 January 2022; Revised 26 January 2022; Accepted 27 January 2022; Published 21 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Li Beiwei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

China's rural energy consumption structure has dominated by coal, and carbon dioxide emissions are relatively large. Under the dual historical responsibilities of the carbon peak, the carbon neutral dual-carbon, and the rural revitalization strategy, the rural energy transition is imminent. This paper conducts an in-depth study on the green entrepreneurship of migrant workers returning to their hometowns and carefully analyzes the factors affecting the return of migrant workers to green entrepreneurship. This paper establishes a regression model to clarify that the green entrepreneurial guidance mechanism can effectively increase entrepreneurial opportunities, promote the rational allocation of regional resources, increase the success rate of entrepreneurship, and promote the construction of new rural areas in China with proper healthcare. The implementation of the scientific guidance mechanism for migrant workers' green entrepreneurship should adhere to the scientific development concept, correctly guide farmers to return to their hometowns to start businesses, and provide them with a green channel for fiscal taxation. Green entrepreneurship can improve the entrepreneurial ability of migrant workers in order to effectively increase the success rate of migrant workers in entrepreneurship, ensure the rational use of regional resources, and promote the harmonious and stable development of society.

1. Introduction

As global climate change poses a major threat to human society, more and more countries have elevated "carbon neutrality" as a national strategy and proposed a vision for a carbon-free future. The "dual-carbon" target proposed has a profound domestic and foreign development background, which will definitely have a profound impact on the economy and society. The realization of the "dual-carbon" goal should have been comprehensively considered and dealt with in the overall strategic of promoting high-quality development and comprehensive modernization. China has become the world's second largest economy, a leader in green economy and technology, and its global influence continues to expand. Facts have proved that only by making the development mode greener can we adapt to the laws of nature. To this end, in 2020, based on the inherent requirements of promoting sustainable development and the responsibility, China announced the goal of achieving

carbon peaks and carbon neutrality. General Secretary Xi Jinping emphasized that carbon peaking and carbon neutrality should have been incorporated into the overall layout of ecological civilization construction that can help in improving the healthcare of the citizens [1, 2]. It is necessary to promote major breakthroughs in green and low-carbon technologies, promptly deploy low-carbon cutting-edge technology research, accelerate the promotion and application of pollution reduction and carbon reduction technologies, and establish and improve green and low-carbon technology evaluation, trading system, and technological innovation service platform. In the future, China will form a greener, more efficient, and sustainable consumption and productivity model for sustainable development and jointly composing the ecology in civilization.

As China's economic structure and industrial development are in an important period of transformation, Premier Li Keqiang proposed the development strategy of "mass entrepreneurship and innovation." Many migrant workers

with knowledge, vision, and skills have chosen to return and set up entrepreneurship through the funds and some experience they have accumulated during their migrant work period, making migrant workers the group with the most potential and enthusiasm in entrepreneurship and innovation. However, looking at the living conditions of migrant workers who have not yet returned to their hometowns, they still have many worries about the behavior of returning to set up entrepreneurship, which have been seriously affected by many factors of themselves and the outside world. The return of migrant workers to set up entrepreneurship in their hometown is conducive to proposing more targeted countermeasures, which plays a vital role in enhancing migrant workers' enthusiasm for returning to their hometown, promoting successful entrepreneurship, and promoting rural modernization. Entrepreneurship is an important engine for economic growth and social development [3]. Since 2015, China has increased its strategic support for rural revitalization and issued a series of policy documents to support entrepreneurship. China Council has also incorporated migrant workers' return to their hometowns to set up entrepreneurship, which promotes economic development in the unified deployment of the "double entrepreneurship" strategy. As a result, the "returning goose economy" driven by the return of migrant workers to their hometowns to set up entrepreneurship has quietly risen across the country [4]. The migrant workers returning to their hometowns are large in scale, rich in social experience, and highly skilled. Many migrant workers have seen the potential for future development in their hometowns and have returned to their hometowns to set up entrepreneurship. The return of migrant workers is an important part of China's "double innovation" and an important way to accelerate the construction of new urbanization and fully promote the "village revitalization" strategy [4].

Previous studies mainly focus on the determinants of migrant workers returning to set up entrepreneurship in the structural reform of the agricultural supply side, continue to activate the rural economy, and ultimately achieve rural revitalization [5]. But, quantitative analysis is lacking, and the green entrepreneurial behavior index of migrant workers is also very small. Therefore, it is an urgent task to describe the influencing factors in detail. In this research, migrant workers return to their hometown to set up green entrepreneurship. Under the current background of building an ecologically civilized society and achieving sustainable development, the enterprise adopts advanced technology and reformed management methods to implement green in the entire production and operation process. This paper establishes a measurement model of the green entrepreneurial behavior index of migrant workers returning home. Based on this, this paper studies the factors that affect the green entrepreneurial behavior of migrant workers returning to their hometowns, which can create favorable conditions for green entrepreneurship, further promote the development of green entrepreneurial activities, and help achieve the goal of China's ecological civilization society. Therefore, this paper contributes to study the green entrepreneurial behaviors and influencing factors of migrant workers returning

to their hometowns. This article has certain reference value for formulating relevant policies to promote the green entrepreneurship and achieve green development.

2. Related Theories

Green entrepreneurship has gradually merged with other disciplines and themes, and theoretical research on the connotation and types of green entrepreneurship has been greatly enriched. Some scholars have combined green entrepreneurship with social practice research and proposed a green entrepreneurial operation model to provide guidance for companies to achieve green production and increase green benefit. The research on green entrepreneurship in western countries started early and has produced certain results in the research of green entrepreneurship [6]. Domestic scholars mainly use different perspectives and methods to study the theories and types of green entrepreneurship based on the theoretical results of foreign green entrepreneurship research. In-depth analysis was carried out, and at the same time, some scholars discussed the current problems facing enterprises' green entrepreneurship and provided solutions for the realization of sustainable development [7]. From the emergence of green entrepreneurship, to the construction of green entrepreneurship theory and its combination with practical research, relevant literature has begun enriching. Table 1 shows related works as follows.

2.1. The Background of Dual-Carbon Construction. On the international front, the world's economic development has accelerated the conversion of old and new kinetic energy, and the development of clean and low-carbon energy has become a global consensus. In July 2020, before the EU announced its carbon neutral plan, more than 30 countries had announced carbon neutral targets, including Mexico and the Maldives. Since then, China, Japan, and South Korea have successively proposed carbon neutral goals. US President-Elect Biden also proposed in his speech that the United States should return to the "Paris Agreement." The basic requirement is that the United States should propose a timetable and roadmap for carbon neutrality. In this way, the world's important economies, which account for 75% of global GDP and 65% of global carbon emissions, become carbon neutral [8]. China is also a major energy consumer. Its commitment to carbon peaking and carbon neutrality fully reflects the responsibility and mission of a major country, and it will become a powerful promoter of promoting the world's carbon peaking goal as soon as possible. Domestically, China has achieved positive results in controlling greenhouse gas emissions, promoting energy conservation and emission reduction in key areas, developing renewable energy, and accelerating ecological governance and land greening [9, 10]. Han et al. analyzed the development trends and research focus of industry 5.0 in the future [11]. Jagatheesaperuma et al. focused on the research perspective towards deployment of Industry 5.0 [12]. However, problems such as unbalanced and inadequate

TABLE 1: Related works.

| No. | Content |
|-----|---|
| 1 | Factors affecting the return of migrant workers to green entrepreneurship |
| 2 | Establishing a regression model |
| 3 | Influence on the green entrepreneurial behavior |

economic development, structural imbalances in high-emission industries, large-scale manufacturing, and uncoordinated high-quality development are still prominent. The country needs to continue to promote low-carbon energy consumption, industrial restructuring, and improving manufacturing eco-efficiency.

Combining with the series of green carbon sink actions that China has carried out in recent years to address climate change at both the international and domestic levels, the proposed carbon peak and carbon neutral targets have richer connotations and significance. The proposal of the carbon peak and carbon neutral vision coincides with the major node. The construction of ecological civilization has achieved new progress and the efficiency of national governance has improved. The main economic and social development provides an important opportunity for the realization of the vision. In addition, the overall vision of carbon peaking and carbon neutrality also provides an effective driver for continuing to force China to deepen the green economy transformation, accelerate the adjustment of the energy structure, promote the construction of the carbon market, and coordinate with the construction of ecological civilization to form a joint force for joint realization.

2.2. Green Entrepreneurship. As an emerging research topic, the definition of the concept of green entrepreneurship has not yet been unified, and the terms used by scholars to describe green entrepreneurship are different. Academic names such as “environmental entrepreneurship,” “green entrepreneurship,” and “sustainable entrepreneurship” have appeared. In the definition of the concept of green entrepreneurship, scholars have defined it from different angles. Based on the perspective of entrepreneurial opportunity identification, Yuniriyanti et al. [13] believe that green entrepreneurship is a process by which companies identify, evaluate, and utilize business opportunities under the principles of sustainable development goals. Such business opportunities arise when the market fails. It is closely related to the environment [13]. Based on the perspective of entrepreneurial output, Hornsby et al. believe that, in a narrow sense, green entrepreneurship is a process of enterprise creation in which entrepreneurs play their subjective initiative and creatively provide environmentally friendly products [14]. In a broad sense, green entrepreneurship can be described as the development of enterprises. At each stage, through environmental innovation, an innovative, market-oriented, individual-driven form of value creation is present for providing environmentally friendly products [15]. Wang and Peng believe that green entrepreneurship refers to innovation or the creation of a green-oriented organization and emphasizes the need to create a green

benefit (GVA) to protect the natural environment and increase the interests of corporate stakeholders. Green entrepreneurship is an entrepreneurial behavior for protecting the environment [16]. Chinese scholar Šneiderienė et al. combed and summarized the concept of green entrepreneurship by predecessors and pointed out that the connotation of green entrepreneurship is divided into broad and narrow sense [17]. The narrow sense of green entrepreneurship is only a short-term and partial entrepreneurship, which refers to the cost, innovative or marketing advantages are the motivation to achieve green, and in a broad sense, green entrepreneurship is comprehensive and of long term. It is a form of value creation, based on innovative environmental protection methods, market-oriented, driven by individuals, and “sustainable” as the goal.

Green entrepreneurship is a new way of entrepreneurship, and the internal operating mode of green entrepreneurial enterprises is different from traditional entrepreneurship. According to the research on green entrepreneurship they believe that green entrepreneurship is a process of creating green benefit (GVA), which must protect the natural environment and increase the interests of corporate stakeholders [18]. The realization of green benefit requires the use of the company’s own advantages and the external favorable environment to achieve green operations in all aspects of the company’s production and operation. Generally, the green entrepreneurial operation model includes five major aspects [19]. (1) Green input logistics: Green enterprises should avoid water, air, and soil pollution during the procurement process, use healthy and environmentally friendly materials, clean energy, and use energy-saving technologies or processes to improve their greenness in warehousing and logistics to achieve ecological benefits. (2) Green production: Green production is the strict use of green technology and production processes in the production of green products or the provision of green services. (3) Green export logistics: Green output is mainly in the final link of production, strictly controlling and recycling by-products, using green environmental protection materials for packaging, and harmless treatment of waste to achieve green products and services. (4) Green marketing: Green marketing refers to the design of marketing channels and the utilization of environmental protection technologies. It is necessary to consider and meet the environmental demands of consumers and society and realize the greening of the marketing network and the recycling of resources, to create profits while achieving sustainable operation. (5) Green service: The service itself and service-oriented green constitute the green service link. The first purpose is to avoid negative impacts on the environment during the use of products or the provision of services, while the latter purpose is to affect consumers’ non-green consumption

behaviors. Change and design some green services to guide consumers to develop the habit of green consumption. The basic framework and objectives of the green entrepreneurial operation model are shown in Figure 1.

In summary, green entrepreneurship, as an emerging research topic, is a new research field that has emerged at a certain stage of social and economic development and has not yet formed a unified theoretical system. In terms of research methods, qualitative analysis is mostly used, and most of the literature uses cases. There are few special studies on the green entrepreneurial behavior of agricultural enterprises. Regarding the research on the driving factors of corporate green entrepreneurship, most scholars tend to analyze from the perspective of a single factor, such as personal values, consumer needs, and institutional regulations [20]. In terms of corporate green entrepreneurial behavior research, there is a lack of specific quantitative indicators and very few empirical studies. The relevant conclusions, especially the research on behavioral influence factors, are not sufficiently persuasive, and they are still in the low-paradigm research stage. Based on this, this paper adopts the sample data of the green entrepreneurial behavior of migrant workers to construct a green entrepreneurial behavior evaluation index, which establishes a measurement model for the green entrepreneurial behavior index.

3. Analysis of the Influencing Factors of Green Entrepreneurship by Returning Migrant Workers

3.1. Data Sources. In order to better promote migrant workers to return to their hometowns to set up entrepreneurship, the migrant workers' return to set up entrepreneurship and implement the rural revitalization strategy; this paper designs a questionnaire on the factors affecting migrant workers' return to set up entrepreneurship. Statistics and analysis were carried out. 480 questionnaires were distributed to the central region in this survey, and 460 valid questionnaires were returned. The survey subjects included the new generation of migrant workers and the first generation of migrant workers, and the distribution was relatively even. The proportions of men and women are roughly equal. The selected migrant workers' industries include agricultural product sales, construction, clothing, handicrafts, service industries, business management, and helpers for farms. Therefore, the samples selected in this study have a certain level of representativeness, which provides an important basis for determining the practical obstacles for migrant workers to return to their hometowns to set up entrepreneurship and for this article to propose strategies to promote migrant workers to return to their hometowns to set up entrepreneurship. As an important force in promoting social development, whether the group of migrant workers will return to their hometown to set up entrepreneurship after graduation has received great attention from the society. According to the questionnaire survey data, the largest proportion of migrant workers is engaged in the construction industry, as high as 37%. There is also a lot of

agricultural products sales, accounting for 12.2%. Handicrafts, service industries, business management, and helpers who help farms account for 9.8%, 7%, 4.3%, and 4.5%, respectively. In addition, this part is the high-quality returning entrepreneurial talents that the rural areas urgently need [21]. According to investigations and studies, most of the motives for migrant workers to return to their hometowns to set up entrepreneurship are due to the high pressure of survival in the city, while fewer people return to their hometowns to build their hometowns. Men's willingness to set up entrepreneurship is higher than women's. The biggest difficulty faced by migrant workers returning to set up entrepreneurship is insufficient funds. The willingness of high-educated migrant workers to start business is lower than that of low-educated migrant workers. The statistical results show that there are varying levels of obstacles and difficulties for migrant workers to return to set up entrepreneurship. On the other hand, these also provide us with some ideas for solving the practical obstacles for migrant workers to return to set up entrepreneurship as shown in Table 2.

3.2. Characteristics of Migrant Workers' Willingness to Set Up Green Entrepreneurship. After migrant workers have accumulated a certain amount of capital and experience, their willingness to set up entrepreneurship is very high, and the willingness of green entrepreneurship is relatively large in Figure 2. In addition, it can be seen from Table 3 that migrant workers and entrepreneurs prefer the green entrepreneurial model of family eco-farms and eco-tourism. It shows that the green management concept of migrant workers has continuously improved.

It can be seen from Table 4 that as the awareness of environmental protection continues to increase, consumers' demand for ecological products increases, and entrepreneurs' green entrepreneurial willingness increases, they will prefer environmentally friendly entrepreneurial projects. In addition, the government supervision is becoming more and stricter, the environmental investment is also increasing, and the penalties for environmental pollution behaviors have constantly strengthened, thus promoting the green entrepreneurship of migrant workers.

The source of resistance to the willingness of migrant workers starts a green entrepreneurship. It can be seen from Table 5 that migrant workers have limited abilities, no available superior resources, and insufficient funds for their own businesses. The lack of the relevant support policies and the poor ability to use the policies has led to their inability to obtain sufficient financial support. Therefore, some migrant workers are hindered from green entrepreneurship.

Through a descriptive analysis of 460 questionnaires on the status quo of migrant workers' green entrepreneurship, the status of migrant workers' green entrepreneurship is divided into three categories: migrant workers' green entrepreneurship human capital characteristics analysis, migrant workers' green entrepreneurship willingness characteristics, and migrant workers' green entrepreneurship issues. In this part, the analysis of the personal

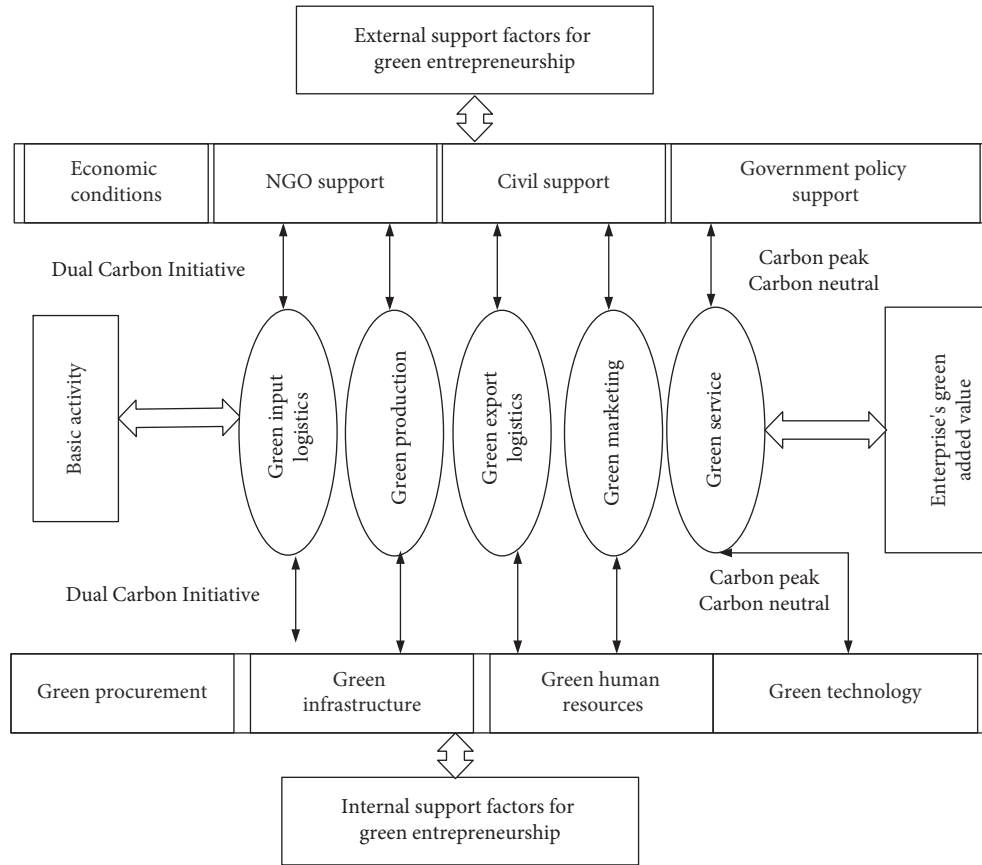


FIGURE 1: The operating model of green entrepreneurship.

TABLE 2: Staff composition of green entrepreneurship.

| Index | Code | Sample | Percentage (%) |
|------------|----------------------------|--------|----------------|
| Gender | Male | 240 | 52.2 |
| | Female | 220 | 47.8 |
| | Sum | 460 | 100 |
| Profession | Agricultural product sales | 80 | 17.4 |
| | Clothing industry | 160 | 34.8 |
| | Handicraft | 80 | 17.4 |
| | Service industry | 70 | 15.2 |
| | Business management | 20 | 4.3 |
| | Farm helper | 50 | 10.9 |
| | Total | 460 | 100 |

characteristics of migrant workers' green entrepreneurial human capital shows that migrant workers are mainly young and middle-aged, their education level is concentrated in junior high school, and they have few skills and training opportunities. The work income is low, and relatives and fellow villagers dominate the network. After analyzing the characteristics of migrant workers' green entrepreneurial willingness, it is found that migrant workers have a higher willingness to green entrepreneurship and prefer green entrepreneurial projects such as family ecological farms, eco-tourism, and resource recycling. Among them, the increase of migrant workers' awareness of environmental protection and the increase in market demand for ecological products have promoted the increase of migrant workers' willingness

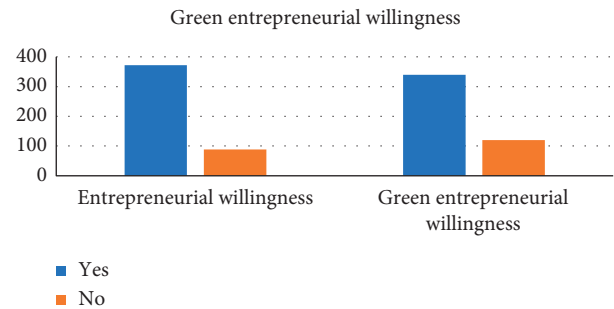


FIGURE 2: The survey of green entrepreneurial willingness.

to start green entrepreneurship. As more and more attention has been paid to environmental pollution, the government has increased its investment in environmental protection and strengthened penalties for environmental pollution behaviors, thereby increasing the initiative and enthusiasm of migrant workers' green entrepreneurial willingness. In addition, the education level is low, and the ability to learn new things is poor, which hinders the willingness of migrant workers to set up green entrepreneurship. Although China has given certain loans, subsidies, and other preferential policies to encourage migrant workers to start green entrepreneurship, the conditions of the preferential policies are harsh, and fewer migrant workers can enjoy preferential policies. In addition, because the social network of migrant workers has dominated by fellow villagers and relatives, the

TABLE 3: Project preference of green entrepreneurship.

| Entrepreneurship type | No. | Percentage (%) |
|--|-----|----------------|
| Waste recycling | 23 | 5 |
| Comprehensive utilization of resources | 46 | 10 |
| Family eco-farm | 184 | 40 |
| Eco-tourism | 138 | 30 |
| Dual-carbon farming and forestry | 23 | 5 |
| Others | 46 | 10 |
| Sum | 460 | 100 |

TABLE 4: Power source of green entrepreneurship.

| Power source | No. | Percentage (%) |
|---|-----|----------------|
| Increasing awareness of environmental protection | 253 | 55 |
| Strong desire for green entrepreneurship | 230 | 50 |
| Project investment with dual-carbon background | 207 | 45 |
| Government investment in environmental protection | 207 | 45 |
| Penalties for environmental pollution | 184 | 40 |
| Green product demand | 184 | 40 |

TABLE 5: Resistance source of green entrepreneurship.

| Limited understanding of dual-carbon protection | No. | Percentage (%) |
|---|-----|----------------|
| I do not understand the loan policy | 322 | 70 |
| Poor natural conditions | 253 | 55 |
| Policy utilization | 184 | 40 |
| Fund acquisition | 184 | 40 |
| Limited understanding of dual-carbon protection | 138 | 30 |

channels for obtaining information are single, and policies cannot be correctly understood and used flexibly. Market-oriented analysis is derived from accumulated experience and cannot adapt to the development of the times, which makes its green entrepreneurial path difficult.

3.3. Empirical Model. According to the previous relevant analysis, it is concluded that the willingness of migrant workers to start a green business is affected by factors such as the characteristics of ecological capabilities, the characteristics of the external environment, and the characteristics of policy support. Therefore, the function expression can be set as follows:

$$\hat{Y} = \hat{\beta} + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_9 X_9. \quad (1)$$

$$\text{Among them, } \hat{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_1 \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_9 \end{bmatrix}.$$

When migrant workers have no ecological entrepreneurial willingness, $\hat{Y} = 0$. When migrant workers are willing to start an ecological business, $\hat{Y} = 1$. Among them, X_1 , respectively, represents 9 variables in the three types of characteristics of migrant workers' green entrepreneurship capabilities, external environment characteristics, and policy support characteristics. The specific variables of each influencing factor in the regression model are explained in Table 6.

Most of the variables in the regression analysis are the same as those in the correlation analysis. Some of the

variables are explained as follows. First, the variable of education level is divided into "below elementary school = 1, junior high school and above = 3, bachelor's level and above = 5," among which "junior high school," "senior high school (secondary school)," and "junior college" in the questionnaire options are classified as "junior high school and above = 3." Because it is an unfavorable cause, the option is classified as an influencing factor that hinders entrepreneurial willingness, and the answer is "yes." It is 1 and the others are 0. The variable "work" involves more occupations in the questionnaire, and it is divided into primary, secondary, and tertiary industries. Among them, the options are "company management, sales staff," "private small business owners," and "business owner" and "contractor" are classified as "industry" with a value of 2. "Waiter," "cashier," and "security, cleaning, and driver" are classified as "service industry" with a value of 3. "Agricultural" is classified as the value of "agriculture" is 1. The value of "others" is 3. Secondly, the variable "social capital" is replaced by the number of friends, and more migrant workers interact with each other. The social capital is higher, and the opportunity for green entrepreneurship is greater [22]. The willingness of green entrepreneurship is stronger. In addition, in the variables, according to the opinion measurement method, the options have assigned a value of 1–5 points. For example, for the variable "protect environmental awareness," the questionnaire asks "Your awareness of protecting the environment is increasing," and the answer options are "very nonconforming," "nonconforming," "general," "basically conforming," and

TABLE 6: Characteristic of migrant workers' green entrepreneurial ability.

| Variable name | Explanation | Mean value | Standard deviation | Direction |
|------------------------------------|--|------------|--------------------|-----------|
| Education | Junior high school and below 1, junior high school and above 3, high school and above 5. | 2.8 | 0.7 | +/- |
| Jobs | Agriculture = 1, service industry = 3, others = 5. | 2.7 | 1.2 | +/- |
| Social capital | None = 1, yes = 3. | 3.0 | 0.9 | +/- |
| Educational investment | No investment = 1, investment = 3. | 3.5 | 0.8 | +/- |
| Bring up | No = 1, yes = 3. | 1.6 | 0.7 | +/- |
| Self-protection | Whether one's own ability hinders entrepreneurship, yes = 1, no = 0. | 0.7 | 0.5 | +/- |
| Environmental protection awareness | Yes = 3, no = 1. | 3.6 | 0.9 | +/- |
| Entrepreneurial resources | Up to standard = 3, not up to standard = 1. | 3.2 | 1.1 | +/- |
| Ability to integrate resources | Junior high school and below 1, junior high school and above 3, high school and above 5. | 3.5 | 0.9 | +/- |

“very conforming.” The score is higher, indicating that migrant workers' awareness of environmental protection has continuously enhanced, thereby promoting their green entrepreneurial willingness.

3.4. Analysis of Influencing Factors of Peasant Workers' Green Entrepreneurship Willingness

- (1) From the perspective of social science research, the model fits well. The awareness of environmental protection is stronger and the ability to integrate resources is stronger; the willingness of migrant workers is stronger to set up green entrepreneurship. In terms of the characteristics of its own green entrepreneurial ability, environmental awareness, ability to integrate resources, and its own ability have passed the significance test and are basically consistent with the previous expected direction, while the education level, work, training, educational skills investment, and social capital are not significant. Because migrant workers have been working abroad for a long time, their ideology and their own abilities are constantly changing, and they have certain green entrepreneurial resources and abilities. According to the survey, although migrant workers work harder and have low incomes, their environmental protection awareness and ideological awareness continue to increase. For further improvement, therefore, migrant workers are willing to make full use of their own resources to set up entrepreneurship and constantly improve their own resources and strengthen their own resource integration capabilities, to promote a stronger willingness to green entrepreneurship.
- (2) The limited ability and poor natural conditions hinder the process of migrant workers' green entrepreneurial willingness.

According to the survey, the limited ability and poor natural conditions have become the main obstacles to set up green entrepreneurship. As migrant

workers are mainly educated in junior high school, they have relatively little knowledge and insights, and their ability to accept new things and learn is weak. The ability to integrate resources and optimize their own resources is relatively weak, and the channels for obtaining information are single. In addition, some companies do not pay attention to environmental protection and waste of resources, which makes the natural conditions worse and worse, and the migrant workers have fewer and fewer resources and fewer and fewer entrepreneurial resources, which hinders the process of green entrepreneurship for migrant workers.

- (3) From the characteristics of the external environment, the green entrepreneurial environment and poor natural conditions have passed the significance test. Among them, the green entrepreneurial environment is highly significant. However, the awareness of purchasing environmental products and the demand for ecological products are not obvious. Continuously improve relevant laws and systems, resolve the risks of migrant workers' green entrepreneurship, and provide guarantees for migrant workers' green entrepreneurship, thereby enhancing their willingness to engage in green entrepreneurship.
- (4) The more the government invests in environmental protection, the greater the subsidy tax support is. In terms of government policy support, environmental protection investment and assistance organizations have passed the significance test, while the significance of loans, environmental pollution penalties, publicity, and subsidy taxation is not obvious. According to the survey, the lack of entrepreneurial funds hinders migrant workers from starting green entrepreneurship. The government should increase investment in environmental protection and reduce the pressure on migrant workers' green entrepreneurial funds. Moreover, as the government's investment in environmental protection increases, migrant workers' awareness of environmental

TABLE 7: Variables in the equation.

| | B | SE | Ws | df | Sig | Exp |
|--|-------|------|------|------|------|------|
| Education | -0.26 | 0.15 | 3.2 | 1.00 | 0.07 | 0.74 |
| Jobs | -0.05 | 0.1 | 0.32 | 1.00 | 0.60 | 0.95 |
| Social capital | 0.15 | 0.1 | 1.6 | 1.00 | 0.12 | 1.1 |
| Educational skills investment | -0.1 | 0.12 | 0.8 | 1.00 | 0.23 | 1.2 |
| Environmental protection awareness | 0.26 | 0.11 | 6.2 | 1.00 | 0.37 | 0.9 |
| Own entrepreneurial resources | -0.16 | 0.1 | 2.95 | 1.00 | 0.10 | 0.8 |
| Ability to integrate resources | 0.25 | 0.13 | 4.7 | 1.00 | 0.03 | 1.3 |
| Entrepreneurial environment | 0.48 | 0.12 | 9.6 | 1.00 | 0.01 | 1.6 |
| Poor natural conditions | 0.00 | 0.1 | 2.8 | 1.00 | 0.1 | 1.5 |
| Environmental product purchase awareness | 0.16 | 0.1 | 0.00 | 1.00 | 0.97 | 1.1 |
| Ecological product demand | 0.18 | 0.12 | 2.86 | 1.00 | 0.09 | 1.2 |
| Loan | 0.15 | 0.11 | 1.83 | 1.00 | 0.18 | 1.20 |
| Environmental protection investment | 0.28 | 0.12 | 5.83 | 1.00 | 0.03 | 1.27 |
| Subsidy | -0.08 | 0.18 | 3.16 | 1.00 | 0.56 | 1.64 |
| Environmental pollution penalty | 0.04 | 0.15 | 0.08 | 1.00 | 0.83 | 1.05 |

TABLE 8: Model test results.

| Variable | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
|--|-------|-------|-------|-------|-------|
| Enterprise size | -0.26 | -0.22 | -0.21 | -0.14 | -0.41 |
| Whether the company is listed | -0.20 | -0.06 | -0.13 | -0.15 | -0.22 |
| Corporate assets | -0.05 | -0.04 | -0.02 | 0.01 | 0.04 |
| Enterprise nature (state-owned/collective) | -0.23 | 0.21 | 0.15 | 0.18 | 0.09 |
| Enterprise nature (private) | -0.04 | 0.08 | 0.14 | 0.15 | 0.08 |
| Industry (plantation) | 0.1 | 0.06 | 0.09 | 0.01 | 0.03 |
| Industry (aquaculture) | -0.04 | -0.01 | -0.07 | -0.01 | -0.08 |
| Life cycle (growth period) | -0.21 | -0.09 | -0.11 | 0.01 | 0.14 |
| Life cycle (mature period) | -0.23 | -0.11 | -0.17 | 0.00 | 0.08 |
| Life cycle (transition period) | -0.14 | -0.09 | -0.13 | 0.04 | 0.11 |
| Policy support | 0.19 | 0.22 | 0.21 | 0.21 | 0.14 |
| Environmental regulation | 0.17 | 0.18 | 0.22 | 0.27 | 0.23 |

protection has also increased. It can be seen from Table 7 that the support organization passed the significance test, indicating that the support organization is a significant factor influencing the green entrepreneurship of migrant workers. The establishment of a support organization will unite the government, universities, and scientific research institutions, leading enterprises and green entrepreneurial experts to provide assistance. It has made up for the shortcomings of migrant workers' lack of ability, improved their green entrepreneurial skills, enhanced the competitiveness and success rate of green entrepreneurship, and drove migrant workers to engage in green entrepreneurship.

4. Results and Discussion

This paper puts the four environmental variables of policy support, environmental regulation, stakeholder pressure, and corporate environmental awareness into the model for quantile regression testing and gets the model results, as shown in Table 8. The table lists the regression results of each variable at the 0.1, 0.25, 0.5, 0.75, and 0.9 quantiles. Through the differential performance of each variable at different distribution points, it is possible to have a more

comprehensive and in-depth understanding of the entrepreneurial environment and corporate characteristics for corporate green entrepreneurship [23]. This paper uses stata12 software to estimate, and the results obtained are shown in Table 8.

The variable of enterprise scale has a negative impact on the level of greening of enterprise production and operation, and the impact is more significant. The quantile regression coefficient of this variable shows a trend of rising first, reaching the maximum value at the 75% quantile, and then rapidly decreasing, passing significance at the 10%, 25%, 50%, and 90% quantiles. There is a greater level of complexity and inefficiency in environmental improvement, resulting in a low level of greenness in its production and operations. The variable of the nature of the enterprise has an impact on the level of greening of the production and operation of the enterprise. Among them, the nature variable of the private enterprise has a positive effect on the level of greening of the production and operation of the enterprise. The life cycle of an enterprise has a significant impact on the green level of production and operation of the enterprise. Among them, the mature stage enterprise variables have a negative impact on the level of green production and operation of the enterprise, and the quantile regression coefficient shows an increase first and then a downward and

rapid upward trend. This variable has passed the significance test at the 50% quantile, and the coefficient is negative, indicating that the production and operation of the start-up enterprises are greener. The possible reason is that when the company is in the initial stage, it faces instability and urgently needs the support of shareholders and the affirmation of consumers. This makes managers have stricter requirements on the company in terms of raw material procurement, production management, and staff quality and skill training. Therefore, it shows a higher level of greenness in terms of product quality and service. The regression coefficient of the policy support variable is positive, and it has passed the significance test at each quantile point. The quantile regression coefficient shows a trend of first rising, reaching the maximum value at the 25% quantile, and then slowly decreasing. At the 25% quantile, the policy support environment will improve the level of green production and operation of the company every time the environment is improved. The government can provide financial support, financing support, and tax relief for enterprises that implement environmental protection activities. It has a good influence on the green entrepreneurial behavior of enterprises and enhances the green level of enterprise production and operation.

5. Conclusions

Green entrepreneurship is conducive to improving the forest ecological environment and stimulating rural areas to get rid of poverty and becoming rich, thereby reducing the government's pressure on ecological compensation. The green entrepreneurship of migrant workers has promoted the construction of rural infrastructure and public utilities and accelerated the transformation of agricultural development methods. The contributions of this research are as follows. Migrant workers encouraged starting their businesses; it will help solve the problems of empty old people, left-behind children, and left-behind women. Secondly, the regression coefficient of the environmental regulation variable is positive, and it has passed the significance test at each quantile point. Hence, strict environmental management rules and regulations can force enterprises to implement green entrepreneurship and promote the improvement of the green level. The quantile regression coefficient shows a trend of rising first, reaching the maximum value at the 75% quantile, and then slowly decreasing. Based on the above, the migrant workers' green entrepreneurial willingness is related to environmental protection awareness, green entrepreneurial environment, their own resource integration ability, environmental protection investment, and other factors. Among them, the green entrepreneurial environment and assistance organizations are significant influencing factors. This article combines the effective ways of encouraging migrant workers to green entrepreneurship in different regions. It proposes practical and reasonable relevant countermeasures to encourage migrant workers to green entrepreneurship and enhance their green entrepreneurship willingness in the future.

Data Availability

All data included in tables are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] M. Zid, A. T. Alkhudri, A. R. Casmana, A. Marini, and A. Wahyudi, "Ex migrant workers of international women and social entrepreneurship: study at kenanga village in Indramayu Regency in West Java Province in Indonesia," *International Journal of Advanced Science and Technology*, vol. 29, 2020.
- [2] J. Fitra and D. Rizana, "Strategi pemberdayaan kewirausahaan mantan buruh migran perempuan di kabupaten kebumen," *Fokus Bisnis: Media Pengkajian Manajemen dan Akuntansi*, vol. 18, no. 1, pp. 36–42, 2019.
- [3] J. Chen, R. Liu, and X. Luo, "Research on incentive mechanism of returning migrant workers to set up entrepreneurship under the background of mass entrepreneurship and innovation," in *Proceedings of the 5th International Conference on Economic and Business Management*, Jilin, China, February 2020.
- [4] C. Mukonza, "An analysis of factors influencing green entrepreneurship activities in South Africa," *Advances in African Economic, Social and Political Development*, United Nations University Institute for Natural Resources in Africa, Accra, Ghana, pp. 47–67, 2020.
- [5] B. Ranasinghe and R. Ajward, "Factors affecting green entrepreneurial intention among small and medium enterprise owners in western province, Sri Lanka," in *Proceedings of the 12th International Research Conference of General Sir John Kotelawala Defence University*, Dehiwala-Mount Lavinia, Sri Lanka, September 2019.
- [6] Al-Dmour, H. Rand, T. Mohammed, and H. H. Al-Dmour, "Factors influencing students' intentions towards entrepreneurship: comparative study," *International Journal of Sustainable Entrepreneurship and Corporate Social Responsibility (IJSECSR)*, vol. 4, no. 1, pp. 1–26, 2019.
- [7] M. X. Wang, "Research on the route of entrepreneurship education for migrant workers returning home from the perspective of new urbanization," in *Proceedings of the 3rd International Conference on Economy, Management and Education Technology (ICEMET 2017)*, Jinan, China, October 2017.
- [8] J. E. Amorós, R. Basco, and G. Romání, "Determinants of early internationalization of new firms: the case of Chile," *The International Entrepreneurship and Management Journal*, vol. 12, no. 1, pp. 283–307, 2016.
- [9] A. Davari, A. Emami, V. Ramadani, and S. Taherkhani, "Factors influencing academic entrepreneurship: a case-based study," *Journal of Science and Technology Policy Management*, vol. 9, no. 3, pp. 284–295, 2018.
- [10] P. N. Srinivasu, A. K. Bhoi, R. H. Jhaveri, G. T. Reddy, and M. Bilal, "Probabilistic deep Q network for real-time path planning in censorious robotic procedures using force sensors," *Journal of Real-Time Image Processing*, vol. 18, no. 5, pp. 1773–1785, 2021.

- [11] J. Q. Han, L. Hong-Qiang, and D. X. Zhong, "Analysis on internal rules of industrial revolution development and trends of future industry 5.0," *Information Technology and Informatization*, vol. 8, no. 4, 2016.
- [12] S. K. Jagatheesaperuma, M. Rahouti, K. Ahmad, M. Guizani, and A. Al-Fuqaha, "The duo of artificial intelligence and big data for industry 4.0: review of applications, techniques, challenges, and future research directions," *IEEE Internet of Things Journal*, 2021, <https://arxiv.org/abs/2104.02425>.
- [13] E. Yuniriyanti, R. Sudarwati, and B. Nurdewanto, "Philanthropy as a form of social entrepreneurship in an effort to empower ex migrant women workers: study at Malang district, Indonesia," *The International Journal of Humanities & Social Studies*, vol. 8, no. 9, 2020.
- [14] J. S. Hornsby, J. Messersmith, M. Rutherford, and S. Simmons, "Entrepreneurship everywhere: across campus, across communities, and across borders," *Journal of Small Business Management*, vol. 56, no. 1, pp. 4–10, 2017.
- [15] M. O. Faruk, N. Hassan, and N. Islam, "Factors influencing the development of social entrepreneurship in Bangladesh," *SSRN Electronic Journal*, 2016, <https://ssrn.com/abstract=2856210> or <http://dx.doi.org/10.2139/ssrn.2856210>.
- [16] J. Wang and C. Peng, "Factors influencing university students' coastal ecology and environmental-friendly entrepreneurship in coastal universities," *Journal of Coastal Research*, vol. 109, no. sp1, 2020.
- [17] A. Šneiderienė, R. Viederytė, and L. Abele, "Green growth assessment discourse on evaluation indices in the European Union," *Entrepreneurship and Sustainability Issues*, vol. 8, no. 2, pp. 360–369, 2020.
- [18] E. Wahyono, L. M. Kolopaking, M. C. T. Sumarti, and A. V. S. Hubeis, "Jaringan digital dan pengembangan kewirausahaan sosial buruh migran perempuan," *Jurnal ILMU KOMUNIKASI*, vol. 16, no. 1, pp. 57–76, 2019.
- [19] M. Talić, M. Ivanović-Đukić, and T. Radenović, "Sustainable entrepreneurship: creating opportunities for green products development," *Economics of Sustainable Development*, vol. 4, no. 2, pp. 1–13, 2020.
- [20] R. Yuniarto, "From entrepreneurship to social activist: the role of Indonesian migrant entrepreneurs in taiwan and socio- economic functions of return-migrant entrepreneurship in Malang, east java," *Entrepreneurship-Trends and Challenges*, vol. 2, 2018.
- [21] A. Tleuberdinova, Z. Shayekina, D. Salauatova, and S. Pratt, "Macro-economic factors influencing tourism entrepreneurship: the case of Kazakhstan," *Journal of Entrepreneurship*, vol. 30, no. 1, pp. 179–209, Article ID 097135572098143, 2021.
- [22] V. O. Alejandra, G. E. Laura, Z. H. Manuel, and G. L. Cruz, "Specification of a Local Entrepreneurship Model," *Revista de Investigación Académica Sin Frontera: División de Ciencias Económicas y Sociales*, vol. 33, pp. 1–16, 2020.
- [23] T. Khan, K. Singh, M. H. Hasan et al., "ETERS: a comprehensive energy aware trust-based efficient routing scheme for adversarial WSNs," *Future Generation Computer Systems*, vol. 125, pp. 921–943, 2021.

Research Article

An Experimental and Modeling Study on the Combustion of Gasoline-Ethanol Surrogates for HCCI Engines

Peng Yin,¹ Wenfu Liu,¹ Yong Yang,¹ Haining Gao,¹ and Chunhua Zhang^{1,2} 

¹School of Energy Engineering, Huanghuai University, Zhumadian, 463000, China

²School of Automobiles, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Chunhua Zhang; 2017022008@chd.edu.cn

Received 15 January 2022; Revised 25 January 2022; Accepted 27 January 2022; Published 21 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Peng Yin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As an effective clean fuel, ethanol has the characteristics of improving antiknock quality and reducing emissions. It is an ideal antiknock additive for Homogeneous Charge Compression Ignition (HCCI) engines. The oxidation of gasoline-ethanol surrogates in HCCI engines is a very complex process which is dominated by the reaction kinetics. This oxidation process directly determines the performance and emissions of HCCI engines. Coupling the computational fluid dynamic (CFD) model with the gasoline-ethanol surrogate mechanism can be used for fuel design, so the construction of a reduced mechanism with high accuracy is necessary. A mechanism (278 species, 1439 reactions) at medium and low temperatures and experiments in a HCCI engine for the oxidation of gasoline-ethanol surrogates were presented in this paper. Directed relation graph with error propagation (DRGEP) method and quasi-steady-state assumption (QSSA) method were used in order to get a reduced model. Then, the kinetics of the vital reactions related to the formation and consumption of H and OH were adjusted. To validate the model, the HCCI experiments for the oxidation of gasoline-ethanol surrogates were conducted under different operating conditions. The verification result indicated that the present model can predict the oxidation process of gasoline-ethanol effectively.

1. Introduction

Due to rapid urbanization and industrialization, pollution levels are increasing at an alarming rate recently. Processing the fuel efficiently plays an important role in reducing the pollution due to fuel emissions, which can play an important role in improving the health of the citizens, especially in urban areas. Gasoline-ethanol is a multicomponent substance of thousands of macromolecular hydrocarbons and it is time-consuming and costly to develop a mechanism for the real fuel. Moreover, the application of the detailed model for complex gasoline-ethanol surrogate fuels in HCCI engine simulations is not practical with current computing resources, due to the large scale and the stiffness of the detailed mechanism. Therefore, the representative components of gasoline should be selected reasonably and the model of multicomponent gasoline surrogates should be reduced while maintaining its good performance.

Primary reference fuel (PRF), the two-component (isooctane/n-heptane) mixture, is generally considered to be the most common surrogates for gasoline. In recent years, more and more experiments were conducted in HCCI engines under high pressure, medium and low temperatures, and low equivalent ratio conditions, which provided a basis for the application of PRF mechanism in HCCI engine simulations. The PRF oxidation process of “the first oxygen addition → the first isomerization → the second oxygen addition → the second isomerization” is the key section during the autoignition process. In addition, as a commonly used additive, ethanol has become an important component for gasoline surrogate fuels due to its good antiknock performance and low emissions. The chemical kinetic mechanism of ethanol-PRF coupled with CFD software helps to understand the oxidation phenomena of mixture such as autoignition, flame propagation, flameout, combustion stability, and emissions. This is of great importance for

further improving combustion efficiency and reducing emissions.

The PRF models given by Halstead et al. [1] and Cox and Cole [2] were empirical models, which were still widely applied in the simulation of autoignition process. Then Li et al. [3] proposed a reduced model for predicting PRF oxidation behaviors, including ignition delay (τ), heat release rate (HRR), and molarity of vital species. This model can well predict the oxidation behaviors of PRF in the low and medium temperature, but the predictions at high temperature phase were not satisfactory. The reduced mechanism developed by Tanaka et al. [4] can be applicable for predicting τ , HRR , and knock in HCCI engines in a wide range, but it was difficult to predict the emission characteristics. However, the PRF mechanism constructed [5] by using the hierarchical expansion method can be used to calculate emissions of PHAs and other pollutants, although it was not accurate in predicting τ under the intake temperature (T_{in}) range of 300 K ~ 434 K and the pressure (P) of 4.0 MPa.

The model of Curran et al. [6] can well predict the ignition process on a wider scale of T_{in} s, P s, and ϕ s. According to the model [6], Ra and Reitz [7] proposed a reduced model, involving 41 species and 130 reactions, which may predict in-cylinder pressure (P), τ , and HRR accurately. Then, in order to solve the problem of cross-reactions, Kirchen et al. [8] added the cross-reactions to the models of Tanaka et al. [4] and Marinov [9]. In 2013 and 2015, two mechanisms developed by Liu et al. [10] and Wang et al. [11] were also presented for the combustion of gasoline surrogates.

A detailed three-component (iso-octane/n-heptane/ethanol) model [12] may predict laminar flame speeds (S_L s) accurately under high temperature. Then Zheng and Zhong [13] developed a reduced three-component model (50 species, 193 reactions). Its calculated τ s were highly consistent with the experimental values. Based on this model, a three-component model [14] was proposed by adding some elementary reactions related to H and updating relevant kinetic parameters, which can predict S_L s and τ s accurately. Moreover, Lemaire et al. [15] analyzed the effect of the additive (ethanol) to gasoline on the formation of soot. They pointed out that adding 10% ~ 30% (by volume) ethanol can reduce the production of soot precursor significantly; the amount of reduction for soot was 25% ~ 81%. In 2019, Li et al. [16] developed a highly reduced four-component gasoline-ethanol model, which may predict the experimental data for PRF, toluene primary reference fuel (TRF), and PRF-ethanol surrogates.

In recent years, we have conducted in-depth studies on the HCCI test and chemical reaction kinetics of related mixtures (Energy, 2019, 169:572–579. Tehnički Vjesnik Technical Gazette, 2020, 27(5):1571–1578. (SCI); Acta Microscopica, 2020, 29(2):720–731. (SCI)). In the field of chemical reaction kinetics, the mechanism of linear alkane and iso-alkanes and the chemical reaction kinetics of paraffin fuels have been constructed. In addition, preliminary achievements have been made in the research of primitive reactions and active groups sensitive to ignition delay.

In summary, the combustion process of gasoline-ethanol blend has attracted more and more attention recently. Many gasoline-ethanol mixture mechanisms have been constructed. However, due to the stiffness caused by long simulation time scale, the existing models are too large in scale and have poor accuracy under the current computing resources. Furthermore, HCCI validation experiments fueled with a hydrocarbon blend or a hydrocarbon-oxygen blend are rare and more experimental data is needed to compare with the calculated value, in order to further verify the reduced model. Therefore, the objective of this paper is to perform HCCI experiments on the combustion of gasoline-ethanol surrogates and to develop a smaller size mechanism for the lean gasoline-ethanol surrogates ($\phi < 1$) by implementing a reduction and merge scheme using DRGEP and QSSA methods. In order to carry out extensive validation of this reduced model, not only the calculation used by the proposed model should be compared with the results of the HCCI experiments, but also the new model should be compared with the previous literature models.

2. Kinetic Modeling

In this section, the processes of the initial mechanism construction, the automatic chemistry mechanism reduction for gasoline-ethanol surrogates, and the determination of the final mechanism are presented, as displayed in Figure 1. The reduced gasoline-ethanol mechanism was constructed by first reducing and then merging.

Firstly, in order to get a mechanism of smaller size, DRGEP method was used to eliminate the insignificant species efficiently, and QSSA method was used to identify the species that were in quasi-steady-state. Following the above process, the two submechanisms (the models of ethanol [9] and PRF [6]) were reduced. Taking the reduction of PRF submechanism as an example, according to the PRF reaction path given by Ra and Reitz [7], initial reactants (iso-octane/n-heptane/ O_2), intermediate components (C_2H_3 , CH_3 , CH_2CHO), and final products (CO_2 , H_2O) were selected as target substances. The larger the error threshold was set, the smaller the reduced mechanism scale would be and the prediction accuracy would decrease. In the process of reduction, a smaller threshold is set and reduction is carried out several times to ensure the prediction accuracy. Subsequently, a powerful and accurate merge for the reduced submechanisms from three disparate fuels was conducted.

Secondly, several relevant reactions involving H and OH in the products or reactants were revised, and the parameters of these reactions were adjusted. Finally, the final mechanism (278 species and 1439 reactions) was proposed.

Thirdly, the repeated reactions and components of PRF and ethanol mechanisms are mainly small molecule reactions of $C_1 \sim C_3$, H, and O_2 . However, rate constants of the same reactions in these two mechanisms are different, leading to great changes in the generation or consumption rates of many free radicals (H, OH, H_2O_2 , HO_2 , etc.). The rate constants of the small molecular reactions of $C_1 \sim C_3$, H, and O_2 in PRF submechanism were selected.

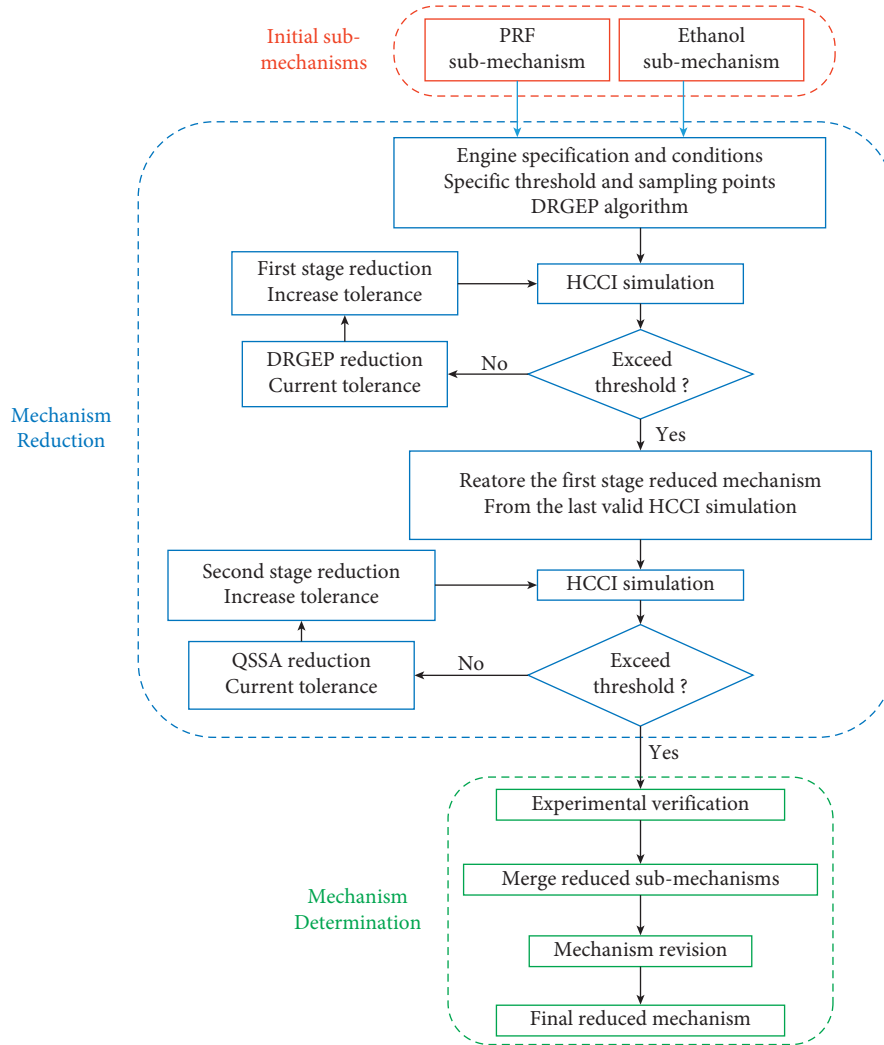


FIGURE 1: Flowchart of mechanism construction.

2.1. Base Mechanism and Case Settings. In order to keep the consistency between predictions and experimental results on reaction rate, transport, and thermodynamic data, the detailed chemical kinetic mechanism of PRF (1034 species, 4236 reactions) [6] and the semidetained ethanol model (57 species, 383 reactions) [9] were taken as base mechanisms.

Mechanism [6] has been used in the simulations of HCCI engines in [17, 18]. The ethanol submechanism has been validated by comparing with the experimental S_L s and τ_s . The predictions by ethanol submodel were also compared with mole fractions of species measured in stirred and flow reactors. In this paper, by combining complex calculation with recently obtained rate constants, the ethanol sub-mechanism was updated by adjusting the Arrhenius coefficients of some elementary reactions, as shown in Table 1 and [19–22].

The reduction work of the mechanism for gasoline-ethanol surrogates was carried out on the conditions targeted for HCCI engines. HCCI engines generally operate at low ϕ s, so the flame temperature should be lower compared with conventional internal combustion engines. To reduce the chemical model, 18 conditions were chosen at different

temperatures (500 K~1000 K) and different ϕ s (0.25, 0.3 and 0.5). The three-component fuel (iso-octane/n-heptane/ethanol = 62%: 18%: 20% by volume) should be considered as a substitute for ethanol-gasoline fuels [23], and it was used as gasoline-ethanol surrogates in this paper.

All the related calculations were performed by using CHEMKIN package. Assuming that mass transport is infinitely fast, the gas phase reaction is controlled solely on the nature of each species, not by transport constraints. The whole domain has uniform thermodynamic and transport properties.

For nonadiabatic cases, heat transfer between the cylinder and the wall was noticed. Related parameters of heat transfer were set according to the Woschni formula [24]. The relevant data in the model were set according to the specification of the test engine bench and the running conditions. Specifications for the test engine were presented by Zhang and Wu [25].

2.2. Mechanism Reduction. The reduction of the mechanism was implemented by DRGEP and QSSA approaches.

The DRGEP method is used for initial reduction. For a detailed mechanism, when specific species (reactants, reaction

TABLE 1: Updated reactions for ethanol.

| Reaction | A | n | E | Reference |
|--------------------------------------|---------|-----|-------|-----------|
| $C_2H_5OH + OH = C_2H_4OH + H_2O$ | 6.20E3 | 2.7 | -576 | [17] |
| $C_2H_5OH + OH = CH_3CHOH + H_2O$ | 1.31E5 | 2.4 | -1457 | [17] |
| $C_2H_5OH + H = C_2H_4OH + H_2$ | 1.88E3 | 3.2 | 7150 | [16] |
| $C_2H_5OH + H = CH_3CHOH + H_2$ | 1.79E5 | 2.5 | 3420 | [16] |
| $C_2H_5OH + O = C_2H_4OH + OH$ | 9.69E2 | 3.2 | 4658 | [19] |
| $C_2H_5OH + O = CH_3CHOH + OH$ | 1.45E5 | 2.4 | 876 | [19] |
| $C_2H_5OH + O = CH_3CH_2O + OH$ | 1.46E-3 | 4.7 | 1727 | [19] |
| $C_2H_5OH + CH_3 = C_2H_4OH + CH_4$ | 3.30E2 | 3.3 | 12291 | [18] |
| $C_2H_5OH + CH_3 = CH_3CHOH + CH_4$ | 1.99E1 | 3.4 | 7635 | [18] |
| $C_2H_5OH + CH_3 = CH_3CH_2O + CH_4$ | 2.04 | 3.6 | 7722 | [18] |

products, and important intermediate species) are set as the target component, then a series of species are strongly coupled to the target component. When an error threshold is set, DRGEP method can identify unimportant components and thus remove those components and the reactions associated with them. The DRGEP method was implemented efficaciously by removing the species and reactions whose target variable error in the worst case exceeds the threshold. The worst-case error was considered to be the maximum relative error of the original and reduced mechanisms in the 24 target cases. When the worst-case error was within the threshold, the tolerance would be increased and the first stage reduction would continue. The reduction of the first phase was terminated until the worst-case error of simulation results exceeded the preset threshold (5%). At this stage, some species (ethanol, $i-C_8H_{18}$, $n-C_7H_{16}$, $s-C_2H_4OH$) was chosen to be the starting species. Relative tolerance of mole-fractions for CO, CO₂, and H was 0.009, respectively, and threshold for iso-octane, $s-C_2H_4OH$, CH₃CHO, n-heptane, and ethanol was set to 5%, respectively. After reduction, the first-stage model (296 species, 1691 reactions) was developed.

Subsequently, QSSA method was also applied. The QSSA method [26–28] can be used to identify some intermediate species whose production and consumption rates were nearly equal. For these species, the change in concentration was almost negligible. After setting a threshold, these intermediate quasi-steady-state (QSS) species would be processed by a nonlinear algebraic system. In this paper, the relative tolerances of *HRR* and ignition delay were set to 0.1 compared with the initial mechanism; the relative tolerance of mole fractions for CO, iso-octane, and n-heptane were set to 7%. iC_4H_8 and C_5H_3 were identified under quasi-steady-state. It not only saved computational time, but also greatly reduced the stiffness of the ordinary differential equation system. After second-stage reduction, the model (277 species, 1437 reactions) was constructed.

3. Experimental Setting

In this paper, the HCCI experiments were performed to validate the mechanism of gasoline-ethanol surrogates and to provide more basic data on combustion characteristics of the test fuel. The three-component fuel (iso-octane/n-heptane/ethanol = 62%: 18%: 20% by volume) (a surrogate for 95 RON gasoline) was used as the test fuel in HCCI experiments. The selected HCCI operating conditions are shown in Table 2.

TABLE 2: HCCI operating conditions.

| Test | n (r/min) | ϕ | T_{in} (K) | Fuel quantity per cycle (mg/cyc) |
|------|-------------|--------|--------------|----------------------------------|
| OP1 | 1200 | 0.3 | 433 | 11.19 |
| OP2 | 1200 | 0.3 | 423 | 11.45 |
| OP3 | 1200 | 0.3 | 413 | 11.74 |
| OP4 | 1200 | 0.3 | 403 | 12.03 |
| OP5 | 1200 | 0.3 | 393 | 12.35 |
| OP6 | 1200 | 0.4 | 433 | 15.66 |
| OP7 | 1200 | 0.4 | 423 | 16.04 |
| OP8 | 1200 | 0.4 | 413 | 16.43 |
| OP9 | 1200 | 0.4 | 403 | 16.85 |
| OP10 | 1200 | 0.4 | 423 | 17.29 |
| OP11 | 1200 | 0.35 | 423 | 13.19 |
| OP12 | 1200 | 0.25 | 393 | 10.29 |

3.1. Experimental Setup. The test engine was retrofitted based on a water cooled, direct injection, naturally aspirated, original engine, CT2100Q. The first cylinder maintained the conventional diesel engine mode, while the second cylinder operated in HCCI mode.

In order to meet the requirements of HCCI operating mode, the intake system, exhaust system, and fuel system were modified. The details and schematic of the test HCCI engine system can be found in [25]. In order to control the intake temperature (T_{in}), an independent port-fuel-injection system and an electric heating system were installed on the second intake pipe. When the injection pulse width was bigger than 2.5 milliseconds, the test fuel was injected into the HCCI cylinder. The injection timing was set to 30°CA BTDC and duration can vary from 5 milliseconds to 8 milliseconds under different operating conditions. The P of the second cylinder was recorded by Kistler 6052A. This piezoelectric pressure makes it possible to record digital signals (TTL: >4, 5 V high, <1 V low level). These signals are then transmitted, either by ECU or statically from the test cylinder, and then it was analyzed by Kibox 283A. The accuracy for the instruments employed can be seen in [25].

3.2. Experimental Procedure. To start the test engine in HCCI mode smoothly, at the beginning, the engine worked in diesel mode. When the water temperature reached 95°C and oil temperature reached 85°C, the diesel supply to the first cylinder was stopped and the fuel injection for the other cylinder was started simultaneously. As a result, the operating mode was successfully switched. When the HCCI

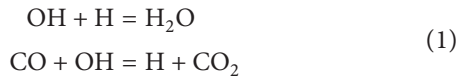
engine run stably, the in-cylinder pressures were recorded, averaged, and analyzed based on 100 consecutive cycles.

4. Results and Discussion

τ , S_L , HRR , P , and molarity for vital species are the key parameters of the prediction.

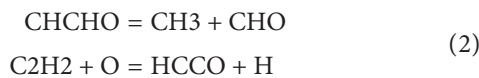
4.1. Laminar Flame Speed. The simulation analysis was performed in flame speed reactor. Figure 2 gives S_L s of three single-component fuels at $T_{in} = 298$ K and $P_{in} = 0.1$ MPa. The comparison of S_L s between the results of calculation obtained by using the second-stage model and experimental results under different ϕ s can be seen in Figure 2(a). It can be found that the trends of the calculated S_L s with ϕ s are consistent with the experimental values in [14]. That is, as ϕ increases from 0.6 to 1.3, the S_L s of the three fuels first increase and then decrease. However, there is a significant gap between the two curves. In other words, the calculated S_L s for the three single-component fuels cannot match with the experimental values. Therefore, the second-stage mechanism should be revised.

By analyzing the sensitivity of the base model on S_L at selected operating conditions, the reactions that affect S_L obvious were picked out. These highly sensitive main-chain branching reactions were revised and subsequently several reaction rate constants were also adjusted to improve the prediction. In addition, the two following reactions keeping highly sensitive were also added, based on the detailed discussion shown by Westbrook et al. [29]:



Any elementary reaction that produces hydrogen (H) radical increases the rate of branch reaction R9. R21 not only increases the production of H, but also affects the destruction of hydroxyl (OH) radical, so these two elementary reactions take vital parts in the autoignition process. Moreover, this stage reduced model cannot predict the reactions related to H accurately. As a result, the parameters of these reactions related to H require revision.

Based on detailed model developed by Mehl et al. [30], some Arrhenius coefficients were adjusted. The elementary reactions associated with H radical may accurately reproduce the characteristics of iso-octane flame, so vital reactions that affect S_L were extracted, as shown below:



In addition, OH radical has an obvious effect on auto-ignition of gasoline-ethanol surrogates [14]; the parameters of the reactions involving H or OH in the products or reactants should be subjected to sensitivity analysis, and then the kinetics of the vital reactions need to be adjusted.

As a result, the final model including 278 species and 1439 reactions was proposed. To determine the prediction accuracy of this model, S_L s of the three initial fuel

components were calculated. Figure 2(b) shows the calculated value and the experimental data under different ϕ s. It is obvious that final mechanism can predict S_L s of the three initial components more accurately than second-stage model. This also indicates that the revision of the model is effective.

4.2. Pressure and Heat Release Rate. To further study the oxidation process of gasoline-ethanol surrogates in the HCCI engine, in-cylinder pressure (P) and heat release rate (HRR) in HCCI mode were calculated by coupling CHEMKIN with CFD software. The given HCCI operating conditions were as follows: $n = 1200$ r/min, $P_{in} = 0.1$ MPa, $T_{in} = 423$ K, and different ϕ s ($\phi = 0.25, 0.30, 0.35, 0.40$).

The HCCI combustion chamber model based on HCCI engine parameters is given in Figure 3.

Coupling CHEMKIN with CFD, calculated values and experimental results for in-cylinder pressures and HRR s were compared, when $P_{in} = 0.1$ MPa, $n = 1200$ r/min, $\phi = 0.25, 0.3, 0.35, 0.4$, as shown in Figure 4.

As can be seen from Figure 4, curves of experimental values and the calculated data show the same trend. Firstly, under the above conditions, when equivalent ratio ϕ rises, the two peak values of P and HRR obtained by simulation show an increasing trend, and the corresponding timing of the peak value is advanced. Obviously, variation trend of the two parameters (P and HRR) obtained by simulation is consistent with the experimental data. Secondly, the calculated data are in good agreement with the experimental values.

P rises slightly as the piston moves up. After auto-ignition, P rises sharply until it reaches the peak in-cylinder pressure (P_{max}), and then P gradually decreases as the piston moves down during the power stroke. Inevitably, there are two factors that may cause the predicted P s to be higher than the experimental values. Firstly, the calculation is performed by a zero-dimensional model that ignores crevices and the inhomogeneity of mixture concentration and temperature in the cylinder. Secondly, the assumptions are closed, constant volume and adiabatic.

In addition, when T_{in} rises from 413 K to 433 K, the formation and combustion speed up, the formation of OH and related active groups are accelerated, so the ignition delay is shortened. However, the P_{max} at high T_{in} (433 K) is smaller than that at low temperature (413 K). This is because T_{in} is too high, resulting in a reduction in the density of the mixture, which in turn reduces the quantity of injection fuel per cycle.

4.3. Heat Release Rate. The calculated HRR at two different ϕ s (0.3, 0.4) using this model and the model by Li et al. [16] are compared with HCCI experimental data in Figures 5(a) and 5(b).

The given conditions were $P_{in} = 0.1$ MPa, $T_{in} = 433$ K and $n = 1200$ r/min.

The simulation curve (blue line) obtained in this study follows the trend: when ϕ increases, the heat release

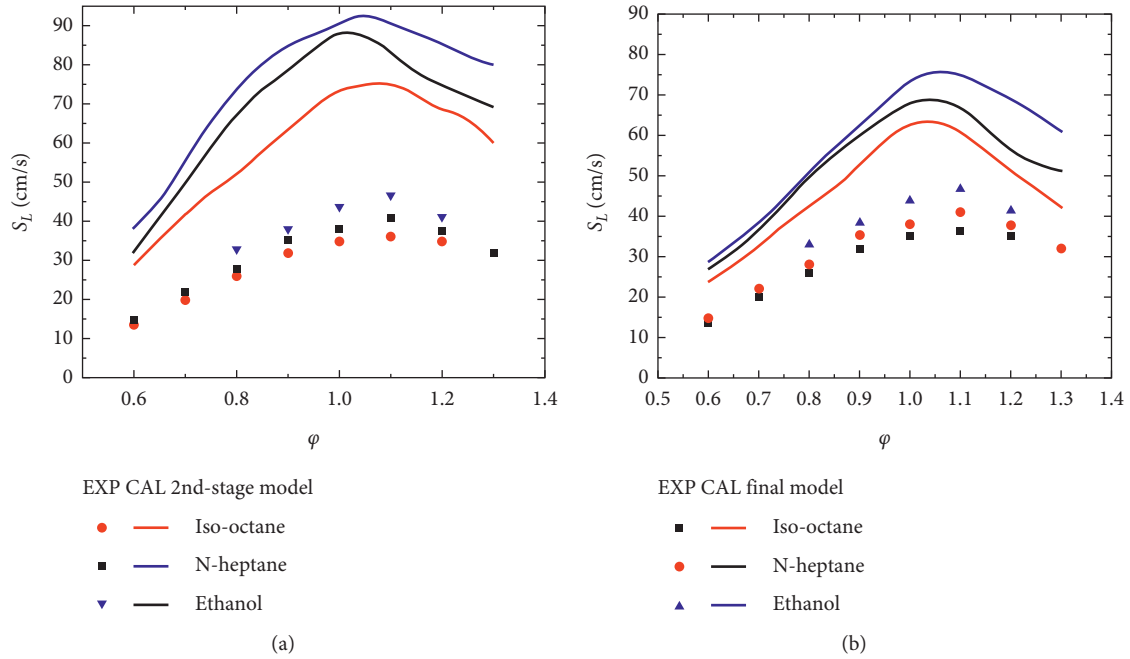


FIGURE 2: S_L s for the calculated data and experimental values under different ϕ s: (a) predictions of the second-stage model; (b) predictions of the final model.

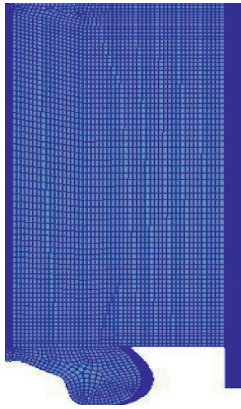


FIGURE 3: HCCI combustion chamber 3D-CFD model.

in the cylinder is more concentrated, and the shape of the HRR curve will change to a narrow and high trend. That is, when ϕ increases, the peak heat release rate increases accordingly, and the time of occurrence is advanced. This is mainly because, as ϕ increases, the number of activated molecules increases and the thermal energy in the cylinder is more sufficient, which not only leads to faster overall reaction, shorter ignition delay period, and concentrated heat release, but also increases the heat release in the cylinder.

Due to the inhomogeneity, the HRR s calculated by this model and the model of Li et al. [16] are significantly greater than the HCCI experimental data. As mentioned earlier, for the same reasons, the calculated P s are also higher than the experimental results.

Above all, the predicted trends using this model are in consistency with the HCCI experimental results.

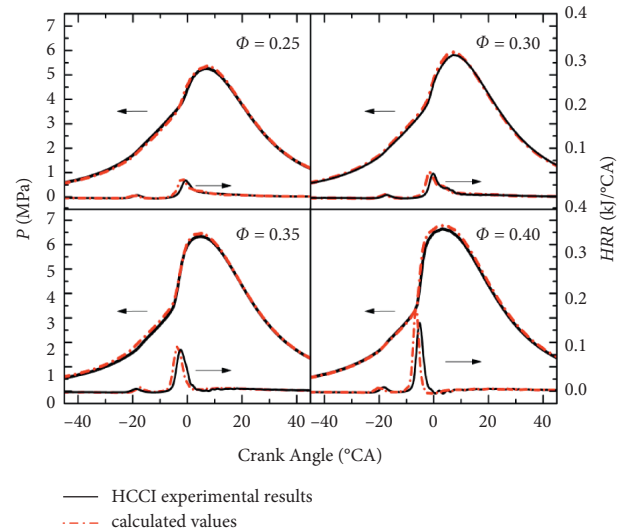


FIGURE 4: Calculated values and experimental results for P and HRR .

4.4. CA10 and CA50. In engine experimental research, τ is the duration from the opening of the injector needle valve until the moment when the P curve starts to separate from the pure compression curve at the compression process. This moment when P rises sharply refers to the CA corresponding to the heat release percentage being 10% (CA_{10}). Moreover, when 50% of the heat has been released is usually considered as the midpoint of the combustion process per engine cycle, which was marked as CA_{50} .

To further study the oxidation process in HCCI cylinder, CA_{10} s and CA_{50} s at different T_{in} s by using the final model, the HCCI results are compared in Figures 6 and 7. As can be

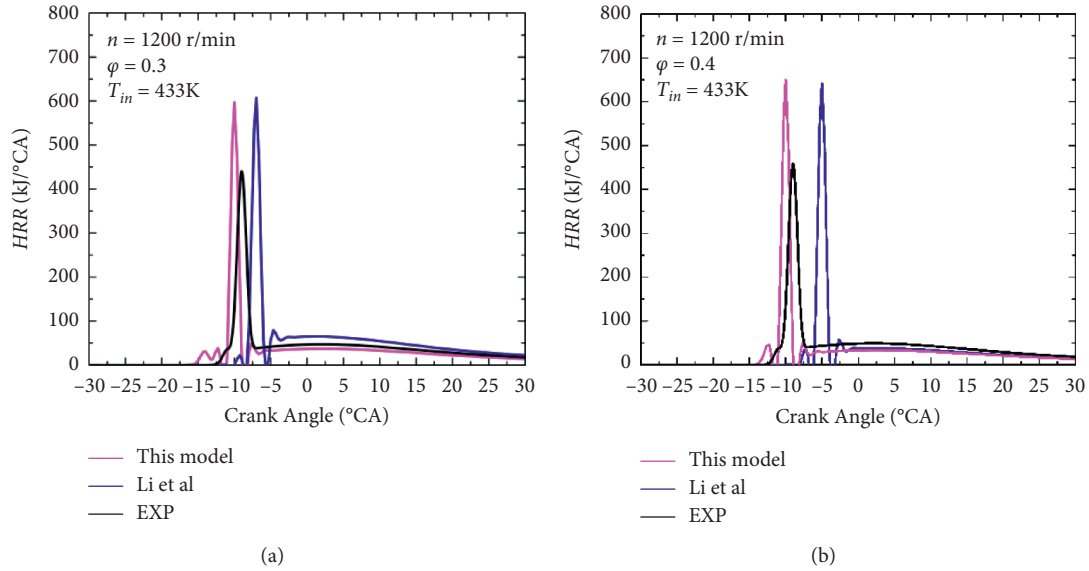


FIGURE 5: Comparison of the HCCI experimental and calculated HRRs as functions of CA at $P_{in} = 0.1$ MPa, $T_{in} = 433$ K, $n = 1200$ r/min, and two different ϕ s: (a) $\phi = 0.3$, (b) $\phi = 0.4$.

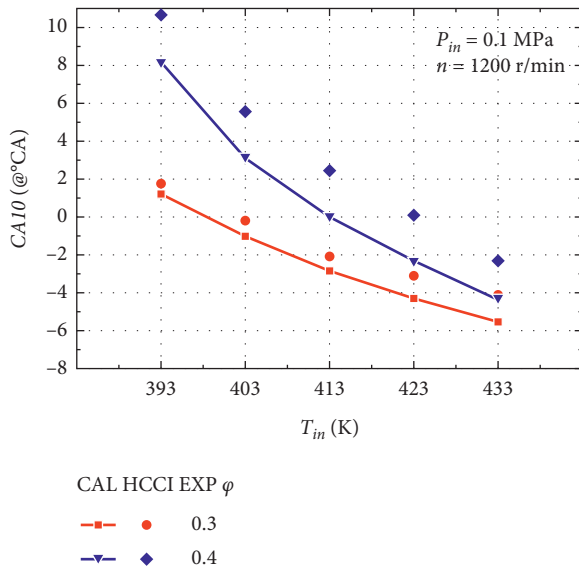


FIGURE 6: HCCI experimental CA10s and calculated results under different T_{in} s, when $P_{in} = 0.1$ MPa, $n = 1200$ r/min.

seen from Figure 6, when T_{in} decreases from 433 K to 393 K, CA10 is in advance about 6°CA ~12°CA. Figure 7 shows that the midpoint of combustion process (CA50) at T_{in} of 393 K is 6°CA ~12°CA later than at T_{in} of 433 K.

Above all, the results indicate that T_{in} play a vital part in oxidation process of the test fuel. Good agreements can be achieved between the calculation and the experimental data.

Comparison of the experimental data (scatters) [31] and calculated results (lines) for PRF100/PRF90/PRF0 in a rapid compression machine (RCM) can be seen in Figure 8. It shows the relationship between τ and ϕ . As expected, the higher the proportion of n-heptane in PRFs is, the shorter the τ is.

Moreover, the higher T_{in} and the higher T during the compression stroke cause the mixture to be more homogeneous.

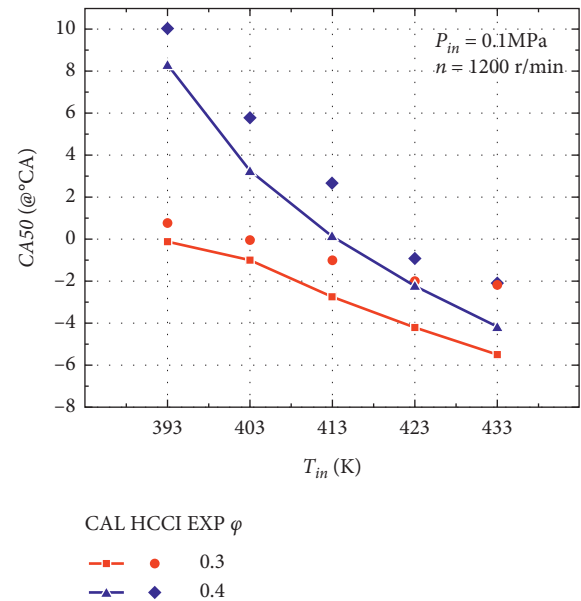


FIGURE 7: HCCI experimental CA50s and predictions under different T_{in} s, when $P_{in} = 0.1$ MPa and $n = 1200$ r/min.

As a result, more active radicals may be generated and the reaction rate may be accelerated, which lead to a shorter τ .

Overall, a higher intake air temperature will advance the phase of the peak heat release rate, which means that the heat release of the fuel in the HCCI cylinder will be more concentrated; that is, when T_{in} is appropriately increased, CA10 and CA50 advanced and the HRR curve tends to be narrow and high.

4.5. Mole Fractions of the Vital Species. Comparison of the predicted mole fractions for iso-octane, heptane, and ethanol by our developed model and Li et al. model [16] at different T_{in} s can be seen in Figure 9.

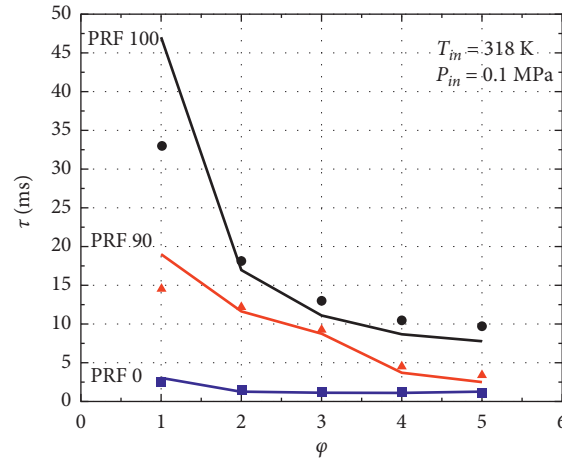


FIGURE 8: Comparison of experimental τ [31] in RCM and modeling results.

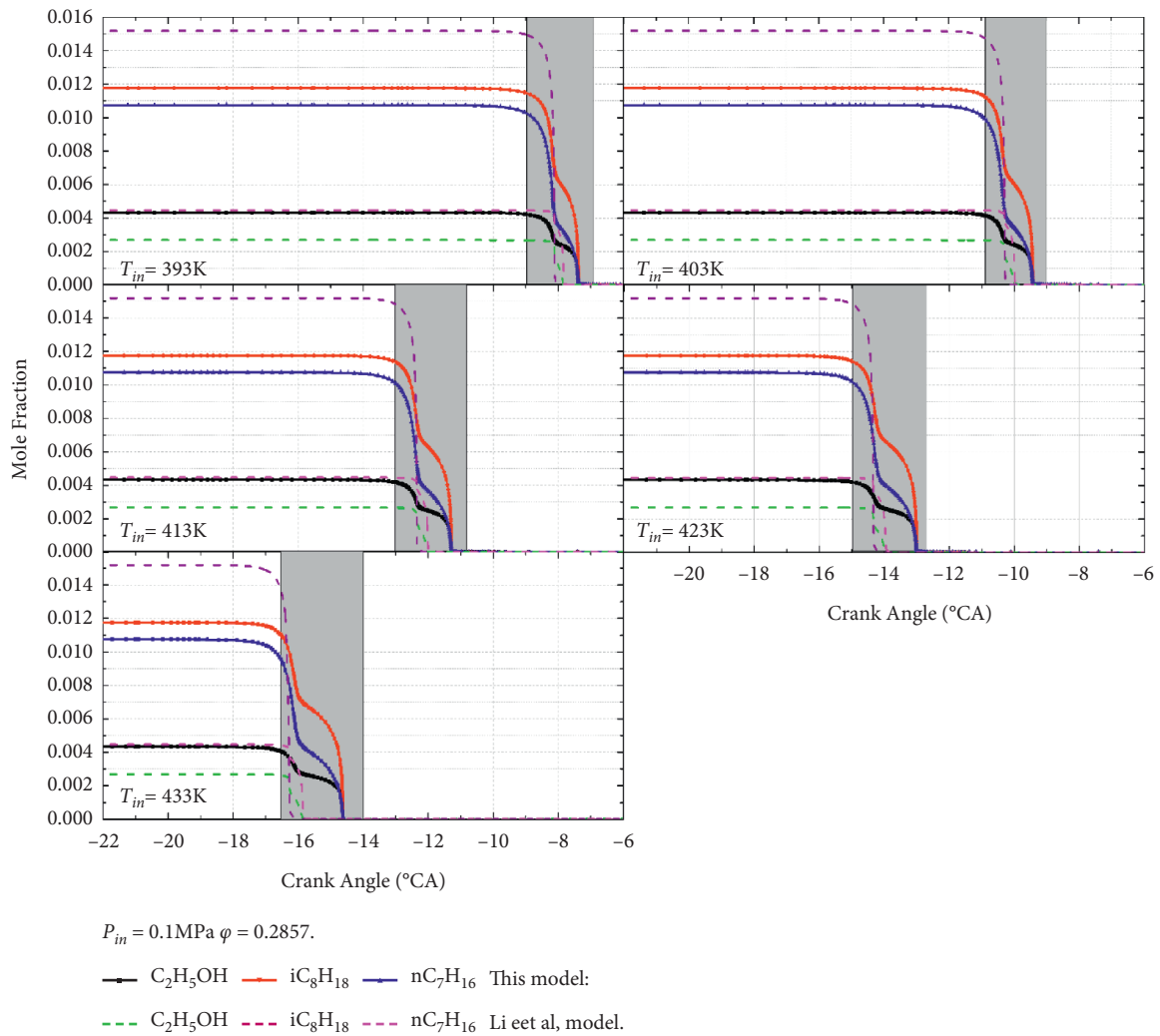


FIGURE 9: Mole fractions of iso-octane, heptane, and ethanol predicted by our developed model and Li et al. model [16] at different T_{in} s.

The following findings can be drawn from Figure 9. Firstly, under the same operating condition, the moments at which the three initial components begin to decrease sharply predicted by the two models are relatively close, as shown by

the shaded area in Figure 9. This indicates that the τ predicted by the two mechanisms are roughly the same. Secondly, when T_{in} increases from 393 K to 433 K, the moments at which the three initial components begin to

decrease sharply are advanced from -8°CA to -14°CA . This is because the formation of free radicals may be accelerated at high T_{in} , thus leading to the advance of autoignition.

In summary, the final model is predictive at the autoignition phase by comparing the predicted τ , P , S_L , HRR , T , CA_{10} , CA_{50} , and mole fractions of vital species under selected conditions.

5. Conclusions

This work developed a mechanism (278 species, 1439 reactions) for gasoline-ethanol surrogates by implementing the reduction and merge scheme. DRGEP and QSSA methods were used to efficiently reduce the mechanism. Moreover, the kinetic parameters of the relevant reactions related to the formation and consumption of H and OH were adjusted.

HCCI experiments were conducted on the combustion of gasoline-ethanol surrogates. More HCCI experimental data were provided to validate the models for the oxidation.

The proposed mechanism was validated as well as the predictions of the previous literature model. Since the calculation is based on ideal assumptions, there is a gap between the simulated curve and the experimental curve. Overall, the prediction of this developed model was found satisfactory in terms of certain characteristic parameters involving S_L , P , T , CA_{10} , CA_{50} , τ , HRR , and mole fractions of species under the selected HCCI conditions.

Based on the reduced mechanism and HCCI engine model developed in this paper, we will further analyze the influence of other boundary conditions (such as intake pressure, engine speed, and EGR rate) on the flame structure and combustion flow field of fuel combustion in the future.

Nomenclature

CA_{10} : Timing at which 10% of the heat has been released (@ $_{\text{QCA}}$)

CA_{50} : Crank angle at which 50% of the heat has been released (@ $_{\text{QCA}}$)

T_{in} : Intake temperature (K)

φ : Equivalence ratio

Abbreviation

BTDC: Before top dead center

CFD: Computational fluid dynamics

TRF: Toluene primary reference fuel.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Special Fund Chang'an University (300102228509, 300102228403, 300102228505, 300102229502, and 300102229202).

References

- [1] M. P. Halstead, L. J. Kirsch, and C. P. Quinn, "The autoignition of hydrocarbon fuels at high temperatures and pressures-fitting of a mathematical model," *Combustion and Flame*, vol. 30, pp. 45–60, 1977.
- [2] R. A. Cox and J. A. Cole, "Chemical aspects of the autoignition of hydrocarbon-air mixtures," *Combustion and Flame*, vol. 60, no. 2, pp. 109–123, 1985.
- [3] H. L. Li, D. L. Miller, and N. P. Cernansky, "Development of a reduced chemical kinetic model for prediction of preignition reactivity and autoignition of primary reference fuels," in *Proceedings of the SAE World Congress & Exhibition*, April 1996.
- [4] S. Tanaka, F. Ayala, and J. C. Keck, "A reduced chemical kinetic model for HCCI combustion of primary reference fuels in a rapid compression machine," *Combustion and Flame*, vol. 133, no. 4, pp. 467–481, 2003.
- [5] N. A. Slavinskaya and O. J. Haidn, "Modeling of n-heptane and iso-octane oxidation in air," *Journal of Propulsion and Power*, vol. 19, no. 6, pp. 1200–1216, 2003.
- [6] H. J. Curran, W. J. Pitz, C. K. Westbrook, G. V. Callahan, and F. L. Dryer, "Oxidation of automotive primary reference fuels at elevated pressures," *Symposium (International) on Combustion*, vol. 27, no. 1, pp. 379–387, 1998.
- [7] Y. Ra and R. D. Reitz, "A reduced chemical kinetic model for IC engine combustion simulations with primary reference fuels," *Combustion and Flame*, vol. 155, no. 4, pp. 713–738, 2008.
- [8] P. Kirchen, M. Shahbakhti, and C. R. Koch, "A skeletal kinetic mechanism for PRF combustion in HCCI engines," *Combustion Science and Technology*, vol. 179, no. 6, pp. 1059–1083, 2007.
- [9] N. M. Marinov, "A detailed chemical kinetic model for high temperature ethanol oxidation," *International Journal of Chemical Kinetics*, vol. 31, no. 3, pp. 183–220, 1999.
- [10] Y. D. Liu, M. Jia, M. Z. Xie, and B. Pang, "Development of a new skeletal chemical kinetic model of toluene reference fuel with application to gasoline surrogate fuels for computational fluid dynamics engine simulation," *Energy & Fuels*, vol. 27, no. 8, pp. 4899–4909, 2013.
- [11] H. Wang, M. Yao, Z. Yue, M. Jia, and R. D. Reitz, "A reduced toluene reference fuel chemical kinetic mechanism for combustion and polycyclic-aromatic hydrocarbon predictions," *Combustion and Flame*, vol. 162, no. 6, pp. 2390–2404, 2015.
- [12] S. Jerzembeck, A. Sharma, and N. Peters, "Laminar burning velocities of nitrogen diluted standard gasoline-air mixture," in *Proceedings of the SAE World Congress & Exhibition*, June 2008.
- [13] D. Zheng and B. J. Zhong, "Chemical kinetic model for ignition of three-component fuel comprising iso-Octane/n-Heptane/Ethanol," *Acta Physico-Chimica Sinica*, vol. 28, no. 9, pp. 2029–2036, 2012.
- [14] B.-J. Zhong and D. Zheng, "Chemical kinetic mechanism of a three-component fuel composed of iso-octane/n-heptane/ethanol," *Combustion Science and Technology*, vol. 185, no. 4, pp. 627–644, 2013.

- [15] R. Lemaire, E. Therssen, and P. Desgroux, "Effect of ethanol addition in gasoline and gasoline-surrogate on soot formation in turbulent spray flames," *Fuel*, vol. 89, no. 12, pp. 3952–3959, 2010.
- [16] Y. Li, A. Alfazazi, B. Mohan et al., "Development of a reduced four-component (toluene/n-heptane/iso-octane/ethanol) gasoline surrogate model," *Fuel*, vol. 247, no. 1, pp. 164–178, 2019.
- [17] M. Sjoberg and J. E. Dec, "An investigation into lowest acceptable combustion temperatures for hydrocarbon fuel in HCCI engines," *Proceedings of the Combustion Institute*, vol. 30, no. 2, pp. 2719–2726, 2004.
- [18] M. Sjoberg and J. E. Dec, "Isolating the effects of fuel chemistry on combustion phasing in an HCCI engine and the potential of fuel stratification for ignition control," in *Proceedings of the SAE World Congress & Exhibition*, Chicago, IL, USA, June 2004.
- [19] J. Park, Z. F. Xu, and M. C. Lin, "A computational study of the kinetics and mechanism for the $C_2H_3 + CH_3OH$ reaction," *International Journal of Chemical Physics*, vol. 47, no. 12, pp. 764–772, 2003.
- [20] S. Xu and M. C. Lin, "Theoretical study on the kinetics for OH reactions with CH_3OH and C_2H_5OH ," *Proceedings of the Combustion Institute*, vol. 31, no. 1, pp. 159–166, 2007.
- [21] Z. F. Xu, J. Park, and M. C. Lin, "Thermal decomposition of ethanol. III. A computational study of the kinetics and mechanism for the $CH_3 + C_2H_5OH$ reaction," *The Journal of Chemical Physics*, vol. 120, no. 14, pp. 6593–6599, 2004.
- [22] C.-W. Wu, Y.-P. Lee, S. Xu, and M. C. Lin, "Experimental and theoretical studies of rate coefficients for the reaction $O(3P) + C_2H_5OH$ at high temperatures," *The Journal of Physical Chemistry A*, vol. 111, no. 29, pp. 6693–6703, 2007.
- [23] M. Fikri, J. Herzler, R. Starke, C. Schulz, P. Roth, and G. T. Kalghatgi, "Autoignition of gasoline surrogate mixtures at intermediate temperatures and high pressures," *Combustion and Flame*, vol. 152, no. 1-2, pp. 276–281, 2008.
- [24] J. B. Heywood, *A Textbook for Internal Combustion Engines Fundamentals*, McGraw-Hill Science/Engineering/Math, New York, NY, USA, 1988.
- [25] C. Zhang and H. Wu, "Combustion characteristics and performance of a methanol fueled homogenous charge compression ignition (HCCI) engine," *Journal of the Energy Institute*, vol. 89, no. 3, pp. 346–353, 2016.
- [26] Y. Chen, M. Mehl, Y. Xie, and J.-Y. Chen, "Improved skeletal reduction on multiple gasoline-ethanol surrogates using a Jacobian-aided DRGEP approach under gasoline compression ignition (GCI) engine conditions," *Fuel*, vol. 210, pp. 617–624, 2017.
- [27] E. A. Tingas, D. J. Diamantis, and D. A. Goussis, "Issues arising in the construction of QSSA mechanisms: the case of reduced n-heptane/air models for premixed flames," *Combustion Theory and Modelling*, vol. 6, no. 22, pp. 1049–1083, 2018.
- [28] J. C. G. Andrae, T. Brinck, and G. T. Kalghatgi, "HCCI experiments with toluene reference fuels modeled by a semi-detailed chemical kinetic model," *Combustion and Flame*, vol. 155, no. 4, pp. 696–712, 2008.
- [29] C. K. Westbrook, Y. Mizobuchi, T. J. Poinso, P. J. Smith, and J. Warnatz, "Computational combustion," *Proceedings of the Combustion Institute*, vol. 30, no. 1, pp. 125–157, 2005.
- [30] M. Mehl, T. Faravelli, E. Ranzi, D. Miller, and N. Cernansky, "Experimental and kinetic modeling study of the effect of fuel composition in HCCI engines," *Proceedings of the Combustion Institute*, vol. 32, no. 2, pp. 2843–2850, 2009.
- [31] S. Tanaka, F. Ayala, J. C. Keck, and J. B. Heywood, "Two-stage ignition in HCCI combustion and HCCI control by fuels and additives," *Combustion and Flame*, vol. 132, no. 1-2, pp. 219–239, 2003.

Research Article

A New V-Net Convolutional Neural Network Based on Four-Dimensional Hyperchaotic System for Medical Image Encryption

Xiaowei Wang,¹ Shoulin Yin,¹ Muhammad Shafiq ,² Asif Ali Laghari,³ Shahid Karim,⁴ Omar Cheikhrouhou ,^{5,6} Wajdi Alhakami,⁷ and Habib Hamam^{8,9,10}

¹Software College, Shenyang Normal University, Shenyang, China

²Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China

³Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan

⁴Faculty of Science and Technology, ILMA University, Karachi, Pakistan

⁵CES Laboratory, National School of Engineers of Sfax, University of Sfax, Sfax 3038, Tunisia

⁶Higher Institute of Computer Science of Mahdia, University of Monastir, Mahdia 5111, Tunisia

⁷Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

⁸Faculty of Engineering, Moncton University, Moncton, NB E1A3E9, Canada

⁹Spectrum of Knowledge Production & Skills Development, Sfax 3027, Tunisia

¹⁰School of Electrical Engineering, Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2006, South Africa

Correspondence should be addressed to Muhammad Shafiq; srsshafiq@gmail.com

Received 7 December 2021; Revised 31 December 2021; Accepted 10 January 2022; Published 14 February 2022

Academic Editor: Celestine Iwendi

Copyright © 2022 Xiaowei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the transmission of medical images, if the image is not processed, it is very likely to leak data and personal privacy, resulting in unpredictable consequences. Traditional encryption algorithms have limited ability to deal with complex data. The chaotic system is characterized by randomness and ergodicity, which has advantages over traditional encryption algorithms in image encryption processing. A novel V-net convolutional neural network (CNN) based on four-dimensional hyperchaotic system for medical image encryption is presented in this study. Firstly, the plaintext medical images are processed into 4D hyperchaotic sequence images, including image segmentation, chaotic system processing, and pseudorandom sequence generation. Then, V-net CNN is used to train chaotic sequences to eliminate the periodicity of chaotic sequences. Finally, the chaotic sequence image is diffused to change the raw image pixel to realize the encryption processing. Simulation test analysis demonstrates that the proposed algorithm has better effect, robustness, and plaintext sensitivity.

1. Introduction

At present, there are two main ways for information transmission, one is text and the other is image. Therefore, in addition to words, images also contain a lot of important and confidential information. In the current era of computer network, images are mostly stored in the form of digital images, which is simple, quick, and easy to find. However, at the same time, it also increases the risk of information

leakage, especially when images are transmitted on the network; they are easy to attack. In this context, image information encryption is an important means to prevent information leakage [1–5].

There are many research studies on image encryption protection. The traditional encryption algorithms mainly consist of randomly disturbing the row or column of image information to encrypt, randomly disturbing the image pixel information for encryption and decryption, zooming in and

out the image information of the pixel point, and so on [6–8]. They are easy to crack. To solve the problems in the above methods, chaos-based encryption algorithm appears, which is the most widely used image encryption algorithm at present [9, 10]. Although it has higher encryption effect, this method has two defects. One is that all the image information is encrypted into ciphertext image, resulting in a sharp increase in the amount of information after image encryption, occupying a large amount of storage space. Secondly, the generated chaotic sequence by pure use of the chaotic system shows local linearity and strong correlation, that is, it will show a certain degree of periodicity and so on. The existence of this feature makes the image security relatively lower [11–15].

In this article, aiming at the periodicity shortcoming of chaotic encryption algorithm, V-net CNN is used to learn chaotic sequence to break the periodicity of chaotic sequence to improve the confidentiality of image encryption. The validity and practicability of the new scheme are proved by testing, which provides a reference for image encryption.

2. 4D Hyperchaotic System

The 4D hyperchaotic system [16, 17] studied in this paper is as

$$\dot{w} = dx. \quad (1)$$

When parameters $a = 35$, $b = 3$, $c = 33$, and $d = 8$, a typical hyperchaotic attractor exists in system (1). The phase diagram is shown in Figure 1. Figure 1(a) is the x - y - z three-dimensional projection phase diagram. Figure 1(b) is the y - z - w three-dimensional projection phase diagram.

2.1. Analysis of Chaos Characteristics. The dissipative property of the new system (1) is analyzed. Dissipation value is $\nabla V = (\partial \dot{x} / \partial x) + (\partial \dot{y} / \partial x) + (\partial \dot{z} / \partial x) + (\partial \dot{w} / \partial x)$; when $\nabla V < 0$, the system is wasteful. If the system parameters are substituted, $\nabla V = -a - b = -38 < 0$; if the dissipation condition is satisfied, the trajectory of the system eventually contracts asymptotically to a particular limit set of zero volume at an exponential rate and is eventually fixed to an attractor.

Four Lyapunov exponents are obtained, $LE1 = 0.343$, $LE2 = 0.052$, $LE3 = -0.305$, and $LE4 = -36.640$, of which two Lyapunov exponents are greater than zero, that is, system (1) is a hyperchaotic system.

2.2. Stability Analysis. Adding the time-delay term τ to the second nonlinear formula of hyperchaotic system (1), the time-delay model equation is shown as

$$\dot{y} = cx(t - \tau) - xz - w. \quad (2)$$

When the hysteresis term $\tau = 0$, system (1) is locally asymptotically stable at $E_0 = (0, 0, 0, 0)$, and the Jacobi matrix is

$$J = \begin{pmatrix} -a & a & 0 & 0 \\ ce^{-\lambda\tau} & 0 & 0 & -1 \\ 0 & 0 & -b & 0 \\ d & 0 & 0 & 0 \end{pmatrix}. \quad (3)$$

The feature equation is as follows:

$$(\lambda + b)(\lambda^3 + a\lambda^2 - ac\lambda e^{-\lambda\tau} + a d) = 0. \quad (4)$$

According to the substitution law, $P_1 = a$, $P_2 = -ac$, and $P_3 = ad$, if only the virtual root is considered, when $\tau = 0$, the characteristic equation of system (1) is

$$\lambda^3 + P_1\lambda^2 + P_2\lambda + P_3 = 0. \quad (5)$$

According to Routh–Hurwitz criterion, if $P_1, P_2, P_3 > 0$ and $P_1P_2 - P_3 > 0$, then the real parts of the characteristic roots of equation (3) are all negative. By substituting corresponding parameters into the above inequalities, it can be seen that the time-delay system (1) is locally asymptotically stable at $E_0 = (0, 0, 0, 0)$.

3. Proposed Image Encryption

Computer technology is developing day by day; image storage is mostly realized in the form of the digital image. The image has many information, especially in some special fields (national defense, military, finance and personal privacy, etc.), the information in the image is confidential and not allowed to be disclosed. So, how to ensure the safety of image information is very important. Image information encryption is the main solution at present [18]. Among them, chaotic encryption is the most commonly used method. Its principle is to superimpose one or more chaotic signals on the useful signals to be transmitted at the sending end so that the signals in the transmission channel have the shape of random noise and then achieve the purpose of encryption and secure communication. This method has high encryption speed, lossless compression, and high security, but the generated chaotic sequence still shows a certain degree of periodicity. This paper improves and optimizes a chaotic sequence image encryption algorithm based on V-Net CNN. The proposed method is shown in Figure 2.

3.1. Chaotic Sequence Generation. Chaotic sequence generation is the first step in image encryption, which aims to transform plaintext image into random sequence. The specific process includes three parts: image segmentation, chaotic system processing, and pseudorandom sequence generation.

3.1.1. Plaintext Image Segmentation. The function of plaintext image segmentation is convenient for chaotic system processing. In general, the monitoring image of the target sequence is divided into any one of the ten different sizes in Table 1. The size of each round block is determined

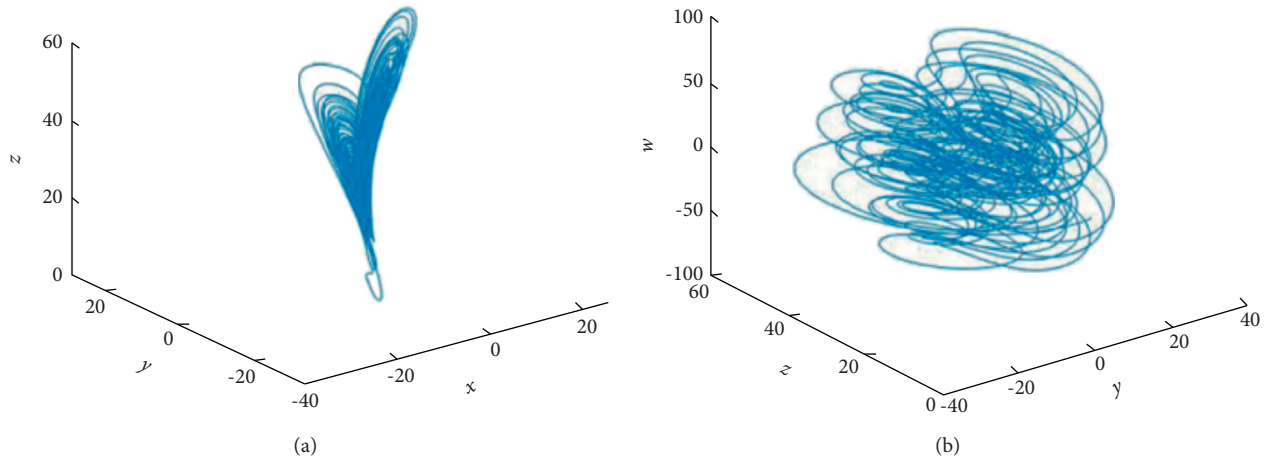


FIGURE 1: System (1) chaotic attractor phase diagram.

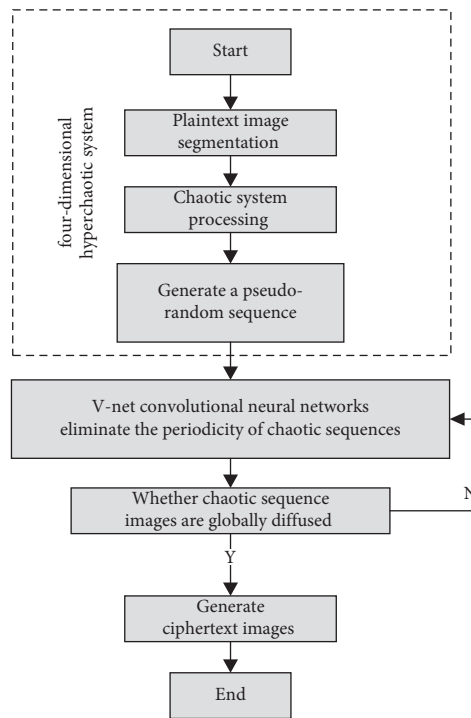


FIGURE 2: Proposed medical image encryption.

TABLE 1: Block size table of each round plaintext image

| Mod10 | Block size | Mod10 | Block size |
|-------|------------|-------|------------|
| 0 | 16 | 5 | 56 |
| 1 | 24 | 6 | 64 |
| 2 | 32 | 7 | 72 |
| 3 | 40 | 8 | 80 |
| 4 | 48 | 9 | 96 |

Note that there is no overlap between these subblocks. In the postprocessing, these subblocks are used as units for correlated operation.

by the session key used during the encryption of that particular round.

3.1.2. Chaotic System Processing. A chaotic system is used to generate real number sequence for plaintext image subblock. There are five commonly used chaotic systems, namely, logistic chaotic system, Chebyshev chaotic system, Skew Tent chaotic system, Henon chaotic system, and Lorenz chaotic system [19–23]. The description of the above five chaotic systems is as follows:

(A) Logistic chaotic system:

$$r_{n+1} = ar_n(1 - r_n), \quad (6)$$

here $r_n \in [0, 1], a \in (1, 5), n = 0, 1, 2, \dots$

(B) Chebyshev chaotic system:

$$x_n = \cos(k \arccos(x_n)), \quad (7)$$

where $x_i \in (-1, 1), i = 1, 2, \dots, N$, and k is the system control parameter. When $k \geq 2$, the Chebyshev map enters the chaotic state:

(C) Skew Tent chaotic system:

$$f(x) = \begin{cases} \frac{x}{\delta}, & x \in (0, \delta), \\ \frac{(1-x)}{(1-\delta)}, & x \in (\delta, 1). \end{cases} \quad (8)$$

When $\delta \in (0, 1)$, the system is chaos.

(D) Henon chaotic system:

$$\begin{cases} x_{n+1} = -\alpha x_n^2 + y_n + 1, \\ y_{n+1} = \beta x_n, \end{cases} \quad (9)$$

where α and β are the system control parameters; when $0.54 < \alpha < 2$ and $0 < |\beta| < 1$, the system is in the chaos state.

(E) Lorenz chaotic system:

$$t = ex - zx - y. \quad (10)$$

3.1.3. Pseudorandom Sequence Generation. After chaotic system processing, the sequence generated is real number sequence, which also needs to be converted into pseudorandom sequence, namely, chaotic sequence. There are three generation methods for pseudorandom sequence, namely, threshold method, binary sequence method, and quantitative extraction method [24]. The following is a specific analysis.

(A) Threshold method define a threshold function $\Theta_t(w)$ as

$$\Theta_t(w) = \begin{cases} 0, & \text{when } w < t, \\ 1, & \text{when } w \geq t. \end{cases} \quad (11)$$

Its complement is $\Theta'_t(w) = 1 - \Theta_t(w)$. t is the set threshold. w is the value of chaotic sequence. Equation (11) is applied to the real number sequence to obtain the pseudorandom sequence.

(B) Binary sequence method: the chaotic sequence value $w(|w| \leq 1)$ can be written as the binary form of

$$|w| = 0, A_1(w), A_2(w), \dots, A_i(w), \dots, \quad (12)$$

$$A_n(w) \in \{0, 1\},$$

$A_i(w)$ can be expressed as

$$A_i(w) = \sum_{r=1}^{2^i-1} (-1)^{r-1} \{\Theta_{r/2^i}(w) + \Theta'_{-r/2^i}(w)\}. \quad (13)$$

So, it can get a pseudorandom sequence.

(C) Quantitative extraction method: if the obtained chaotic sequence is not in the range of $[0, 1]$, the chaotic sequence $\{w_i\}_{i=0}^{\infty}$ is normalized to the interval $[0, 1]$ to obtain $\{x_i\}_{i=0}^{\infty}$. In the representation of x_i as a binary number, it takes the lowest or middle N bits as required. The binary bits corresponding to each x_i value are combined to obtain the key sequence used for encryption.

3.2. V-Net Convolutional Neural Networks Eliminating the Periodicity of Chaotic Sequences. 3D V-Net full convolutional neural network [25, 26] is used in this paper. 3D convolutional neural network can convolve 32 layers of medical images at the same time. Besides learning image features, the 3D convolutional neural network can also learn the position change information of images between different layers. 3D convolutional neural network is a network model with the huge parameter system. In order to make the model perform better, the overall flow of V-Net CNN is shown in Figure 3.

In the process of downsampling, the high-level feature map contains semantic category information, while the low-level feature map retains image details. In the process of downsampling, convolutional neural network will lose important category information. As the downsampling process goes on, the image gradient gradually disappears. To preserve the semantic information of high-level images, the convolution results of high-level images are sent to the upsampling process through the connection layer. However, the complete upsampling process undoubtedly increases the training difficulty. In this paper, the feature maps in the lower sampling process are connected to the upper sampling process by multiplying certain weight values through the global average weight module. The specific approach is to first average pool the feature maps output by the first four layers in the downsampling process and then calculate the corresponding weight values by Softmax function. The formula for calculating the weight is as follows:

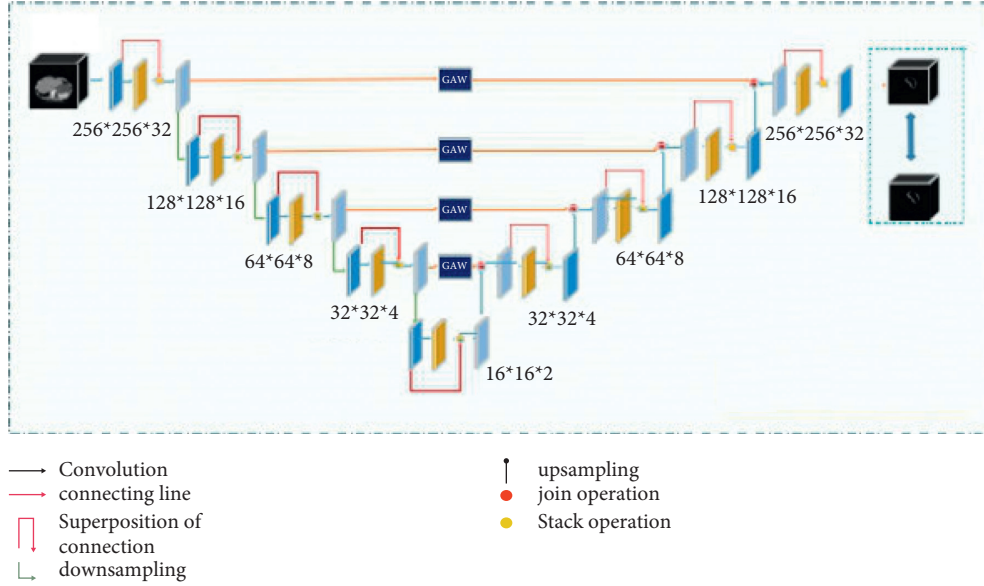


FIGURE 3: V-Net CNN structure.

$$W_i = \frac{\exp(\text{GAP}(F_i))}{\sum_{i=1}^4 \exp(\text{GAP}(F_i))}, \quad (14)$$

where F_i represents the convolution output result of the i th layer. The purpose of global average pooling (GAP) is to eliminate the influence of different scales on weight values in the process of downsampling. Global average weight (GAW) module is adopted to effectively utilize multiscale feature information to improve the learning efficiency of deep learning. The weight acquisition process is shown in Figure 4.

Level set (LS) loss function is a loss function based on the level set method [27], which is the first application of the level set method in loss function of deep learning network. LS loss is denoted as

$$\begin{aligned} F(c1, c2, \varphi) = & \mu \cdot \text{Length}(\varphi) + \nu \cdot \text{Area}(\varphi) \\ & + \lambda_1 \int_{\Omega} |\mu_0(x, y) - c1|^2 H(\varphi(x, y)) dx dy \\ & + \lambda_2 \int_{\Omega} |\mu_0(x, y) - c2|^2 (1 - H(\varphi(x, y))) dx dy, \end{aligned} \quad (15)$$

where $\mu \geq 0$, $\nu \geq 0$, and $\lambda_1, \lambda_2 > 0$ is the fixed value parameter. Ω is the whole image area. φ is the level set function.

Length(φ) and Area(φ) represent the curve length and area regularization terms, respectively. $\mu_0(x, y)$ is the pixel at (x, y) in the image. H is a differentiable step function, where α is a hyperparameter used to improve the gradient of the function, which is set to 2.5 in the experiment:

$$H(x) = \frac{1}{\pi} \left(\arctan \alpha x + \frac{\pi}{2} \right). \quad (16)$$

The idea of LS loss is to first use step function to set all the inside edges of the outer wall of prediction results and ground truth to 1 and the outside edges to 0. When calculating the loss, multiply the predicted result and the ground truth and then sum to calculate the loss and perform the same operation after taking the reverse. The purpose of this is to give enough weight to the edges. This loss function is suitable for the segmentation of single outer edge objects, but not for the segmentation of medical objects with both inner and outer edges. Based on the level set, we propose a regularized level set loss function (LSR Loss), which can optimize the edge through LS loss and constrain the internal details of gastric wall through regularization, giving full play to the respective advantages of the level set method and the deep learning method. LSR loss is defined as

$$\text{LSR loss} = \frac{1}{\lambda_1} \sum_{\Omega} |G_I(x, y) - c_{11}|^2 H(\varphi(x, y)) + \frac{1}{\lambda_2} \sum_{\Omega} |G_I(x, y) - c_{12}|^2 (1 - H(\varphi(x, y))) + \frac{1}{\lambda_3} \sum_{\Omega} |G_I(x, y) - \varphi(x, y)|^2, \quad (17)$$

where Ω represents the entire image region, $G_I(x, y)$ represents the pixel value in ground truth, and $\varphi(x, y)$ represents the pixel value of the image predicted by the network. Here,

$$\begin{aligned} c_{11} &= \frac{\sum_{\Omega} G_I(x, y) H(\varphi(x, y))}{\sum_{\Omega} \Omega H(\varphi(x, y))}, \\ c_{12} &= \frac{\sum_{\Omega} G_I(x, y) (1 - H(\varphi(x, y)))}{\sum_{\Omega} (1 - H(\varphi(x, y)))}. \end{aligned} \quad (18)$$

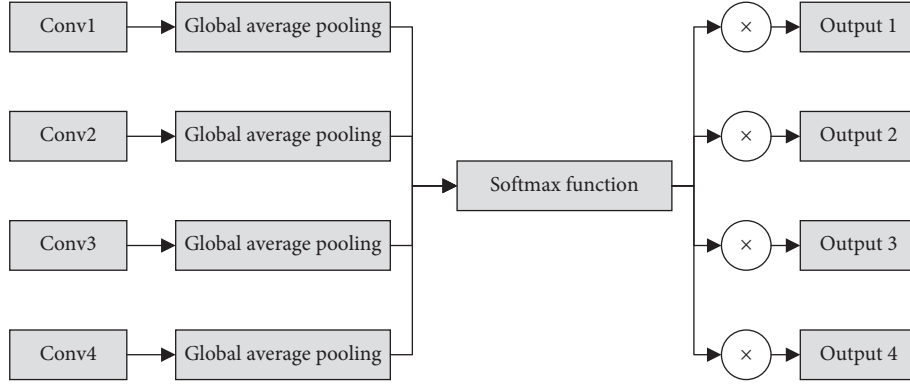


FIGURE 4: Global average weight module.

When the neural network predicted value $\varphi(x, y)$ and the corresponding position of the image were more accurate, the values of c_{l1} and c_{l2} would be closer to 1. Then, the difference between the GT and the predicted value will be equivalent to taking the opposite, and then, it will be close to 0 when multiplied by $\varphi(x, y)$ corresponding to the predicted value. However, when the image boundary error, the loss value will be very large. Therefore, λ_1 and λ_2 are added in this paper to constrain the size of the loss function to normalize it.

Step 1: input the plaintext images generated in the previous into the V-net convolutional neural network structure as training samples.

Step 2: the first layer is conducted convolution operation on the plaintext image, namely, the weighted sum.

Step 3: downsampling the plaintext image after convolution operation, that is, pooling.

Step 4: repeat Step 2 and Step 3 to extract key features of plaintext images and reduce the amount of data processing.

Step 5: enter the full connection layer and connect all key features of plaintext images together.

Step 6: output the training results, and judge whether the error between the convolution result and the actual output is less than the set threshold. If the result is less than the threshold value, the V-net convolutional neural network training has been completed. If it is greater than the threshold value, error backpropagation is required to adjust the thresholds and weights of each layer until the convolutional neural network training is completed.

3.3. Chaotic Sequence Image Diffusion. After the above processing, the periodicity of chaotic sequence is eliminated. However, its pixel value has not changed, so there are still certain security risks. In this case, chaotic images need to be diffused [28]. Diffusion treatment is as follows.

Formula (19) is used to replace each component of chaotic sequence:

$$\begin{aligned} FR(j) &= FR(j) \oplus PX(j) \oplus FR(j-1) \oplus FJ(j-1) \oplus FB(j-1) \\ FG(j) &= FG(j) \oplus PY(j) \oplus FG(j-1) \oplus FR(j-1) \oplus FB(j-1) \\ FB(j) &= FB(j) \oplus PZ(j) \oplus FB(j-1) \oplus FG(j-1) \oplus FR(j-1), \end{aligned} \quad (19)$$

where FR, FG, and FB are the RGB components of chaotic sequences. The matrices PX, PY, and PZ are pixel matrices. \oplus is XOR operation.

It does the substitution again and changes the pixel value further. A diffusion function needs to be introduced here. The expression of the diffusion function is as follows:

$$S_i = (Y_i + S_{i-1}) \bmod U \oplus e, \quad (20)$$

where S_i represents the ciphertext of the current pixel point, Y_i is the plaintext of the current pixel and the ciphertext of the previous pixel, U represents the maximum value of pixel point, \oplus represents XOR operation, and e represents a random value.

Again the substitution formula is as follows:

$$\begin{aligned} FR(j) &= (FB(j) + FG(j))S_i \oplus FR(j), \\ FG(j) &= (FR(j) + FB(j))S_i \oplus FG(j), \\ FB(j) &= (FR(j) + FG(j))S_i \oplus FB(j). \end{aligned} \quad (21)$$

4. Experiments and Analysis

Two images of Lena and Skull with a size of 512×512 pixels are selected for simulation and analysis experiment. Setting logistic chaos system $\mu = 3.9999$, two-dimensional logistic chaos system have $\mu_1 = 0.9$, $\mu_2 = 0.9$, and $r = 0.1$. Experimental hardware environment is 64 GB memory, Windows10 operating system environment. The software simulation platform is Matlab 2017a. To evaluate the overall effect of the encryption algorithm, the following security performance analysis is made from the histogram, information entropy, correlation coefficient analysis, robustness, key space and sensitivity, and antidifferential attack. Figure 5 is the original image. Figures 5(a) and 5(b) are Skull image and Lena image, respectively. Figure 5 is from this paper Vaseghi et al. 2021 (Under the Creative Commons Attribution License/Public Domain) [29].

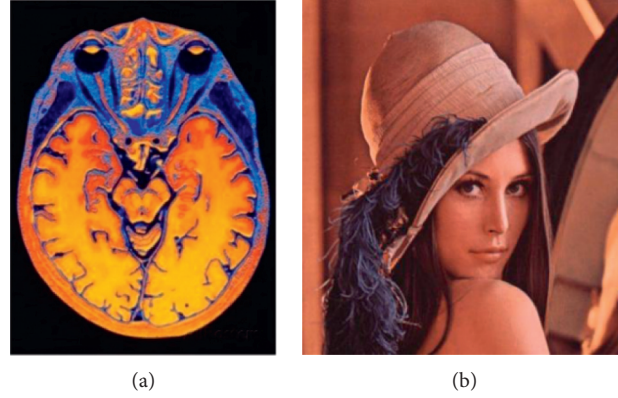


FIGURE 5: Original image.

4.1. Image Gray Histogram. Generally speaking, the histogram distribution is relatively uniform, which can effectively prevent attackers from analyzing the histogram to obtain plaintext information. Figures 6(a)–6(d) show the plaintext image of Lena, ciphertext image of Lena, skull plaintext image, and skull ciphertext image, respectively. It is observed that the distribution of ciphertext histogram is uniform. Therefore, this new algorithm can resist histogram analysis attacks and conceal the statistical characteristics of plaintext images.

4.2. Information Entropy. The information entropy is mainly used to measure the uncertainty or randomness of information source. Conversely, chaos has a high entropy. It is reflected in the image; the more uniform distribution of the pixel gray value denotes the higher entropy value and the stronger randomness. For 8-bit images, the entropy value should be as close as possible to the ideal value 8. The calculation of information entropy is as follows:

$$H(R) = \sum_{i=0}^{255} p(R_i) \text{lb}(p(R_i)), \quad (22)$$

where $p(R_i)$ is the occurrence frequency of pixel value i in the ciphertext image R . Table 2 is the comparison of global information entropy with other methods including FRFT [29], ASFS [30], and CENN [31].

However, there are some deficiencies in the global information entropy and the measurement of the image before and after encryption is not accurate. Therefore, based on the global information entropy, Lin et al. [3] proposed a more rigorous local information entropy test. The core idea is to randomly select nonoverlapping subblocks in the target image, represented as L_1, L_2, \dots, L_k . Each subblock contains T_b pixels, and then, calculate the global information entropy of each subblock. The local information entropy \bar{H}_{k,T_b} of the image is

$$\bar{H}_{k,T_b} = \sum_{i=1}^k \frac{H(L_i)}{k}. \quad (23)$$

It selects $k=30$ and $T_b = 1936$ to carry out local information test on the gray image. Through the local information entropy test, this algorithm is compared. Table 3 shows that the local information entropy test of the proposed algorithm has a relatively high pass rate.

4.3. Correlation Coefficient Analysis. If its value is close to 0, the correlation between image pixels is weaker. If its value is close to 1, the pixels are more relevant. The lower correlation coefficient denotes that it can better avoid the attacker obtaining the meaningful information from the ciphertext image [5].

The correlation coefficient is calculated by the following formula:

$$r_{xy} = \frac{\text{cov}(u, v)}{\sqrt{D(u)}\sqrt{D(v)}}$$

$$\text{cov}(u, v) = \frac{1}{N} \sum_{i=1}^N (x_i - E(u))(y_i - E(v)) \quad (24)$$

$$D(u) = \frac{1}{N} \sum_{i=1}^N (u_i - E(u))^2$$

$$E(u) = \frac{1}{N} \sum_{i=1}^N u_i.$$

The test results of correlation coefficients between test images and ciphertext are shown in Tables 4 and 5. It can be observed that the adjacent pixels of the plaintext test image have a strong correlation, while the adjacent pixels of the ciphertext image basically have no correlation.

Figure 7 and 8 show the plaintext and ciphertext images of Lena and Skull in three directions. As can be seen from the figures, the relationship between the adjacent pixels of the plaintext image in each direction is linear, while the relationship between the adjacent pixels of the ciphertext image is relatively discrete, with basically no correlation and good encryption effect.

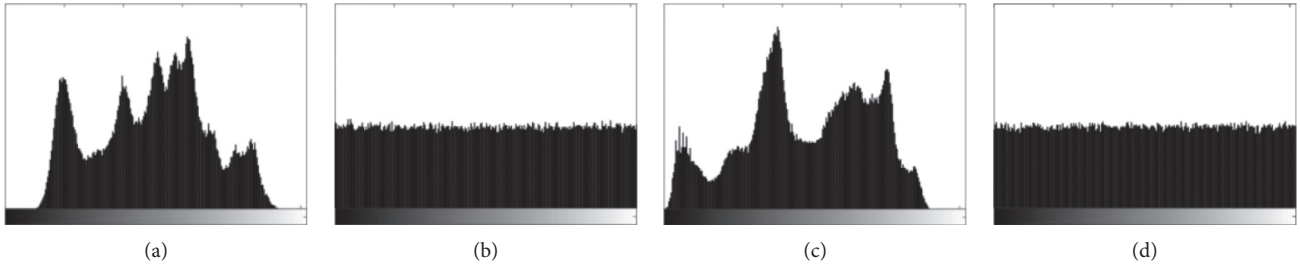


FIGURE 6: Plaintext and ciphertext histograms.

TABLE 2: Comparison of information entropy.

| Method | Lena | Skull |
|----------|--------|--------|
| FRFT | 7.9972 | 7.9994 |
| ASFS | 7.9992 | 7.9995 |
| CENN | 7.9992 | 7.9993 |
| Proposed | 7.9996 | 7.9997 |

TABLE 3: Local IE comparison.

| Image | Size | Plaintext entropy | FRFT | ASFS | CENN | Proposed |
|-------|-----------|-------------------|--------|--------|--------|----------|
| Lena | 256 × 256 | 6.7094 | 7.9031 | 7.9034 | 7.9032 | 7.9021 |
| Skull | 256 × 256 | 7.3118 | 7.9028 | 7.9031 | 7.9029 | 7.9024 |

TABLE 4: The correlation coefficient of adjacent pixels of Lena image

| Direction | Proposed | | FRFT | ASFS | CENN |
|------------|-----------|------------|---------|---------|---------|
| | Plaintext | Ciphertext | | | |
| Horizontal | 0.97423 | -0.0006 | 0.0020 | -0.0003 | -0.0009 |
| Vertical | 0.98592 | 0.0024 | 0.0099 | 0.0105 | 0.0139 |
| Diagonal | 0.96275 | 0.0002 | -0.0049 | 0.0078 | -0.0006 |

TABLE 5: The correlation coefficient of adjacent pixels of Skull image

| Direction | Proposed | | FRFT | ASFS | CENN |
|------------|-----------|------------|---------|---------|---------|
| | Plaintext | Ciphertext | | | |
| Horizontal | 0.97762 | 0.0001 | 0.0018 | 0.0029 | 0.0195 |
| Vertical | 0.97742 | 0.0035 | -0.0098 | -0.0007 | -0.0092 |
| Diagonal | 0.96126 | -0.0026 | 0.0012 | 0.0009 | 0.0165 |

4.4. Robustness Analysis. With the rapid development of computer and password cracking technology, attackers can intercept ciphertext images and add or modify them to attack ciphertext and images, causing interference to decryption. A new encryption algorithm should have strong robustness after encrypting the plaintext image and be able to resist various attacks and decrypt successfully. Clipping attack, noise attack, and JPEG compression are carried out on the ciphertext image. Figures 9(a)–9(c) are clipping 25%, Gaussian noise mean square error 50, JPEG compression, clipped 25% decryption, Gaussian noise decryption, and JPEG compression decryption, respectively. We also conduct data loss and noise attack for Skull image. Figure 10 is the data cut with 64×64 . Figure 11 is the 4% salt and pepper

noise. Figures 12 and 13 are the corresponding decrypted image for Figures 10 and 11, respectively. The results show that our proposed encryption method has good robustness.

4.5. Key Space and Sensitivity. Large key space can resist key exhaustive blasting effectively. According to [5], only when the key space is greater than or equal to 2^{100} can it better provide reliable security guarantee for the algorithm. The proposed algorithm uses seven groups of keys, and each group of keys has a floating point accuracy of 10^{16} . Therefore, the key space is $(10^{16})^7 = 10^{112}$, which is larger than 2^{100} , so it can resist the explosive attack.

For different keys, the decrypted image should not contain any information about the plaintext image, which requires the encryption algorithm to be sensitive to the key. The sensitivity of the key is tested below. Minor changes are made to one of the 7 groups of keys $k_i = k_i \pm \delta$ ($\delta = 10^{-16}$), other keys remain unchanged, and the plaintext image is compared with the mean square error. R is the ciphertext image to be compared:

$$\overline{M} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |I(i, j) - R(i, j)|^2. \quad (25)$$

As shown in Figure 14 and Table 5, Figure 14(a) is the unchanged image and Figures 14(b)–14(h) are the image decrypted with the wrong key including (b) $k_1 + \delta$, (c) $k_2 + \delta$, (d) $k_3 + \delta$, (e) $k_4 + \delta$, (f) $k_5 + \delta$, (g) $k_6 + \delta$, and (h) $k_7 + \delta$. By comparing with the image decrypted correctly in Figure 14(a), it can be seen that the plaintext image cannot be recovered and the information related to plaintext cannot be obtained after minor changes in the key. It can be observed from Table 6 that the mean square error values are above 0.8, which is almost the same as the mean square error values of ciphertext images and plaintext images, and the entropy values are also above 7.99 (close to 8), which proves that the image decrypted with the error key is very different from the plaintext image and further indicates that the new algorithm in this paper has a high key sensitivity.

4.6. Differential Attack Resistance. If the method is more sensitive to plaintext, it is more resistant to differential attacks. The sensitivity to plaintext can be measured by the indexes NPCR [32, 33] and UACI [34, 35]. When there is only one pixel difference between two plaintext images, the ciphertext image obtained after encryption changes greatly.

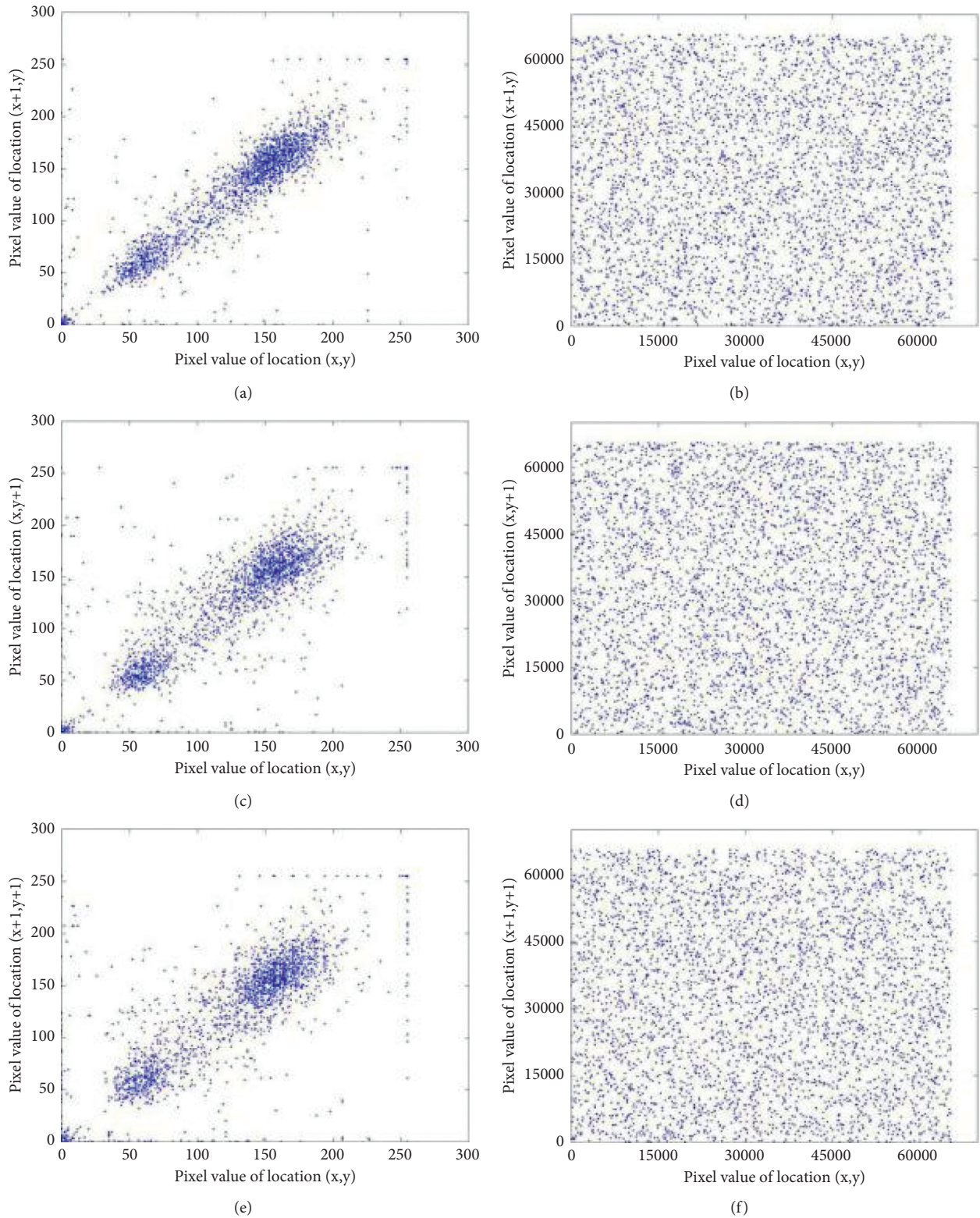


FIGURE 7: Analysis of the correlation coefficient between Lena plaintext and ciphertext.

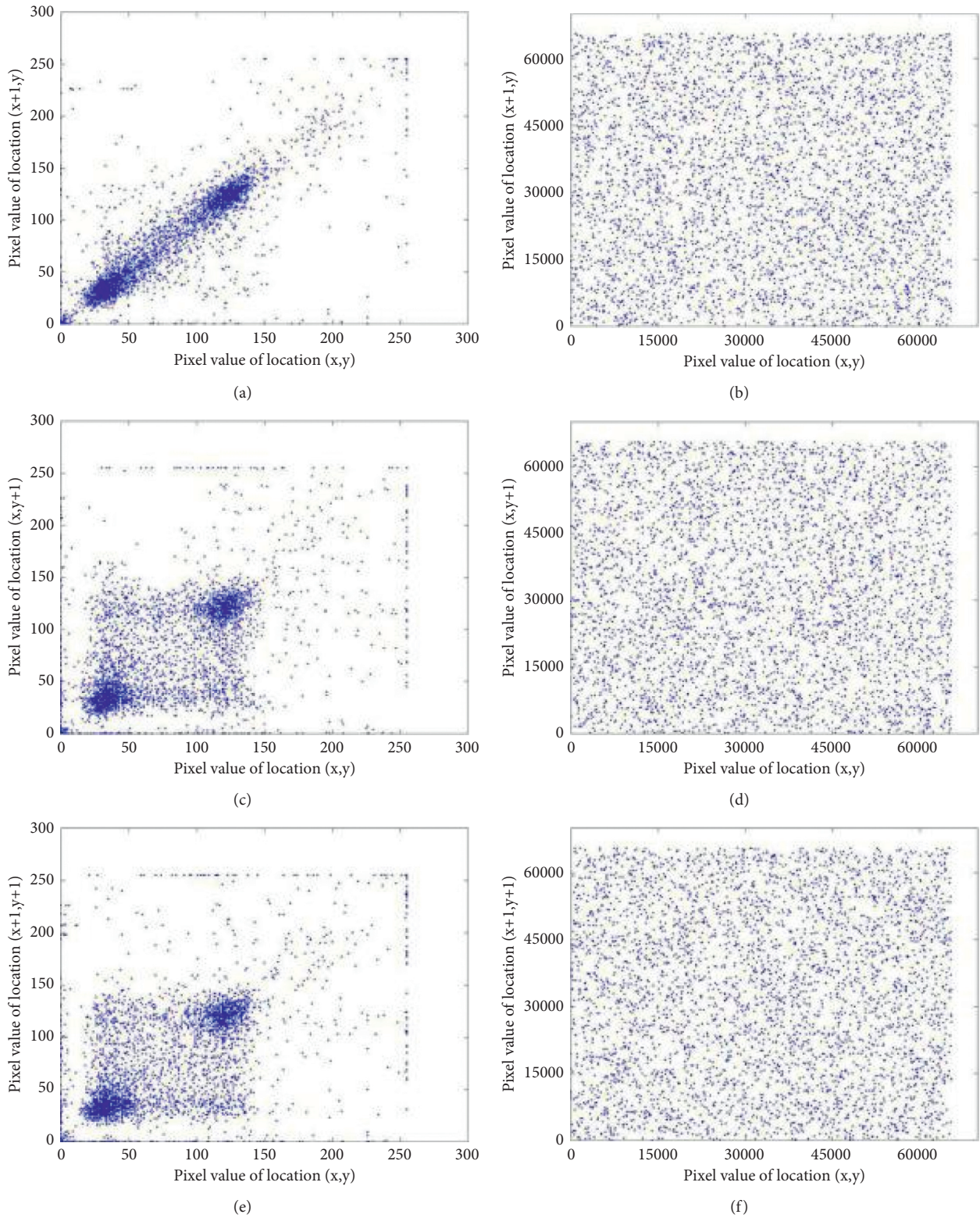


FIGURE 8: Analysis of the correlation coefficient between Skull plaintext and ciphertext.

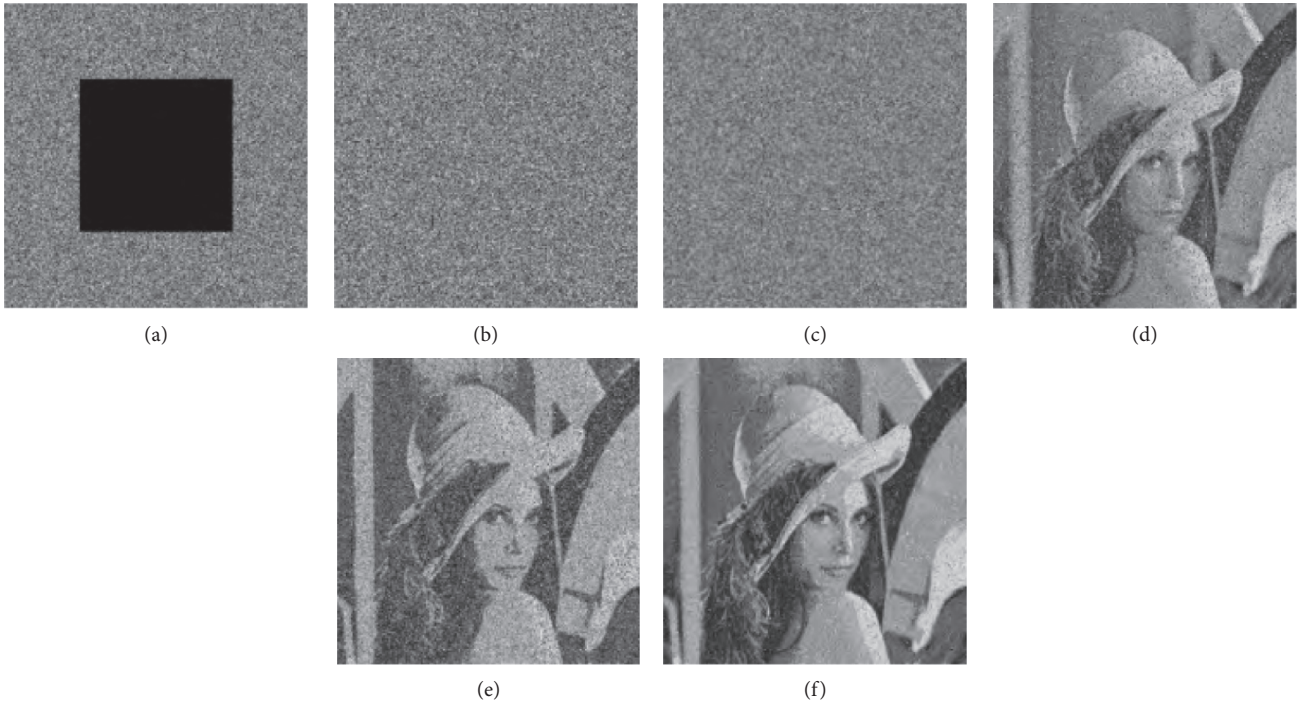


FIGURE 9: Robustness test of Lena image.

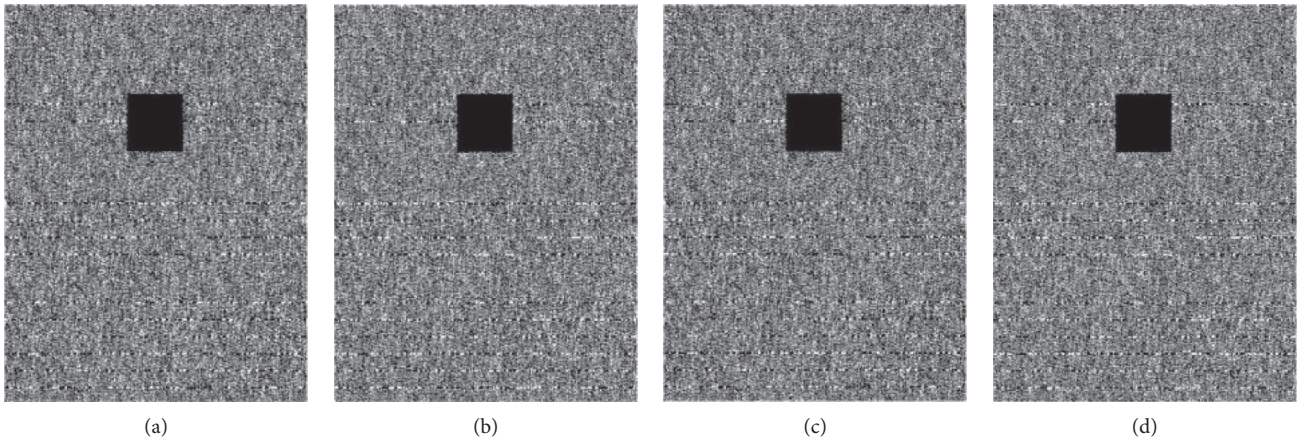


FIGURE 10: The data loss in encrypted images.

Let the pixel of point (i, j) in the two ciphertext images be $u_1(i, j)$ and $u_2(i, j)$; then, NPCR and UACI are calculated as

$$\begin{aligned}
 \text{NPCR} &= \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N D(i, j) \times 100\% \\
 D(i, j) &= \begin{cases} 0, & u_1(i, j) = u_2(i, j) \\ 1, & u_1(i, j) \neq u_2(i, j) \end{cases} \quad \text{UACI} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N \frac{|u_1(x, y) - u_2(i, j)|}{255} \times 100\%.
 \end{aligned}
 \tag{26}$$

$N=99.6094070$ and $U=33.4635070$ are the expected values of the two indicators. In this study, one hundred

groups of Lena images are selected for testing, and each group contains the image with one bit value changed. We

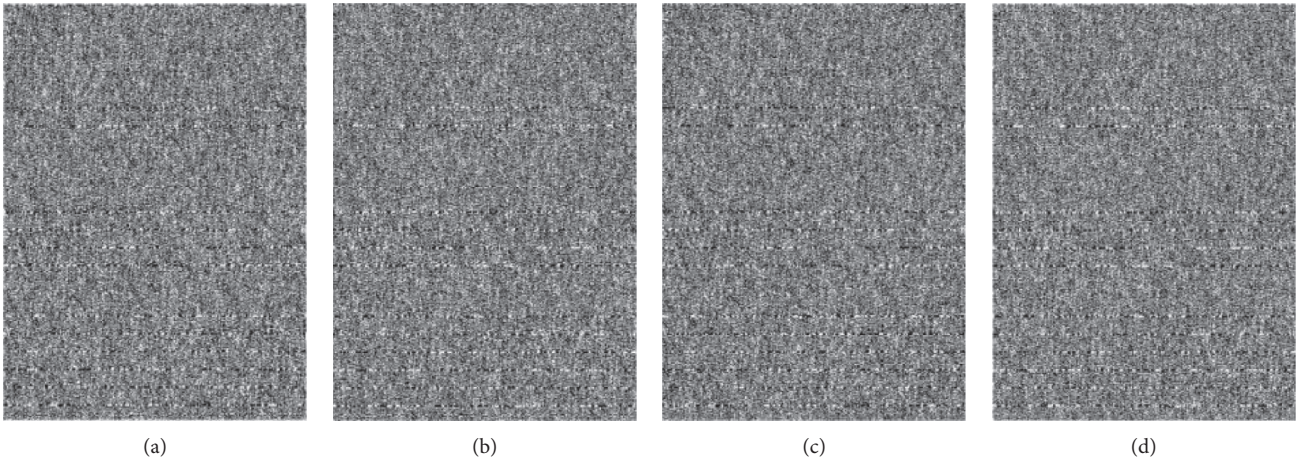


FIGURE 11: The encrypted images with adding 4% salt and pepper noise.

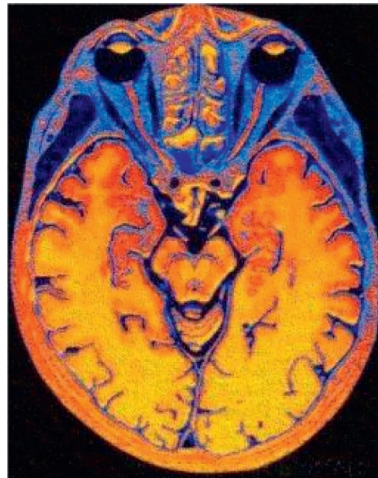


FIGURE 12: Decrypted image with data loss.

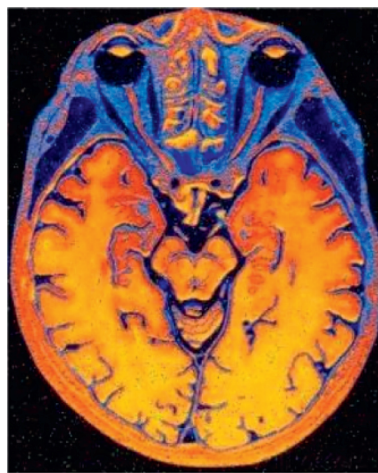


FIGURE 13: Decrypted image with noise.

take the average values of the two indicators, and the test results are shown in Table 7. The obtained NPCR and UACI by the proposed algorithm are closer to the ideal value. And

the algorithm is highly sensitive to plaintext, which can effectively resist differential attack and selective plaintext attack.

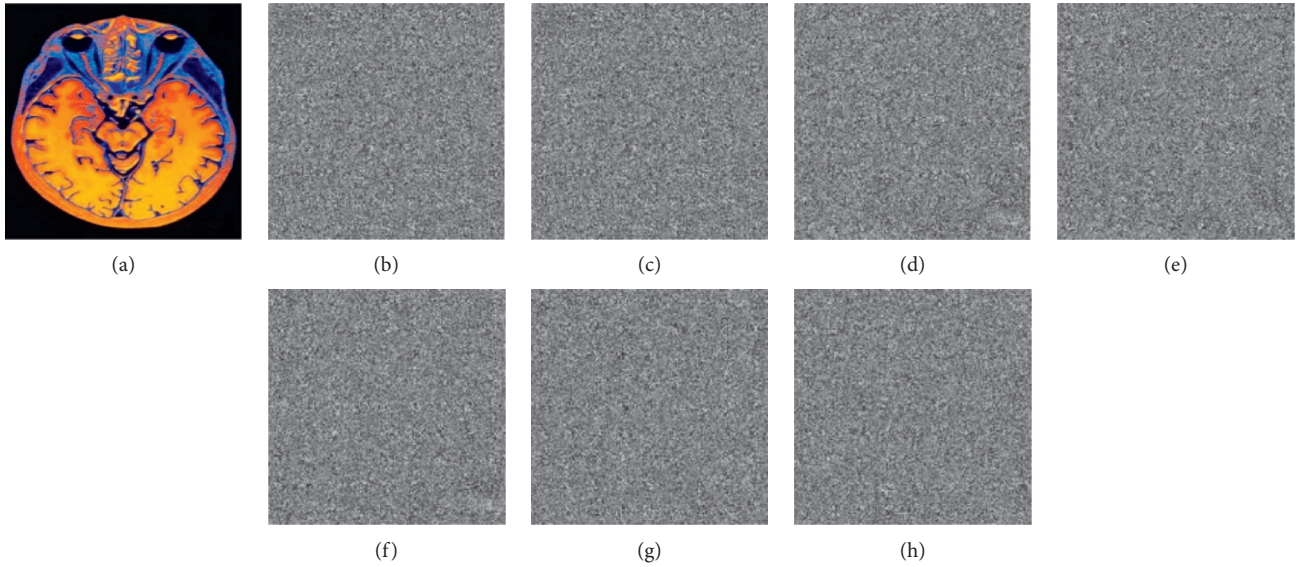


FIGURE 14: Decrypted image with correct key and wrong key.

TABLE 6: Mean square error of decryption and plaintext with correct key and wrong key.

| Key | Global information entropy | Mean square error |
|----------------|----------------------------|-------------------|
| Unchanged | 7.5968 | 0.8428 |
| $k_1 + \delta$ | 7.9993 | 0.8427 |
| $k_2 + \delta$ | 7.9989 | 0.8117 |
| $k_3 + \delta$ | 7.9988 | 0.8073 |
| $k_4 + \delta$ | 7.9991 | 0.8150 |
| $k_5 + \delta$ | 7.9989 | 0.8135 |
| $k_6 + \delta$ | 7.9991 | 0.8353 |
| $k_7 + \delta$ | 7.9991 | 0.8372 |

TABLE 7: Plaintext sensitivity test results.

| Method | NPCR | UACI |
|----------|---------|---------|
| FRFS | 99.5651 | 30.9132 |
| ASFS | 99.5998 | 33.4602 |
| CENN | 99.6188 | 33.4823 |
| Proposed | 99.6102 | 33.4659 |

5. Conclusion

In summary, with the rapid development of computer network, images are mostly presented in the form of digital images, which are not only convenient to save but also fast to transmit. However, at the same time, image information is more likely to be leaked and stolen due to the openness of the network. Therefore, V-Net convolutional neural network is used to improve and optimize the general chaotic encryption algorithm, which has a certain degree of periodicity. Simulation results show that the proposed method improves the encryption effect.

Data Availability

The data that support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This project was supported by Taif University researchers, supporting project no. TURSP-2020/107, Taif University, Taif, Saudi Arabia. The authors would like to acknowledge the Scientific Research Funds of Education Department of Liaoning Province in 2021 (General Project) (LJKZ1311) for its support, which facilitated the publication of this paper.









References

- [1] X. Wang, X. Qin, and C. Liu, "Color image encryption algorithm based on customized globally coupled map lattices," *Multimedia Tools and Applications*, vol. 78, 2019.
- [2] R. Ren, Z. Li, L. Deng, and X. Shan, "Non-orthogonal polarization multiplexed metasurfaces for tri-channel polychromatic image displays and information encryption," *Optics Express*, vol. 10, 2021.
- [3] T. Lin, H. Li, J. Liu, and S. Yin, "An efficient and secure Cipher-Text retrieval scheme based on mixed homomorphic encryption and Multi-Attribute Sorting Method under Cloud Environment," *International Journal on Network Security*, vol. 20, no. 5, pp. 872–878, 2018.
- [4] H. Wen-Wen, Z. Ri-Gui, and J. Shexiang, "Quantum image encryption algorithm based on Arnold scrambling and wavelet transforms," *Quantum Information Processing*, vol. 19, no. 3, pp. 1–29, 2020.
- [5] L. Teng, H. Li, and S. Yin, "Im-MobiShare: An improved privacy preserving scheme based on asymmetric encryption and bloom filter for users location sharing in social network," *Journal of Computers*, vol. 30, no. 3, pp. 59–71, 2019.
- [6] Z. Yong, C. Aiguo, and T. Yingjun, "Plaintext-related image encryption algorithm based on perceptron-like network - ScienceDirect," *Information Sciences*, vol. 526, pp. 180–202, 2020.

- [7] R. Wang, G.-Q. Deng, and X.-F. Duan, "An image encryption scheme based on double chaotic cyclic shift and Josephus problem," *Journal of Information Security and Applications*, vol. 58, no. 2, Article ID 102699, 2021.
- [8] R. G. Zhou and Y. B. Li, "Quantum image encryption based on Lorenz hyper-chaotic system," *International Journal of Quantum Information*, vol. 18, 2020.
- [9] G. Li-Hua, C. Shan, and H. Xiang-Tao, "Quantum image encryption algorithm based on quantum image XOR operations," *International Journal of Theoretical Physics*, vol. 55, 2016.
- [10] Z. H. Gan, X. L. Chai, and D. J. Han, "A chaotic image encryption algorithm based on 3-D bit-plane permutation," *Neural Computing & Applications*, pp. 1–20, 2018.
- [11] G. d. Li and L. I. Wang, "Double chaotic image encryption algorithm based on optimal sequence solution and fractional transform," *The Visual Computer*, vol. 35, no. 9, pp. 1267–1277, 2019.
- [12] S. Yin, H. Li, and L. Teng, "A novel proxy Re-encryption scheme based on identity property and stateless broadcast encryption under cloud environment," *International Journal on Network Security*, vol. 21, no. 5, pp. 797–803, 2019.
- [13] S. Yin, J. Liu, and L. Teng, "Improved elliptic curve cryptography with homomorphic encryption for medical image encryption," *International Journal on Network Security*, vol. 22, no. 3, pp. 419–424, 2020.
- [14] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "CorrAUC: a malicious bot-IoT traffic detection method in IoT network using machine-learning techniques," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3242–3254, 2021.
- [15] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "IoT malicious traffic identification using wrapper-based feature selection mechanisms," *Computers & Security*, vol. 94, Article ID 101863, 2020.
- [16] J. Yang, Z. Wei, and I. Moroz, "Periodic solutions for a four-dimensional hyperchaotic system," *Advances in Difference Equations*, vol. 1, 2020.
- [17] S. Gu, B. Du, and Y. Wan, "A new four-dimensional non-Hamiltonian conservative hyperchaotic system," *International Journal of Bifurcation and Chaos*, vol. 30, 2020.
- [18] I. Jemal, M. A. Haddar, C. Omar, and A. Mahfoudhi, "Performance evaluation of convolutional neural network for web security," *Computer Communications*, vol. 175, pp. 58–67, 2021.
- [19] Z. Wang, X. Huang, and N. Li, "Image encryption based on a delayed fractional-order chaotic logistic system," *Chinese Physics B*, vol. 21, no. 5, pp. 107–112, 2012.
- [20] H. Lai, M. A. Orgun, J. Xiao, J. Pieprzyk, L. Xue, and Y. Yang, "Provably secure three-party key agreement protocol using Chebyshev chaotic maps in the standard model," *Nonlinear Dynamics*, vol. 77, no. 4, pp. 1427–1439, 2014.
- [21] S. S. Jamal, A. Anees, M. Ahmad, M. F. Khan, and I. Hussain, "Construction of cryptographic S-boxes based on mobius transformation and chaotic tent-sine system," *IEEE Access*, vol. 7, Article ID 173273, 2019.
- [22] T. S. Ali and R. Ali, "A novel medical image signcryption scheme using TLTS and Henon chaotic map," *IEEE Access*, vol. 8, Article ID 71974, 2020.
- [23] S. Kanwal, S. Inam, O. Cheikhrouhou, K. Mahnoor, A. Zaguia, and H. Hamam, "Analytic study of a novel color image encryption method based on the chaos system and color codes," *Complexity*, vol. 2021, no. 1, 19 pages, Article ID 5499538, 2021.
- [24] S. Gbashi, P. B. Njobeh, and N. E. Madala, "Parallel validation of a green-solvent extraction method and quantitative estimation of multi-mycotoxins in staple cereals using LC-MS/MS," *Scientific Reports*, vol. 10, no. 1, 2020.
- [25] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," 2016, <https://arxiv.org/abs/1606.04797>.
- [26] C. Zhao, J. H. Keyak, and J. Tang, "ST-V-Net: incorporating shape prior into convolutional neural networks for proximal femur segmentation," *Complex & Intelligent Systems*, 2021.
- [27] Y. Kim, S. Kim, T. Kim, and C. Kim, "CNN-based semantic segmentation using level set loss," in *Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1752–1760, WACV, Waikoloa Village, HI, USA, January, 2019.
- [28] L. Meng, S. Yin, C. Zhao, H. Li, and Y. Sun, "An improved image encryption algorithm based on chaotic mapping and discrete wavelet transform domain," *International Journal on Network Security*, vol. 22, no. 1, pp. 155–160, 2020.
- [29] B. Vaseghi, S. Mobayen, S. S. Hashemi, and A. Fekih, "Fast reaching finite time synchronization approach for chaotic systems with application in medical image encryption," *IEEE Access*, vol. 9, Article ID 25911, 2021.
- [30] C. T. Selvi, J. Amudha, and R. Sudhakar, "Medical image encryption and compression by adaptive sigma filtered synorr certificateless signcryptive Levenshtein entropy-coding-based deep neural learning," *Multimedia Systems*, pp. 1–16, 2021.
- [31] S. J. Sheela, K. V. Suresh, and D. Tandur, "Cellular neural network-based medical image encryption," *SN Computer Science*, vol. 1, no. 6, 2020.
- [32] M. Shafiq, Z. Tian, A. A. Bashir, A. Jolfaei, and X. Yu, "Data mining and machine learning methods for sustainable smart cities traffic classification," *A Survey. Sustainable Cities and Society*, vol. 60, 2020.
- [33] I. Ahmad, T. Rahman, A. Zeb et al., "Analysis of security attacks and taxonomy in underwater wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1444024, 15 pages, 2021.
- [34] M. Shafiq, Z. Tian, Y. Sun, X. Du, and M. Guizani, "Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city," *Future Generation Computer Systems*, vol. 107, pp. 433–442, 2020.
- [35] M. Sajjad, T. Safdar Malik, S. Khurram et al., "Efficient joint key authentication model in E-healthcare," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2739–2753, 2022.

Review Article

Traditional and Hybrid Access Control Models: A Detailed Survey

Muhammad Umar Aftab ^{1,2}, Ali Hamza ¹, Ariyo Oluwasanmi ³, Xuyun Nie ^{2,3},
Muhammad Shahzad Sarfraz ¹, Danish Shehzad ¹, Zhiguang Qin ^{2,3},
and Ammar Rafiq ⁴

¹Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan

²Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu, China

³School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

⁴Department of Computer Science, NFC Institute of Engineering and Fertilizer Research, Faisalabad, Pakistan

Correspondence should be addressed to Xuyun Nie; xynie@uestc.edu.cn and Zhiguang Qin; qinzg@uestc.edu.cn

Received 21 October 2021; Revised 11 December 2021; Accepted 30 December 2021; Published 7 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Muhammad Umar Aftab et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Access control mechanisms define the level of access to the resources among specified users. It distinguishes the users as authorized or unauthorized based on appropriate policies. Several traditional and hybrid access control models have been proposed in previous researches over the last few decades. In this study, we provide a detailed survey of access control models and compare the traditional and hybrid access control models based on their access control criteria. This survey focuses on the growing literature of access control models and summarizes it through comparative analysis, identifying limitations and illustrating the advantages of both traditional and hybrid models. This study will help the researchers to get a deep understanding of the traditional and hybrid access control models.

1. Introduction

Information is the most important asset of any organization that must be secure. The security of information can be ensured with the help of confidentiality, integrity, and availability [1, 2]. Furthermore, an organization's information can be secured with different approaches or technologies such as intrusion detection, steganography, cryptography, and access control [3–5]. These approaches are used according to the goal and objective of the information and organization.

Access control (AC) is one of the best approaches that is used to secure the information from inside and outside attacks of the organization and decisions of granting and revoking access to any user [6]. The access control gives access to those who are authorized to organizations, i.e.,

persons, processes, and systems. The access control models define its mechanisms and security policies first, and then, these models are implemented in organizations according to goals and objectives [7]. There are several traditional and hybrid access control models that have various pros and cons.

The traditional access control models are discretionary access control (DAC), mandatory access control (MAC), role-based access control (RBAC), and attribute-based access control (ABAC). In the DAC model, the owner of the object has the authority to give and deny access to others without a system administrator mechanism [5]. The DAC model is divided into two types: strict DAC and liberal DAC. In the strict DAC model, only the owner has the authority to permit and deny access to created resources, but in the liberal DAC model, the authority of the owner can be

transferred to another individual who will be able to permit and deny the access. In the MAC model, the centralized mechanism is used to permit and deny the access of resources to users [8]. The MAC model is more secure, flexible, and efficient for commercial and military use due to its centralized behaviour. The RBAC model is prominent due to the least privilege and tight security that makes it more powerful than all other models [9]. The ABAC model has dynamic behaviour that is the most suitable model for changing environments [10]. There are some disadvantages of both RBAC and ABAC models. So, researchers proposed some hybrid models as an extension of RBAC and ABAC.

The existing surveys on access control provide a review of basic access control models, i.e., MAC, DAC, RBAC, and ABAC, or focus on access control trends, i.e., IoT, cloud, and fog computing, but there is no comprehensive survey that explains advanced access control models with their framework and applications along with pros and cons. So, this study presents access control models and advanced hybrid access control models with their framework and applications in a comprehensive manner. The access control models are used in small and large organizations according to the pros and cons of the model and the requirements of the organization. This survey encourages the researchers to propose new hybrid access control models according to the problem.

There are some existing survey studies on access control models that tried to explain access control policies with few models in specific contexts, i.e., IoT, cloud, and fog computing. Bertin et al. [11] conduct a survey paper that explains the basic access control model in detail, but this study does not include advanced hybrid access control models. The studies [12, 13] conduct surveys that focus on IoT security and challenges, and they proposed solutions based on a trust-based access control model. Zhang et al. [14] present a survey paper that explains some access control models and trusted system computing in the IoT domain. The author proposed a novel method for IoT that includes access control, network attack, and trusted computing, but this study does not explain the applications, limitations, pros, and cons of each model.

The rest of the study is organized as follows and also described in Figure 1. The second section describes the access control and its traditional and hybrid models. The third section makes comparisons of the access control models, and the fourth section concludes the study.

2. Access Control

The access control (AC) mechanism is used to permit or deny the access of resources within the organization to secure the data [6]. The AC permits the access of resources only to authorized personnel of the organization and denies the access of resources to unauthorized and other users. The access control is normally consisting of identification, authentication, and authorization. The access control grants access to authorized users according to user privilege level after authentication [15]. The access control is classified into traditional and hybrid models as shown in Figure 2. The traditional access control is further divided into four types:

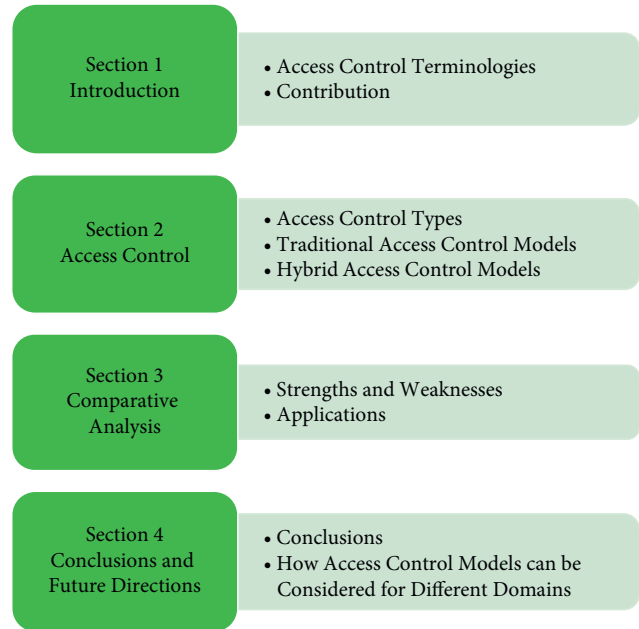


FIGURE 1: Organization of the work.

MAC, DAC, RBAC, and ABAC. The hybrid access control has also several types. Each traditional and hybrid access control model has its pros and cons. So, organizations use access control models according to their objectives and goals.

2.1. Traditional Access Control Models. There are different traditional models of access control, i.e., MAC, DAC, RBAC, and ABAC. Each model has its pros and cons. The traditional access control models are classified into two categories: DAC and non-DAC. The non-DAC is further divided into MAC, RBAC, and ABAC [16]. The traditional access control models are also compared with each other based on criteria; the principle of least privilege, dynamic behaviour, safety of models, separation of duties, capability delegation, configuration flexibility, and auditing as shown in Table 1.

2.1.1. Discretionary Access Control (DAC). The DAC is a model that allows owner-based access where the owner is the creator of a resource or object. The owner of the object decides the access granting or revoking policy for the subjects or users as shown in Figure 3. In this manner, there is no need for the administrator to provide its services regarding access rights. DAC is divided into two different types: liberal and strict DAC. According to the liberal DAC, the owner can transfer the access rights or ownership to other individuals so that they can also work as an owner of the resource. On the contrary, the access rights are limited to the owner of the resource, and ownership is restricted for that individual, in the strict DAC [17, 18]. It can be assumed that the DAC model works according to the choice or discretion of the owner. The enforcement of access control policies is made on three different categories: resource

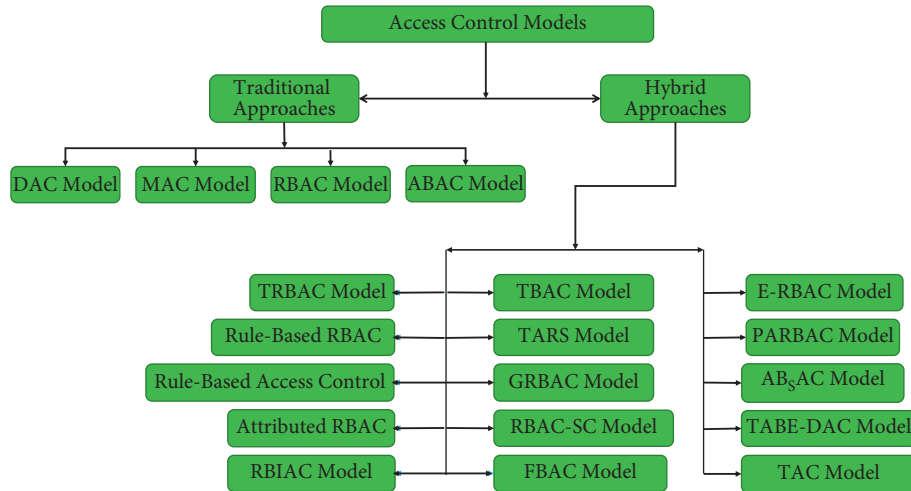


FIGURE 2: Types of traditional and hybrid access control.

TABLE 1: Comparison of traditional access control models.

| Criteria | DAC | MAC | RBAC | ABAC |
|------------------------------|-----|-----|------|------|
| Principle of least privilege | X | X | ✓ | ✓ |
| Dynamic behaviour | X | X | X | ✓ |
| Safety of models | X | ✓ | ✓ | ✓ |
| Separation of duties | X | X | ✓ | ✓ |
| Capability delegation | ✓ | X | X | X |
| Configuration flexibility | ✓ | X | ✓ | X |
| Auditing | ✓ | ✓ | ✓ | ✓ |

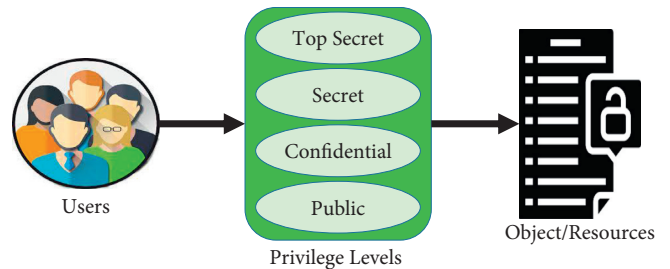


FIGURE 4: Abstract view of mandatory access control (MAC).

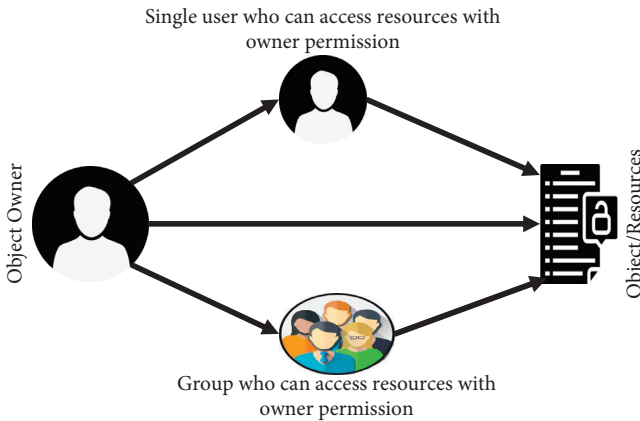


FIGURE 3: Abstract view of discretionary access control (DAC).

ownership, user identities, and permission delegation. DAC is not an appropriate model for commercial and government organizations due to the deficiencies or limitations because it allows the users to set or deploy the access rights that might lead it towards Trojan horse attacks [19]. Moreover, DAC is popular due to its integration quality with different types of computer systems.

2.1.2. Mandatory Access Control (MAC). MAC works on the basis of security labels that can be either taken as a hierarchy model. It controls the access rights of users or processes against the resources of the system. The users are assigned to various

security levels, while the objects are assigned to security labels as shown in Figure 4. The user access is affiliated with the security levels of resources that are equal or lower than their hierarchy [20]. The access control rights are strictly controlled by the administrator, who can also set the permissions in the access control. MAC is effectively used for military and commercial systems due to its high-level security [21,22]. There are some limitations of MAC such as difficult to manage the MAC systems because the system puts all burden on the administrator to set permissions, manage configurations, and future maintenance. This complexity may increase as the size of the system increases [23]. Furthermore, the MAC operating systems are costly to set up and hard to operate due to the dependence on the trusted parts [24].

2.1.3. Role-Based Access Control (RBAC). The RBAC model made a revolutionary change in the field of access control due to its strictness and tight security. This model is based on five different entities: objects, actions, permissions, roles, and users, as shown in Figure 5. Objects are considered as the resources such as directories, files, or folders. In addition, actions are the tasks or operations that can be performed on the objects. The examples of the actions are write, edit, and delete. The permissions are the combined form of an object and action; such one permission can be considered as “Edit (action) and File.doc (object).” Any change in the action or object will be considered as new permission. The intermediate and one of the key entity of RBAC is the role that connects users and permissions. The

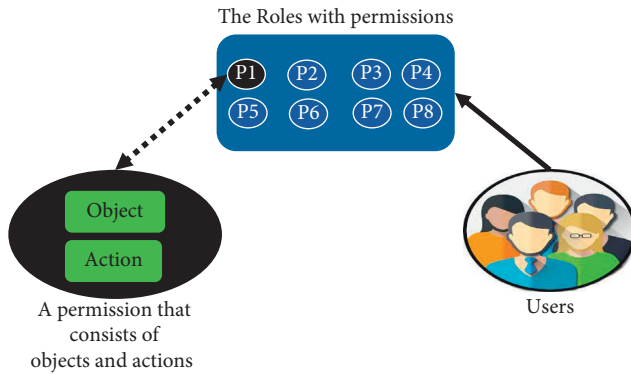


FIGURE 5: Abstract view of role-based access control (RBAC).

roles are the containers that have various permissions. For example, a role named “deputy manager” contains all necessary permissions to fulfil or perform the tasks of the deputy manager. Furthermore, the roles are assigned to users according to their designated positions. After assigning the roles, the permissions inside every role are automatically assigned to users [25]. RBAC provides the least privilege with the usage of roles that is the central entity between users and permissions. In this way, RBAC is not allowing users to deal with the permissions directly and it eradicates the ownership rights. So, it behaves significantly better as compared to DAC because the ownership rights of a resource owner may lead to a Trojan horse attack [26]. RBAC implements the least privileges using the concept of roles because a user can only access those permissions that are assigned to the role, not more than that. This is one of the reasons that makes it popular. On the contrary, RBAC puts a lot of burden on the administrator by managing all the tasks related to permission creation, permission and user assignment to roles, role designing, etc. As the size of the organization increases, the workload of the administrator will also increase [27]. RBAC also violates the rules of separation of duties provided in the NIST standard [25]. The violations are discussed in detail by some researchers [9, 26, 28].

(1) *RBAC Model Components.* The RBAC model is most suitable for healthcare centers and especially for the hospital to make sure the security features of all the records and information details of a patient [29]. Interestingly, RBAC is implemented in the dialysis department for kidney disease due to flexibility and security. The sessions are used to connect users. A user may have more than one session at one time. The RBAC model is classified into three components or modules: core RBAC, hierarchical RBAC, and constrained RBAC. The constrained component of the RBAC model is further divided into two parts: dynamic separation of duty (DSD) and static separation of duty (SSD). The main reason behind this tight security is the implementation of dynamic and static separation of duty [25].

(i) *Core RBAC.* The core RBAC is an essential and fundamental component of the RBAC model that is implemented first in any organization, and then, advanced components of the RBAC model are considered to implement [30]. A user is described as a person, and the role of user denotes functionality and authority. The permission represents a permit

to do any operation on more than one object. The permission can be read and write. The object is anything that is holding some information or receiving information. The object can be a row, table, directory, view, or file. Also, the object might be CPU cycles, printer, or disk storage space.

The main concern of the core RBAC model is to assign users and permissions to roles in many-to-many fashions. It is possible to assign one role to one or more users and vice versa. It is also possible to assign permission to one or more roles and vice versa. There is a lack of research on permission, roles, and their relation. Some authors proposed the symmetric RBAC model that applies constraints on permissions using role hierarchies and separation of duty (SOD) [25].

(ii) *Hierarchical RBAC.* The hierarchical RBAC is the second component of the RBAC model that is constructed on the basis of core RBAC component [30]. The roles are implemented using the role hierarchy (RH) concept that is based on the firm’s authoritative structure [31]. In RBAC, the roles faced some common standard permission again and again, which is not a better choice. The RH is used to link the same permission so that the security admin can face the same permission in few roles. Hence, every role will contribute common permissions and will lie in RH [32]. There are some roles that standalone separately with the RH approach.

In role inheritance (RI), all permissions of juniors can be assigned to senior roles and junior roles cannot have permission as having senior roles. The system cannot manage the situation when junior needs to access the permissions of senior role. The security admin has to permit and deny the same permissions again and again without RI that is a very hard job. This thing needs to be a hierarchy feature in the form of a tree with respect to different categories such as a senior, junior, junior-most, and senior-most. The role inheritance is the best choice for such type of situation; from one side, a role may inherit some permission, and on the other side, another role can inherit some permission [33].

(iii) *Constrained RBAC.* The constrained RBAC has some specific constraints along separation of duty (SOD) to implement. These constraints can be either location-based or time-based. The main theme of these types of constraints is to grant access based on specific time slots and locations. The RBAC constraints enable RBAC with the implementation of information security, which protects the whole system from both external and internal threats. Same as RBAC, the safety conditions are confirmed for access control models [34].

2.1.4. *Attribute-Based Access Control (ABAC).* ABAC is a model that is capable to provide fine-grained access control, flexibility, and dynamicity. The main story revolves around

the attributes allocated by the attribute authorities. The Boolean formula is used to define an access control policy using the set of attributes so that an authorized and valid access can happen. There is no need to create and assign numerous roles. Moreover, there is no need to make or design access control lists for everyone in the organization [35, 36].

The attributes provide the facility to automatically perform access control decisions. Examples of attributes are citizenship, IP address, identity, location, and user-name. ABAC works on the evaluation rules of the attributed entities such as objects and subjects, environment related to a request, and operations. If the attributes, as well as attribute values, match, then the access is granted to a user; otherwise, access is denied [37, 38]. The benefit of this facility is the dynamic behaviour as shown in Figure 6. In this manner, any change in the attribute values or user identities will be dynamically detected and the decision has been made. Previously, the RBAC model was unable to deal with this issue. On the other hand, the ABAC model has complexity issues. If the number of the attributes increases, then the complexity of the system will also increase [27, 28].

Figure 6 shows that each subject and object has its own attribute. The attribute-based access control allows the subject to access objects by checking attributes. In Figure 6, *desig*, *locat*, *categ*, and *AR* stand for designation, location, category, and access rules, respectively.

The user of system will define as subject by the administrator to access the file management system. The characteristics of user will capture as subject attributes. The attributes of subject can be name, designation, organizational affiliation, gender, age, nationality, or security clearance. The identity information of subject is maintained by administrator or authorities in file management system. The proper management and assignment of subject attributes on a regular basis are required as member leave or joins the organization on a regular term [39].

The required functionality of ABAC is based on device policy, documents, or procedural rules on which a business operates. The object may have a policy or rule on which it allows access to the subject. For example, only physician is permitted to access the patient record or information for treatment and prescription in a medical emergency setup. The nonmedical person is not allowed to access the information recorded in the file of a patient. This case also defines access privileges for a specific subject [40].

The ABAC protects the objects as object, subject, attributes, and policies are defined. The access control method gathers information related to the subject, object, and policy to render the logical decision for the execution of the requested operation. Access control mechanism (ACM) must be smart enough to recognize information, policy, attributes, and their chronology and source along with necessary computations for decision-making [41].

The policies related to ABAC depend upon the richness of computational languages and the degree to which attributes are available. The system is flexible when subjects can access more objects. A subject can have

maximum access to maximum objects and can perform a number of operations on the object under the established policies or rules. It is not required to create a new additional role in the system with new members because a new member shares the same attributes that are already defined. For example, a nurse wants to access patient information in medical emergency, and there is no need to set a new rule set or policy as it shares the same attributes defined earlier.

The four basic access control models are compared with each other on the basis of parameters, i.e., least privilege, dynamic behaviour, safety, separation of duties, capability delegation, configuration flexibility, and auditing as shown in Table 1. The principle of least privilege means that the user should have access to only the necessary resources when needed to do a specific operation or task. The dynamic behaviour means that the operations and tasks should be performed automatically using different access rules rather than manual instructions. The safety of models means preventing permission leakage of access control models from unauthorized users. The separation of duties means permitting the access of resources only to authorize users and denying the access request of unauthorized users. The capability delegation means the ability of a user to revoke their own features to other users that have already been granted. Configuration flexibility means providing an easy way to users for installation and uninstallation like the wizard menu. Auditing means monitoring the access control model by recording requests from users.

2.2. Hybrid Access Control Models. In this section, we explained various hybrid models that are extensions of traditional AC models.

2.2.1. Temporal Role-Based Access Control (TRBAC). The TRBAC [42] is an advanced form of the RBAC model that eliminates non-permanent limitations on the on/off switching of roles. The TRBAC braces up seasonal role enabling and disabling and transitive dependencies on those types of activities. Those forms of dependencies that are stated using role triggers can also be utilized to limit the series of roles that a specific user can make operative at a particular period. The release of a trigger can lead to the switching on or off of a role that can happen instantly or after a specified period of time. The enabling and disabling activities can be assigned for resolving disputes, for instance, the constant switching on and off of a role. In this case, the activity that has the highest assigned priority will always be performed [43].

To enhance the capacity of the security officer (SO) to react in emergency circumstances, the authors give the access to manipulate the state of role and the series of users that have the control to perform that specific role by giving run time requests [44]. The run time requests are those requests that are not attached with other events or the validation of stated conditions. For example, a run time request can be used to temporarily delay the user from making a role operative. This is useful, especially when a user

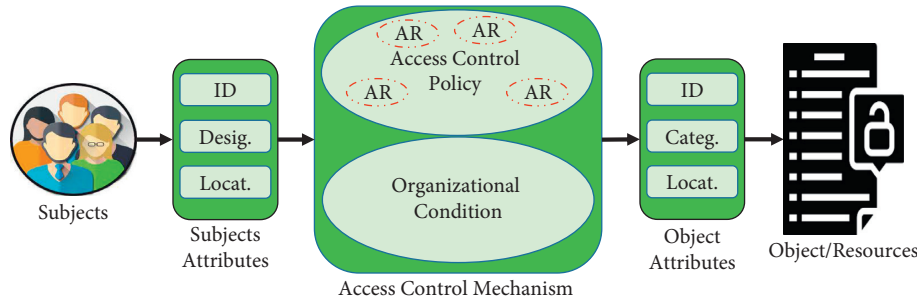


FIGURE 6: Abstract view of attribute-based access control (ABAC).

utilizes a specific role to execute an activity that could be detrimental to the system. In this situation, the SO can react by releasing a run time request that will cause a temporary denial for the user and prevent him from executing the role. Just like triggers, run time requests can be performed immediately or after a specified period of time.

2.2.2. Rule-Based RBAC. The rule-based RBAC is basically a modification in RBAC. Kahtani and Sandhu [45] proposed a model that works like the traditional RBAC model. They made a different set of rules for the enterprise to define its access policy. The rules are activated automatically for the assignment of users to roles. The permission creation and assignment of the permissions to roles are working the same as the traditional RBAC model. The modification was done in between user role assignments. The authors made the user role assignment portion dynamic. The system will verify the attributes of the users with attributes of roles. If attributes on both ends match with their attribute values, then the assignment will be done automatically, otherwise not. For example, a user from country of India, with age of 19, can view the adult sites. It means a user should qualify the attributes of age and country, with the values of their attributes; then, he/she can access those particular roles with the same access rule. The working of rule-based RBAC model is very good because it decreased a load of an administrator by automating the concept of user role assignment. The efficiency of the model can be increased by giving the idea of a fully dynamic RBAC model that can make reliability and ease of management [46].

2.2.3. Rule-Based Access Control. The rule-based access control model is used for Web-based social network (WBSN). It permits access to resources that are located online. In this framework, authorized subjects are expressed based on the relationship form, depth, and degree of trust that exist among the network users with attribute-based RBAC. Access to resources is given based on distinct access rules. In rule-based models, protocols are given by resource owners and they indicate the profile of authorized users by one or more access conditions. The access conditions include limitations on the type, depth, and trust level of their associations with other network users. The access control needs a particular object that can be clearly stated by a series of conditions [47].

For instance, for an object created by v_o (node that has a relationship with requester), the series of access conditions applicable to the object is given by an access rule that is determined by v_o . This type of concept is usually described as follows: the access rule is always in the form of (oid, cset), where oid represents the identifier of the object and cset represents a series of conditions (cond1..., condn). For instance, assume that Tom is the one who created an object that is associated with the identifier obj1 and he wants users who are his direct pals and whose trust level is up to 0.5 to have access to his object. Also, he wishes to give access to all his direct friends that are his colleagues provided that their trust level is up to 0.5.

2.2.4. Attributed RBAC. The RBAC model is famous due to its strictness in terms of security, and the ABAC model is famous due to its dynamic behaviour [27, 48]. Some studies proposed a hybrid model that used basic entities of RBAC such as actions, objects, permissions, roles, and users. They introduced the concept of attributes for the creation of permissions, permission assignment to roles, and role assignment to users. This sort of addition makes the RBAC model a dynamic model. Most of the work in the hybrid model is done automatically, which made it different from the existing models and covered some of the deficiencies of the RBAC and ABAC. All the objects of the system have some attributes such as time, IP address, and location. These attributes of objects are automatically granted to permissions after the creation permissions.

This model also creates permissions automatically with the merger of object containers and action-level containers. So, this kind of merger creates more than one permission at a time and creates it automatically. After that, the permission is assigned to roles by matching their attributes. If the attributes of roles and permissions are matched, the permissions will be added to those roles automatically. In last, the user's attributes are matched with roles and automatic role assignment will be done with the help of attributes. If a user's time, location, and IP address matched with the role's same attributes, then that user can access that particular role. If one of the attributes does not match, then the user cannot access that role. The model idea was good, but it only supports the basic working of RBAC. If the administrator wants to do the whole access control working through this model, then the model is not useful. The reason is that this model does not support conflicts

of interest, separation of duty, and role hierarchy concepts [49]. So, these are some limitations of attributed RBAC model. Some authors proposed various models to resolve this issue by extending this work. The proposed techniques are capable to support separation of duty in various ways. Furthermore, the hybrid models proposed different methods to generate permissions [26, 28].

2.2.5. Role-Based Integrated Access Control (RBIAC). Reliability and security are the most important concerns in multi-domain service-based systems, where data are used to flow from one domain to another domain. There are many access control models. The data provenance methods are developed for service-based systems. On the other hand, there was not a single mechanism that provides an integrated model with data provenance and access control. The role-based data provenance scheme was developed to track originator's and contributor's roles. Moreover, data reliability can be evaluated using the information of data objects from the roles. The proposed [50] model is better for the applications of multi-domain services with respect to reliability and security. This model provided a new way in the field of integrated or hybrid models. In addition, RBAC is used for the evaluation of data security and reliability. Moreover, the extended version of typical RBAC is used to control data usage and flow of information in multi-domain systems. The developed model is also capable of using information about newly added roles and implementing data quality derivation [51].

2.2.6. Trust-Based Access Control (TBAC). The threat level is comparatively more when users interact with online social networks (OSN). Several users download and upload data from the OSN that may lead to different data security risks and access control. The trust-based access control was proposed as a solution or strategy for users and their friends for restricting them through a proper trust rule in accessing the data from OSN. The proposed [52] model works on the concept of roles such as the owner, contributor, and stakeholder. These roles are associated with users to play during the usage of OSN. There are different security levels introduced with the help of different roles. The concept of a multi-role environment is also introduced. In this way, more than one security parameter can be applied by the users. The user and his friends can make the decision of access grant or revoke for the other users on the OSN. So, policy conflicts do not occur between various users. The model was proposed for the OSN, but it is not suitable for other fields such as wireless sensor networks, IoT, and cloud computing. Moreover, the access decision is placed between users and their friends, but there is no role of the administrator that can make sure security issues. If the administrator wants to delete some unethical photographs or material, then how can an administrator remove it? Even the role of the administrator is not discussed, and this is a question or research gap in this model [53].

2.2.7. Trust-Aware RBAC. During the communication process, there are certain threats in breaching the security from the malicious users. The reason behind the threat is the absence of some access control mechanism. The trust-aware RBAC system (TARAS) [54] model was proposed to solve the security issues in IoT devices communication. The users with similar roles are considered to respond in the same manner so that a trust level can be established between IoT and smart devices, and users. The TARAS is capable of detecting unauthorized and malicious users. Moreover, TARAS performed dynamic trust estimation and increased the integrity of data. The TARAS also increased the availability, detection of accuracy, robustness, and provided better performance under high attack density. The model is specifically designed for IoT, but the model can be implemented only for wireless sensor networks and cloud computing devices. In addition, some researches are proposed regarding the privacy of IoT environments for cloud and blockchain [55, 56].

2.2.8. Garbled RBAC. Data outsourcing originates different security issues in the cloud and IoT environment. Moreover, security threats and privacy risks are leading problems in the fields of military, health care, and intelligent organizations that are associated with the task assignment. As a solution to the problems, the garbled RBAC (GRBAC) [57] model was proposed. The model is a fine-grained security model that adopted a garbled function. The proposed model is specifically designed for those organizations where roles are not disclosed with the servers and for the users. Moreover, the main contributions of the model are that a user cannot activate more than one garbled role set. The data of organization is secret from everyone, but the algorithm is not secret. The model can be implemented in the IoT environment as an extension. On the other hand, the model is not flexible. Moreover, one more disadvantage is restricting the server from the user's roles. In this way, the server is unable to keep the record of roles and the server cannot make the necessary steps for controlling the access control system.

2.2.9. RBAC Using Smart Contract. The open blockchain platform Ethereum provides flexibility, adaptability, and security. In this model, smart contract is used with the typical RBAC model. The RBAC smart contract (RBAC-SC) [58] model is proposed to verify users' role ownership in small organizations. In this model, RBAC-SC is deployed on Ethereum's testnet blockchain and the design of RBAC-SC is also provided with performance analysis. The proposed model is efficient, secure, and minimizes the costs, but it is only suitable for small organizations. In this way, we cannot consider this model for large organizations. This is the drawback and limitation of the model; that is, it is restricted to small organizations only. Some other authors also proposed a lightweight technique for blockchain-based systems for the authentication process [59].

2.2.10. Feasible Fuzzy-Extended ABAC (FBAC). The ABAC model is becoming a mature model day by day, and it is famous due to the dynamic authorization technique. The ABAC model can even dynamically perform in complex environments, but it is unable to provide flexible, exceptional approval. The limitation of ABAC model is that it is unable to perform efficiently resource usability and business timeliness. The proposed FBAC [60] model is comparatively efficient and flexible for granting exceptional critical authorization. The FBAC model is better by increasing the utilization of resources and business suitability. The FBAC is also tested for the audit mechanism and the credit system at high-risk requests. Moreover, the proposed model is analysed for risks, usability, and evaluated for its effectiveness by different experiments. The FBAC model is comparatively better than the traditional ABAC model due to its time efficiency and flexibility. On the other hand, the model is the extended version of ABAC, and it is unable to provide tight security and least privilege.

2.2.11. Emergency Role-Based Access Control (E-RBAC). Nazerian [61] proposed the emergency role-based access control (E-RBAC) model to increase the flexibility of RBAC model in emergency situations. Because the RBAC model is failed to achieve better results in emergency situations. The proposed E-RBAC model is based on break the glass (BTG) policy and separation of duty (SOD) constraint. The BTG policy was proposed to override access control and give maximum responsibility to users, and SOD constraints are used to restrict the users. The proposed E-RBAC model can achieve better results in normal, emergency, and exception situations. The normal situation is the same as RBAC in which the access of user is known. In the emergency situation, the events are predictable except their time and access are not given to users due to privilege contradicts. In an exceptional situation, the user access is unknown and policies are not predefined. This model improves the flexibility of RBAC model in normal, emergency, and exception situations.

2.2.12. Priority-Attribute-Based RBAC (PARBAC). Thakare [62] proposed a priority-attribute-based RBAC (PARBAC) model for medical based on authentication mechanism to increase the consistency and flexibility of RBAC model. Because the RBAC model is failed to handle large number of requests from user in large organizations that cause overloading on the cloud server, the proposed PARBAC works in seven steps. In the first step, the users get token that consists of individual's details. In the second step, user calls to API. In the third step, the Azure resource manager (ARM) accepts or denies assignments of users based on priority. In the fourth step, ARM

advises to user based on role assignment. In the fifth step, ARM verifies the activity and privileges of users. In the sixth step, logging is not allowed to user if he has no role with activity. In the last step, access is blocked if a denial assignment is applied. This PARBAC model is able to handle problems in large organizations with dynamic scenarios.

2.2.13. Attribute-Based Access Control Model Supporting Anonymous Access (ABSAC). Zhang [63] proposed attribute-based access control model supporting anonymous access (AB_sAC) model that is used to protect user data for Internet of things (IoT) in small cities. The models of attribute-based access control (ABAC) are not protected and efficient to work in large organizations properly. According to researcher, anonymous access is able to protect user data and it is not stored in authentic place. This proposed model is more secure for the transaction of user data in public place with minimum risk factors.

2.2.14. Traceable Attribute-Based Encryption Scheme with Dynamic Access Control (TABE-DAC). Guo [64] proposed an efficient traceable attribute-based encryption scheme with dynamic access control (TABE-DAC) model to share secret data on cloud servers based on blockchain technology. The confidentiality of secret data can be protected using attribute-based encryption (ABE), but the ABE scheme is not flexible and efficient to fulfil access control policies. The TABE-DAC model can control illegal sharing of secret data on cloud by tracing malicious users using accountability method. This model provides flexibility to data owners to modify access control policy. The proposed TABE-DAC model is efficient and flexible to share secret data on cloud without illegal sharing.

2.2.15. Time-Based Access Control. Wang [65] proposed time-based access control (TAC) model to secure user data in Internet of things (IoT). The user data are divided into two directional subspaces that represent attribute and time generation of data. Access control and privacy are achieved by sending encrypted data before transmission. The data owner or data source has authority to give access to anyone using sub-key. The TAC model is able to generate sub-key of data within minimum time and memory space for each subspace. The proposed TAC model is efficient and flexible to share secret data on IoT.

3. Comparative Analysis of Traditional and Hybrid Access Control Models

This section contains a summarized comparison and information of traditional and hybrid AC models in tabular form as shown in Table 2.

TABLE 2: Comparative analysis of traditional and hybrid access control models.

| Model name | Strengths | Weaknesses |
|------------------------------|---|--|
| DAC [18] | Flexible to implement, customize access policies, and read/write access for users | Lack of security, no dynamicity, access management is not centralized and inefficient for government use |
| MAC [24] | Confidentiality, data protection, suitable for military use, and strictly enforced by OS | No dynamic alteration in policies, not user-friendly, load on administrator, maintainability, and scalability |
| RBAC [25] | Easy administration, least privileges, best for local domains, and tight security | Mobility problem, no flexibility, role explosion, and no dynamic behaviour |
| ABAC [37] | Flexible for big systems, dynamic behaviour, and global agreement of attributes | Complex implementation and maintenance, time constraints to define attributes, and difficult policy specification |
| TRBAC [42] | Periodic role enable/disable and temporal dependencies for actions | No dynamic behaviour, role explosion, and limited specification language |
| Rule RBAC [45] | Dynamicity, less load on administrator, and induced role hierarchy | Not consistent for conflicts of interest and policy specification complexity |
| Rule BAC [47] | Use certificates for authenticity, good for social networks, and dynamic environment | Only useful for WSBNs and difficult to manage |
| Attributed RBAC [49] | Dynamic behaviour, tight security, and decrease load of administrator | Limited features of RBAC, role explosion, and complexity in designing access policy |
| RBIAC [50] | Enhanced data security, trustworthiness, and data provenance for multi-domain service applications | Execution time overhead due to the addition of various elements of data provenance. |
| TBAC [52] | Automated access control model designed for multi-role implementation | Reliability and scalability problems, and not secure because users also decide access rights |
| TARAS [54] | Enhanced detection accuracy, robustness, and service availability against malicious users | Designed for smart objects and not suitable for military and government organization due to unknown users' run time access |
| GRBAC [57] | Enhanced security, flexibility, and user's identity and task information is secret | SOD is implemented on the level of roles |
| RBAC-SC [58] | Efficient and secure for the verification of user's ownership for role | Specification language is not provided and is limited to basic functionalities |
| FBAC [60] | Enhanced business timeliness and resource usability for unpredictable scenarios | Hard to manage access control policies and security risks |
| E-RBAC [61] | Efficient and flexible for large systems, effective behaviour for normal, emergency, and exception situations | SOD violations can occur, limit the user access according to situations |
| PARBAC [62] | Secure, consistent, flexible, and efficient to handle dynamic scenario problems in large organizations | Denial-of-service occurrence, high system execution, and third-party reliance |
| AB_sAC [63] | Efficient, secure, minimum risk factor, and support to anonymous access | Increase in number of policies will affect the execution time, third-party reliance |
| TABE-DAC [64] | Protect to illegal data sharing, trace to malicious users by accountability method | The authorities can be dishonest, no dynamic access policies |
| TAC [65] | Protect data using encryption in IoT, different sub-keys of each subspace | The sub-keys are not secure and the data owner cannot manage its privacy |

TABLE 3: Applications of traditional and hybrid access control models.

| Model name | Applications |
|-------------------------|--|
| DAC [18] | The most appropriate applications of DAC are Web applications and operating systems such as Unix and Linux |
| MAC [24] | MAC is used in operating systems and database management systems. Furthermore, it is used in the organizations such as government departments and military |
| RBAC [25] | The applications of RBAC are banking and education systems |
| ABAC [37] | The application of ABAC is for companies such as telecommunications, insurance, and airlines |
| TRBAC [42] | The TRBAC is an extension of the RBAC model to achieve dynamic behaviour for activation and deactivation of role |
| Rule RBAC [45] | The rule RBAC model is an extension of the RBAC model to achieve dynamic behaviour of user role assignment |
| Rule BAC [47] | The application of rule BAC is Web-based social networks |
| Attributed RBAC [49] | The attributed RBAC model is a hybrid model of RBAC and ABAC to achieve strict security and dynamic behaviour |
| RBIAC [50] | The RBIAC model is extension of the RBAC model to provide integrity of user data |
| TBAC [52] | The applications of TBAC are online social networks (OSN) and websites |
| TARAS [54] | The application of TARAS is communication of IoT devices |
| GRBAC [57] | The application of GRBAC is IoT environment where roles are not disclosed |
| RBAC-SC [58] | The application of RBAC-SC is blockchain-based smart contract |
| FBAC [60] | The applications of FBAC are auditing, business environment |
| E-RBAC [61] | The E-RBAC is an extension of the RBAC model to work in emergency situations |
| PARBAC [62] | The application of PARBAC is cloud server-based authentication mechanism for medical domain |
| AB _s AC [63] | The application of AB _s AC is IoT-based user data protection |
| TABE-DAC [64] | The application of TABE-DAC is sharing of secret data on cloud servers based on blockchain and also control illegal sharing of secret data |
| TAC [65] | The application of TAC is IoT-based user data protection |

3.1. Applications of Traditional and Hybrid Access Control Models. The access control models are classified into traditional and hybrid models. The basic traditional access control models are DAC, MAC, RBAC, and ABAC. The hybrid access control models are proposed as extension of traditional access control models on the basis of pros and cons. Each traditional and hybrid access control model has its own application as described in Table 3.

4. Conclusions and Future Directions

The access control (AC) mechanism is used to control the access level of resources among legitimate users. The main purpose of access control mechanism is to ensure the security of data by limiting the access of data to only authorized users. The access control is classified into traditional and hybrid models. Due to several limitations of traditional access control models, hybrid access control models were proposed as an extension of traditional access control models. The hybrid access control models are more efficient, flexible, scalable, and secure. The hybrid access control models are used generally in both small and large organizations according to the objective of the organization.

In the future, the access control models also can be designed using fog computing instead of cloud computing. The fog computing stores data over the fog in the form of chunks. Suppose user wants to update the stored data, then user will download only specific chunk of data for modification instead of downloading whole data. The access control model can be made more secure using fog computing due to data chunk mechanism. Moreover, the access control models also can be designed using artificial intelligence (AI) to achieve some key characteristics such as detecting malicious code in resources, identifying illegal sharing of resources, and distinguishing unauthorized users. AI will also be used to

permit and deny the access of resources among users and will limit the users so that they can perform tasks up to the specified role. In short, the access control models can be fully automated with the help of artificial intelligence.

Data Availability

All the data used to support the findings of this study are available in this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was funded by a Faculty Research Support Grant (FRSG-21) of FAST-NUCES, Pakistan, under Project ID “11-71/NU-R/20” and by International Scientific and Technological Innovation Cooperation Project in Sichuan Province (Project ID 2020YFH0062).

References

- [1] H. Huang, F. Shang, J. Liu, and H. Du, “Handling least privilege problem and role mining in RBAC,” *Journal of Combinatorial Optimization*, vol. 30, no. 1, pp. 63–86, 2015.
- [2] J. Hassan, D. Shehzad, I. Ullah et al., “A lightweight proxy Re-encryption approach with certificate-based and incremental cryptography for fog-enabled E-healthcare,” *Security and Communication Networks*, vol. 202117 pages, 2021.
- [3] S. Latif, Z. E. Huma, S. S. Jamal et al., “Intrusion detection framework for the internet of things using a dense random neural network,” *IEEE Transactions on Industrial Informatics*, vol. 99, p. 10, 2021.

- [4] A. Hamza, D. Shehzad, M. S. Sarfraz, U. Habib, and N. Shafi, "Novel secure hybrid image steganography technique based on pattern matching," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 3, pp. 1051–1077, 2021.
- [5] N. W. Hundera, C. Jin, D. M. Geressu, M. U. Aftab, O. A. Olanrewaju, and H. Xiong, "Proxy-based public-key cryptosystem for secure and efficient IoT-based cloud data sharing in the smart city," *Multimedia Tools and Applications*, 2021.
- [6] H. Zhang, J. Wang, and J. Chang, "An access control model for multi-level security in multi-domain networking environments," in *Proceedings of the in 2017 ninth International Conference On Modelling, Identification and Control (ICMIC)*, pp. 809–814, Kunming, China, July 2017.
- [7] T. Sultana, A. Almogren, M. Akbar, M. Zuair, I. Ullah, and N. Javaid, "Data sharing system integrating access control mechanism using blockchain-based smart contracts for IoT devices," *Applied Sciences*, vol. 10, no. 2, Article ID 488, 2020.
- [8] S. Kausar, A. Rahman, A. M. Khan, and T. Ahmad, "Attribute-based Access Control in Web Applications," in *Applications Of Artificial Intelligence Techniques In Engineering*, Springer, New York City, NY, USA, pp. 385–393, 2019.
- [9] M. A. Habib, N. Mahmood, M. Shahid, M. U. Aftab, U. Ahmad, and C. M. N. Faisal, "Permission based implementation of dynamic separation of duty (DSD) in role based access control (RBAC)," in *Proceedings of the 2014 eighth International Conference on Signal Processing and Communication Systems (ICSPCS)*, pp. 1–10, Gold Coast, QLD, Australia, December. 2014.
- [10] M. Liu, C. Yang, H. Li, and Y. Zhang, "An efficient attribute-based access control (ABAC) policy retrieval method based on attribute and value levels in multimedia networks," *Sensors*, vol. 20, no. 6, p. 1741, 2020.
- [11] E. Bertin, D. Hussein, C. Sengul, and V. Frey, "Access control in the Internet of Things: a survey of existing approaches and open research questions," *Annals of Telecommunications*, vol. 74, no. 7-8, pp. 375–388, 2019.
- [12] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of things: the road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015.
- [13] H. A. Khattak, M. A. Shah, S. Khan, I. Ali, and M. Imran, "Perception layer security in internet of things," *Future Generation Computer Systems*, vol. 100, pp. 144–164, 2019.
- [14] Y. Zhang and X. Wu, "Access control in internet of things: a survey," 2016, <http://arxiv.org/abs/1610.01065>.
- [15] P. Samarati and S. C. de Vimercati, "Access control: policies, models, and mechanisms," *Foundations of Security Analysis and Design FOSAD 2000 LNCS*, Springer, vol. 2171, pp. 137–196, Bertinoro, Italy, 2001.
- [16] M. A. Habib, *Secure RBAC with Dynamic, Efficient & Usable DSD*, Johannes Kepler University Linz, Linz, Austria, Ph.D, 2011.
- [17] J. Moffett, M. Sloman, and K. Twidle, "Specifying discretionary access control policy for distributed systems," *Computer Communications*, vol. 13, no. 9, pp. 571–580, 1990.
- [18] R. S. Sandhu, "Role-based access control," *Advances in Computers*, vol. 46, pp. 237–286, 1998.
- [19] D. Ferraiolo, D. R. Kuhn, and R. Chandramouli, *Role-based Access Control*, Artech House, NY, USA, 2003.
- [20] L. Bo, C. ShuhuiB, and D. Jinsheng, "Survey of bell-LaPadula model," *Application Research of Computers*, vol. 30, pp. 656–660, 2013.
- [21] R. R. Jueneman, "Integrity controls for military and commercial applications," in *Proceedings of the 1988 Fourth Aerospace Computer Security Applications*, pp. 298–322, Orlando, FL, USA, September 1988.
- [22] D. D. Clark and D. R. Wilson, "A comparison of commercial and military computer security policies," in *Proceedings of the 1987 IEEE Symposium on Security and Privacy*, pp. 184–194, Oakland, CA, USA, April 1987.
- [23] D. Rountree, "Chapter 2—what Is Federated Identity?" *D. B. T.-F. I. P. Rountree*, Syngress, Rockland, MA, USA, pp. 13–36, 2013.
- [24] E.-B. Choi and S.-J. Lee, "Access control mechanism based on MAC for cloud convergence," *Journal of the Korea Convergence Society*, vol. 7, no. 1, pp. 1–8, 2016.
- [25] American National Standard for Information Technology, *ANSI INCITS 359-2004*, American National Standards Institute, Washington, D.C, USA, 2004.
- [26] M. U. Aftab, Z. Qin, N. W. Hundera et al., "Permission-based separation of duty in dynamic role-based access control model," *Symmetry*, vol. 11, no. 5, Article ID 669, 2019.
- [27] M. U. Aftab, A. Oluwasanmi, A. Alharbi et al., "Secure and dynamic access control for the Internet of Things (IoT) based traffic system," *PeerJ Computer Science*, vol. 7, Article ID e471, 2021.
- [28] M. U. Aftab, Y. Munir, A. Oluwasanmi et al., "A hybrid access control model with dynamic COI for secure localization of satellite and IoT-based vehicles," *IEEE Access*, vol. 8, pp. 24196–24208, 2020.
- [29] K. Z. Bijon, R. Krishnan, and R. Sandhu, "A Framework for Risk-Aware Role Based Access Control," in *Proceedings of the in 2013 IEEE Conference on Communications and Network Security (CNS)*, pp. 462–469, National Harbor, MD, USA, October 2013.
- [30] A. Anderson, *Core and Hierarchical Role Based Access Control (RBAC) Profile of XACML V2. 0*, OASIS Stand, 2005.
- [31] C.-J. Moon, D.-H. Park, S.-J. Park, and D.-K. Baik, "Symmetric RBAC model that takes the separation of duty and role hierarchies into consideration," *Computers and Security*, vol. 23, no. 2, pp. 126–136, 2004.
- [32] N. Solanki, Y. Huang, I.-L. Yen, F. Bastani, and Y. Zhang, "Resource and role hierarchy based access control for resourceful systems," *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 480–486, 2018.
- [33] R. Ghazal, A. K. Malik, N. Qadeer, B. Raza, A. R. Shahid, and H. Alquhayz, "Intelligent role-based access control model and framework using semantic business roles in multi-domain environments," *IEEE Access*, vol. 8, pp. 12253–12267, 2020.
- [34] T. Jaeger and J. E. Tidswell, "Practical safety in flexible access control models," *ACM Transactions on Information and System Security*, vol. 4, no. 2, pp. 158–190, 2001.
- [35] M. K. Hedayati, A. Abdipour, R. Sarraf Shirazi et al., "Challenges in on-chip antenna design and integration with RF receiver front-end circuitry in nanoscale CMOS for 5G communication systems," *IEEE Access*, vol. 7, Article ID 43190, 2019.
- [36] S. Jha, S. Sural, V. Atluri, and J. Vaidya, "Enforcing Separation of Duty in Attribute Based Access Control Systems," *Information Systems Security. ICISS 2015. Lecture Notes in Computer Science*, Springer, Cham, Switzerland, 2015.
- [37] V. C. Hu, D. R. Kuhn, and D. F. Ferraiolo, "Attribute-based access control," *Computer*, vol. 48, no. 2, pp. 85–88, 2015.
- [38] Y. Zhang, D. Zheng, and R. H. Deng, "Security and privacy in smart health: efficient policy-hiding attribute-based access control," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2130–2145, 2018.

- [39] M. Sookhak, F. R. Yu, M. K. Khan, Y. Xiang, and R. Buyya, "Attribute-based data access control in mobile cloud computing: taxonomy and open issues," *Future Generation Computer Systems*, vol. 72, pp. 273–287, 2017.
- [40] V. C. Hu, D. Ferraiolo, R. Kuhn et al., *Guide to Attribute Based Access Control (ABAC) Definition and Considerations*, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2014.
- [41] C.-W. Liu, W.-F. Hsien, C. C. Yang, and M.-S. Hwang, "A survey of attribute-based access control with user revocation in cloud data storage," *International Journal on Network Security*, vol. 18, no. 5, pp. 900–916, 2016.
- [42] E. Bertino, P. A. Bonatti, and E. Ferrari, "TRBAC: a temporal role-based access control model," *ACM Transactions on Information and System Security*, vol. 4, no. 3, pp. 191–233, 2001.
- [43] J. B. D. Joshi, E. Bertino, U. Latif, and A. Ghafoor, "A generalized temporal role-based access control model," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 1, pp. 4–23, 2005.
- [44] E. Uzun, V. Atluri, S. Sural, and J. Vaidya, "Analyzing temporal role based access control models," in *Proceedings of the 17th ACM Symposium on Access Control Models and Technologies*, pp. 177–186, New Jersey, NJ, USA, June 2012.
- [45] M. A. Al-Kahtani and R. Sandhu, "Induced role hierarchies with attribute-based RBAC," in *Proceedings of the eighth ACM symposium on Access control models and technologies - SACMAT '03*, pp. 142–148, Como, Italy, June 2003.
- [46] A. Rashid, I. K. Kim, and O. A. Khan, "Providing authorization interoperability using rule based HL7 RBAC for CDR (Clinical Data Repository) framework," in *Proceedings of the 2015 12th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 343–348, Islamabad, Pakistan, January 2015.
- [47] B. Carminati, E. Ferrari, and A. Perego, "Rule-based access control for social networks, On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops," in *Proceedings of the in OTM Confederated International Conferences On the Move to Meaningful Internet Systems, LNCS*, vol. 4278, pp. 1734–1744, Montpellier, France, November 2006.
- [48] M. U. Aftab, M. A. Habib, N. Mehmood, M. Aslam, and M. Irfan, "Attributed role based access control model," in *Proceedings of the 2015 Conference on Information Assurance and Cyber Security (CIACS)*, pp. 83–89, Rawalpindi, Pakistan, Dec 2015.
- [49] J. Yong, E. Bertino, and M. T. D. Roberts, "Extended RBAC with role attributes," *PACIS 2006 Proc.* vol. 8, 2006.
- [50] W. She, W. Zhu, I.-L. Yen, F. Bastani, and B. Thuraisingham, "Role-based integrated access control and data provenance for SOA based net-centric systems," *IEEE Transactions on Services Computing*, vol. 9, no. 6, pp. 940–953, 2016.
- [51] Q. M. Rajpoot, C. D. Jensen, and R. Krishnan, "Integrating attributes into role-based access control," in *Proceedings of the in IFIP Annual Conference On Data And Applications Security And Privacy*, pp. 242–249, Fairfax, VA, USA, July 2015.
- [52] V. Takalkar and P. N. Mahalle, "Trust-based access control in multi-role environment of online social networks," *Wireless Personal Communications*, vol. 100, no. 2, pp. 391–399, 2018.
- [53] O. Folorunso and O. A. Mustapha, "A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments," *Applied Computing and Informatics*, vol. 11, no. 2, pp. 116–129, 2015.
- [54] B. Gwak, J.-H. Cho, D. Lee, and H. Son, "TARAS: trust-aware role-based access control system in public internet-of-things," in *Proceedings of the 2018 Seventeenth IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 74–85, New York, NY, USA, September 2018.
- [55] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for IIoT devices," *IEEE Transactions on Industrial Informatics*, 2021.
- [56] T. Wang, Y. Quan, X. S. Shen, T. R. Gadekallu, W. Wang, and K. Dev, "A privacy-enhanced retrieval technology for the cloud-assisted Internet of Things," *IEEE Transactions on Industrial Informatics*, 2021.
- [57] M. Alam, N. Emmanuel, T. Khan, Y. Xiang, and H. Hassan, "Garbled role-based access control in the cloud," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 4, pp. 1153–1166, 2018.
- [58] J. P. Cruz, Y. Kaji, and N. Yanai, "RBAC-SC: role-based access control using smart contract," *IEEE Access*, vol. 6, pp. 12240–12251, 2018.
- [59] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and PUF-Based Lightweight Authentication Protocol for Wireless Medical Sensor Networks," *IEEE Internet Things J*, 2021.
- [60] Y. Xu, W. Gao, Q. Zeng, G. Wang, J. Ren, and Y. Zhang, "A feasible fuzzy-extended attribute-based access control technique," *Security and Communication Networks*, vol. 201811 pages, 2018.
- [61] F. Nazerian, H. Motameni, and H. Nematzadeh, "Emergency role-based access control (E-RBAC) and analysis of model specifications with alloy," *Journal of Information Security and Applications*, vol. 45, pp. 131–142, 2019.
- [62] A. Thakare, E. Lee, A. Kumar, V. B. Nikam, and Y.-G. Kim, "PARBAC: priority-attribute-based RBAC model for azure IoT cloud," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2890–2900, 2020.
- [63] R. Zhang, G. Liu, S. Li, Y. Wei, and Q. Wang, "ABSAC: attribute-based access control model supporting anonymous access for smart cities," *Security and Communication Networks*, vol. 2021, pp. 1–11, 2021.
- [64] L. Guo, X. Yang, and W.-C. Yau, "TABE-DAC: efficient traceable attribute-based encryption scheme with dynamic access control based on blockchain," *IEEE Access*, vol. 9, pp. 8479–8490, 2021.
- [65] B. Wang, W. Li, and N. N. Xiong, "Time-based access control for multi-attribute data in internet of things," *Mobile Networks and Applications*, vol. 26, no. 2, pp. 797–807, 2021.

Research Article

A Blockchain-Assisted Electronic Medical Records by Using Proxy Reencryption and Multisignature

Xiaoguang Liu ^{1,2}, Jun Yan,³ Shuqiang Shan,¹ and Rongjun Wu^{1,2}

¹School of Mathematics, Southwest Minzu University, Chengdu, Sichuan 610 041, China

²Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi 541 004, China

³Faculty Affairs Office, Southwest Minzu University, Chengdu, Sichuan 610 041, China

Correspondence should be addressed to Xiaoguang Liu; dtcr-gg@163.com

Received 2 December 2021; Revised 30 December 2021; Accepted 10 January 2022; Published 1 February 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Xiaoguang Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electronic medical records (EMR) have been commonly used in medical institutions in recent years. In particular, the combination of EMR and the cloud server has significantly improved the work efficiency and therapeutic level of the hospital. It also raises some security concerns, e.g., the information leaks. Blockchain has features including decentralization, traceability, openness, and tamper resistance. Therefore, the technology may be used to overcome the above flaws. In this paper, we introduce a new blockchain-assisted EMR in the cloud environment by using proxy reencryption and sequential multisignature. Firstly, blockchain makes the scheme have high-security performance without a trusty center. Secondly, we use proxy reencryption to protect personal medical data while helping doctors to access patients' historical medical records. Moreover, the doctors have used a sequential multisignature, which is practical and can effectively improve security performance. The analysis results show that the proposed scheme can satisfy various security features of EMR and has an ideal computational and communication cost. Finally, the scheme is implemented to show its performance.

1. Introduction

With the full application of modern information technologies such as big data, cloud computing, and artificial intelligence in the medical field, medical informatization has exerted a significant influence on the optimal allocation of medical resources [1, 2]. EMR has emerged from this context, and it uses electronic devices (such as computers and smartphones) to store, manage, and transmit digitized medical records [3]. It can significantly enhance the work efficiency and therapeutic level of the hospital [4]. Also, EMR provides a judgment basis for dealing with medical malpractice [5]. When a patient goes to see a doctor, his/her medical history can help the doctor make an accurate diagnosis. However, most patients are often unable to detail their medical history due to long-time intervals and a lack of relevant expertise. It will affect the current diagnosis and increase the fiscal burden. Therefore, an ideal EMR should be

able to help doctors timely obtain complete and accurate historical medical information. Furthermore, security and privacy preservation are crucial in EMR since medical information is sensitive and personal [6, 7].

EMR has developed significantly in recent years of its remarkable advantages, such as transmitting fast and easy to use [8, 9]. Notably, the emergence of cloud storage is a new milestone in the development of EMR [10]. They move medical data from the traditional data center to a cheaper and safer cloud server. It can improve work efficiency and allow hospitals to invest more time and resources in diagnosis and care. Thus, researchers proposed many cloud-assisted EMR architectures in recent years. However, the privacy, confidentiality, and integrity of medical data will face more threats since the data is outsourced to a third party, i.e., the cloud [11]. For example, doctors can collude with the cloud server to modify their erroneous diagnoses in medical malpractice. So, how to improve the efficiency of

data storage and data sharing while ensuring data security and protecting patient privacy is the focus of research [12, 13]. It is necessary to design a lightweight, efficient, and secure EMR system.

Blockchain technology was introduced in 2008 [14]. It is a decentralized distributed (distributed in multiple locations and able to work together) database system. Blockchain has features of decentralization, tamper-resistant, openness, autonomy, and traceability. It can effectively overcome the adverse effects of centralization and reduce the cost of trust [15]. Therefore, blockchain may be a promising assisted technology of EMR, and it has received attention [16]. However, when blockchain technology is applied to the medical industry, it must be measured between improving efficiency and reducing cost. Only when appropriate blockchain technology is adopted and the system efficiency and operating cost are well balanced, can the business model be established. In addition, many problems are still unsolved before satisfying the practical application in recurrent [17, 18]. For example, (1) the data owner usually encrypts the data with the public key of the user or the session key of both parties, which leads to weak data sharing; (2) only the diagnosis of a single doctor was considered, regardless of a situation in which multiple doctors consult; (3) the cost of computing, communication, and storage is too high.

In this paper, we propose a blockchain-assisted EMR in the cloud environment by using proxy reencryption and sequential multisignature, and we call it BC-EMR. In BC-EMR, a group key and a sequential multisignature are utilized to enhance data security (a diagnosis may be made by a doctor or multiple doctors) [19]. Proxy reencryption helps doctors to access patients' historical medical records while protecting data [20]. Especially, blockchain technology has enabled BC-EMR to overcome many flaws in general cloud-assisted EMR and dramatically improve security [21]. BC-EMR has an ideal computational and communication cost. The main contributions are listed as follows:

We establish a group key between a hospital's server and the doctors of one team utilizing a lightweight one-to-many authentication protocol. It can protect the patient's information.

We propose a blockchain-assisted EMR in the cloud environment by using proxy reencryption and sequential multisignature. The proposed scheme not only can realize the safe storage of data but also make secure data sharing between doctors at different hospitals.

The security analysis of BC-EMR is given. The results show that BC-EMR can satisfy various security features. It also can fend off some specific threats, such as illegal cooperation between doctors and the cloud. Finally, we compare the computational and communication cost of BC-EMR with three existing schemes and then have implemented BC-EMR.

The remainder of the paper is organized as follows. In Section 2, we introduce related works. The preliminaries are presented in Section 3. In Section 4, we introduce the details of BC-EMR. In Section 5, the security analysis of BC-EMR is

given. In Section 6, we evaluate the performance of BC-EMR and implement it. Finally, we conclude our paper in Section 7.

2. Related Works

Ekblaw et al. [22] used the Ethereum platform to realize MedRec that is a medical information sharing platform combining medical blockchain and big data. The system makes use of blockchain, the embedded authentication system, the security system, and an accountability system, which can provide users with powerful security technology when dealing with sensitive information. Xia et al. [23] proposed a blockchain-based health data sharing architecture that only allows the invited (verified) users to access. Thus, it solves many of the access control challenges associated with sensitive data. They have also come up with a system called MeDShare in [24]. The scheme deals with the problem of sharing medical data with big data custodians in untrusted environments. It uses smart contracts and access control mechanisms to track data behavior. Xue et al. [25] proposed a medical blockchain system by combining the medical server and auditing server. Zhang et al. [26] used a hospital-owned private blockchain to store patients' health data, and the consortium blockchain to store safety indexes for personal health data. In particular, the authors have described the details and implemented the scheme on JUICE. In [27], Ivan analyzed the feasibility of using blockchain to protect health data, the implementation barriers, and specific plans for transitioning from current technology to blockchain solutions. Cao et al. [28] proposed a secure cloud-assisted EMR. This scheme utilizes Ethereum platform-based blockchain to protect outsourced medical data. Because every operation of the EMR is put into the blockchain as a transaction, it has excellent security. Esposito et al. [29] comprehensively analyzed the potential of blockchain to protect medical data in the cloud. They also pointed out the practical challenges and future works. Israa et al. [30] elaborated on the benefits and threats of blockchain technology in healthcare. Abdellatif et al. [31] introduced a new smart and safe healthcare system, which takes advantage of edge computing and blockchain to allow for epidemic detection and remote monitoring. The system also allows for the secure exchange of medical data between local medical entities. Shen et al. [32] analyzed the topological relationship among participants in the process of income distribution and established some Shapley value models from simple to complex. Based on the analysis of distribution rules, the incentive effect of secure data sharing and the rationality of the design scheme is discussed. Patil et al. [33] proposed an efficient blockchain authentication protocol for the Internet of Things based on the secret computational model of a physically unclonable function, which can guarantee data provenance and data integrity. Based on the elliptic curve digital signature algorithm, Xiong et al. [34] introduced an efficient and large-scale batch verification scheme with group testing technology for blockchain-enabled IoMT. Zhang et al. [35] proposed a reliable and efficient system based on edge computing and blockchain.

Simulation results show that the proposed method has better computational efficiency and higher reliability than the existing methods. Cheng et al. [36] designed a blockchain-based data-sharing network model for medical cyber-physical systems and used BAN logic to analyze security protocols. Saini et al. [37] built an access control framework based on smart contracts, which is built on top of distributed ledger (blockchain) to ensure EMR sharing between different entities involved in smart healthcare systems.

In Table 1, we give a comparison between the different schemes introduced above. For convenience, we let F1, F2, F3, F4, and F5 denote payment for the blockchain platform, consensus mechanism and reduce the pressure for the main chain, the demand for calculating power, and the private blockchain.

3. Preliminaries

3.1. Blockchain. Blockchain is a novel application of distributed data storage, peer-to-peer transmission, and consensus mechanism, etc. [38]. As shown in Figure 1, a blockchain system consists of many blocks, and each block contains a block header and a block body. The block header includes the hash value of the current block, the timestamp, the hash value of the previous block, and so forth. Block body stores some transaction records. Its main characteristics are listed as follows:

- (1) Decentralization: there are numerous nodes distributed in the blockchain network, which can be freely connected to exchange information without any third institution.
- (2) Tamper resistance: after the information is added to the blockchain by consensus mechanism, all nodes will record it. Each block contains the hash value of the previous block. If a block's data is modified, all the blocks behind that block need to be changed, which is almost impossible.
- (3) Traceability: blockchain stores all data through the block data structure, and any data stored in the blockchain can trace its origin through the chain structure.
- (4) Openness: any node can get the ledger of the whole network. Except for the information of the parties directly related to the data being encrypted by asymmetric encryption technology, other data is open to all nodes.
- (5) Autonomy: the use of a consensus mechanism in blockchain enables all nodes in the whole system to freely and securely exchange, record, and update data.

3.2. The Basic Requirements of EMR

- (1) *Security and Privacy Preservation.* (a) The system can resist malicious attacks on medical data such as forgery attacks, modification attacks, replay attacks, guess attacks, man-in-the-middle attacks, and

trackable attacks. (b) *Nonrepudiation.* The participants cannot deny the historical data generated by themselves. (c) *Confidentiality.* The data transmitted and stored in the network is sensitive personal information, so the system needs to resist data leakage. (d) *Authenticity.* The data cannot be illegally modified. For example, in the cloud environment, the system can prevent doctors and the cloud from conspiring to tamper with medical data.

- (2) *Data Sharing.* Authorized third parties such as other doctors can access the patient's historical medical data with the consent of the patient. In particular, these data may be generated from different doctors at different hospitals.
- (3) *Patient Control.* Patients can control other people's access to their historical medical records.
- (4) *Uniform Standard.* There are uniform data standards and sharing principles among all participants in the system, which is conducive to improving the efficiency and stability of the system.

3.3. Bilinear Map. Let \mathbb{G}_1 and \mathbb{G}_2 denote two multiplicative groups, respectively, and they have the same prime order p . If a map $e: \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ satisfies the following three properties, then e is called the bilinear map [39]:

- (1) Bilinear: for any points $A, B \in \mathbb{G}_1$ and any points $a, b \in \mathbb{Z}_p^*$, $e(A^a, B^b) = e(A, B)^{ab}$ is satisfied.
- (2) Nondegeneracy: there is a point $A \in \mathbb{G}_1$ so that $e(A, A) \neq 1$, 1 is \mathbb{G}_2 's an identity element.
- (3) Computability: for any points $A, B \in \mathbb{G}_1$, $e(A, B)$ can be computed within polynomial time.

3.4. Intractable Problems

- (1) Discrete Logarithm Problem (DLP): knowing two points A and B in \mathbb{G}_1 and $A = B^n$, it is hard to find $n \in \mathbb{Z}_p^*$ so that $A = B^n$.
- (2) Computational Diffie–Hellman Problem (CDH): knowing point A in \mathbb{G}_1 , for a given (A, A^m, A^n) , it is hard to compute A^{mn} , where $m, n \in \mathbb{Z}_p^*$.

3.5. Proxy Reencryption. Proxy reencryption means that a delegatee A generates a proxy reencryption key $PK_{A \rightarrow B}$ of a delegatee B and then sends $PK_{A \rightarrow B}$ to the agent. The agent uses $PK_{A \rightarrow B}$ to convert the ciphertext encrypted with A 's public key PK_A to the ciphertext encrypted with B 's public key PK_B . It does not need to use A 's private key to decrypt the ciphertext, and we will list the details as follows [40]:

- (1) A encrypts the plaintext M with PK_A , i.e., $C_A = E_A(PK_A, M)$.
- (2) A generates the proxy reencryption key $RK_{A \rightarrow B}$ for B and sends C_A and $RK_{A \rightarrow B}$ to the agent.
- (3) The agent converts C_A into C_B utilizing $RK_{A \rightarrow B}$, where C_B is M 's ciphertext encrypted with PK_B .

TABLE 1: Comparison between existing schemes.

| Schemes | F1 | F2 | F3 | F4 | F5 |
|---------|----|---------------|----|-------|----|
| [22] | √ | POW | × | Big | × |
| [23] | × | DPOS | √ | Small | √ |
| [24] | × | DPOS | √ | Small | √ |
| [25] | × | Improved DPOS | √ | Small | √ |
| [26] | × | DBFT | √ | Big | √ |
| [28] | √ | Improved DPOS | √ | Small | √ |
| [36] | × | Improved DPOS | √ | Small | × |
| [37] | × | POW | √ | Small | × |
| Ours | × | Improved DPOS | √ | Small | √ |

√/Support; ×not-support.

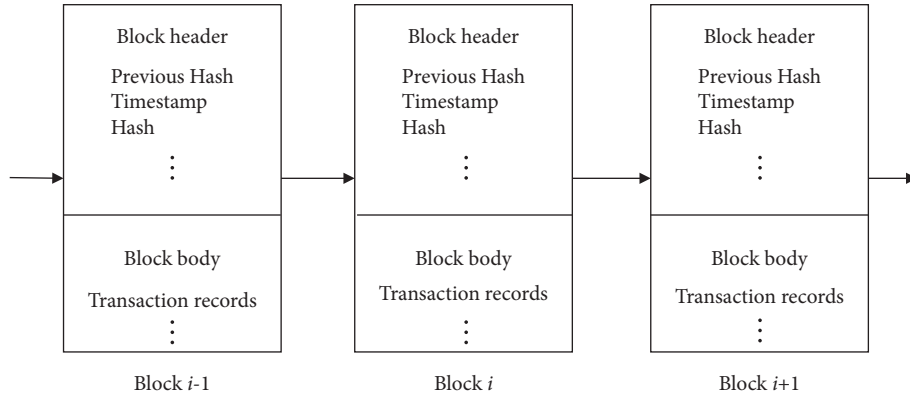


FIGURE 1: The basic structure of blockchain.

Notably, the agent only makes the transformation service of ciphertext and does not know M .

- (4) The agent sends C_B to B that decrypts it using its private key to get M .

3.6. Sequential Multisignature. Sequential multisignature is a particular digital signature scheme. It means that multiple users sign the message in a specific order [19, 41]. A general sequential multisignature usually needs to execute the following four algorithms, i.e., Setup, Key Generation, Sign, and Verify:

- (1) Setup: the key generation center (KGC) inputs a security parameter and generates system parameter para, and system master key.
- (2) Key Generation: given para, users $N_i (i = 1, 2, \dots, n)$ generate their private key SK_i and then compute their own public key PK_i by inputting SK_i .
- (3) Sign: the signer $N_i (i = 2, \dots, n)$ orderly verifies the partial signature s_{i-1} of the previous signer N_{i-1} . If it is valid, N_i outputs own partial signature s_i signed by SK_i .
- (4) Verify: the verifier verifies the signatures by inputting $(m, ID_i, PK_i, s_n) (i = 1, 2, \dots, n)$ and the order of signature.

4. The Proposed BC-EMR

4.1. System Model. In this section, the details of BC-EMR will be given. We use the sequential multisignature of [41] and the proxy reencryption of [42] to construct the scheme. As shown in Figure 2, BC-EMR mainly consists of four entities, i.e., a doctor team, a patient, a hospital, and a cloud server. In BC-EMR, a patient first registers at the hospital. If the identity is approved, the hospital server assigns a medical team to the patient based on the initial condition. Then, the server and members of the team establish a group key that is used to protect the patient's diagnosis results. In the diagnosis, when a doctor receives a message from the former doctor, he/she first verifies previous all doctors' signatures. If it passes, the doctor will make the diagnosis and broadcast his/her signature in the blockchain. Otherwise, he/she requests the former doctor to resend the message. When the last doctor has finished the signature, he/she encrypts the result by using the patient's public key and sends the ciphertext to the cloud server. The ciphertext and signatures will be stored in the cloud and blockchain, respectively, if the signatures pass the verification. Different doctors at different hospitals have the right to access the patient's medical history with the patient's consent. BC-EMR includes the following five phases, i.e., Initialization, Group key generation, Diagnose, Data storage, and Data sharing. In Table 2, we give the used notations in the paper.

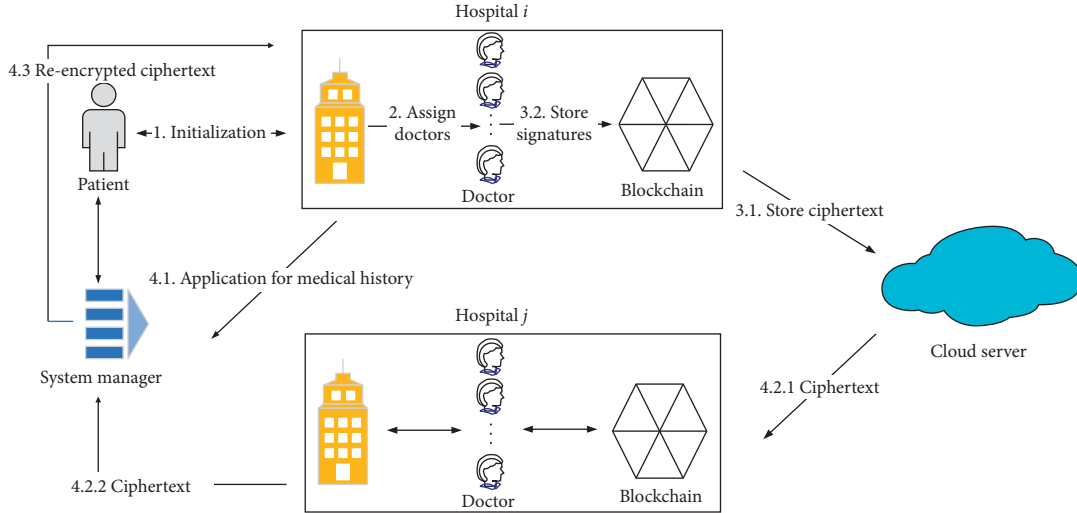


FIGURE 2: The basic structure of BC-EMR.

TABLE 2: Notations.

| Notation | |
|------------|---------------------------------|
| p, q | Two prime numbers |
| SM | The system manager |
| H_i | The i th hospital |
| P_j | The j th patient |
| D_k | The k th doctor |
| $PK_{(.)}$ | The public key |
| $SK_{(.)}$ | The private key |
| $ID_{(.)}$ | The identity |
| s | The diagnosis order |
| g | The generator of \mathbb{G}_1 |
| KGC | The key generation center |
| $E_{(.)}$ | Encryption |
| $D_{(.)}$ | Decryption |
| e | The bilinear map |
| $H_{(.)}$ | The hash function |
| MAC | The message authentication code |
| γ | The security parameter |

4.2. Initialization

- (1) SM inputs a security parameter 1^γ , selects the bilinear map $e: \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ and a random number $\bar{g} \in \mathbb{G}_1$, where \mathbb{G}_1 and \mathbb{G}_2 are two multiplicative groups with the same prime order p . g is a generator of \mathbb{G}_1 . Four hash functions are defined as follows: $H_0: \{0, 1\}^* \rightarrow \mathbb{G}_1$, $H_1: \{0, 1\}^{\leq l} \rightarrow \mathbb{G}_1$, $H_2: \{0, 1\}^{\leq l} \rightarrow \mathbb{G}_1$, and $H_3: \mathbb{G}_2 \rightarrow \{0, 1\}^\gamma$, where l is the length of the verification keys [42]. Besides, SM selects a random number $x \in \mathbb{Z}_p^*$ as the system master key, and the public key $Y = g^x$. The public parameters of BC-EMR are $\{p, g, Y, \bar{g}, H_0, H_1, H_2, H_3, e, \mathbb{G}_1, \mathbb{G}_2\}$. In BC-EMR, we limit the number of the signer reissues the signature to no more than N .
- (2) Hospital H_i selects a random number $h_i \in \mathbb{Z}_p^*$ as its private key and the public key $PK_i = g^{h_i}$.

- (3) Patient P_j selects a random number $p_j \in \mathbb{Z}_p^*$ as the private key and sets $PK_j = g^{p_j}$ as the public key.
- (4) Doctor D_k randomly selects $d_k^1, d_k^2, d_k^3 \in \mathbb{Z}_p^*$, computes $A_k = g^{d_k^1}$, $B_k = g^{d_k^2}$, and $C_k = g^{d_k^3}$. The private key is (d_k^1, d_k^2, d_k^3) and the public key $PK_k = (A_k, B_k, C_k)$.

4.3. Group Key Generation. When a patient P_j sees a doctor in the hospital H_i , P_j sends an identity ID_j and symptoms to HO_i 's server securely. If the identity is legal, the server first selects a random number $\lambda_j \in \mathbb{Z}_p^*$, computes P_j 's pseudo-identity $PID_j = E_{H_i}(ID_j \oplus \lambda_j \| \lambda_j)$ and sends it to P_j . It also assigns initial doctors D_k ($k = 1, \dots, n$) to make a diagnosis according P_j 's condition, sends the evidence $\alpha \in \{0, 1\}^*$ and a diagnosis order s to P_j , and sends a signature timestamp T , α , and s to D_k ($k = 1, \dots, n$) securely. Especially, as in Figure 3, a group key between H_i and D_k ($k = 1, \dots, n$) will be set to protect medical information. The details are given as follows:

- (1) H_i chooses a random number $l_i \in \mathbb{Z}_p^*$, computes $U_i = g^{l_i}$, and sends (ID_i, U_i) to D_k .
- (2) D_k randomly selects a number $l_k \in \mathbb{Z}_p^*$, computes $V_k = g^{l_k}$, and sends (ID_k, V_k) to H_i .
- (3) H_i computes $s_i = V_k^{l_i}$, $MAC_i = MAC_{s_i}(ID_k, V_k, U_i)$, and sends MAC_i to D_k .
- (4) D_k computes $s_k = U_i^{l_k}$ and $MAC_k = MAC_{s_k}(ID_k, V_k, U_i)$. If $MAC_i = MAC_k$, D_k computes $MAC_k^\dagger = MAC_{s_k}(ID_i, U_i, V_k, s_k)$ and sends MAC_k^\dagger to H_i . Otherwise, \perp .
- (5) H_i computes $MAC_i^\dagger = MAC_{s_i}(ID_i, U_i, V_k, s_i)$ and checks $MAC_i^\dagger = MAC_k^\dagger$. If not, \perp . Otherwise, H_i computes $K = V_1^{l_i} \dots V_n^{l_i}$ and $M = E_{s_i}(K)$, and then sends M to D_k .
- (6) D_k decrypts M using s_k to get the group key K .

The correctness of the above protocol is based on the following equation

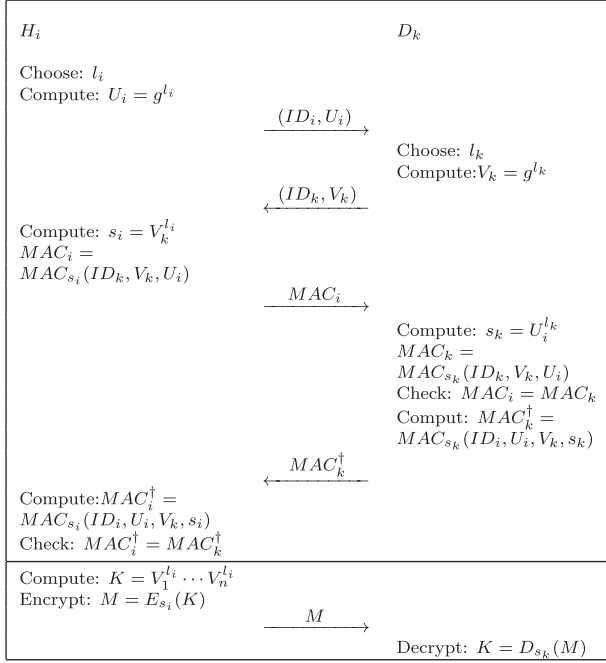


FIGURE 3: The group key generation.

$$s_i = V_k^{l_i} = g^{l_i l_k} = U_i^{l_k} = s_k. \quad (1)$$

4.4. Diagnosis

- (1) P_j shows α and PID_j to D_1 as the evidence, so that D_1 makes a diagnosis or accesses the history records of P_j . If it is legal, D_1 first generates a diagnosis m_1 , randomly selects $r_1 \in \mathbb{Z}_p^*$, computes $R_1 = g^{r_1}$, $X_1 = R_1^{d_1 + d_1^*}$, $W_1 = H_0(m_1, T)$, $Q_1 = W_1^{d_1} X_1$ and $c_1 = E_K(m_1)$. Then D_1 sends the signature message $(c_1, PID_j, (R_1, Q_1))$ to D_2 . Meanwhile, D_1 broadcasts signature (PK_1, PID_j, R_1, Q_1) in the blockchain; please see Figure 4 for the structure of block. In BC-EMR, each block is used to store one patient's information such as all doctors' signatures.
- (2) P_j shows α to $D_k (k = 2, \dots, n)$. If it is legal, D_k confirms whether he/she received $(c_{k-1}, PID_j, (R_{k-1}, Q_{k-1}))$ before $T_k = kT$. If not, D_k requests D_{k-1} to resend the message. Then, D_k decrypts c_{k-1} to get m_1, \dots, m_{k-1} and verifies the following:

$$e(Q_{k-1}, g) = \prod_{i=1}^{k-1} e(W_i, A_i) e\left(\prod_{i=1}^{k-1} B_i C_i^i, R_{k-1}\right). \quad (2)$$

If it is true, D_k first randomly selects $r_k \in \mathbb{Z}_p^*$, generates diagnosis m_k , computes $R_k = R_{k-1} g^{r_k}$, $X_k = R_k^{d_k + kd_k^*}$, $Z_k = \left(\prod_{i=1}^{k-1} B_i C_i^i\right)^{r_k}$, $W_k = H_0(m_1 \| m_2 \dots \| m_k, T)$, and $Q_k = W_k^{d_k} Q_{k-1} X_k Z_k$. Then D_k encrypts the results $m_1 \| m_2 \dots \| m_k$ as $c_k = E_K(m_1 \| m_2 \dots \| m_k)$ and sends the signature $(c_k, PID_j, (R_k, Q_k))$ to D_{k+1} . Meanwhile, D_k

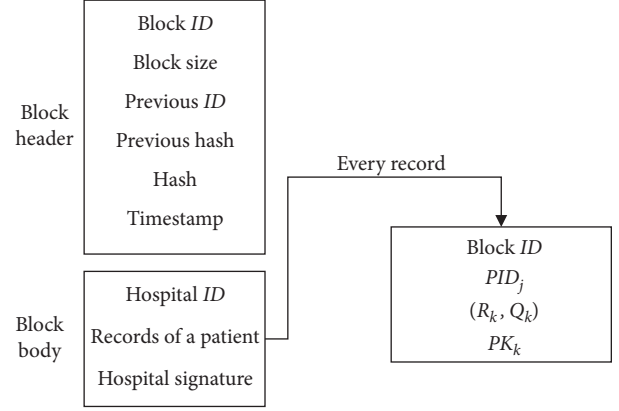


FIGURE 4: The structure of block in the hospital's blockchain.

boardcasts signature (PK_k, PID_j, R_k, Q_k) in the blockchain. Thus, the final diagnosis is $m = m_1 \| m_2 \dots \| m_n$ and the signature message is $(c_n, PID_j, (R_n, Q_n))$.

- (3) D_n encrypts the results m using P_j 's public key PK_j to generate the ciphertext C_j . We will give the details as follows:
 - (a) D_n selects a general signature key pair (PK, SK) and sets $PK = A$
 - (b) D_n randomly selects a number $r \in \mathbb{Z}_p^*$ and computes $B = PK^r$, $C = e(g, H_1(A))^r \oplus m$, $D = H_2(A)^r$, $E = \bar{g}^r$, $F = e(PK_j, H_0(\alpha))^r$, and $G = H_3(F)$
 - (c) D_n signs the message (C, D, E, G) using SK and outputs the ciphertext $C_j = (S, A, B, C, D, E, G)$, where S is the signature
 - (d) D_n sends ciphertext C_j and PID_j to the cloud server

4.5. Data Storage. In BC-EMR, every doctor is the general node of the blockchain. The cloud server and HO_i 's server are the supernodes, and they are responsible for verifying the signature message. That is, if the signature message passes their verification, all nodes will put the current signatures about the patient P_j in a block and update their stored records. The verification scheme is that the supernodes check the following equation:

$$e(Q_n, g) = \prod_{i=1}^n e(W_i, A_i) e\left(\prod_{i=1}^n B_i C_i^i, R_n\right). \quad (3)$$

If it is true, the supernodes send a confirmation message in the blockchain so that all nodes accept the signatures about P_j , put them in a block, and update the stored records. Otherwise, \perp .

4.6. Data Sharing. When the doctor D_k in the hospital H_d makes a diagnosis for the patient P_j , P_j 's medical history in other hospitals H_i may help D_k . Therefore, if D_k wants, he/she can obtain these records with the consent of P_j . The details are as follows:

- (1) D_k and P_j send their identities and request to SM . If it is passed, SM sends a notice to H_i , and the server of H_i extracts the ciphertext $C_j = (S, A, B, C, D, E, G)$ from the cloud server and sends it to SM . In addition, P_j sends a trapdoor $T_\alpha = H_0(\alpha)^{P_j}$ to D_k .
- (2) D_k and P_j send the private keys d_k^1 and p_j to SM , respectively. Then, SM outputs the reencryption key $rk_{j \leftrightarrow k} = d_k^1 / p_j$.
- (3) SM checks the signature S on (C, D, E, G) , i.e., $e(B, H_2(A)) = e(PK_j, D)$, and $e(B, \bar{g}) = e(PK_j, E)$. If any of them fails, \perp . Otherwise, SM computes $B_I = B^{rk_{j \leftrightarrow k}} = PK_j^{rk_{j \leftrightarrow k} \times r} = (g^{P_j r})^{d_k^1 / p_j} = g^{d_k^1 r} = A_k^r$ and sends the ciphertext (S, A, B_I, C, D, E, G) to D_k .
- (4) D_k checks the signature S , i.e., $e(B_I, H_2(A)) = e(A_k, D)$, $e(B_I, \bar{g}) = e(A_k, E)$, and $G = H_3(e(B_I, T_\alpha)^{1/d_k^1})$. If any of them fails, \perp . Otherwise, D_k recovers the message $m = C \oplus e(B_I, H_1(A))^{1/d_k^1}$.

5. Solutions to the Basic Requirements

BC-EMR has provided for the advantages described in Subsection 3.1 since it uses blockchain. Especially, the scheme is based on the sequential multisignature of [41] and the proxy reencryption of [42]. They are proven secure in the random oracle model, which is based on the hardness of the CDH problem and the modified DBDH problem, respectively, and please see [41, 42] for the full formal proof.

In this subsection, we will show why BC-EMR satisfies the basic requirements of BMR. In Table 3, we list the comparison results about BC-EMR and the other three blockchain-based EMR schemes ZL, CZ, and AS. Here, the schemes in [26, 28, 37] are denoted as ZL, CZ, and AS, respectively.

5.1. Security and Privacy Preservation (SP)

(a) Malicious attacks (MA)

Forgery attack (M1): to get the doctor's private key to generate a legal signature, the adversary must solve DLP intractable problem. Especially, it is not feasible to falsify the diagnosis by the cloud server or the collaboration between the patient and the cloud server. The reason is that they also can not get the doctors' legal signatures from the hospital's server or SM can detect any forged information by verifying the signatures stored in the blockchain. So, BC-EMR can resist the forgery attack.

Modification attack (M2): in BC-EMR, the last doctor encrypts the diagnosis results with the patient's public key and outsources them to the cloud. If an adversary wants to modify them, it first needs to obtain the patient's private key and the cloud's permission. To get the private key illegally, the adversary must solve DLP intractable problem. Therefore, it is not possible. More importantly, BC-EMR stores the doctors' signatures of their diagnoses in the blockchain. Then, it is easy to detect the

TABLE 3: Comparison of the basic requirements.

| | SP | | | | | | | | | DS | PC | US |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|
| | M1 | M2 | M3 | M4 | M5 | M6 | NR | CO | AU | | | |
| ZL | √ | √ | √ | × | √ | √ | √ | √ | √ | √ | √ | √ |
| CZ | √ | √ | √ | × | √ | √ | √ | √ | √ | × | × | √ |
| AS | √ | √ | √ | × | √ | √ | √ | √ | √ | √ | √ | √ |
| BC-EMR | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |

√/Support; ×not-support.

modification to the diagnosis results, even if the patient's private key is leaked or the patient (or doctor) cooperates with the cloud. Thus, BC-EMR can resist the modification attack.

Replay attack (M3): BC-EMR has introduced the timestamp T . D_k will confirm whether he/she received the message before $T_k = kT$. It is impossible to change timestamp T since the signatures stored in the blockchain contain it. Any modification to T is easily detected, and thus BC-EMR can resist the replay attack.

Guess attack (M4): in BC-EMR, the system sets the number of resigning as N . If the number of resigning exceeds N , the signature terminates. It can be to limit the number of attacks effectively and resist the guessing attack.

Man-in-the-middle attack (M5): protection against the man-in-the-middle attack follows from the protection against the forgery attack, modification attack, and replay attack.

Trackable attack (M6): the patient and hospital generate different random numbers including l_i , l_k , and r_i in each execution of BC-EMR. Thus, there is no constant value in the transmitted or stored messages, and the adversary can not trace the action. Therefore, our BC-EMR can resist the trackable attack.

- (b) **Nonrepudiation (NR):** blockchain technology makes BC-EMR satisfy traceability. We can search the origin of any record stored in the blockchain by the chain structure and the doctor's signature. Thus, no participant can deny the data generated by himself/herself.
- (c) **Confidentiality (CO):** before diagnosis, the hospital server will set a pseudoidentity for each patient. During the diagnosis, the patient will use the pseudoidentity to interact with doctors. The doctors will encrypt the diagnosis results with the group key, and only members of the medical team can get them. When the diagnosis is over, the last doctor will encrypt the result with the patient's public key before storing it in a cloud server. If the adversary wants to get the patient's private key to decrypt the ciphertext, he/she needs to solve the DLP intractable problem. So BC-EMR has ideal confidentiality.
- (d) **Authenticity (AU):** no one but the patient can decrypt the ciphertext of diagnosis since the final

doctor encrypts diagnosis results with the patient's public key. Doctors have stored the signatures of diagnoses in the blockchain, and these violations are easy to spot.

5.2. Data Sharing (DS). The scheme has utilized proxy reencryption technology. If the doctor has obtained the consent of the patient, he/she will get the ciphertext encrypted by their public key. Then the doctor accesses the patient's historical medical records by decrypting the ciphertext. That is, BC-EMR realizes data sharing between different doctors at different hospitals.

5.3. Patient Control (PC). When a doctor needs to know the patient's historical records of another hospital, the doctor must get the ciphertext that is encrypted by his/her public key. However, the original ciphertext is encrypted by the patient's public key. The transformation of ciphertext needs to be completed by *SM* using the reencryption key that is computed by *SM* utilizing the patient's private key. So, the patient can control the doctor to access the historical data.

5.4. Uniform Standard (US). In hospitals' BC-EMR systems, we can use a uniform standard such as the same encryption algorithm. It is beneficial to implement data sharing and other functions.

In Table 3, we give the comparison results of BC-EMR, ZL, CZ, and AS according to the basic requirements. We can know that ZL and AS can not resist the guess attack. CZ not only can not resist the guess attack but also can not satisfy patient control and make the essential data sharing.

Remarks. Without the blockchain, if a doctor tries to forge EMR that has been outsourced to a cloud server in a medical accident, he/she can incentivize the cloud server to forge or modify the existing EMR at will. This is consistent with reality. The introduction of blockchain makes the scheme resistant to threats such as doctor-cloud collusion to forge or modify EMR without additional security mechanisms, strong hypotheses, and trusted entities. In other words, blockchain plays a key role in ensuring security in BC-EMR. Moreover, in BC-EMR, blockchain only stores some lightweight information such as the doctors' signatures, and diagnostic results are stored in the cloud, thus reducing the burden of blockchain and facilitating the future implementation of the scheme.

6. Performance Evaluation

In this section, we will evaluate BC-EMR from the following two aspects: (1) computational and communication cost; (2) the implementation of BC-EMR.

6.1. Computational and Communication Cost. In this subsection, we will compare the main computational cost and communication cost of BC-EMR, ZL, CZ, and AS. *SM* usually has sufficient computational power and storage capacity, so we consider the burden on the patient and the

last doctor of the team (the last doctor is responsible for outsourcing data, he/she needs to pay for more cost than other doctors). The comparison results of computational cost are shown in Table 4. Here, m denotes the scale multiplication operator in \mathbb{G}_1 , e denotes the exponentiation operation in \mathbb{G}_1 , b denotes the bilinear pairing operation. We ignore the remaining operations because of their low computational cost.

In Table 4, λ is the size of the disease keyword set [26]. On the doctor's side, ZL has the highest computational cost of $(17 + \lambda)m + 7e + 4b$ since λ is usually large such as 1000 in ZL. The cost of CZ is lower than BC-EMR, but it can not satisfy a critical feature of EMR, i.e., the data sharing between doctors. CZ also omits the authentication between the server and doctors in generating the treatment key, and the current doctor only verifies the previous doctor's signature. But BC-EMR will make the authentication in creating the group key, and the current doctor verifies previous all doctors' signatures. The cost of AS is also lower than BC-EMR on the doctor's side, but the scheme needs to decrypt the ciphertext using the patient's private key and then encrypt the EMR using the shared secret key in the process of data sharing. Every time generation and management of the shared secret key both are consuming cost and there is a risk of data leakage. So BC-EMR has higher security, and the additional computational cost is worthy. In addition, the results in [43] show that the computational cost of the operation e is approximately two times that of the operation m . So, on the side of the patient, we can find that BC-EMR has the lowest computational cost of $2e$. It is worthy to note that doctors often have relatively reliable computational power, and it is vital to reduce the computational burden on patients.

In Table 5, we will give the main communication cost of the three schemes. $|x|$ and $|t|$ denote the size of the element in the ciphertext space and the size of the timestamp, respectively. $|n_p|$ is the number of the private blockchain's verifiers in ZL, and $|D_p|$ is the size of personal data in AS. We use the supersingular curve $E(F_q)$ with order p over the finite field \mathbb{GF}_q . To give a more explicit comparison of communication cost, we assume the prime number p is 160 bits, the prime number q is 1024 bits, the point in \mathbb{G}_1 is 1024 bits, the point in \mathbb{G}_2 is 512 bits, the point in the ciphertext space, the hash value, and α all are 160 bits, the timestamp, the order message s , and the identity all are 32 bits, the security parameter γ is 512 bits, $n_p = 3$ in ZL, l is 512 bits, and $|D_p|$ is 1024 bits. In Figure 5, we give the comparison diagram of communication cost versus λ (without loss of generality, we assume $n = 3$). Besides, we provide a comparison diagram of communication cost versus the number of doctors n in Figure 6. In ZL, $\lambda = 1000$, but we set $\lambda = 200$ for clearly showing the comparison results in Figure 6.

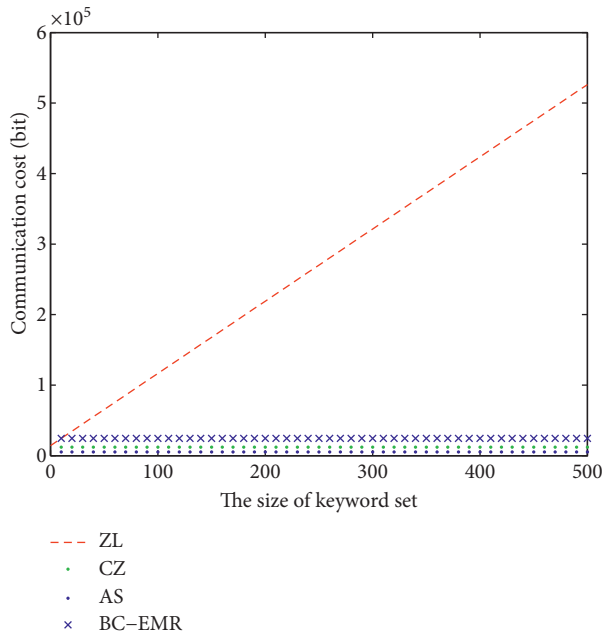
We can see from Figure 5 that the communication cost of ZL linearly increases with λ . The communication cost of CZ, AS, and BC-EMR is constant versus λ . Since λ is usually relatively large such as 1000 in ZL. So, ZL has the highest communication cost. In Figure 6, the communication cost of CZ and BC-EMR both linearly increase with n , and the communication cost of BC-EMR is higher than CZ's. AS has

TABLE 4: Comparison of the computational cost.

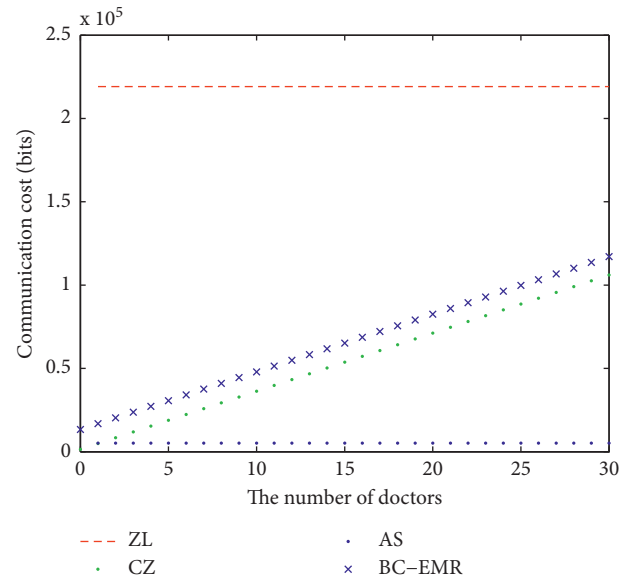
| Scheme | Patient | Doctor |
|--------|-----------|-----------------------------|
| ZL | $7m$ | $(17 + \lambda)m + 7e + 4b$ |
| CZ | $5e + nm$ | $2b$ |
| AS | $5m$ | $9m$ |
| BC-EMR | $2e$ | $(11 + n)e + (n + 9)b$ |

TABLE 5: Comparison of the communication cost.

| Scheme | Communication cost |
|--------|---|
| ZL | $(\lambda + 12) \mathbb{G}_1 + \mathbb{G}_2 + 5 Z_p^* + 13\lfloor 2/3n_p \rfloor + t + x + I D + 2 \text{Hash} $ |
| CZ | $(3n + 1) \mathbb{G}_1 + n Z_p^* + 2(n + 1) I D + n \text{Hash} + 2 x + (n + 1) t $ |
| AS | $3 \mathbb{G}_1 + I D + 4 x + \text{Hash} + Z_p^* + t + D_p $ |
| BC-EMR | $(3n + 8) \mathbb{G}_1 + 2 \mathbb{G}_2 + 2 Z_p^* + 5 I D + (n + 6) x + 2 \text{Hash} + (n + 2) \alpha + n t + 2 \gamma + 2 l + (n + 1) s $ |


 FIGURE 5: Communication cost comparison versus λ .

the lowest communication cost, and an important reason is that the scheme does not consider the case that multiple doctors generate the EMR for a patient. In addition, as mentioned in the analysis of computational cost, CZ can not satisfy some features such as data sharing. AS requires decrypting the ciphertext using the patient's private key and then encrypting the EMR using the shared key to make data sharing. Every time the shared key is generated and managed, there is cost, and it will also increase the risk of EMR data leakage. Thus, the above results show that BC-EMR can achieve a better balance among security, basic requirements, computational cost, and communication cost. So it is an ideal EMR scheme.


 FIGURE 6: Communication cost comparison versus n .

6.2. Implementation of BC-EMR. In this subsection, we will give some implementation results of BC-EMR. The experiment used *Python* 3.7.4 as the programming language. Specially, we used the C language library PBC (v0.5.14) for bilinear pairing calculations, pypbc to call the PBC library in *Python*, and *Python*'s encryption algorithm library pycryptodome (v3.9.0) to implement AES encryption. We used a computer equipped with an Intel Core i7-9750H CPU @ 2.60 GHz and 15.6 GB of memory to run our experimental program. The operating system is Manjaro Linux 64 bit, the desktop environment is KDE (v5.61.0), and the kernel version is 4.19.69-1-MANJARO. During the experiment, we started multiple processes on the experimental computer. Each procedure was bound to a separate port and communicated with each other using sockets. In this way, we

TABLE 6: Phased time costs when $n = 3$ (ms).

| Phases | Security levels | | |
|----------------------|-----------------|---------|---------|
| | Level 1 | Level 2 | Level 3 |
| Initialization | 62.78 | 280.17 | 704.80 |
| Group key generation | 83.30 | 405.80 | 905.90 |
| Diagnose | 296.60 | 1366.90 | 2220.30 |
| Data storage | 30.10 | 167.14 | 307.40 |
| Data sharing | 193.30 | 971.50 | 2125.50 |

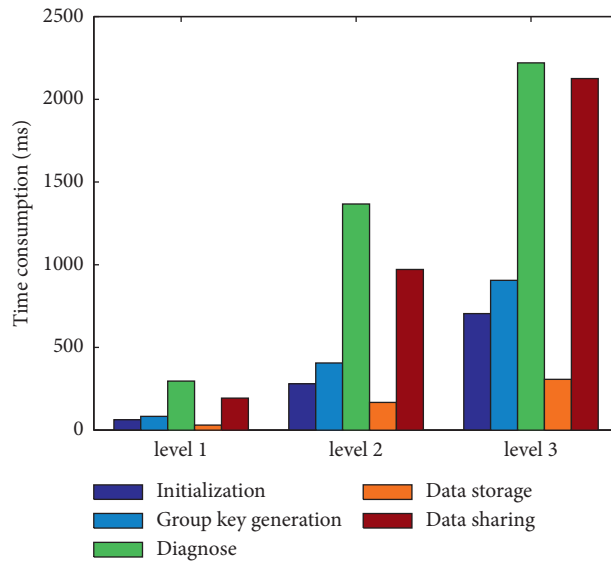


FIGURE 7: Phased time costs when $n = 3$.

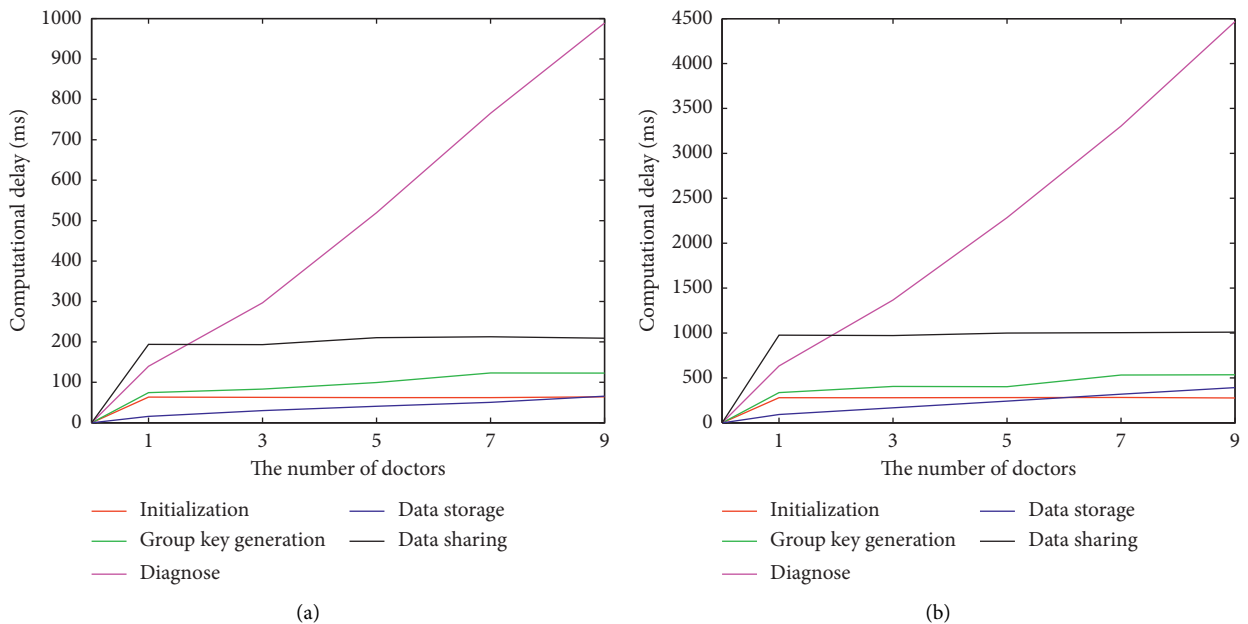


FIGURE 8: Continued.

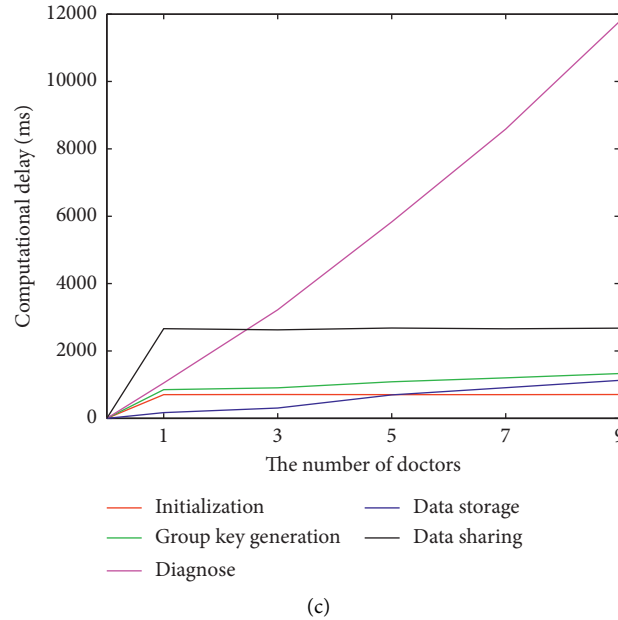


FIGURE 8: Computational delay for three security levels. (a) Level 1. (b) Level 2. (c) Level 3.

simulated the behavior of different roles in each phase of the scenario. All values are the average result of 100 times experiments.

We consider three security levels for BC-EMR, i.e., level 1 ($p = 160$ bits, $q = 1024$ bits), level 2 ($p = 224$ bits, $q = 2048$ bits), and level 3 ($p = 256$ bits, $q = 3072$ bits). In Table 6, we assume $n = 3$ and summarize the phased computational costs for three security levels. The corresponding histogram is given in Figure 7. The result shows that the computational costs of the five stages increase as the security level increases. Besides, in Figure 8, we consider the computational delay versus n . We can find that the computational delays of the initialization phase and data sharing phase have no significant change. With the increase in the number of doctors, the other three stages' computational delays all significantly increase. The computational delay of the diagnostic phase is the fastest growing. The reason is that as the number of doctors increases, BC-EMR needs to perform more time-consuming exponential operations and bilinear pairing operations.

7. Conclusion

In this paper, we proposed a blockchain-based EMR in the cloud environment. A lightweight one-to-many authentication protocol is given to set a group key, which is used to protect the patient's diagnosis results before storing them in the cloud. The proxy reencryption is utilized to make secure data sharing between doctors at different hospitals. Blockchain and sequential multisignature technologies ensure that the stored medical information is safe. Especially, BC-EMR can resist threats such as doctor-cloud collusion to forge or modify EMR. The analysis shows that BC-EMR has a lower computational cost on the side of the patient. It is very important for EMR since the patients usually rely on

resource-limited mobile devices. Besides, BC-EMR can satisfy more basic requirements and security features. So the extra computational cost on the side of the doctor and the extra communication cost compared with CZ both are worthy. That is, BC-EMR is a practical EMR. Of course, as the analysis results show, the method presented in this paper has some shortcomings, such as a slightly higher cost of communication. Since blockchain is a massive ledger backup measure, these deficiencies will directly reduce the system performance. So the balance between efficiency, security, and cost remains at the heart of what we do next.

Data Availability

Some or all data, models, or codes generated or used during the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities of Southwest Minzu University (No. 2020NQ21) and the Fund of Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS202121).

References

- [1] D. V. Dimitrov, "Medical internet of things and big data in healthcare," *Healthcare Informatics Research*, vol. 22, no. 3, pp. 156–163, 2016.

- [2] M. Alizadeh, S. Abolfazli, M. Zamani, S. Baharun, and K. Sakurai, "Authentication in mobile cloud computing: a survey," *Journal of Network and Computer Applications*, vol. 61, pp. 59–80, 2016.
- [3] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and puf-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, pp. 1–9, 2021.
- [4] C. Y. Weng, "Data accuracy in electronic medical record documentation," *Jama Ophthalmology*, vol. 135, no. 3, pp. 232–233, 2017.
- [5] S. S. Mangalmurti, L. Murtagh, and M. M. Mello, "Medical malpractice liability in the age of electronic health records," *New England Journal of Medicine*, vol. 363, no. 21, pp. 2060–2067, 2010.
- [6] J. Song, Z. Han, W. Wang, J. Chen, and Y. Liu, "A new secure arrangement for privacy-preserving data collection," *Computer Standards & Interfaces*, vol. 80, Article ID 103582, 2022.
- [7] F. Li, Y. Han, and C. Jin, "Cost-effective and anonymous access control for wireless body area networks," *IEEE Systems Journal*, vol. 12, no. 1, pp. 747–758, 2018.
- [8] R. Pivovarov, D. J. Albers, J. L. Sepulveda, and N. Elhadad, "Identifying and mitigating biases in ehr laboratory tests," *Journal of Biomedical Informatics*, vol. 51, pp. 24–34, 2014.
- [9] A. Sheth, U. Jaimini, and H. Y. Yip, "How will the internet of things enable augmented personalized health?" *IEEE Intelligent Systems*, vol. 33, no. 1, pp. 89–97, 2018.
- [10] J. Haskew, G. Rø, K. Saito et al., "Implementation of a cloud-based electronic medical record for maternal and child health in rural Kenya," *International Journal of Medical Informatics*, vol. 84, no. 5, pp. 349–354, 2015.
- [11] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. J. Lou, "Privacy-preserving public auditing for secure cloud storage," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 362–375, 2013.
- [12] V. Casola, A. Castiglione, K. R. Choo, and C. Esposito, "Healthcare-related data in the cloud: challenges and opportunities," *IEEE Cloud Computing*, vol. 3, no. 6, pp. 10–14, 2016.
- [13] J. L. Fernandez-Aleman, I. C. Senior, P. A. Lozoya, and A. Toval, "Security and privacy in electronic health records: a systematic literature review," *Journal of Biomedical Informatics*, vol. 46, no. 3, pp. 541–562, 2013.
- [14] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," 2009, <http://bitcoin.org/bitcoin.pdf>.
- [15] W. Z. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Y. Han, and C. H. Su, "Blockchain-based reliable and efficient certificateless signature for iiot devices," *IEEE Transactions on Industrial Informatics*, pp. 1–9, 2021.
- [16] F. T. Stafford and T. Horst, "Characteristics of a blockchain ecosystem for secure and sharable electronic medical records," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1340–1362, 2020.
- [17] I. Makhdoom, M. Abolhasan, H. Abbas, and W. Ni, "Blockchain's adoption in iot: the challenges, and a way forward," *Journal of Network and Computer Applications*, vol. 125, pp. 251–279, 2019.
- [18] R. Z. Yang, F. R. Yu, P. B. Si, Z. X. Yang, and Y. H. Zhang, "Integrated blockchain and edge computing systems: a survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1508–1532, 2019.
- [19] Y. L. Qin and X. P. Wu, "Efficient certificateless sequential multi-signature scheme," *Journal on Communications*, vol. 34, no. 7, pp. 105–110, 2013.
- [20] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," *ACM Transactions on Information and System Security*, vol. 9, no. 1, pp. 1–30, 2006.
- [21] L. C. Wang, X. Y. Shen, J. Li, J. Shao, and Y. X. Yang, "Cryptographic primitives in blockchains," *Journal of Network and Computer Applications*, vol. 127, pp. 43–58, 2019.
- [22] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: using blockchain for medical data access and permission management," in *Proceedings of the 2016 IEEE of International Conference on Open and Big Data (OBD)*, pp. 25–30, Vienna, Austria, August 2016.
- [23] Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. S. Zhang, "Bbds: blockchain-based data sharing for electronic medical records in cloud environments," *Information*, vol. 8, no. 2, 2017.
- [24] Q. Xia, E. B. Sifah, K. O. Asamoah, J. B. Gao, X. J. Du, and M. Guizani, "Medshare: trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, Article ID 14757, 2017.
- [25] T. F. Xue, Q. C. Fu, C. Wang, and X. Y. Wang, "A medical data sharing model via blockchain," *Acta Automatica Sinica*, vol. 43, no. 9, pp. 1555–1562, 2017.
- [26] A. Q. Zhang and X. D. Lin, "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain," *Journal of Medical Systems*, vol. 42, 2018.
- [27] D. Ivan, "Moving toward a blockchain-based method for the secure storage of patient records," 2018, https://www.healthit.gov/sites/default/files/9-16-drew_ivan_20160804_blockchain_for_healthcare_final.pdf.
- [28] S. Cao, G. X. Zhang, P. F. Liu, X. S. Zhang, and F. Neri, "Cloud-assisted secure ehealth systems for tamper-proofing ehr via blockchain," *Informance Science*, vol. 485, pp. 427–440, 2019.
- [29] C. Esposito, A. D. Santis, G. Tortora, H. Chang, and K. R. Choo, "Blockchain: a panacea for healthcare cloud-based data security and privacy?" *IEEE Cloud Computing*, vol. 5, no. 1, pp. 31–37, 2018.
- [30] A. Israa, H. Asma, N. Anjanarani, H. Mowafa, and A. Alaa, "The benefits and threats of blockchain technology in healthcare: a scoping review," *International Journal of Medical Informatics*, vol. 142, pp. 1–9, 2020.
- [31] A. A. Abdellatif, Z. A. Abeer, M. Amr, E. Aiman, C. F. Carla, and R. Ahmed, "sshealth: toward secure, blockchain-enabled healthcare systems," *IEEE Network*, vol. 34, no. 4, pp. 312–319, 2020.
- [32] M. Shen, J. X. Duan, L. H. Zhu, J. Zhang, X. J. Du, and M. Guizani, "Blockchain-based incentives for secure and collaborative data sharing in multiple clouds," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 6, pp. 1229–1241, 2020.
- [33] A. S. Patil, R. Hamza, H. Yan, A. Hassan, and J. Li, "Blockchain-puf-based secure authentication protocol for internet of things," in *Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing*, pp. 331–338, New York, USA, January 2020.
- [34] H. Xiong, C. J. Jin, M. Alazab et al., "On the design of blockchain-based ecdsa with fault-tolerant batch verification protocol for blockchain-enabled iomt," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [35] L. J. Zhang, Y. F. Zou, W. Z. Wang, Z. L. Jin, Y. S. Sue, and H. L. Chen, "Resource allocation and trust computing for blockchain-enabled edge computing system," *Computers & Security*, vol. 105, 2021.

- [36] X. Cheng, F. L. Chen, D. Xie, H. Sun, and C. Huang, "Design of a secure medical data sharing scheme based on blockchain," *Journal of Medical Systems*, vol. 44, no. 52, pp. 1–11, 2020.
- [37] A. Saini, Q. Y. Zhu, N. Singh, Y. Xiang, L. X. Gao, and Y. S. Zhang, "A smart-contract-based access control framework for cloud smart healthcare system," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5914–5925, 2021.
- [38] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka, "Security services using blockchains: a state of the art survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 858–880, 2019.
- [39] T. Okamoto, "Cryptography based on bilinear maps," in *Proceedings of the International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes*, pp. 35–50, Las Vegas, NV, USA, February 2006.
- [40] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," in *Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 127–144.
- [41] A. Boldyreva, C. Gentry, A. O. Neill, and D. H. Yum, "Ordered multisignatures and identity-based sequential aggregate signatures, with applications to secure routing," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, pp. 276–285.
- [42] J. Shao, Z. F. Cao, K. liang, and H. Lin, "Proxy re-encryption with keyword search," *Information Sciences*, vol. 180, no. 13, pp. 2576–2587, 2010.
- [43] J. W. Liu, Z. H. Zhang, X. F. Chen, and K. S. Kwak, "Certificateless remote anonymous authentication schemes for wireless body area networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 332–342, 2014.

Research Article

iReTADS: An Intelligent Real-Time Anomaly Detection System for Cloud Communications Using Temporal Data Summarization and Neural Network

Gotam Singh Lalotra ¹, Vinod Kumar ², Abhishek Bhatt,³
Tianhua Chen,⁴ and Mufti Mahmud ^{5,6,7}

¹Department of Computer Science, Govt. Degree College Basohli, University of Jammu, Jammu, India

²Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

³Department of Electronics and Telecommunication Engineering, College of Engineering, Pune, India

⁴Department of Computer Science, School of Computing and Engineering, University of Huddersfield, UK

⁵Department of Computer Science, Nottingham Trent University, Nottingham, UK

⁶Medical Technologies Innovation Facility, Nottingham Trent University, Nottingham, UK

⁷Computing and Informatics Research Centre, Nottingham Trent University, Nottingham, UK

Correspondence should be addressed to Gotam Singh Lalotra; singh.gotam@gmail.com and Vinod Kumar; drvinodkumar2019@kluniversity.in

Received 20 September 2021; Revised 31 October 2021; Accepted 23 December 2021; Published 22 January 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Gotam Singh Lalotra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new distributed environment at less financial expenditure on communication over the Internet is presented by cloud computing. In recent times, the increased number of users has made network traffic monitoring a difficult task. Although traffic monitoring and security problems are rising in parallel, there is a need to develop a new system for providing security and reducing network traffic. A new method, iReTADS, is proposed to reduce the network traffic using a data summarization technique and also provide network security through an effective real-time neural network. Although data summarization plays a significant role in data mining, still no real methods are present to assist the summary evaluation. Thus, it is a serious endeavor to present four metrics for data summarization with temporal features such as conciseness, information loss, interestingness, and intelligibility. In addition, a new metric time is also introduced for effective data summarization. Finally, a new neural network known as Modified Synergetic Neural Network (MSNN) on summarized datasets for detecting the real-time anomaly-behaved nodes in network and cloud is introduced. Experimental results reveal that the iReTADS can effectively monitor traffic and detect anomalies. It may further drive studies on controlling the outbreaks and controlling pandemics while studying medical datasets, which results in smart healthy cities.

1. Introduction

In the last one and half decade, computer technology has significantly overpowered the conventional ways of handling the daily routines in almost every walk of life. All the daily routine activities like reading newspapers, shopping, running a business, studying, and a lot of official works have taken a shift over to the computer networks like LAN, WAN, MAN, Internet, and cloud computing. As these computer networks offer daily routine services to people across the

globe, at the same time, attackers have also joined the international community on the same platform, but to disturb the streamlined activities over the networks. These kinds of regular attackers or hackers not only affect the daily routine activities but also disturb the business or government networks [1]. Countering hackers and ensuring the smooth working of computer networks need the construction of a new security mechanism to provide security to the network users and their own secret stored data. For safe and secure communication services for exponential growing e-business

and electronic transactions, the use of intrusion detection and prevention systems, encryption, firewalls, authentication, and effective security mechanisms has been done [2]. Data have a significant role in every domain. The storage requirements over networks, in addition to the analysis of data, are of utmost importance to obtain knowledge [3]. A new secure arrangement method is presented, which is based on matrix eigenvalue. The aim of these arrangements is to generate a secret position for each user for storing data, which is called a secure arrangement [4]. Sometimes, the input data could be faster and easier to examine for attaining similar knowledge. For instance, a network administrator over the computer network requires surveillance and supervising the activities of the network [5]. Yet, for a small corporation network like HTTP, FTP, e-mail, and P2P applications, the quantity of data generated is enormous and difficult to analyze [6]. Moreover, the network traffic is increasing at a very rapid rate, which in turn becomes infeasible to monitor a network in real time by administrator [7]. Therefore, to analyze the current scenario of the network, a summary of the network traffic is quite useful to immediately review the situation of the network.

For a large volume of different kind of data generated from different resources like wireless sensors and cloud [8, 9], a summary is essential [10]. The intent of summarizing is to present a crisp dataset as input [11, 12]. Summarization is extensively traversed in different domains such as network data streams [11], intrusion detection systems (IDS) [13, 14], point of sales (POS) data [15], and natural text processing [16]. The summarization has been applied to various domains like healthcare, transport, security, logistics, and daily life [17] and has been demonstrated to be efficient in getting useful data out of huge datasets generated through IoT (Internet of Things) and cloud applications, which is easier to understand or interpret. It becomes even more important to summarize when the whole world is facing a pandemic outbreak and each and every sector of human life is affected.

In the same manner, a network sniffer protects and collects packets in an indiscriminate way, and an intrusion detection system (IDS) does the same. An IDS has the capability to detect different kinds of network attacks in the presented environment. The malicious network activities are identified by analyzing the packets collected by IDS, which gives alert signals to the system administrator and attack connections are blocked to avoid additional destruction from attacks. In general, intrusion detection algorithms are categorized as misuse detection and anomaly detection [18]. Misuse detection algorithm identifies attacks on the basis of known attack signatures. These algorithms are efficient in identifying known attacks with low errors. These algorithms are unable to identify newly created attacks, which do not have similar properties to the known attacks. On the other hand, the anomaly detection method relies on the hypothesis that the attackers have different behavior than a normal user. This paper, being part of *iReTADS*, presents the following contributions:

- (i) An existing metric named information loss has been modified that is biased towards recurring attributes and proposed a novel summarization technique that

is based on a newly defined time metric for data summarization purpose

- (ii) The newly proposed metric has been employed to split the dataset into different time intervals
- (iii) A novel method named modified synergetic neural network (MSNN) has been designed for effective anomaly detection

Further, the paper is organized as follows: Section 2 offers the related works. Section 3 discusses the overall system architecture. Section 4 explains the proposed work. Section 5 gives the results and discussion. Section 6 has the conclusion and future enhancements.

2. Related Works

The techniques of data summarization and anomaly detection have already been extensively researched. For association rule mining and clustering various data, summarization techniques are used, and different metrics have been proposed to improve the technique of data summarization. The authors in [19] have demonstrated that there are no universally excepted standards on the subject of what is a good summarization technique or a good summary. The aim of their technique is to represent a transactional database, implementing the notion of hyperrectangles, the Cartesian product of a set of transactions, and a set of items. To define the effectiveness of hyper and hyper+ techniques. They calculate the conciseness and the quality as the ratio of coverage per cost of each hyperrectangle so that the final summary cannot be compared to another summary as there is no measure to evaluate it. In [11], the authors explored the technique of compacting the specified number of transactions to a smaller set of summaries so that every summary entitles a subset of the input transactions in such a way that every transaction is represented in the summary. The original dataset is considered as the summary by the Bottom-Up Summarization (BUS) algorithm. In the beginning, frequent item set mining is employed on the input dataset, and then item sets are searched greedily, replacing those with minimum information loss and maximum compaction gain, and represented data points are replaced with them in the summary. The process is repeated till the desired compaction gain is achieved. Here, the metric used is the same for all the techniques of compaction gain. Information loss is also used by the authors to evaluate the results for measuring the amount of information not present in the summary of the original data. Table 1 represents some of the summarized contribution towards anomaly detection. The problem of summarization of a dataset of transactions, where two objective functions, compaction gain and information loss, were used with categorical attributes, is an optimization problem by authors in [11]. In order to describe the output of any summarization algorithm, a new metric was presented by them, and for addressing this problem, they investigated two approaches. In the first approach, clustering was implemented, and for the second approach, the frequent

TABLE 1: Summarized contribution towards the anomaly detection.

| SL | Title | Method | Dataset | Pros | Cons |
|----|--|---|--|--|---|
| 1 | RADS: Real-time anomaly detection system for cloud data centres [20] | OpenStack-based cloud data centre, one-class classification (RF, SVM, and naïve Bayes), and window-based time series analysis | Twitter dataset | Achieved 90–95% accuracy with a low FPR of 0–3% | Two metrics, precision and recall, need to be investigated for proper evaluation of the system |
| 2 | Real-time anomaly detection using ensembles [21] | Base learner (i) perceptron; (ii) ML-OzaBagadWin; (iii) binary class SVM | KDD CUP 99 | Accuracy attained 89.9% by MLP | Only a few base learners were used |
| 3 | A real time anomalies detection system based on streaming technology [22] | Spout architecture | Data are one-hour (22:00–23:00) exported flow data in <i>L</i> province, China | Real-time anomalies detection from mass stream data in a scalable manner | Up to 4 GB dataset is tested |
| 4 | Adapted K-nearest neighbors for detecting anomalies on spatio-temporal traffic flow [23] | K-nearest neighbors | Urban traffic flow Beijing dataset | Able to detect the real distribution of flow outliers. Outperforms the baseline algorithms for high-urban traffic flow | Does not work well with high dimensions because of the inherent feature of k-NN |
| 5 | Real-time anomaly detection based on long short-term memory and Gaussian mixture model [24] | LGMAD, based on long short-term memory (LSTM) and Gaussian mixture model (GMM) | NAB public dataset and synthetic dataset | A novel idea of the health factor <i>alpha</i> is proposed additionally to describe the health level of the system | Evaluated on precision, recall, F1-measure, and overlooked accuracy |
| 6 | Malware traffic classification using convolutional neural network for representation learning [25] | Convolution Neural Network (CNN) | USTC-TRC2016 traffic dataset | Malware traffic classification | Study the CNN parameters tunings |
| 7 | Adaptive real-time anomaly detection in cloud infrastructures [26] | Robust PCA and SVD | Amazon CloudWatch and Yahoo! | Accuracy: 87.24%; F-measure: 86% | Precision, recall, and metrics are ignored in evaluation |
| 8 | ADSaS: comprehensive real-time anomaly detection system [27] | Seasonal autoregressive integrated moving average (SARIMA) model and seasonal trend decomposition using loess (STL) | Numenta Anomaly Benchmark (NAB) | ADSaS performed well in terms of precision, recall, and F1-score | Error range variation is large in precision: 2.5%–97%; F1-score: 4.9%–95.1%; recall: 22.2%–100% |

item sets from the association analysis domain were used. In their work of summarization, they proposed one of the applications in the field of network data in which they showed how their technique could be efficiently used for summarizing network traffic into a meaningful and compact representation.

In [28], a recent investigation was done to find the possibility of anomaly detection in the context of real-time big data preprocessing and machine learning techniques. This survey includes the essential components of real-time processing of big data for anomaly detection, taxonomy of real-time big data processing, and various research challenges.

The authors in [29] studied the concept of information gain for network summarization and put forward a measure called information entropy for measuring the quality of a resolution. A probabilistic model of the information contained in a network was developed, and a formula is derived based on this model for information entropy. The network summarization method determines the computational complexity of computing network entropy; for simple deterministic node-reduction summarizations, they developed

an O(E) algorithm. In order to decide the most appropriate level of summarization, analysts use network entropy. With the help of information entropy, the information is measured over the network; this network information is combined with information provided by other attributes like geospatial labels for providing a complete scenario of the information enclosed in a particular network resolution.

In [30], a hierarchical data summarization data structure is presented; it was labeled hierarchical as the data structure implemented the concept of subcomponents to systematically attain conceptually larger components. The methods proposed herein acquire a bigger component repeatedly induced by the domain understanding of the users. So, for hierarchical data summarization, the rules implemented in the creation of data structure like $B+$ trees were also considered, and various data structures were implemented in hierarchical data summarization. Authors in [31] proposed estimation and a real-time loss performance monitoring scheme. Asymptotic relationship between the buffer size for both Markovian and self-similar traffic and Common Language Runtime was used in the proposed scheme. Results obtained by implementing the

proposed scheme showed that it required less monitoring time and obtained improved accuracy in comparison to the existing schemes.

The paper [32] presented a group of techniques and methods for traffic data collection, preprocessing, transformation, and integration till the data is forwarded for processing and transfer further for integration or fusion. Real-time data is very imperative for encouraging model accuracy, comprehensive use of assignment models, and historical traffic data for assisting online services and operations. The reliability of information and output from data fusion and processing as proposed by authors in [33] proposed the concern for analyzing the large network data for packet loss in real time, irrespective of any device installed at the network node in advance at monitoring place. However, it is found that the proposed system needs some sort of training in advance for adopting the features of the traffic to be monitored. The time series models are used for training which efficiently represent the high-speed traffic; with the help of these models, important conclusions like how to sample the data can be drawn by simulating the similar behavior shared by traffic.

This work [34] proposed four metrics, conciseness, information loss, interestingness, and intelligibility. These could be used for characterizing data summarization results. However, they modified the information loss metric because of its biased nature towards the recurring attributes. With the use of these four metrics, they assessed the existing summarization techniques on renowned network traffic datasets. A summarization method based on an already existing method was proposed, but here it is taken as an objective function; further, the classification of summarized datasets is carried out to reveal the usability of the metrics. Authors [35], with the help of the Bayesian Network, explained anomaly detection and getting learned by the real world automated identification system (AIS) data and from the additional data, resulting in the production of static and dynamic Bayesian network model. In their finding, they proposed that learned networks were pretty easy to inspect and verify in spite of the large number of variables being incorporated. In order to improve the anomaly detection performance, they confirmed the combination of both static and dynamic modeling approaches for improving the coverage of the overall model.

This work focused [36] on reducing security risk and presented two techniques of the network traffic anomaly detection in cloud communication, and these techniques, with the assistance of synergetic neural networks and the catastrophe theory, understand the dynamic behavior of the network traffic. In synergetic neural networks, a synergetic dynamic equation along with a set of ordered parameters is implemented for describing the complex nature of the network traffic system over cloud communications. Once this equation is solved, the ordered parameters confirmed by the primary factors can converge to 1, which results in anomaly detection. Catastrophe theory makes use of catastrophe potential function to explain the catastrophe dynamic process of the network traffic in cloud communications.

State of the network traffic derives from the normal one; whenever there is an anomaly in the network, the catastrophe distance index is used to assess the derivation, which helps in detecting the anomaly. They assessed the two approaches by implementing these techniques over standard Defense Advanced Research Projects Agency datasets, and it proved to be effective in detecting the network traffic anomaly and accomplished the high detection probability and low false alarm rate.

This contribution [37] presented a new increasing mapping-based hidden Markov model (IMHMM) in order to monitor the dynamic traffic efficiently. An increasing mapping is set up between the observation sequence and possible state sequence. In spite of FB variables, these mappings are used for obtaining the reestimation formulas for the model parameters. The IMHMM can be used to make fault detection and process monitoring framework to deal with large-scale dynamic processes. The IMHMM needs less storage space, and it is simple in comparison to HMM. Kim et al. [3] presented a new hybrid intrusion detection method that integrates hierarchically an anomaly detection model and a misuse detection model. Based on the C4.5 decision tree algorithm, a misuse detection model is built; later, using this model, normal training data is crumbled into smaller subsets. After that, these subsets are used to make multiple one-class SVM models. This method considerably optimizes the high time complexity of training and testing processes.

In this paper [38], a new technique SVM-L is given for anomaly detection in network traffic. Based on the concept of the dual formulation of kernel SVM and Linear Discriminant Analysis, an optimization model was proposed to adjust the hyperparameter of the classifier. Experimentally, 99% accuracy was claimed over network traffic dataset.

The work [39] proposed the ANN-based techniques for anomaly detection in Apache Spark, which works effectively for complex scenarios with multiple types of anomalies, like CPU contention, cache thrashing, and context switching anomalies, and showed 98–99% *F*-scores. Also, they claimed that a random duration, random start instant, and overlapped anomalies do not cause a significant influence on the performance of the proposed method.

The authors of [40] demonstrated the Convolution Neural Network features with bidirectional long short-term memory for real-time anomaly detection in the surveillance system. They claimed a 3.41% and 8.09% upsurge in accuracy on UCF-Crime and UCF-Crime2Local databases when compared to the newest methods.

The authors of [41] did a multiperspective review over smart anomaly detection in sensor systems and discussed the potential of computing (machine learning models), efficiency in communications medium, and engineering (constraints) in development of a smart anomaly detection system.

The research work [42] proposed a novel multistage anomaly detection ensemble technique named BFA-PDBSCAN for the incessant execution of computations on IoT-based applications. This selection of relevant features from the dataset is carried out by the Boruta method and extended k-medoid with a firefly-inspired strategy for

performing partitioning. They claimed the effectiveness of the proposed model over several datasets.

Blockchain and smart contract-based dependable and efficient lightweight certificateless signature (CLS) scheme, which is more secure than CLS protocol alone (Susceptible to security risks), is popular for resource-constrained Industrial Internet of Things (IIoT) protocol design [43].

This paper presents a new model for anomaly detection in the cloud using data summarization and neural network. Network traffic in the cloud is monitored with the help of data summarization; it also collects the necessary data. This summarized data further can be sent to the proposed Modified Synergetic Neural Network for anomaly detection.

3. Proposed Work

In this paper, a new model called *iReTADS* for detecting a real-time anomaly in the cloud during communications using data summarization and neural network with temporal features is introduced. Temporal Data summarization monitors the network traffic in the cloud in real time and collects the necessary data, and the summarized data can be sent to the proposed Modified Synergetic Neural Network for anomaly detection.

3.1. System Architecture of *iReTADS*. This proposed system architecture comprises seven key components, namely, data collection agent, cup dataset, network trace data, data summarization module, anomaly detection module, temporal information manager, and knowledge base, as shown in Figure 1.

3.1.1. Data Collection Agent. The network data are collected from the network layer or from the KDD'99 cup dataset by the data collection agents. This data is further sent to the data summarization module for summarizing the data.

3.1.2. Data Summarization Module. Data summarization module is comprised of three chief components as a quality threshold, that is, setting, optimization, and clustering. These components use different algorithms for quality threshold setting, optimization of the data based on the threshold, and clustering the data, and then this summarized data is sent for anomaly detection module.

3.1.3. Anomaly Detection Module. Anomaly detection module for efficient classification of the dataset makes use of proposed classification techniques. This module has a component that sets the rules on the basis of fuzzy concepts by integrating the various combinations of summarized datasets to have efficient classification. For setting the time interval, this module bears the responsibility of remaining in contact with the temporal information manager.

3.1.4. Temporal Information Manager. This module has the accountability of allotting the time interval for summarizing the data; with the assistance of the knowledge base, the time-

based fuzzy rules also formed the knowledge base. The knowledge base has a set of rules to answer the queries being fired by the users and execute efficient decision-making. It has contained rules in order to make a decision regarding the summarization process and classification. The decision manager manipulates and maintains the knowledge base.

3.1.5. Decision Manager. The whole process of this proposed system is monitored by the decision manager. In collaboration with the temporal information manager and knowledge base, the decision manager takes all the decisions regarding the classification and data summarization. It has all the control over data summarization, collection agent, and anomaly detection module.

3.2. Data Summarization Technique. This section is presented with a proposed metric along with four existing metrics in order to serve the purpose of data summarization. Although it is well noted that the data summarization obtains the data as input and outputs the data as well, the output cannot be misinterpreted with information or knowledge. Summary here is simply a precise form of the input data, which is made to use as a replacement for competence reasons. As a result, the whole of the measures dealing with data summarization should be objective in nature. This section is comprised of five objective measures for data summarization, namely, time, conciseness, information loss, interestingness, and intelligibility.

3.2.1. Time. This paper presents a new metric, time, which is used for data summarization. In order to monitor traffic, the time interval is a very significant parameter, as a large number of the users access the network or cloud, and many times a situation arises when the network is not traffic-free. In such conditions, when online traffic is being monitored, it should be managed with the help of time intervals amongst the data groups. Already existing techniques are performed remarkably well, even without concentrating on real-time traffic control. These existing techniques have not considered the retrieval time and randomly retrieving the information from the database. Datasets could be prepared between the particular time intervals, whether it is one week, two weeks, one day, or one hour from the server. From datasets, we can find different numbers of time appearances in various time slots. Based on this, the metric subset has to be broken based on the time intervals; using this approach, data can be retrieved based on the exact data occurrence in various time intervals.

3.2.2. Conciseness. The metric conciseness explains the compact data summary as per the dataset; conciseness is discussed and explained with different terminologies like summarization, compression ratio, and compaction gain in various works and different papers like [11, 34, 44]. In another paper, the conciseness is calculated for a set of records at a specified time interval in accordance with the [25]; the calculation is done in a similar manner as of three

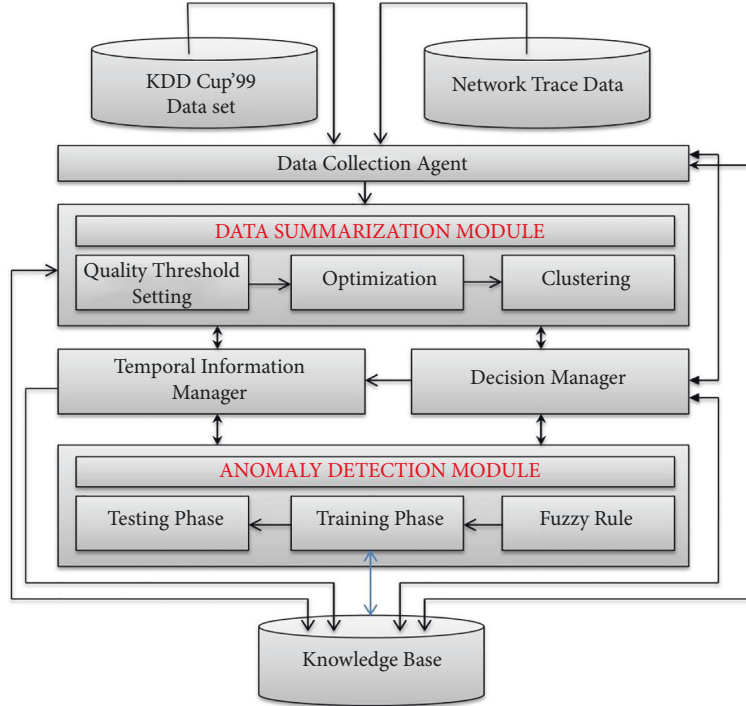


FIGURE 1: System architecture.

previous approaches, which is expressed as the ratio of the input dataset size to the summarized dataset size. Let M be the number of data points in the input dataset and S be the number of tuples in the summary. Starting time and ending time are denoted by t_1 and t_2 .

$$\text{Conciseness}(t_1, t_2) = \frac{M}{S}. \quad (1)$$

3.2.3. Information Loss. Information loss denoted the amount of information loss in the process or in very simple terms the information that is absent in the summary. According to Ha-Thuc et al. [44], the amount of information lost is known as distortion. As per their definition, distortion is the total of the squared Euclidean distance between every data point and centroid of the cluster to which these points belong. It is quite evident from the definition of distortion in this particular work [44] that this method of information loss calculation can only be used in clustering-based summarization, hence not applicable as a general summarization metric. As per Chandola and Kumar [11, 44], information loss is given by an individual summary for a transaction as “the weighted sum of all the features which are absent in the individual summary.” Hence, aggregating the information loss of each transaction results in the total information loss of the summary. In this work, the information loss for the set of records is calculated in accordance with the work of Chandola and Kumar [11, 44] for the specific time interval further; the results are normalized. Sum of all the ratios of attributes not present to the total attributes represented per summary gives the information loss. Let S be the number of individual summaries, t_i be the number of different

attributes represented by summary i , and l_i be the number of different attributes not present in summary i . Starting time and ending time will be indicated by t_1 and t_2 . Then,

$$\text{Information Loss}(t_1, t_2) = \frac{1}{s} \sum_{i=1}^s \frac{l_i}{t_i}. \quad (2)$$

3.2.4. Interestingness. The interestingness metric has been discussed extensively in literature; it has been an area of interest for researchers, particularly for finding interesting classification and association rules in data mining [2, 45]. Interestingness is taken up as a broad concept in the literature. It focuses on conciseness, peculiarity, coverage, diversity, reliability, surprisingness, utility, novelty, and actionability. Hoplaros et al. [34] defined interestingness as follows:

$$\text{IRAE}(t_1, t_2) = \sum_{i=1}^m \frac{n_i(n_i - 1)}{N(N - 1)}. \quad (3)$$

Let n_i be the derived count attribute of a summary tuple, m be the number of tuples in a summary, and N be the number of total input data points. The starting time and the ending time are represented by t_1 and t_2 .

3.2.5. Intelligibility. Intelligibility defines the characteristics of the summary, that is, the level of sense a summary makes out of the data, on account of the number of ANY attributes present in the summary. Every tuple in the summary represents a subset of the input dataset. If a tuple contains the most closely related data, then there will be fewer ANY

attributes present. Let m , be the number of tuples in a summary the i tuple having the attributes n_i , and l_i be the number of non-ANY attributes present in tuple i . As per our approach towards information loss and interestingness, intelligibility should be normalized. We define

$$\text{Intelligibility}(t_1, t_2) = \frac{1}{m} \sum_{i=1}^m \frac{l_i}{n_i}. \quad (4)$$

3.3. Temporal Data Summarization Algorithm. This algorithm is comprised of three phases, namely, threshold setting, optimization, and clustering. All these three phases make use of various algorithms proposed by different authors on the basis of different metrics. These algorithms are modified in our proposed method by using temporal constraints.

3.4. Proposed Temporal Data Summarization Method. This proposed real-time data summarization method according to [34] is the combination of four phases. These four phases also contain different algorithms, namely, modified quality threshold summarization algorithm, BUS algorithm, K-means clustering algorithm, and data summarization algorithm. We used the data summarization technique, which was proposed by Hoplaros et al. [34]. A new metric called time is introduced for handling real-time data. This new metric plays a major role in the quality threshold summarization algorithm and summarization algorithm, which are present in phase 1 and phase 2, respectively. This metric uses a modified data summarization method that plays necessary roles for handling real-time data in cloud. The proposed data summarization method has introduced a new metric called time interval in Algorithm 1. This new metric is used in the quality threshold algorithm.

3.5. Modified Synergetic Neural Network. This section presents a new system for anomaly detection in network and cloud communications which is a combination of temporal data summarization and a Modified Synergetic Neural Network. A Modified Synergetic Neural Network is introduced according to [46]. A new layer based on fuzzy rules for all sets of data is introduced in the framework of SNN in the starting and ending time based on Ganapathy et al. [47]. For anomaly detection, fuzzy intervals will be implemented for making efficient and appropriate decisions; when it is compared with time series, these fuzzy rules have been framed on the basis of different time intervals. There is no denying the fact about the dynamic nature of the network traffic; it is not only dynamic in nature but also complex dynamic. The network traffic has shown nonlinear, non-stationary, and complex dynamic behavior. There are many factors involved in describing the behavior or nature of this network at the broader level [1, 48]. Therefore, the network problem is treated as a high-dimensional problem. Network generation in the cloud communication environment is a task that involves many factors, and with the assistance of all

these factors or parameters, it could be achieved. There are various factors that dominate the network equally, and the changes over the network are normal, but the network traffic reflects large randomness. Attackers or abnormal users dominate the key factor because when anomalies happen, all the above-discussed factors do not contribute to the network traffic at par. The network at an abnormal state shows strong certainty. Synergetic, says primary, factors contribute to the generation of the order parameter. Randomness and similarity in the cloud communication and network traffic are the characteristics of order parameters in the network. Interdisciplinary science named synergetic demonstrates the organization and formulation of structures and patterns in an open system, which are far from thermodynamics equilibrium. This science focuses on bringing temporal, spatial, and functional structures on macroscopic scales of the various individual factors of a dynamic system. According to synergetic [46], a dynamical system can be expressed as follows:

$$\begin{aligned} q &= -\frac{\partial V}{\partial q^+}, \\ q^+ &= -\frac{\partial V}{\partial q}, \end{aligned} \quad (5)$$

where q is the system state; q^+ is the adjoint state of q ; V is the potential function of the system; q is the differential of q ; and the other is the same on the following equations in this paper.

According to the control principle of synergetic, stable models have a dependency on the unstable models. Whenever there is the process of evolution of the system, the number of certain unstable models keeps on increasing; on the other hand, the number of stable models starts decreasing. When the number of unstable models becomes large, they start behaving as the primary factor in the system, which, as a consequence, transforms the high-dimensional problem into the low dimension problem and the values of the unstable models known as order parameters. These unstable models with the highest original order parameter decide the final state of the system. Synergetic science explains the fundamental building principle for pattern recognition and comes up with an opinion: the process of pattern recognition is pattern formation, which is a top-down approach for analyzing a problem or system. How pattern recognition works are explained here, macroqualitative variation of the system corresponds to the pattern formation, and transformation of the process from testing data to the training data is equivalent to pattern recognition, which reflects the similarity between pattern recognition and pattern formation. MSNN is the technique of pattern recognition when it comes to network traffic anomaly detection in cloud communication based on MSNN. Identified patterns and prototype patterns are presented by testing data and training data, respectively, in the process of MSNN. And the technique to identify the testing data is to map the testing data to some already existing training dataset. The

Phase 1: modified quality threshold summarization algorithm.
 Input: dataset D , threshold T , time interval.//Threshold setting will be different for different time interval summarized data.
Output: cluster centroids $\{C_1, C_2, \dots, C_k\}$
 Step 1: initialize the cluster centroid $C = \emptyset$;
 Step 2: initialize the threshold $t = 0$;
 Step 3: initialize the $k_0 = 1$;
 Step 4: choose data item from D and set $I_0 = d$ for the particular time interval.
 Step 5: $(C_t, E_t, \langle t_1, t_2 \rangle) = K\text{-means}(D, k_t, I_t)$;
 Step 6: $K_{t+1} = K_t$;
 Step 7: $I_{t+1} = C_t$;
 Step 8: for $i \leftarrow 1$ to k_t do
 Step 9: if E_t less than T , then
 Step 10: $C = C \cup \{C_{ii}\}$
 Step 11: remove cluster i out of D
 Step 12: $K_{t+1} = K_{t+1} - 1$
 Step 13: $I_{t+1} = I_{t+1} - \{C_{ii}\}$
 Step 14: end
 Step 15: if D equals \emptyset , then
 Step 16: return centroids $\{C_1, C_2, \dots, C_k\}$;
 Step 17: randomly choose a data point d approximately close to the centroid of the largest cluster;
 Step 18: insert d to I_{t+1} ;
 Step 19: $k_{t+1} = k_{t+1} + 1$;
 Step 20: $t = t + 1$;
 Step 21: go to 5;
 Phase 2: apply BUS algorithm
 Phase 3: apply K-means clustering
 Phase 4: apply data summarization algorithm that incorporates all metrics [29].

ALGORITHM 1: Temporal data summarization method.

identification of testing network traffic patterns q can be explained as a dynamic process in cloud communications. After mapping the initial testing data $q(0)$ from intermediate state $q(t_1, t_2)$ to a training data vector v_k , the training data vector v_k is most near to $q(0)$. The process can be described as $q(0) \rightarrow q(t_1, t_2) \rightarrow v_k$, where $q(0)$ is the testing network traffic data, v_k is the stored normal or abnormal traffic network traffic, and the intermediate state $q(t_1, t_2)$ is the order parameter ω_k . To be precise, this process can be represented by a dynamic equation (2). The assumption is made that the number of the training data vectors is M and the dimension of the training data vector is N . For maintaining the linear independence of the M training data vector, $M \leq N$ is required.

$$q = \sum_{k=1}^M \gamma_k V_k (V_k^+ q) - B \sum_{k=1}^M \sum_{k=1, k \neq 1}^M (V_k^+ q) 2 (V_k^+ q) V_k - C (q^+ q) q + F(t_1, t_2), \quad (6)$$

where q is the testing network traffic data vector in cloud communications with the original input data value $q(0)$. Scalar value γ_k is the attention parameter because when it is positive, only testing data can be identified. $F(t_1, t_2)$ is the fluctuation factor for the particular record login and logout time of the network traffic in cloud communications and can be ignored. Scalar values B and C are specified coefficients and must be greater than 0. v_k is the training data vector, $v_k = (v_{k,1}, v_{k,2}, \dots, v_{k,N})^T$, where superscript T is vector transposition. v_k^+ is the adjoint vector of v_k .

$$V_k^+ V_{k'} = \delta_{k,k'} = \begin{cases} 1, & k = k', \\ 0, & k \neq k'. \end{cases} \quad (7)$$

v_k should be prepared with normalization and zero mean:

$$\sum_{l=1}^N V_{k,l} = 0, \quad \sqrt{\left(\sum_{l=1}^N V_{k,l}^2 \right)} = 1. \quad (8)$$

In order to reduce the dimensionality of the system, the order parameters γ_k are features extracted from vector q . q can be represented by the order parameters γ_k , a training data vector v_k , and the remaining vector w :

$$q = \sum_{k=1}^M \gamma_k V_k + W, \quad V_k^+ W = 0. \quad (9)$$

The adjoint vector of q is defined as follows:

$$q^+ = \sum_{k=1}^M \gamma_k V_k^+ + W^+, \quad W_k^+ V_k = 0. \quad (10)$$

There is a relationship:

$$V_k^+ q = q^+ V_k. \quad (11)$$

Typing (5) into (7), according to the orthogonal relationship, the order parameter γ_k is defined as follows:

$$\gamma_k = V_k^+ q. \quad (12)$$

Style described in (2) is a powerful dynamic equation. If we neglect the fluctuation factor $F(t_1, t_2)$ during the particular time interval of the network traffic in cloud communications, according to equations (1) and (2), the potential function can be described as follows:

$$V = -\frac{1}{2} \sum_{k=1}^M \gamma_k (V_k^+ q)^2 + \frac{1}{4} B \sum_{k=1}^M \sum_{k'=1, k' \neq k}^M (V_k^+ q)^2 + \frac{1}{4} C \left(\sum_{k=1}^M (V_k^+ q)^2 \right). \quad (13)$$

According to equations (1), (2), and (8), correspondingly, the dynamic equations and the potential function described by the order parameter are as follows:

$$\omega_k = \gamma_k \omega_k - B \sum_{k'=1, k' \neq k}^M \omega_k^2 \omega_k - C \left(\sum_{k'=1}^M \omega_k^2 \right) \omega_k, \quad (14)$$

$$V = -\frac{1}{2} \sum_{k'=1}^M \gamma_k \omega_k^2 + \frac{1}{4} B \sum_{k=1}^M \sum_{k'=1, k' \neq k}^M \omega_k^2 \omega_k^2 + \frac{1}{4} C \left(\sum_{k'=1}^M \omega_k^2 \right)^2. \quad (15)$$

At the lowest potential energy of a system, the system controlled by the order parameters reaches the most stable state. Here, the order parameters attain their extreme value. The stable state of the network system is described by the following formula:

$$\omega_k = 0, \quad 0 \leq k \leq M. \quad (16)$$

That is,

$$\omega_k = \gamma_k \omega_k - B \sum_{k' \neq k}^M \omega_k^2 \omega_k - C \left(\sum_{k'=1}^M \omega_k^2 \right) \omega_k = 0. \quad (17)$$

If we define

$$D = (B + C) \sum_{k'=1}^M \omega_k^2, \quad (18)$$

then the following equations can be inferred from (10) and (12):

$$\begin{aligned} \omega_k &= \omega_k (\gamma_k - D + B \omega_k^2), \\ \omega_k (\gamma_k - D + B \omega_k^2) &= 0. \end{aligned} \quad (19)$$

As per the (15), there are four layers in the MSNN architecture in Figure 2. The top layer is the input layer in which unit j receives component $q_j(0)$ of need recognized pattern vectors original value $q(0)$. All the order parameter components form the middle layer, where the order parameter ω_k is obtained by summing all angle indexes j through each input value $q_j(0)$, multiplying its joint unit v_{kj}^+ . The active order parameter ω_k recognizes the special training data chosen by the angle index k . According to the dynamical equation, MSNN will be evolved to the end state

that only one of the order parameters survives, and q_j is obtained through reciprocity and competition of D . The down layer is the output layer in which output pattern can be expressed as $q_j(t_1, t_2) = \sum_k \omega_k(t_1, t_2) V_{kj}$, where q_j is active of output unit j and ω_k is the end of the state of the middle layer. There is $\omega_k = 1$ if $k = k_0$; otherwise, $\omega_k = 0$. v_{kj} is the component of j of the training data vector.

Time series of the network traffic in cloud communications are represented by y_1, \dots, y_N ; these are sampled in bytes or bits or packets per time unit. The fuzzy temporal information manager of the proposed system MSNN deals with the fuzzy time interval for detecting anomalies. Fuzzy time intervals are set up as shown in Figure 3. Normal and abnormal network traffic are set for constructing a training dataset that may contain M components from a specific time interval. Network traffic fuzzy time intervals also share the same size N . This proposed anomaly detection method is designed in such a way to make a distinction between normal and abnormal network traffic.

The process of anomaly detection includes two stages as shown in Figure 4: the training stage is to learn the training data of the normal and abnormal network traffic, and the testing stage is to detect the network traffic anomaly. The detailed detection steps are given as follows.

3.5.1. The Training Stage

- (a) Choose the training data vectors $\{y_1, \dots, y_N\}$ from the trained dataset for specific time intervals
- (b) Deal with the training pattern vectors $\{y_1, \dots, y_N\}$ with normalization and zero mean and then compute the training data vectors v_k for the particular time interval records
- (c) Compute the corresponding adjoint v_k^+ of the training data vectors v_k at particular time intervals

3.5.2. The Testing Stage

- (a) Test on the testing data vector $q(0)$ consisted of the testing network traffic data in cloud communications dealt with normalization and zero mean.
- (b) Compute the corresponding order parameter ω_k of each training data according to (8), which is in the particular time interval.
- (c) Evolve by the following order parameter dynamic equation (17) until the order parameters start converging to a specific training data and then to the specific training data vector $q(0)$ in a particular time interval. Thus, the processes of the network traffic anomaly detection in cloud communications based on MSNN have been completed.

$$\omega_k(n+1) - \omega_k(n) = \beta (\gamma_k - D + B \omega_k^2(n)) \omega_k(n), \quad (20)$$

where β is the iterative step.

The steps of the training stage are as follows:

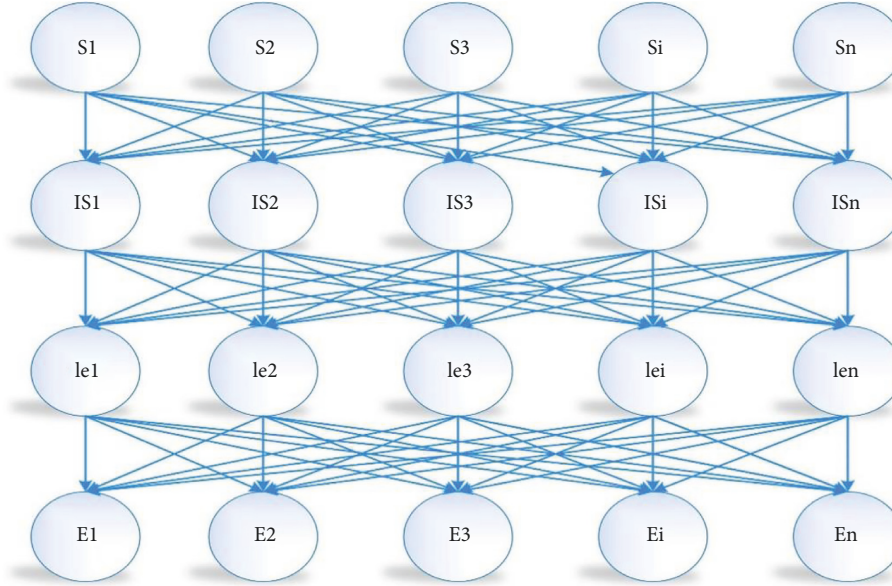


FIGURE 2: The framework of MSSN.

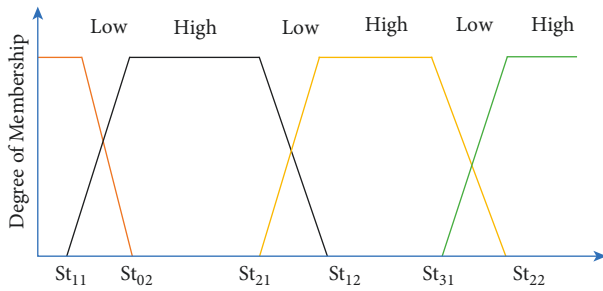


FIGURE 3: Fuzzy time interval.

- (1) Consider the fuzzy time intervals y_1, \dots, y_N of the training data, for each time interval between t_1 and t_2 , and construct the vector set $\{Y_t^p = (y_{t-p+1}, \dots, y_t) \mid t = 1, \dots, N-p+1\}$, which is the time interval window
- (2) Obtain the set of the state variable $\{x_t\}$ and the control variables $\{u_t\}$ and $\{v_t\}$ based on normalized features extracted from each vector Y_{pt}
- (3) Compute the parameters a and b of the cusp catastrophe model using the series $\{x_t\}$, $\{u_t\}$, and $\{v_t\}$

In the testing stage, the main steps are as follows:

- (1) Construct the vector Y_t^p (with the same time window Win_p in the training stage) of the testing data at the observed time I , which is labeled as observed point P_i .
- (2) Extract the selected normalized features to present the state variable x_i and control variables u_i and v_i .
- (3) Compute the catastrophe distance between the observed point $P_i(x_i, u_i, v_i)$ and the bifurcation set $G(x, u, v)$, labeled as D_p . The catastrophe distance D_p is defined as follows: assuming that $P_i(x_i, u_i, v_i)$ is an observed point in the testing data of the traffic in cloud communications and $P_t(x_t, u_t, v_t)$ is a point of the equilibrium surface $G(x, u, v)$, the distance

between two points $P_i(x_i, u_i, v_i)$ and $P_t(x_t, u_t, v_t)$, labeled as $D_E(P_i, P_t)$ is computed by the Minkowski distance. The catastrophe distance D_p between the observed point $P_i(x_i, u_i, v_i)$ and the equilibrium surface $G(x, u, v)$ is defined as

$$D_p(P_i, G(x, u, v)) = \min_{P_t \in G(x, u, v)} \{D_E(P_i, P_t)\}. \quad (21)$$

As catastrophe distance D_p is more than a threshold, then anomaly can be claimed at the observing point $P_i(x_i, u_i, v_i)$. The threshold is obtained by training.

4. Results and Discussion

In this section, the different experimental results carried out for data summarization and anomaly detection have been discussed. For data summarization, different experiments by using the four different metrics in algorithms called quality threshold algorithms (BUS algorithm and K-means clustering algorithm) have been performed. Finally, these algorithms are combined to propose Modified Synergetic Neural Network for providing better classification accuracy.

The proposed method is the combination of the very well-known and state-of-the-art techniques and, while using the strength of the neural network, gives a very good performance.

4.1. Datasets. The dataset for this experiment is taken from the third International Knowledge Discovery and Data Mining Tools Competition (KDD Cup 99) [18, 49]. Every connection record is characterized by 41 attributes. All these attributes are both discrete and continuous in nature; these variables are drastically varying to each other on the basis of statistical distributions, turning it to be a challenging task for intrusion detection [50].

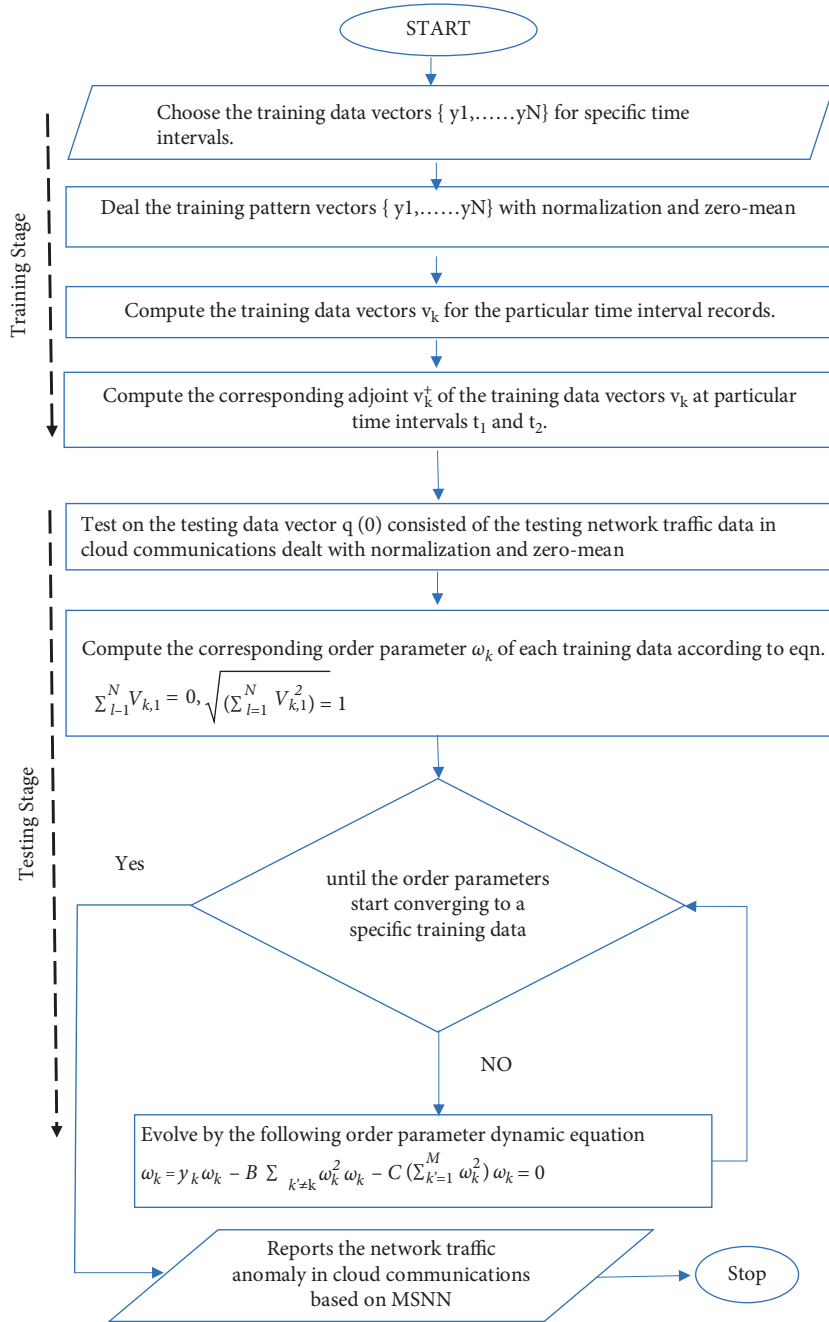


FIGURE 4: MSNN process.

This dataset contains 5 million network connection records like password guess, land attack, Neptune attack, and port scan. The twenty-two categories of attack are from the following four classes: DOS, R2L, U2R, and probe. The 41 features explain the fundamental information regarding network packet, network traffic, host traffic, and content information. Each record has 5 class labels, namely, normal, probe, DOS, R2L, and U2R. It has 391458 DOS attack records, 52 U2R attack records, 4107 probe attack records, 1126 R2L attack records, and 97278 normal records only in 10 percent of this dataset.

4.2. *Experimental Results.* Table 2 shows the performance of quality threshold data summarization. From this table, we can see the different four metrics such as conciseness, information loss, interestingness, and intelligibility values when considering the different threshold values with a number of clusters considered for summarization of data.

Table 3 shows the performance of BUS in data summarization. Here, the different four metrics such as conciseness, information loss, interestingness, and intelligibility values were obtained when considering the different kinds of summary sizes for summarization of data.

TABLE 2: Results for quality threshold summarization phase on the KDD Cup'99 dataset.

| Threshold | Clusters | Conciseness | Information loss | Interestingness | Intelligibility |
|-----------|----------|-------------|------------------|-----------------|-----------------|
| 15000 | 12 | 10493.31 | 0.9891 | 0.14876 | 0.2712 |
| 10000 | 16 | 7869.52 | 0.9893 | 0.088 | 0.29513 |
| 5000 | 24 | 5243.97 | 0.98267 | 0.06621 | 0.3068 |
| 2500 | 43 | 2925.04 | 0.97348 | 0.04912 | 0.3579 |
| 1000 | 75 | 1675.42 | 0.9651 | 0.03017 | 0.3745 |
| 500 | 124 | 1010.128 | 0.9567 | 0.02197 | 0.40649 |
| 250 | 187 | 669.47 | 0.94652 | 0.01662 | 0.4392 |
| 100 | 297 | 420.512 | 0.9262 | 0.0119 | 0.46725 |
| 50 | 437 | 284.17 | 0.89734 | 0.0065 | 0.48343 |

TABLE 3: Results for BUS phase on the KDD Cup'99 dataset.

| Summary size | Conciseness | Information loss | Interestingness | Intelligibility |
|--------------|-------------|------------------|-----------------|-----------------|
| 12 | 10493.31 | 0.9891 | 0.2225 | 0.2624 |
| 16 | 7869.52 | 0.9893 | 0.09351 | 0.29132 |
| 24 | 5243.97 | 0.98267 | 0.12832 | 0.3265 |
| 43 | 2925.04 | 0.97348 | 0.01284 | 0.3768 |
| 75 | 1675.42 | 0.9651 | 0.095721 | 0.2763 |
| 124 | 1010.128 | 0.9567 | 0.03672 | 0.4216 |
| 187 | 669.47 | 0.94652 | 0.07419 | 0.4552 |
| 297 | 420.512 | 0.9262 | 0.2032 | 0.26821 |
| 437 | 284.17 | 0.89734 | 0.05071 | 0.52242 |

TABLE 4: Results for K-means clustering phase on the KDD Cup'99 dataset.

| Clusters | Conciseness | Information loss | Interestingness | Intelligibility |
|----------|-------------|------------------|-----------------|-----------------|
| 12 | 10497.75 | 0.99242 | 0.11231 | 0.30672 |
| 16 | 7873.3125 | 0.98968 | 0.08921 | 0.34128 |
| 24 | 5248.875 | 0.98151 | 0.0762 | 0.39312 |
| 43 | 2929.6046 | 0.9778 | 0.0523 | 0.43345 |
| 75 | 1679.64 | 0.96495 | 0.0348 | 0.49213 |
| 124 | 1015.9112 | 0.95392 | 0.0217 | 0.53279 |
| 187 | 673.6524 | 0.94646 | 0.0108 | 0.5725 |
| 297 | 424.1515 | 0.93889 | 0.0087 | 0.6218 |
| 437 | 288.2677 | 0.93130 | 0.0056 | 0.6617 |

Table 4 shows the performance of the K-means clustering algorithm in data summarization. From this table, we can see the different four metrics such as conciseness, information loss, interestingness, and intelligibility values during the consideration of different thresholds for the different number of clusters for summarization of data.

Figure 5 shows the comparison of performance analysis between the existing quality threshold algorithm and the combination of the proposed MSNN with the quality threshold algorithm. Figure 2 explains that the MSNN framework with quality algorithm has outperformed the existing quality threshold algorithms in terms of classification accuracy. The proposed anomaly detection method provides better anomaly detection accuracy significantly while considering different thresholds for anomaly detection.

Figure 6 shows the comparison of performance analysis between the existing K-means clustering algorithm and the combination of the proposed MSNN with the K-means clustering algorithm.

From Figure 6, it can be observed that the classification accuracy of the proposed MSNN with the K-means

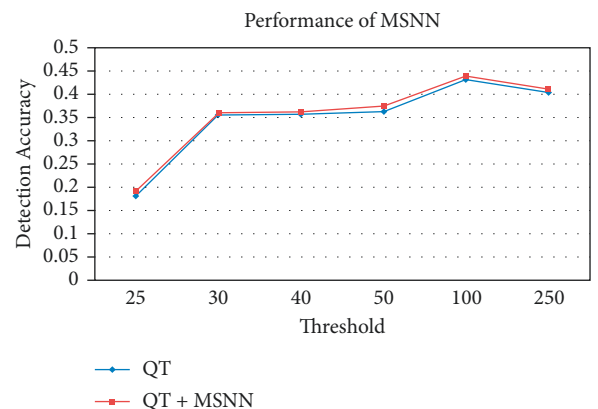


FIGURE 5: Comparison of performance analysis between QT and MSNN with QT algorithm.

clustering algorithm is better than the existing K-means clustering algorithm. The proposed anomaly detection method provides better anomaly detection accuracy quite significantly during the consideration of different thresholds for anomaly detection.

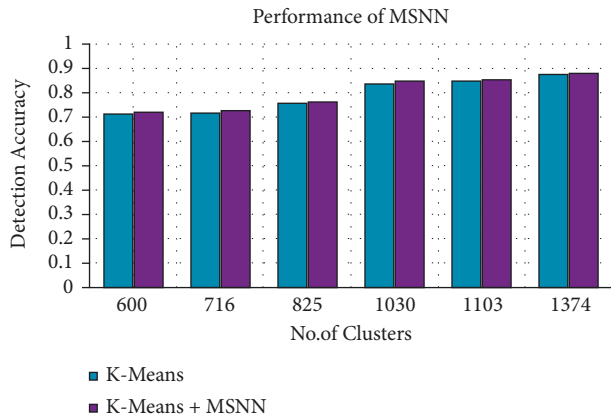


FIGURE 6: Comparison of performance analysis between K-means and MSNN with K-means.

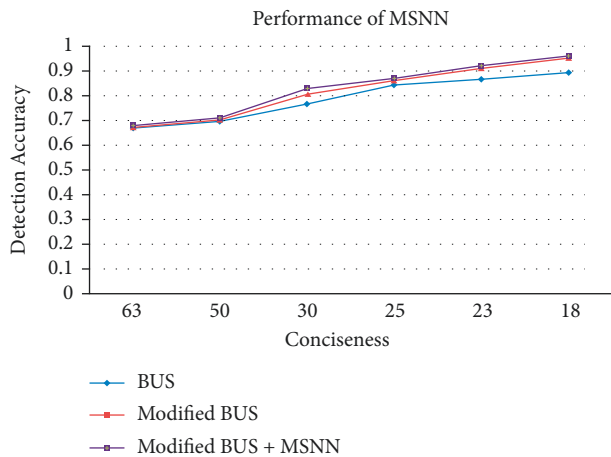


FIGURE 7: Comparison of performance analysis between BUS, modified BUS, and MSNN with modified BUS.

Figure 7 shows the comparison of performance analysis for the existing BUS algorithm, modified BUS algorithm, and the proposed MSNN with modified BUS algorithm. From this figure, it can be observed that the classification accuracy of the proposed MSNN with modified BUS algorithm is better than the existing BUS algorithm and modified BUS algorithm. The proposed anomaly detection method provides better anomaly detection accuracy significantly during the consideration of different thresholds for anomaly detection.

5. Conclusions and Future Work

This paper presents *iReTADS*, an intelligent real-time anomaly detection technique. As a part of it, a new metric time interval is introduced for data summarization in addition to four existing metrics for the same purpose. We proposed a novel neural network framework with fuzzy temporal features comprised of four layers, and this handles the fuzzy time interval for classification. Finally, the demonstration for classification of summarized datasets using the proposed neural network was carried out for assessing its effectiveness. Time taken while using summarized data can

be a fraction of the total time taken over the original dataset, getting approximately the same results, which can save time in a critical application. These methods should be optimized and parallelized. However, the system is tested with large datasets of network traffic, which revealed another necessity. More focus should be given to real-time anomaly detection, and the research efforts will be directed to stream data summarization and anomaly detection methods. In future works, we will explore a new effective real-time anomaly detection method using soft computing techniques.

Data Availability

The data used in this study are available at <https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] C. You and K. Chandra, "Time series models for internet data traffic," in *Proceedings of the 24th Conference on Local Computer Networks. LCN'99*, pp. 164–171, Lowell, MA, USA, October 1999.
- [2] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1690–1700, 2014.
- [3] C. Zins, "Conceptual approaches for defining data, information, and knowledge," *Journal of the American Society for Information Science and Technology*, vol. 58, 2007.
- [4] J. Song, Z. Han, W. Wang, J. Chen, and Y. Liu, "A new secure arrangement for privacy-preserving data collection," *Computer Standards & Interfaces*, vol. 80, Article ID 103582, 2022.
- [5] Q. Tao, G. Xiaohong, L. Wei, and W. Pinghui, "Monitoring abnormal traffic flows based on independent component analysis," in *Proceeding of the 2009 IEEE International Conference on Communications*, Dresden, Germany, June 2009.
- [6] K. Keys, D. Moore, and C. Estan, "A robust system for accurate real-time summaries of internet traffic," *ACM SIGMETRICS - Performance Evaluation Review*, vol. 33, no. 1, pp. 85–96, 2005.
- [7] Z. Lv, L. Wang, Z. Guan et al., "An optimizing and differentially private clustering algorithm for mixed data in sdn-based smart grid," *IEEE Access*, vol. 7, Article ID 45773, 2019.
- [8] H. Patel, D. Singh Rajput, G. Thippa Reddy, C. Iwendi, A. Kashif Bashir, and O. Jo, "A review on classification of imbalanced data for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 4, 2020.
- [9] D. S. Rajput, S. M. Basha, Q. Xin et al., "Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2021.
- [10] P. Karras, "Multiplicative synopses for relative-error metrics," in *Proceedings of the Twelfth International Conference on Extending Database Technology: Advances in Database Technology*, pp. 756–767, Saint Petersburg, Russia, March 2009.

- [11] V. Chandola and V. Kumar, "Summarization - compressing data into an informative representation," *Knowledge and Information Systems*, vol. 12, no. 3, pp. 355–378, 2007.
- [12] R. Saint-Paul, G. Raschia, and N. Mouaddib, "General Purpose Database Summarization," in *Proceedings of the 31st International Conference on Very Large Data Bases*, pp. 733–744, Citeseer, Trondheim Norway, September 2005.
- [13] A. Singhal, *Data warehousing and data mining techniques for cyber security*, Springer Science & Business Media, vol. 31, New York, NY, USA, , 2007.
- [14] R. Zhu, "Intelligent rate control for supporting real-time traffic in wlan mesh networks," *Journal of Network and Computer Applications*, vol. 34, no. 5, pp. 1449–1458, 2011.
- [15] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216, Washington D.C. USA, May 1993.
- [16] L. Yu and F. Ren, "A Study on Cross-Language Text Summarization Using Supervised Methods," in *Proceedings of the 2009 International Conference on Natural Language Processing and Knowledge Engineering*, pp. 1–7, IEEE, Dalian, China, September 2009.
- [17] D. S. Rajput and R. Gour, "An IoT framework for healthcare monitoring systems," *International Journal of Computer Science and Information Security*, vol. 14, no. 5, p. 451, 2016.
- [18] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An intelligent intrusion detection system (ids) for anomaly and misuse detection in computer networks," *Expert Systems with Applications*, vol. 29, no. 4, pp. 713–722, 2005.
- [19] Y. Xiang, R. Jin, D. Fuhry, and F. F. Dragan, "Summarizing transactional databases with overlapped hyperrectangles," *Data Mining and Knowledge Discovery*, vol. 23, no. 2, pp. 215–251, 2011.
- [20] S. Barbhuiya, Z. Papazachos, P. Kilpatrick, and D. S. Nikolopoulos, "Rads: Real-time anomaly detection system for cloud data centres," 2018, <https://arxiv.org/abs/1811.04481>.
- [21] R. R. Reddy, Y. Ramadevi, and K. Sunitha, "Real time anomaly detection using ensembles," in *Proceedings of the 2014 International Conference on Information Science & Applications (ICISA)*, pp. 1–4, IEEE, Seoul, Republic of Korea, May 2014.
- [22] Y. Du, J. Liu, F. Liu, and L. Chen, "A real-time anomalies detection system based on streaming technology," vol. 2, pp. 275–279, in *Proceedings of the Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 2, IEEE, Hangzhou, China, August 2014.
- [23] Y. Djenouri, A. Belhadi, J. C.-W. Lin, and A. Cano, "Adapted K-nearest neighbors for detecting anomalies on spatio-temporal traffic flow," *IEEE Access*, vol. 7, Article ID 10015, 2019.
- [24] N. Ding, H. Ma, H. Gao, Y. Ma, and G. Tan, "Real-time anomaly detection based on long short-term memory and Gaussian mixture model," *Computers & Electrical Engineering*, vol. 79, Article ID 106458, 2019.
- [25] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in *Proceedings of the International Conference on Information Networking (ICOIN)*, pp. 712–717, IEEE, Da Nang, Vietnam, January 2017.
- [26] B. Agrawal, T. Wiktorski, and C. Rong, "Adaptive real-time anomaly detection in cloud infrastructures," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 24, Article ID e4193, 2017.
- [27] S. Lee and H. K. Kim, "Adsas: comprehensive real-time anomaly detection system," in *Proceedings of the International Workshop on Information Security Applications*, pp. 29–41, Springer, Jeju Island, Republic of Korea, August 2018.
- [28] R. A. Ariyaluran Habeeb, F. Nasaruddin, A. Gani, I. A. Targio Hashem, E. Ahmed, and M. Imran, "Real-time big data processing for anomaly detection: a survey," *International Journal of Information Management*, vol. 45, pp. 289–307, 2019.
- [29] J. F. Olson and K. M. Carley, "Summarization and Information Loss in Network Analysis," in *Proceedings of the Workshop on Link Analysis, Counter-terrorism, and Security Held in Conjunction with the SIAM International Conference on Data Mining (SDM)*, Citeseer, Atlanta, Georgia, USA, 2008.
- [30] E. Tanin and M. E. Ali, "Hierarchical Data Summarization," *Encyclopedia of Database Systems*, Boston, MA, USA, Article ID Springer, 2009.
- [31] G. Mao, "A real-time loss performance monitoring scheme," *Computer Communications*, vol. 28, no. 2, pp. 150–161, 2005.
- [32] J. Lopes, J. Bento, E. Huang, C. Antoniou, and M. Ben-Akiva, "Traffic and mobility data collection for real-time applications," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems*, pp. 216–223, IEEE, Funchal, Portugal, September 2010.
- [33] T. Vafeiadis, A. Papanikolaou, C. Ilioudis, and S. Charchalakis, "Real-time network data analysis using time series models," *Simulation Modelling Practice and Theory*, vol. 29, pp. 173–180, 2012.
- [34] D. Hoplaros, Z. Tari, and I. Khalil, "Data summarization for network traffic monitoring," *Journal of Network and Computer Applications*, vol. 37, pp. 194–205, 2014.
- [35] S. Mascaro, A. E. Nicholso, and K. B. Korb, "Anomaly detection in vessel tracks using bayesian networks," *International Journal of Approximate Reasoning*, vol. 55, no. 1, pp. 84–98, 2014.
- [36] W. Xiong, H. Hu, N. Xiong et al., "Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications," *Information Sciences*, vol. 258, pp. 403–415, 2014.
- [37] Z. Li, H. Fang, and L. Xia, "Increasing mapping based hidden Markov model for dynamic process monitoring and diagnosis," *Expert Systems with Applications*, vol. 41, no. 2, pp. 744–751, 2014.
- [38] Q. Ma, C. Sun, B. Cui, and X. Jin, "A novel model for anomaly detection in network traffic based on kernel support vector machine," *Computers & Security*, vol. 104, Article ID 102215, 2021.
- [39] A. Alnafessah and G. Casale, "Artificial neural networks based techniques for anomaly detection in Apache Spark," *Cluster Computing*, vol. 23, no. 2, pp. 1345–1360, 2020.
- [40] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "Cnn features with bi-directional lstm for real-time anomaly detection in surveillance networks," *Multimedia Tools and Applications*, vol. 80, pp. 1–17, 2020.
- [41] L. Erhan, M. Ndubuaku, M. Di Mauro, and W. Song, M. Chen, G. Fortino, O. Bagdasar, and A. Liotta, Smart anomaly detection in sensor systems: a multi-perspective review," *Information Fusion*, vol. 67, 2020.
- [42] S. Garg, K. Kaur, S. Batra, G. Kaddoum, N. Kumar, and A. Boukerche, "A multi-stage anomaly detection scheme for augmenting the security in iot-enabled applications," *Future Generation Computer Systems*, vol. 104, pp. 105–118, 2020.
- [43] W. Wang, H. Xu, M. Alazab, T. Reddy Gadekallu, Z. han, and C. su, "blockchain-based reliable and efficient certificateless

- signature for IIoT devices,” *Journal of latex class files*, vol. 14, no. 8, 2015.
- [44] V. Ha-Thuc, D. C. Nguyen, and P. Srinivasan, “A quality-threshold data summarization algorithm,” in *Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies*, pp. 240–246, IEEE, Ho Chi Minh City, Vietnam, July 2008.
- [45] K. McGarry, “A survey of interestingness measures for knowledge discovery,” *The Knowledge Engineering Review*, vol. 20, 2005.
- [46] H. Haken, *Synergetic computers and cognition: A top-down approach to neural nets*, Vol. 50, Springer Science & Business Media, , New York, NY, USA, 2004.
- [47] S. Ganapathy, R. Sethukkarasi, P. Yogesh, P. Vijayakumar, and A. Kannan, “An intelligent temporal pattern classification system using fuzzy temporal rules and particle swarm optimization,” *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 39, 2014.
- [48] S. Belarouci and M. A. Chikh, “Medical imbalanced data classification,” *Advances in Science, Technology and Engineering Systems*, vol. 2, 2017.
- [49] Q. Yang and X. Wu, “10 challenging problems in data mining research,” *International Journal of Information Technology and Decision Making*, vol. 5, no. 4, pp. 597–604, 2006.
- [50] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and G. T. Reddy, “Canin-telliids: detecting in-vehicle intrusion attacks on a controller area network using cnn and attention-based gru,” *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1456–1466, 2021.

Research Article

Early Detection of Cognitive Decline Using Machine Learning Algorithm and Cognitive Ability Test

A. Revathi ¹, R. Kaladevi ², Kadiyala Ramana ³, Rutvij H. Jhaveri ⁴,
Madapuri Rudra Kumar ³ and M. Sankara Prasanna Kumar ³

¹Department of Computational Intelligence, SRM Institute of Science and Technology, Chennai, India

²Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, India

³Department of Computer Science and Engineering, Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh, India

⁴Department of Computer Science and Engineering, Pandit Deendayal Energy University, Gandhinagar, India

Correspondence should be addressed to Rutvij H. Jhaveri; rutvij.jhaveri@sot.pdpu.ac.in

Received 1 December 2021; Revised 1 January 2022; Accepted 5 January 2022; Published 20 January 2022

Academic Editor: Celestine Iwendi

Copyright © 2022 A. Revathi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Elderly people are the assets of the country and the government can ensure their peaceful and healthier life. Life expectancy of individuals has expanded with technological advancements and survey tells that the elderly population will become double in the year 2030. The noninfectious cognitive dysfunction is the most important risk factor among elderly people due to a decline in their physiological function. Alzheimer, Vascular Dementia, and Dementia are the key reasons for cognitive inabilities. These diseases require manual assistance, which is difficult to provide in this fast-growing world. Prevention and early detection are the wise solution for the above diseases. Diabetes and hypertension are considered as main risk factors allied with Alzheimer's disease. Our proposed work applies a two-stage classification technique to improve prediction accuracy. In the first stage, we train a Support vector machine and a Random Forest algorithm to analyze the influence of diabetes and high blood pressure on cognitive decline. In the second stage, the cognitive function of the person with the possibility of Dementia is assessed using the neuropsychological test called Cognitive Ability Test (CAT). Multinomial Logistic Regression algorithm is applied to CAT results to predict the possibility of cognitive decline in their postlife. We classified the risk factor using the operational definitions: "No Alzheimer's," "Uncertain Alzheimer's," and "Definite Alzheimer's". SVM of stage 1 classifier predicts with an accuracy of 0.86 and Random Forest with an accuracy of 0.71. Multinomial Logistic algorithm of stage 2 classifier accuracy is 0.89. The proposed work enables early prediction of a person at risk of Alzheimer's Disease using clinical data.

1. Introduction

Physical health and mental health carry equal importance in human life. Elderly people are normally affected with cardiovascular disease, cancer, diabetic, arthritis, depression, kidney disease, pulmonary disease, dementia and Alzheimer's disease. Dementia is a cognitive decline in mental ability, which severely affects routine life. A person suffering from dementia is always in need of someone to accomplish his everyday activity since the disease affects cognitive function in multiple domains. Alzheimer's Disease (AD) is one among the overall general neurodegenerative cortical

dementia. The incurable neurodegenerative disorder primarily affects the elderly population. It gradually progresses from mild cognitive impairment to Alzheimer's and other kinds of Dementia. The projections are specifically high in South Asian countries such as India and China. The rise in AD disease is proportionate to the elderly population and it is foreseen that 5% to 7% of elders are affected by dementia. By 2050, 1 in 5 persons of low- and middle-income countries will be above 60 ages which may escalate the disease population [1].

Dementia will be an inevitable result of demographic transition and it causes damage to the brain cells. The stages

of dementia span start with no cognitive decline to severe decline. The different types of dementia are Alzheimer's Disease (AD), Vascular Dementia (VaD), Frontotemporal Dementia, etc. This impairment affects the capacity of synapses to converse with one another which in turn affects person's thinking, emotion, and behavior. Various sorts of dementia align with a specific type of brain cell decay in brain regions. A significant level of specific proteins presents inside and outside of synapses, making it difficult for brain cells to remain healthy and to connect with others. The foremost section to be affected is "*Hippocampus*" region of the brain cell, which is the central point of learning and memory in the cerebrum. This is the reason why cognitive decline is perhaps the initial indication of Alzheimer's Disease. There is no effective handling or treatment available for the disease. The feasible option is to train the population with related risk factors and the defending factors.

People affected by diabetes are growing exponentially and it is expected that 640 million people will be affected by the year 2040 [2,3]. As indicated by the World Alzheimer Report 2014, people who had hypertension in their midlife (individuals age around 40–64 years old) were bound to create vascular dementia in later life [4]. Choked or decreased blood flow to the brain is the basic symptom of dementia. Many people with diabetes have brain changes that are a hallmark of Alzheimer's disease. Hypertension causes hurt on the heart and veins and it happens when the power of blood pushing against within our veins is excessively high. This causes the cells to work tougher, which makes them less effective. A recent exploration in a journal named Neurology Trusted Source shows that elderly people have more average BP that is likely to create tangles and plaques in brain. There exists some evidence for a relation of SBP with AD, specifically tangles [5,6].

Multifactor analysis predicts Alzheimer's disease more precisely by extracting heterogeneous information present in health records. It is feasible to predict AD using administrative, clinical information rather than images. Machine learning algorithms are the ideal alternative to apply to a large volume of health data [7]. The focus of our proposed work is twofold: (i) Predicting people with possibilities of Alzheimer in their late life by doing careful analysis on various risk factors associated with Alzheimer's. (ii) Conducting a neuropsychological test called Cognitive Ability Test (CAT) to assess the cognitive decline of a person [8]. The proposed work considers general health data available in "Data World" repository. We apply 2-stage classifier algorithms in the proposed work. In the first stage, support vector machine learning algorithm and Random Forest algorithm are used to find the associated risk factor of individuals. In the second stage, to enhance the prediction accuracy, cognitive ability test was conducted among the people identified by the stage 1 classifier. The cognitive ability of a person is estimated using CAT test, which contains simple yes or no type questions, values ranging from 0 to 30. The CAT test results are applied to Multinomial Logistic Regression to classify the severity of the disease. The score between 25 and 30 is classified as "No Dementia," between 13 and 24 as "Uncertain Dementia," and less than

13 as "Severe Dementia." The proposed work combines multiple factors associated with Alzheimer's to predict the possibility of disease more accurately.

The paper is ordered as follows: Related work on dementia and Alzheimer's disease is explored in chapter 2. Chapter 3 describes the proposed work that includes the relation between Alzheimer's with type 2 diabetes and hypertension dataset, which relates to our claim. The application of multinomial logistic expression on CAT test results to enhance the prediction process is also explored. Chapter 4 justifies the results and relevant discussions. Conclusion of the present work and extension is mentioned in chapter 5.

2. Literature Survey

Mild Cognitive Impairment (MCI) leads to Alzheimer's and various kinds of Dementia in later life. Exceptional intelligent inability of the Alzheimer's diseased patient weighs more burden on family members and public. It has a physical, psychological, social, and economic impact. Careful review was conducted in various aspects such as cause of disease, the different test applied, clinical diagnosis procedure, statistical techniques used, AI/Machine learning techniques used, and so on in order to find the research gap. The research findings are tabulated in Table 1.

The summary leads to an understanding of the correlation between diseases such as diabetes, hypertension, depression, and cognitive impairment. Few drugs are also identified as the main cause of Alzheimer's and related dementia. Various statistical techniques such as ANOVA, *t*-test, Kaplan–Meier estimates (survival estimation function), and QUADAS-2 (diagnosis test against ref value) are used to analyze the data. The main issue to be addressed by the use of a statistical tool is sampling error present in the dataset. This sampling error in the data set would lead to wrong conception. Application of suitable machine learning algorithm will provide optimal solution. The extensive review shows the importance of analyzing cognitive level of the patient and also the role of machine learning algorithm for prediction and classification. Our proposed research work considers diabetes and hypertension details of the patients and applies suitable machine learning algorithm and cognitive ability test to predict the risk of Alzheimer in person's late life.

3. Proposed Work

The proposed model aims at early prediction of cognitive decline of the people using cognitive data, clinical data, and physical data from his history. Dementia has a lengthy preclinical period during which there are no perceptible cognitive impairments, but neurogenerative changes are happening. Therefore, it is essential to identify individuals at high risk of dementia in an earlier stage to protect them from the possibility of disease in their late life [16]. Population based study among precise age group supports our understanding with fewer possible bias. Studies looking at the mid-age people with chronic diseases are particularly helpful since the chances of dementia development are higher for

TABLE 1: Summary of research findings and problem formulation.

| Author(s) | Journal details, year of publication | Description and findings | Issues and identification of research avenue |
|-------------------------------|---|--|---|
| Subha [9] | Journal of Clinical and Diagnostic Research, 2012 | Dataset contains thirty males and females above 50 years, with and without diabetes. Authors applied unpaired <i>T</i> -test and one-way ANOVA using SPSS to find the correlation between age, sex, diabetes, HbA1C level, and cognitive status. The 3MS-MMSE test was also conducted | SPSS lacking efficient handling of regression data. Data analytics is limited. Appropriate machine learning algorithm would support efficient data analytics |
| Arevalo-Rodriguez et al. [10] | Cochrane Database of Systematic Reviews, 2015 | 1569 MCI patients are considered in the case study. The authors assessed Alzheimer's, Lewy body dementia, vascular dementia, and front to temporal dementia among the patients. They used QUADAS -2 tool and search methods in their research. Finally, an MMSE test was conducted to diagnose whether MCI advanced to dementia or Alzheimer's | Early diagnose of MCI would help to prevent the disease |
| Park et al. [11] | NPJ Digital Medicine, 2020 | Korean national health insurance service data between the year 2002 and 2010 with persons age >65 was used in their research proposal. It includes 4,894 clinical features which includes laboratory values, medication codes, ICD-10 codes, history of illness of the person and his family, and sociodemographics. The authors developed RF algorithm, SVM, and logistic regression models for performance comparison. MMSE test was applied to find the severity of Alzheimer's disease | There exists +ve correlation among level of urine protein and alzheimer incident. Tolfenamic acid and zotepine positively correlated with incident AD. Nicametate citrate, was negatively associated with incident AD |
| Alencar et al. [12] | Diabetology and Metabolic Syndrome, 2010 | Total of 346 outpatients with type 2 diabetes which includes 216 females of average age 58.6 ± 12 is considered for the case study. Authors conducted ANOVA test, ANCOVA test with important features such as diabetes, glycemic control, A1C, drugs used by patients, and smoking | It is perceived that the duration of disease has a direct association with a decline in cognition |
| Zhao et al. [13] | BMC Endocrine Disorders, 2020 | The authors conducted a cohort study among type 2 diabetes patients with age >55 years. The data set is divided into three groups with reference to the level of HbA1C: HbA1c < 7.7%, HbA1c between 7.8% and 8%, and HbA1c greater than 8%. Univariate and multivariate regression analysis was done to find the correlation between the level of HbA1C and cognitive decline | It is noted that HbA1c greater than 8% is an important factor to determine the level of cognitive decline |
| Moore et al. [14] | PLoS One, 2019 | The data set of TADPOLE grand challenge in association with ADNI is taken in their case study. ADAS-13 score and normalized ventricles volume were used to analyze the severity of the Alzheimer's disease. The random forest model is simulated in their study. | The outcome of the model was effective and comparable with other methods. However, image processing adds overhead. |
| Altaf et al. [15] | Biomedical Signal Processing and Control, 2018 | Clinical data and MR imaging available in Alzheimer's disease neuroimaging initiative (ADNI) dataset is analyzed using a multiclass classification algorithm. MMSE test is included finally. | The images are classified into three different classes: AD, normal, and MCI. Overhead due to image processing needs to be addressed. |
| Exalto et al. [16] | Alzheimer's Dement, 2014 | Retrospective cohort study was conducted among 9480 Kaiser Permanente members from 1964 to 1973 of age 40–55(CAIDE). The midlife vascular risk factors are analyzed using C statistic and Kaplan–Meier estimates to predict dementia | The disease prevention strategies need to pinpoint life course perspective on maintaining vascular health |

TABLE 1: Continued.

| Author(s) | Journal details, year of publication | Description and findings | Issues and identification of research avenue |
|------------------------|--|--|--|
| Matioli et al. [17] | Dement Neuropsychol, 2017 | Brazilian aging brain study group between 2004 and 2015 containing 1,037 subjects with diabetes mellitus is considered. Multivariate logistic regression algorithm is used to analyze diabetes. Neuropathological examination is done using immunohistochemistry to identify vascular dementia (VaD) and AD. The CDR value and IQCODE is used as reference metrics | The large clinic pathological study results have shown that diabetes was not directly associated with dementia |
| Cholerton et al. [18] | Diabetes Spectrum 2016 | The level of cognitive decline in older adults with type 2 diabetes is analyzed using preclinical pathophysiological processes | Detected and focused pathological changes in early stages due to type 2 diabetes which cause dementia in later stage |
| Shaji et al. [19] | Indian J Psychiatry, 2010 | Dementia aging, mental health issue, and other late life issues were carefully reviewed in Indian Journal of Psychiatry (IJP) from 1958 to 2009. The authors categorized their review as editorials, research reports, and other articles. They used search strategies to accomplish the task | The authors identified the various key risk factors which cause depression and dementia |
| Lee 2020 [20] | Journal of the American Geriatrics Society, 2020 | A subsample of more than 3,000 respondents aged 60 and higher from LASI- the longitudinal aging study in India is considered. LASI is a nationwide health survey which is conducted among people aged 45 to review their social and economic well-being | Disability assessment for dementia is analyzed using the depth of late-life cognition |
| Cunningham et al. [21] | Ulster med J, 2015 | 19,765 people from northern Ireland were analyzed for VD, AD, dementia with Lewy bodies and frontotemporal lobar dementia | Advanced cognitive decline, pathophysiological processes, and biomarkers are explored in their study. Application of cognitive decline test would give more insights |
| Thunell et al. [22] | The Journal of the Alzheimer's association, 2021 | The authors included 29 drug classes span of 11 therapeutic areas and 404 human studies. They applied drug therapy to find drug classes related to increasing or declining ADRD risk (Alzheimer's disease and related dementias) | Among 13 drug classes, 50% or more reported consistent effects on the risk of ADRD |
| Hane et al. [23] | Journal of Alzheimer's Disease, 2017 | The authors reviewed more than 300 prominent AD research proposals. They divided the work into three parts detailed as Part 1 represents pathogenesis on a molecular and macroscale, part 2 represents genetics and epidemiology, and part 3 includes diagnosis and cure. They applied MMSE and MoCA tests | The medical imaging techniques consider pathogenesis and genetics to identify the risk associated with Alzheimer's disease |

them. Midlife hypertension increases the risk of lacunar infarcts and stroke, which in turn increases the risk VaD.

The existing system requires imaging data or fluid collection, which imposes a delay in early detection. Huge measure of Electronic Health Records available in structured and unstructured manner supports timely diagnosis and decisions. Collection of administrative, electronic medical data requires less amount of time. Viable use of information and attaining precise outcome is the major challenge in different fields, particularly in medical field. Utilization of Machine learning is found in almost all fields like image processing, language automation, computer vision, e-business, etc. The advent of predictive models of machine learning can be applied to these valuable digitized health records for the early risk prediction of VAD and AD.

Chronic diseases like diabetics, blood pressure, heart problems, and kidney infection are increasing worldwide. It was witnessed that diabetics and blood pressure have strong relation with cognitive decline in elderly people [24,25]. The helpless diabetic control and bad adherence to physician instructions are the primary reason for the elevation of AD or dementia in their late life [26, 27]. The early detection aids to prevent AD with the help of proper diabetic control, drugs, cognitive training, and so forth. Reference [28] research finding states that there is a direct connection between glucose dysregulation and neurodegeneration. Diabetes is viewed as a key risk factor for cognitive impairment and few investigations prove that cognitive dysfunction influences both older and younger persons with diabetes [29, 30]. Type 2 diabetes patients ought to be

routinely assessed for their intellectual capacity since a span of infection could be related to a decrease in cognizance.

Physicians do not screen the psychological capacity of the patient until he gets compliant from the patient or from the patient's family. After the beginning indications, patients consult doctor. During this period, dementia is moved to a moderately advanced stage. Timely diagnosis would help the patients to overcome disease progression. The Cognitive Ability Test (CAT) is a brief neuropsychological screening test that provides an outline of cognitive function. The CAT test helps the physician to assess the cognitive function of the patient in the early stage itself [13]. The proposed work applies a support vector machine learning algorithm to identify people with a high risk of cognitive impairment in their late life and they are exposed to CAT screening tests [31, 32]. The test results are analyzed with the help of Multinomial logistic regression to classify them as "Severe dementia," "Uncertain Dementia," and "No Dementia."

3.1. Data and Methodology

3.1.1. Data Set. The primary focus of the proposed work is to provide health care service to the elderly population residing in resource poor areas. People with ages between 40 and 65 years are considered as mid-age people in our case study. The proposed work mainly considers hypertension, diabetes as the most common risk factor for cognitive decline. The appropriate data set available in "Data World" repository is taken and filtered with the required features. Plausible cross-sectional examination provides the best technique for analyzing a causal connection between diabetics, blood pressure (BP), and the occurrence of dementia. To enhance the analysis, we consider two age classifications, namely midlife <65 years and late life >65 years.

The emphasis on midlife is especially relevant for dementia counteraction for two reasons. (i) Midlife is sufficiently early to make an association between risk factors and Alzheimer's before the initiation of neurodegeneration. (ii) A few examinations presented the connection between raised BP in midlife (age 40–64 years) and the beginning of dementia and AD in their late life. This study considers general health data available in "Data world," the world's largest collaborative data community. The database consists of 19 features and 2361 patients records whose snippet is depicted in Table 2.

3.1.2. Feature Selection. The data set contains 19 features describing age, cholesterol, glucose, etc. Since Hemoglobin A1c (HbA1c) is the important measure of long-term control of glucose in our body, it was mainly considered in the early identification of AD [30]. Along with HbA1c, the patient's systolic and diastolic BP was examined. The list of features and their descriptions are given in Table 1. HbA1c value greater than 6.5 is considered diabetes positive. The given data set is separated into two distinctive sets with respect to the age to assess the midlife attributes and their association with AD. Exploratory Data Analytics is performed to summarize the main characteristics of data and to find important features

with the help of visual aids. Multivariate analytics is performed to understand different features and their interaction. Table 3 summarizes the features description.

The correlation factor associated with each pair of features helps to extract relevant attributes for study. The highly influential factors such as age, gender, HbA1c, glucose, systolic pressure, diastolic pressure, and cholesterol are considered in our case study.

3.2. Process Flow. The process flow of the proposed model to predict the level of cognitive decline is shown in Figure 1. The diabetes and pressure data set collected from "Data World" is preprocessed to remove redundant information and missing values. The highly influencing features are extracted with the help of correlation values. We apply 2-stage classification model to determine the cognitive decline more accurately. In the first stage, the selected set of features is applied to the classifier algorithm to identify the associated risk among the population. Support Vector Machine and Random Forest algorithm are used for risk classification. In the second stage, we apply CAT among the people identified in stage 1. The Multinomial Logistic Regression algorithm examines the CAT results and medical care is provided for the people predicted as "Severe Alzheimer". CAT.

3.2.1. Algorithm

Step 1: Input: Patients with Blood Pressure, Diabetes dataset.

Filter 40–65 age group data.

Handle inconsistent and missing data

Output: Preprocessed data

Step 2: Identify correlation between features of the dataset using multivariate analysis

Step 3: Do Initial classification for Alzheimer Disease using stage1 classifier

Step 4: If AD possible perform CAT test on those patients

Step 5: Perform second level classification to find AD Dementia, Uncertain Dementia, No-Dementia using stage 2 classifier

3.2.2. Flowchart. The process flow of the proposed model to predict the level of cognitive decline is represented as a flowchart in Figure 2.

3.2.3. Support Vector Machine. SVM is the commonly used supervised classifier, which classifies data in N-dimensional space using a hyperplane. It has been applied in enormous healthcare applications in predicting diseases from structural data [33]. Figure 3 shows the classification graph of SVM.

The line function $y = ax + b$ helps to easily differentiate linearly separable data. The SVM uses the line equation transformed into a hyperplane which is applied in the prediction process. The model tries to find out optimal bias and variance for both train and test data set. The

TABLE 2: Snippet of Diabetics registry.

| Id | Chols | Stab-glu | Hdl | ... | Glyhb | Age | Height | Weight | Bp-s | Bp-d |
|------|-------|----------|-----|-----|-------|-----|--------|--------|------|------|
| 1000 | 203 | 82 | 56 | ... | 4.31 | 46 | 62 | 119 | 120 | 82 |
| 1001 | 165 | 97 | 24 | ... | 4.44 | 29 | 64 | 218 | 140 | 85 |
| 1002 | 228 | 92 | 37 | ... | 4.64 | 58 | 61 | 243 | 185 | 92 |
| 1003 | 78 | 93 | 12 | ... | 4.63 | 67 | 67 | 121 | 170 | 95 |
| 1005 | 249 | 90 | 28 | ... | 7.72 | 64 | 68 | 179 | 158 | 85 |
| 1008 | 248 | 94 | 69 | ... | 4.81 | 34 | 71 | 186 | 125 | 80 |
| 1011 | 195 | 92 | 41 | ... | 4.84 | 30 | 69 | 180 | 161 | 112 |
| 1015 | 227 | 75 | 44 | ... | 3.94 | 37 | 59 | 170 | 121 | 81 |
| 1016 | 177 | 87 | 49 | ... | 4.84 | 45 | 69 | 166 | 128 | 86 |
| 1022 | 263 | 89 | 40 | ... | 5.78 | 55 | 63 | 198 | 132 | 82 |
| 1024 | 242 | 82 | 54 | ... | 4.77 | 60 | 65 | 164 | 130 | 90 |
| 1029 | 215 | 128 | 34 | ... | 4.97 | 38 | 58 | 191 | 145 | 83 |
| 1030 | 238 | 75 | 36 | ... | 4.47 | 27 | 60 | 167 | 120 | 80 |
| 1031 | 183 | 79 | 46 | ... | 4.59 | 40 | 59 | 173 | 132 | 85 |
| 1035 | 191 | 76 | 30 | ... | 4.67 | 36 | 69 | 179 | 122 | 82 |
| 1036 | 213 | 83 | 47 | ... | 3.41 | 33 | 65 | 154 | 120 | 96 |
| 1037 | 255 | 78 | 38 | ... | 4.33 | 50 | 65 | 187 | 126 | 83 |
| 1041 | 230 | 112 | 64 | ... | 4.53 | 20 | 67 | 159 | 152 | 92 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

TABLE 3: List of features and their descriptions.

| Feature name | Description and values | % Missing | Mean | Standard Deviation |
|--------------|--------------------------|-----------|---------|--------------------|
| ID | Patient ID | 0% | NA | NA |
| Chols | Overall cholesterol | <1% | 207.275 | 44.715 |
| Stab-glu | Stabilized glucose | 0% | 107.338 | 53.798 |
| HDL | High-density lipoprotein | <1% | 50.267 | 17.301 |
| Glyhb | Glycosylated hemoglobin | 0% | 5.59 | 2.243 |
| Age | Patients age (Years) | 0% | 46.774 | 16.436 |
| Ht | Patients height (inches) | <1% | 65.979 | 3.927 |
| Wt | Patients weight (pounds) | 0% | 177.349 | 40.392 |
| Bp-s | Patients systolic BP | = 1.28% | 137.148 | 22.997 |
| Bp-d | Patients diastolic BP | = 1.28% | 83.286 | 13.582 |

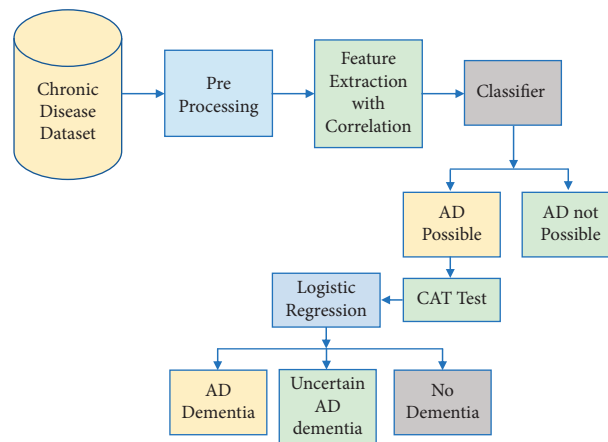


FIGURE 1: Process flow diagram.

comprehensive review has proven that support vector machine provides good performance for big data and healthcare applications.

3.2.4. Random Forest. The Random Forest algorithm trains n different decision trees with different data subset and tuning parameters [34]. It combines the output of all n trees

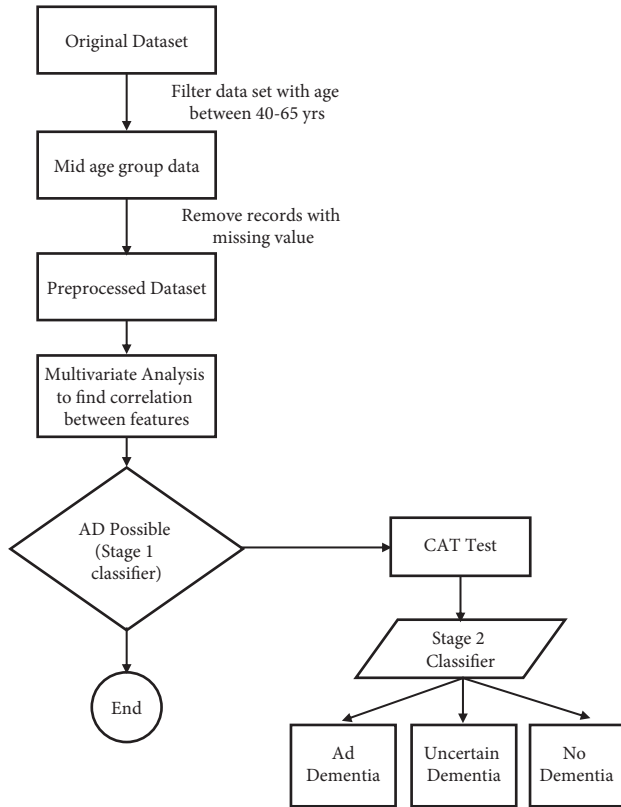


FIGURE 2: Flow chart.

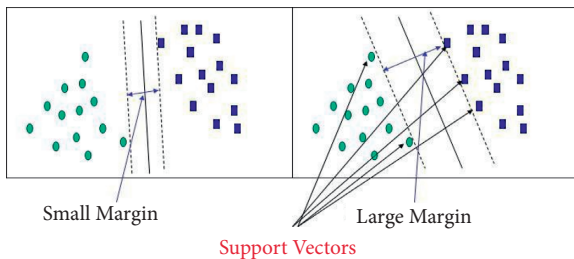


FIGURE 3: SVM classification graph.

with the help of a voting mechanism. Hence, it is also called Ensemble learning. The working principle of Random Forest algorithm is depicted in Figure 4.

3.3. Mental Ability Test. The cognitive decay of a person ranges from mild to severe. The primary causes include medications, disorder among blood vessels, despair, and dementia. Dementia represents a severe loss of mental functioning and the common type is Alzheimer. Cognition of a person includes a blend of processes in the brain involved in all facets of his life. It includes his memory capacity, thinking skill, language, and talent to learn new things. A cognitive ability test is performed to examine the cognitive impairment of a person. With the help of a detailed review conducted to screen the cognitive function, we framed multiple questions to check the decline in mental

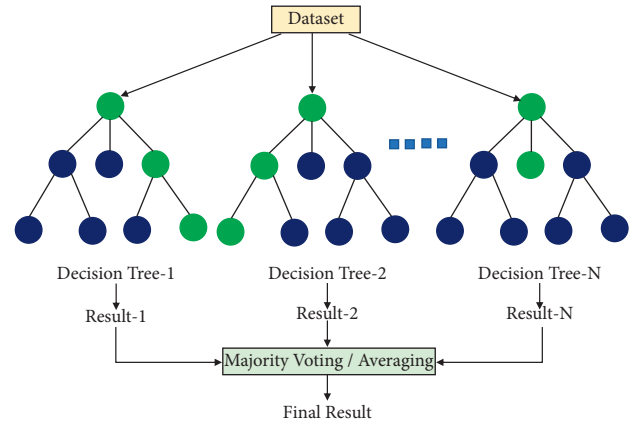


FIGURE 4: Working of random forest algorithm.

function, and the test questionnaire is given in Table 4 [35]. The CAT test scores from 25 to 30 are considered as normal [36]. Items address orientation, memory, attention, recall, naming objects, responding to verbal and written commands, writing a sentence, and copying a figure are the tasks considered in CAT to evaluate the cognitive status of persons. The informant accompanied the patients, and the questions are administered to the informants without unduly alarming the patient.

The maximum CAT score is 30 points. A score of 25 to 30 suggests no cognitive decline, 13 to 24 recommends moderate decline, and less than 12 indicates severe cognitive decline. In every year, the CAT score of Alzheimer’s diseased person declines about two to four points on average. The snippet of CAT dataset is shown in Table 5.

3.3.1. Multinomial Logistic Regression on CAT. The multinomial logistic regression model is applied to predict the severity of illness with respect to the correlation existing among the dependent variables as “Severe dementia,” “Uncertain Dementia,” and “No Dementia.” The multinomial logistic regression is applicable for the class of probe, which has more than two outcomes. Our proposed model owns three different outcomes. For N different outcomes, there are n-1 models developed as a set of independent binary regression. One outcome is referred to as Pivot class, and others are regressed against this reference class.

The probabilities for the N categories are estimated based on dependent variables.

$$\begin{aligned}
 pr(Y_i = k | X_i) &= \beta_1, \beta_2, \dots, \beta_n \\
 &= \frac{\exp(\beta_0 k + X_i \beta' k)}{\sum_{j=1}^n \exp(\beta_0 k + X_i \beta' k)}, \tag{1}
 \end{aligned}$$

where Y is the dependent variable and X is the set of explanatory variables, β_k is the regression coefficient for the kth category of Y.

Based on the estimated probability the output is categorized by the algorithm with reference to the threshold.

TABLE 4: Cognitive ability test-questionnaire.

| S. No. | Question description | Yes/no | Weighted score |
|--------|--|--------|----------------|
| 1 | Does the patient place items properly? | Y | 1 |
| 2 | Does the patient able identify the present date, day, month, year? | Y | 2 |
| 3 | Does the patient comfort level change when they are in new places? | Y | 1 |
| 4 | Does the patient able to manage their medication schedule? | Y | 2 |
| 5 | Does the patient able to manage time while doing tasks? | Y | 1 |
| 6 | Does the patient confuse about certain things? | Y | 2 |
| 7 | Does the patient able to understand context? | Y | 1 |
| 8 | Does the patient confuse to identify known persons? | Y | 2 |
| 9 | Does the patient experience difficulty to recognize people familiar to them? | Y | 1 |
| 10 | Does the patient behavior is different from their earlier stages? | Y | 1 |
| 11 | Does the patient have imaginations? | Y | 2 |
| 12 | Does the patient forget to do regular tasks? | Y | 2 |
| 13 | Does the patient have problem in counting numbers or figures? | Y | 2 |
| 14 | Does the patient able to manage position or direction? | Y | 1 |
| 15 | Does the patient has shown less priory or interest towards hobby or passion? | Y | 1 |
| 16 | Does the patient understand situations or explanations? | N | 1 |
| 17 | Does the patient forget recent activities? | N | 1 |
| 18 | Does the patient have any cognitive issues previously? | Y | 2 |
| 19. | Does the patient not able to recall main or important occasions? | Y | 2 |
| 20 | Does the patient not able to recollect some important days in his life. | Y | 2 |

TABLE 5: Snippet of CAT dataset.

| Subject | Gender | CAT | ageAtEntry | CDR | Memory | dx1 |
|----------|--------|-----|------------|-----|--------|----------------------|
| OAS30124 | Female | 16 | 79.14579 | 1 | 1 | 'AD dementia' |
| OAS30124 | Female | 21 | 79.14579 | 0.5 | 0.5 | 'AD dementia' |
| OAS31129 | Female | 20 | 68.07666 | 1 | 1 | 'AD dementia' |
| OAS31129 | Female | 28 | 68.07666 | 1 | 1 | 'AD dementia' |
| OAS31129 | Female | 29 | 68.07666 | 0.5 | 0.5 | 'AD dementia' |
| OAS31129 | Female | 29 | 68.07666 | 0.5 | 1 | 'AD dementia' |
| OAS30865 | Female | 24 | 73.697464 | 2 | 2 | 'AD dementia' |
| OAS30865 | Female | 28 | 73.697464 | 0.5 | 1 | 'AD dementia' |
| OAS31003 | Male | 30 | 63.92608 | 0.5 | 0.5 | 'AD dementia' |
| OAS31003 | Male | 30 | 63.92608 | 1 | 0.5 | 'AD dementia' |
| OAS31003 | Male | 30 | 63.92608 | 0.5 | 0.5 | 'AD dementia' |
| OAS31003 | Male | 30 | 63.92608 | 0.5 | 0.5 | 'AD dementia' |
| OAS30025 | Female | 23 | 64.81314 | 0.5 | 0.5 | 'AD dementia' |
| OAS31001 | Female | 25 | 62.984257 | 0.5 | 0.5 | 'Uncertain dementia' |
| OAS30331 | Male | 27 | 67.482544 | 0.5 | 0.5 | 'Uncertain dementia' |
| OAS30331 | Male | 17 | 67.482544 | 1 | 1 | 'AD dementia' |
| OAS30331 | Male | 27 | 67.482544 | 0.5 | 0.5 | 'Uncertain dementia' |
| OAS30596 | Female | 29 | 72.29295 | 0.5 | 0.5 | 'Uncertain dementia' |
| ... | ... | ... | ... | ... | ... | ... |

4. Experiments and Results

4.1. Stage 1 Classifier. In stage 1 we train SVM and Random Forest algorithm to diagnose chronic disease and to identify the associated risk. The data set contains 2361 records of mid-age people. The glyhb values in the range between 4% and 5.6% are considered as normal values, and between 5.7% and 6.4% informs more chance of being affected with diabetes. Values above 6.5% mean they have diabetes. Patient's systolic and diastolic pressure are the other important factors to be considered in the development of Alzheimer's. Systolic pressure less than 120 mm Hg and diastolic pressure less than 80 mm Hg is

considered as normal value and the range 120–139 of systolic and 80–89 of diastolic is the prehypertension values. Persons having >140 mm Hg of systolic and >90 mm Hg of diastolic pressure are considered as having hypertension. The chosen data set contains 526 records of persons with no diabetics and pressure, 1187 records of persons having diabetics or pressure, and 648 records of patients having both diabetics and pressure. Since the presence of either pressure or diabetes increases the chance of dementia, the total 1835 patients having either diabetics or pressure or both exposed to CAT test to assess their cognitive power. The details of records are visually represented in Figure 5.

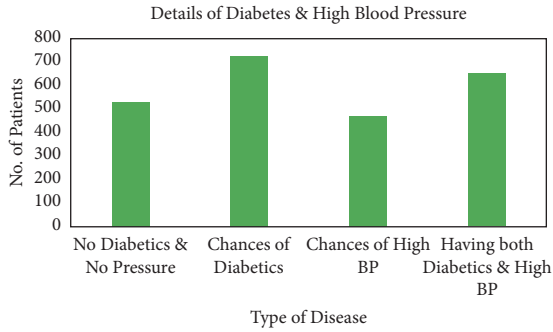


FIGURE 5: Patient count on Diabetes and Pressure.

The performance of the proposed classifier model is analyzed using confusion matrix. The total accuracy, sensitivity and specificity are calculated using the formula given.

(i) Total Accuracy (TC)

$$TC = \frac{TP + TN}{TP + FN + TN + FP} \quad (2)$$

(ii) Sensitivity (SN)

Sensitivity (or) True Positive Rate and Specificity (or) True Negative rate are calculated to assess the performance of the proposed model. True positive rate reveals the details of how effectively the model identifies actual positives present in the data set. Specificity measure is given as follows:

$$SN = \frac{TP}{TP + FN} \quad (3)$$

(iii) Specificity

Specificity (or) True Negative rate measure is given as follows:

$$SP = \frac{TN}{TN + FP} \quad (4)$$

The terms TN, TP, FP, and FN denote True negative (person with no chance of Alzheimer is identified as 'No Alzheimer'), True positive (person subjected to Alzheimer is predicted as 'Alzheimer'), False Positive (healthy person is detected as 'Alzheimer'), and False Negative (person suffering from Alzheimer is identified as healthy), respectively.

4.2. Performance Analysis of Stage 1 Classifier. The proposed SVM classifier outperforms with 0.90 AUC value and for Random Forest AUC is 0.74. The probabilistic classifier shows the tradeoff between sensitivity and specificity. Table 2 shows the performance comparison of SVM and Random Forest algorithm used in our case study. NDP represents patients with No Diabetics and No Pressure, CRD represents patients having either diabetes or pressure, HDP represents patients Having both Diabetics and Pressure. Performance comparison of each measure is given in Table 6. The same is visually represented in Figures 6–8.

In Random Forest algorithm, it is important to consider the subsampling of data points in the tree construction process. More subsampling or no subsampling results in inconsistent effects. It is possible to enhance the accuracy of Random Forest algorithm by varying the parameters. Due to the unavailability of sample data set, it is not probable to fine-tune the parameters for RF in our case study.

4.3. Stage 2 Classifier. The CAT test result dataset contains a minimum age of 47 years and a maximum of 96 years. Clinical Dementia Rating shortly termed as CDR is a numeric scale used to quantify the severity of dementia indications and its score ranges from zero (none) to 3(severe). Summarization of CAT data set is provided in Table 7 and the same is visually represented in Figure 9.

The confusion matrix of multinomial logistic regression is considered for analysis. True Positive Rate and True Negative rate for the three different classes of output is given as follows: $TPR_{No\ Alzheimer} = 98\%$, $TPR_{Uncertain_Alzheimer} = 85\%$ and $TPR_{Alzheimer} = 81\%$. True Negative rates of the three different classes are given as follows. $TNR_{No\ Alzheimer} = 86\%$, $TNR_{Uncertain_Alzheimer} = 73\%$, and $TNR_{Alzheimer} = 75\%$. The results of the proposed model show that the model can predict with improved accuracy provided ample amount of dataset for training.

4.4. Analysis with Bench Mark Models. The study of neurodegenerative diseases caused by the ageing of brain systems necessitates brain banking. The Brazilian Aging Brain Study Group's Brain Bank collects a large number of elderly brains and their related disorders. It encourages researchers to look at a variety of aspects of ageing brain processes and related neurodegenerative illnesses.

Table 8 represents the performance analysis of our proposed model with existing benchmark models. The table explores the detail of the data set used by different authors and the employed machine learning algorithms. Reference [13] considers people above 65 years from 2002 to 2010. The main features considered in their case study include Implantable Cardioverter Defibrillator -10 codes, laboratory results, medication codes, sociodemographics, illness of a person, and his family [37]. They have trained and tested dataset with random forest, logistic regression, and SVM and to predict Alzheimer's incident in 1, 2, 3, and 4 subsequent years. For comparison average of 4 years is taken. The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge is organized in association with Alzheimer's Disease Neuroimaging Initiative (ADNI) to find people at risk of Alzheimer's. The historical measurements of the people were considered to predict future implications. TADPOLE challenge facilitates early identification of Alzheimer disease with the help of appropriate algorithm [14] used data from the TADPOLE grand challenge and claimed their result with benchmark SVM, which produces 62% AUC and classification accuracy of 52%. Reference [15] collected records of Brain Bank of the Brazilian Aging Brain Study Group between 2004 and 2015. Among 1,037 subjects, diabetes was present, with 279 participants (27%). They

TABLE 6: Performance comparison of SVM and Random Forest.

| Algorithm | Accuracy | | | Sensitivity | | | Specificity | | |
|-----------|----------|------|------|-------------|------|------|-------------|------|------|
| | NDP | CRD | HDP | NDP | CRD | HDP | NDP | CRD | HDP |
| SVM | 0.89 | 0.86 | 0.84 | 0.78 | 0.82 | 0.88 | 0.93 | 0.78 | 0.89 |
| RF | 0.81 | 0.78 | 0.79 | 0.62 | 0.83 | 0.71 | 0.82 | 0.75 | 0.78 |

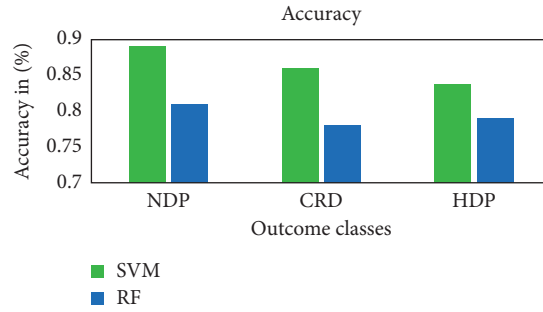


FIGURE 6: Accuracy.

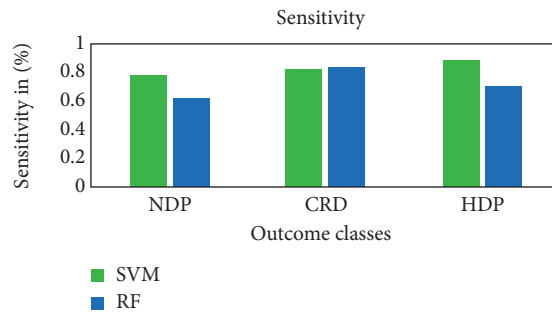


FIGURE 7: Sensitivity.

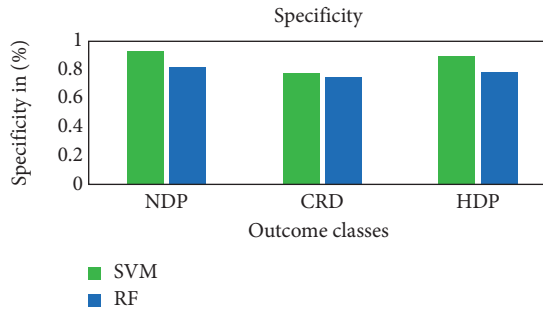


FIGURE 8: Specificity.

TABLE 7: Summarization of CAT dataset.

| CAT score level | Patient count | Cognitive status |
|-----------------|---------------|--------------------|
| 25–30 | 1017 | No dementia |
| 13–24 | 603 | Uncertain dementia |
| 0–12 | 215 | Severe dementia |

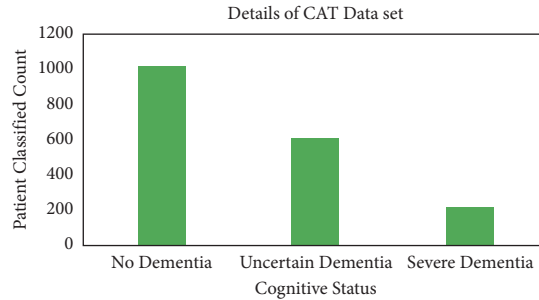


FIGURE 9: Summarization of CAT dataset.

TABLE 8: Analysis with Bench mark models.

| Method | Dataset | ML algorithm | Accuracy (%) | AUC (%) |
|---------------------|--|----------------------------------|--------------|---------|
| Park et al. [11] | Korean National Health Insurance Service database between 2002 and 2010 | Linear regression | 66.8 | 69.5 |
| | | Support vector machine | 67.1 | 70.58 |
| | | Random forest | 70.12 | 76.1 |
| Moore et al. [14] | TADPOLE grand challenge dataset | Random forest | 73 | 82 |
| Matioli et al. [17] | Diabetes data from brazilian aging brain study group between 2004 and 2015 | Multivariate logistic regression | 95 | - |
| | | Support vector machine | 0.86 | 90 |
| Proposed method | Diabetics registry data in kaggle repository and data set generated through CAT test | Random forest | 0.71 | 0.74 |
| | | Multinomial logistic regression | 0.89 | 0.92 |

proved there is no association between diabetes and dementia (OR = 1.22; 95%CI = 0.81–1.82; $p = 0.34$) based on the multivariate analysis.

5. Conclusion

Automated healthcare techniques support physicians in making correct decisions on patient care in resource poor rural areas. The timely identification of risk factors with the help of AI based model's safeguards the person from late life Alzheimer's. The availability of an appropriate dataset with relevant attribute is a cumbersome process in the development of a more accurate model. The proposed method supports the statistically significant diagnosis of persons at risk for Alzheimer's disease simply based on administrative health records. It allows earlier and accurate screening for further clinical testing. Our proposed work analyzes the influence of hypertension and diabetes on Alzheimer's disease. Support Vector Machine algorithm is more suitable when the dataset is not continually distributed. The performance of SVM is relatively good due its convex optimization nature. Survey conducted on the population with chronic disease for cognitive assessment provides the degree of cognitive decline in the community. The CAT test results are analyzed with the help of multinomial logistic regression to exactly identify the possibility of Alzheimer's in patient's late life. To achieve optimum accuracy of the model, a large sample size is essential. In the future, the proposed work may be extended with more classifiers by accumulating a huge

volume of samples and an increased number of surveys on CAT tests. Time series survey among the population for CAT test will further improve the precision of prediction.

Data Availability

The datasets used and/or analyzed during the current study are available in the following repository: <https://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Diabetes.html>.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Authors' Contributions

A. Revathi was responsible for conceptualization, data curation, formal analysis, methodology, software, and writing—original draft; R.Kala Devi was responsible for supervision, writing—review and editing, project administration, and visualization; Kadiyala Ramana was responsible for software, validation, writing—original draft, methodology, and supervision; Rutvij H.Jhaveri was responsible for supervision, writing—review and editing, and visualization; Madapuri Rudra Kumar was responsible for data curation, investigation, resources, and software; M.Sankara Prasanna Kumar was responsible for visualization, investigation, formal analysis, and software.

References

- [1] S. Bhattacharya, P. K. R. Maddikunta, Q.-V. Pham et al., "Deep learning and medical image processing for coronavirus (covid-19) pandemic: a survey," *Sustainable Cities and Society*, vol. 65, Article ID 102589, 2021.
- [2] G. J. Biessels and F. Despa, "Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications," *Nature Reviews Endocrinology*, vol. 14, no. 10, pp. 591–604, 2018.
- [3] C. Qiu, E. V. Strauss, J. Fastbom, B. Winblad, and L. Fratiglioni, "Low blood pressure and risk of dementia in the kungsholmen project: a 6-year follow-up study," *Archives of Neurology*, vol. 60, no. 2, pp. 223–228, 2003 Feb.
- [4] B. Jani and C. Rajkumar, "Ageing and vascular ageing," *Postgraduate Medical Journal*, vol. 82, no. 968, pp. 357–362, 2006.
- [5] J. R. Marden, E. R. Mayeda, E. J. Tchetgen Tchetgen, I. Kawachi, and M. M. Glymour, "High hemoglobin A1c and diabetes predict memory decline in the health and retirement study," *Alzheimer Disease and Associated Disorders*, vol. 31, no. 1, pp. 48–54, 2017.
- [6] P. Ramesh, K. Dunn, W. Eberle, and D. Chaung, "Cognitive health prediction on the elderly using sensor data in smart homes," in *Proceedings of the 31st International FLAIRS Conference*, pp. 317–322, Melbourne, Florida, USA, May 2018.
- [7] K. R. Lakshmana, F. Khan, D. Sadia, S. B. Shahab, M. Amir, and I. Ebuka, "Recurrent neural network and reinforcement learning model for covid-19 prediction," *Frontiers in Public Health*, vol. 9, 2021.
- [8] D. Ngabo, D. Wang, W. Dong, E. Ibeke, and C. Iwendi, "Tackling pandemics in smart cities using machine learning architecture," *Mathematical Biosciences and Engineering*, vol. 18, no. 6, pp. 8444–8461, 2021.
- [9] N. K. Subha, "Assessment of the cognitive status in diabetes mellitus," *Journal of Clinical and Diagnostic Research*, vol. 6, no. 10, pp. 1658–1662, 2012.
- [10] I. Arevalo-Rodriguez, N. Smailagic, M. Roqué i Figuls et al., "Mini-mental state examination (MMSE) for the detection of alzheimer's disease and other dementias in people with mild cognitive impairment (MCI)," *Cochrane Database of Systematic Reviews*, vol. 3, Article ID CD010783, 2015.
- [11] J. H. Park and H. Cho, J. H. Kim, M. M. Wall, Y. Stern et al., "Machine learning prediction of incidence of alzheimer's disease using large-scale administrative health data," *NPJ Digital Medicine*, vol. 3, p. 46, 2020.
- [12] R. C. Alencar, R. A. Cobas, and M. B. Gomes, "Assessment of cognitive status in patients with type 2 diabetes through the mini-mental status examination: a cross-sectional study," *Diabetology & Metabolic Syndrome*, vol. 2, no. 1, p. 10, 2010.
- [13] L. Zhao, C. Han, Z. Zheng, S. L. Xiu, and P. Chan, "Risk of mini-mental state examination (MMSE) decline in the elderly with type 2 diabetes: a Chinese community-based cohort study," *BMC Endocrine Disorders*, vol. 20, p. 129, 2020.
- [14] P. J. Moore, T. J. Lyons, and J. Gallacher, "Random forest prediction of Alzheimer's disease using pairwise selection from time series data," *PLoS One*, vol. 14, no. 2, pp. 1–14, 2019.
- [15] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, vol. 43, pp. 64–74, 2018.
- [16] L. G. Exalto, C. P. Quesenberry, D. Barnes, M. Kivipelto, G. J. Biessels, and R. A. Whitmer, "Midlife risk score for the prediction of dementia four decades later," *Alzheimers Dement*, vol. 10, no. 5, pp. 562–570, 2014.
- [17] M. N. P. D. S. Matioli, C. K. Suemoto, R. D. Rodriguez et al., "Association between diabetes and causes of dementia: evidence from a clinicopathological study," *Dement Neuro-psychol*, vol. 11, no. 4, pp. 406–412, 2017.
- [18] B. Cholerton, L. D. Baker, T. J. Montine, and S. Craft, "Type 2 diabetes, cognition, and dementia in older adults: toward a precision health approach," *Diabetes Spectrum*, vol. 29, no. 4, pp. 210–219, 2016.
- [19] K. S. Shaji, V. P. Jithu, and K. S. Jyothi, "Indian research on aging and dementia," *Indian Journal of Psychiatry*, vol. 52, no. Suppl1, pp. S148–S152, 2010 Jan.
- [20] J. Lee, Y. K. Pranali, J. Banerjee et al., "Design and methodology of the longitudinal aging study in india-diagnostic assessment of dementia (LASI-DAD)," *Journal of the American Geriatrics Society*, vol. 68, 2020.
- [21] E. L. Cunningham, B. McGuinness, B. Herron, and A. P. Passmore, "Dementia," *Ulster Medical Journal*, vol. 84, no. 2, pp. 79–87, 2015.
- [22] J. Thunell, Y. Chen, G. Joyce et al., "drug therapies for chronic conditions and risk of alzheimer's disease and related dementias: a scoping review," *Alzheimer's & Dementia The Journal of the Alzheimer's association*, vol. 17, no. 1, pp. 41–48, 2021.
- [23] F. T. Hane, M. Robinson, Y. L. Brenda, B. Owen, Z. Leonenko, and M. S. Albert, "Recent progress in alzheimer's disease research, Part 3: diagnosis and treatment," *J Alzheimers Dis*, vol. 57, no. 3, pp. 645–665, 2017.
- [24] C. Feng, A. Elazab, P. Yang, T. Wang, B. Lei, and X. Xiao, "3D convolutional neural network and stacked bidirectional recurrent neural network for alzheimer's disease diagnosis," in *Predictive Intelligence in Medicine*, I. Rekić, G. Unal, E. Adeli, and S. Park, Eds., Springer Nature Switzerland AG, Berlin, Germany, pp. 138–146, 2018.
- [25] A. Ramirez, S. Wolfsgruber, C. Lange et al., "Elevated HbA1c is associated with increased risk of incident dementia in primary care patients," *J Alzheimers Dis*, vol. 44, no. 4, pp. 1203–1212, 2015.
- [26] T. R. Gadekallu, N. Khare, S. Bhattacharya et al., "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, 2020.
- [27] C. Tzourio, "Hypertension, cognitive decline, and dementia: an epidemiological perspective," *Dialogues in Clinical Neuroscience*, vol. 9, no. 1, pp. 61–70, 2007.
- [28] S. Ahmed Khan and M. A. Jabbar, "Improved classification techniques to predict the co-disease in diabetic mellitus patients using discretization and apriori algorithm," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN, vol. 8, no. 11, pp. 2278–3075, 2019.
- [29] S. E. Nilsson, S. Read, S. Berg, B. Johansson, A. Melander, and U. Lindblad, "Low systolic blood pressure is associated with impaired cognitive function in the oldest old: longitudinal observations in a population-based sample 80 years and older," *Aging-Clinical & Experimental Research*, vol. 19, no. 1, pp. 41–47, 2007.
- [30] K. R. Kruthika and H. D. Rajeswari, "Maheshappa, multistage classifier-based approach for alzheimer's disease prediction and retrieval," *Informatics in Medicine Unlocked*, vol. 14, pp. 34–42, 2019.
- [31] X. Li, J. Dai, S. Zhao, W. Liu, and H. Li, "Comparison of the value of mini-cog and mmse screening in the rapid identification of chinese outpatients with mild cognitive

- impairment,” *Medicine (Baltimore)*, vol. 97, no. 22, Article ID e10966, 2018.
- [32] P. T. Trzepacz, H. Hochstetler, S. Wang, B. Walke, and A. Saykin, “Relationship between the montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults,” *BMC Geriatrics*, vol. 15, no. 107, 2015.
- [33] G. Battineni, N. Chintalapudi, and F. Amenta, “Machine learning in medicine: performance calculation of dementia prediction by support vector machines (SVM),” *Informatics in Medicine Unlocked*, vol. 16, Article ID 100200, 2019.
- [34] I. Celestine, B. Ali Kashif, P. Atharva, and R. Sujatha, “Chatterjee jyotir moy, pasupuleti swetha, mishra rishita, pillai sofia, jo ohyun, covid-19 patient health prediction using boosted random forest algorithm,” *Frontiers in Public Health*, vol. 8, 2020.
- [35] Y. Maki, T. Yamaguchi, and H. Yamaguchi, “Symptoms of early dementia-11 questionnaire (SED-11Q): a brief informant-operated screening for dementia,” *Dement Geriatr Cogn Disord Extra*, vol. 3, pp. 131–142, 2013.
- [36] M. F. Folstein, S. E. Folstein, and P. R. McHugh, “Mini-mental state”. a practical method for grading the cognitive state of patients for the clinician,” *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.
- [37] J. C. De la Torre, “Alzheimer disease as a vascular disorder,” *Stroke*, vol. 33, no. 4, pp. 1152–1162, 2002.

Research Article

Analysis and Improvement of Blockchain-Based Multilevel Privacy-Preserving Location Sharing Scheme for Telecare Medical Information Systems

Zhenjie Huang ^{1,2}, Yafeng Guo ³, Hui Huang ⁴, Runlong Duan ^{1,4}
and Xiaolong Zhao ^{1,2}

¹Fujian Key Laboratory of Granular Computing and Application, Minnan Normal University, Zhangzhou 363000, China

²School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China

³Department of Electronic and Informatics, Zhangzhou City College, Zhangzhou 363000, Fujian, China

⁴School of Computer Science, Minnan Normal University, Zhangzhou 363000, China

Correspondence should be addressed to Zhenjie Huang; zjhuang@mnnu.edu.cn

Received 17 October 2021; Revised 17 November 2021; Accepted 29 November 2021; Published 18 January 2022

Guest Editor: Thippa Reddy G

Copyright © 2022 Zhenjie Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Patient location sharing is an important part of modern smart healthcare and mobile medical services. Blockchain has many attractive properties and is suitable for managing patient locations in telecare medical information systems (TMIS). Recently, Ji et al. proposed a blockchain-based multilevel privacy-preserving location sharing (BMPLS) scheme for TMIS. In this paper, we show that Ji et al.'s BMPLS scheme does not achieve confidentiality and multilevel privacy-preserving. An adversary outside the system can use an ordinary personal computer to completely break the system within a dozen hours and obtain the location of any patient at any time. The adversary inside the system can use an ordinary personal computer to obtain the location of the designated patient within tens of seconds. Using salting technology, we propose an improved BMPLS scheme to fix our attacks. We also optimized the BMPLS scheme to make it correct and executable. The security analysis shows that the improved BMPLS scheme achieves decentralization, untamperability, confidentiality, multilevel privacy-preserving, retrievability, and verifiability. The simulation shows that the improved BMPLS scheme is practical, the computational overhead of the location record phase is within 10 ms, and the computational overheads of the location sharing and location extraction phases are both within 30 ms.

1. Introduction

Blockchain is a new decentralized infrastructure and distributed computing paradigm, and it is one of the most revolutionary emerging technologies [1]. In a narrow sense, blockchain is a decentralized shared general ledger that combines data blocks in a chain into a specific data structure in chronological order and uses cryptography technology to ensure that data are untamperable and unforgeable. In a broad sense, blockchain technology is a new decentralized computing paradigm. It uses an encrypted chain block structure to store and verify data, consensus algorithms to update data, and smart contracts to manipulate data. Blockchain has the advantages of decentralization, trustless, anonymity, and untamperability. It can break through the

limitations of traditional centralized systems and find important applications in a wide range of fields. More meaningful applications will appear with the further integration of blockchain with cloud computing, edge computing, and the Internet of Things [2–6].

With the rapid development of information technology, medical management has become more intelligent and real-time. Wireless mobile networks and wearable technology enable mobile medical services and telemedicine to be realized. For example, IBM has integrated a real-time asset locator (RTAL) for mobile medical and telecare medical to track the location of patients, equipment, and medical staff. Location management plays a significant role in remote patient service and monitoring, such as monitoring particular patients, handling emergencies, and analyzing

epidemic distribution. Blockchain has many attractive properties and is suitable for managing patient locations in telecare medical information systems (TMIS).

Based on blockchain technology, Zyskind et al. [7] proposed a data storage scheme to protect sensitive data such as user's location. In their scheme, user data needs to be encrypted before being stored on the blockchain to achieve confidentiality. Amoretti et al. [8] proposed a blockchain-based location proof scheme. Different from [7], in this scheme, the location information is signed before being put on the blockchain to achieve verifiability. Ji et al. [9] comprehensively analyzed the security requirements of blockchain-based location sharing for TMIS and proposed a blockchain-based multilevel privacy-preserving location sharing scheme (BMPLS). They claim that their scheme achieves decentralization, untamperability, confidentiality, multilevel privacy-preserving, retrievability, and verifiability. Unfortunately, we found that their scheme is insecure in practice. The adversary can recover any user's location effectively. Recently, Lee et al. [10] proposed a blockchain-based medical data preservation scheme for TMIS. Their scheme consists of a medical sensor area authentication protocol and a social network information transfer protocol.

1.1. Related Works. With the development of Internet technology, telemedicine has gradually replaced the traditional treatment model. Electronic medical records (EMR) and electronic health records (EHR) are generated in large quantities and frequently exchanged and shared among legitimate users. The protection of electronic medical information is related to patients' privacy and related to their life safety. Therefore, the security and privacy protection of electronic medical records are significant for the development of telemedicine.

Blockchain technology has the potential to improve the medical ecosystem. It provides a novel, efficient, and secure model for exchanging EMRs and EHRs and enhances medical data security, privacy, and interoperability [11, 12]. Using blockchain technology, Cao et al. [13] proposed a secure cloud-assisted electronic health system to protect outsourced electronic health records from illegal modification. Wang et al. [14] proposed a blockchain-based eHealthcare system interoperating with wireless body area networks. Using proxy reencryption, Huang et al. [15] proposed a blockchain-based decentralized medical data sharing scheme with privacy-preserving. Zhuang et al. [16] proposed a patient-centric health information exchange framework. Shamshad et al. [17] proposed a novel blockchain-based privacy and security preserving EHR sharing protocol. Huang et al. [18] proposed a blockchain-based eHealth system, in which the manipulation of EHRs can be audited. Zhu et al. [19] proposed an improved convolution Merkle tree-based blockchain electronic medical record secure storage scheme. Uddin et al. [20] proposed a blockchain leveraged decentralized eHealth architecture. Tanwar et al. [21] explored several solutions that use blockchain technology to improve the current limitations of medical systems, including frameworks and tools for measuring the performance of such systems. Using off-chain and on-chain blockchain

system design, Miyachi et al. [22] proposed a modular hybrid privacy-preserving framework. Chen et al. [23] proposed a complete medical information system model based on blockchain technology to realize the goal of safe storage and sharing of medical data. Hossein et al. [24] proposed a novel blockchain-based privacy-preserving architecture for IoT healthcare applications. A summary of related works is shown in Table 1.

1.2. Motivation and Contributions. Among the previous blockchain-based location sharing schemes, Zyskind et al.'s scheme only provides decentralization, untamperability, and confidentiality [7], while Amoretti et al.'s scheme only provides decentralization, untamperability, and verifiability [8]. These are far from enough. Ji et al. [9] considered decentralization, untamperability, confidentiality, multilevel privacy protection, retrievability, and verifiability, but their scheme is insecure in practice. The adversary can recover any user's location effectively. Thus, it is of great significance to propose secure and practical blockchain-based multilevel privacy-preserving location sharing schemes. The comparison of previous blockchain-based location sharing schemes is shown in Table 2, where “√” means satisfied, “×” means dissatisfied, and “-” means uninvolved.

The main contributions of this paper are as follows:

- (1) We analyze the security of Ji et al.'s BMPLS scheme [9] and show that it has fatal flaws in confidentiality and multilevel privacy-preserving. An adversary outside the system can use an ordinary personal computer to completely break the system within a dozen hours and obtain the location information of any patient at any time. The adversary inside the system can use an ordinary personal computer to obtain the location information of the designated patient within tens of seconds. In addition, in some cases, their scheme cannot be executed.
- (2) Using salting technology, we propose an improved BMPLS scheme to fix our attacks. We add **Setup** and **Key generation** phases to the scheme to provide the foundation for other phases and replace the **Location verification** phase with the **Location extraction** phase. We also optimized the BMPLS scheme to make it correct and executable. The security analysis shows that the improved BMPLS scheme achieves decentralization, untamperability, confidentiality, multilevel privacy-preserving, retrievability, and verifiability. The simulation shows that the improved BMPLS scheme is practical, the computational overhead of the location record phase is within 10 ms, and the computational overheads of the location sharing and location extraction phases are both within 30 ms.

1.3. Organization. The rest of this paper is organized as follows. Section 2 presents preliminaries. Section 3 presents the architecture of BMPLS and reviews Ji et al.'s BMPLS scheme. Section 4 analyzes the scheme of Ji et al. Section 5

TABLE 1: Summary of related works.

| Ref. | Contribution | Technologies used | Key features |
|------|--|---|---|
| [11] | A blockchain-based data preservation system for medical data | Blockchain and Ethereum | Ensuring the primitiveness and verifiability of stored data while preserving privacy for users |
| [12] | An APP for health data sharing based on blockchain | Blockchain and secure multiparty computing | Patients own and control their healthcare data and use the indicator centric schema to organize personal healthcare data |
| [13] | A cloud-assisted secure eHealth systems | Blockchain and cloud storage | Every operation of the outsourced EHRs is recorded on the blockchain |
| [14] | An eHealthcare system interoperating with WBANs | Blockchain and wireless body area network | Providing a secure and low-power healthcare solution; utilizing the WBAN and blockchain technology |
| [15] | A blockchain-based privacy-preserving scheme | Blockchain, smart contract, zero-knowledge proof, and proxy reencryption | Achieving the data availability and consistency between patients and research institutions; using zero-knowledge proof to protect patient's privacy |
| [16] | A patient-centric health information exchange framework | Blockchain, data segmentation, and smart contract | Utilizing the smart contract feature to protect data security and patients privacy, ensure data provenance, and provide patients full control of their health records |
| [17] | A secure blockchain-based eHealth records storage and sharing scheme | Private and consortium blockchain and proxy reencryption | All EHRs are public-key encrypted and searchable, using private blockchain to store EHRs and consortium blockchain to store secure indexes |
| [18] | A blockchain-based eHealth system | Blockchain, cloud computing, and attributes-based proxy reencryption | Each legitimate manipulation will be written into the blockchain, and any threatening behavior will be discovered |
| [19] | Improved convolution Merkle tree | Blockchain and convolution operation | Using the convolutional layer structure to replace the original binary tree structure |
| [20] | An IoT eHealth framework based on blockchain | Blockchain, fog computing, edge computing, fuzzy inference, and task offloading | A patient agent (PA) software processes medical data, executes consensus mechanism, and utilizes a task-offloading algorithm to ensure patient's privacy |
| [21] | Blockchain-based electronic healthcare record system | Hyperledger fabric and Wireshark capture engine | Adopting chain code to ensure proper operation of blockchain ledger |
| [22] | Blockchain framework using on-chain and off-chain design | Blockchain and on-chain and off-chain design | Unpluggable components in the face of different data types to cope with different policy requirements |
| [23] | A blockchain-based preserving and sharing system | Proxy reencryption and hyperledger fabric | Real-time patient data collection and managing data with chain code |
| [24] | A blockchain-based architecture for IoT healthcare applications | Blockchain | Using a dual-chain architecture to develop access control policies that isolate data and policies |

TABLE 2: Comparison of previous schemes.

| Schemes | Zyskind et al.'s [7] | Amoretti et al.'s [8] | Ji et al.'s [9] | Our scheme |
|-----------------------|----------------------|-----------------------|-----------------|------------|
| Decentralization | √ | √ | √ | √ |
| Untamperability | √ | √ | √ | √ |
| Confidentiality | √ | × | × | √ |
| Multilevel protection | × | × | × | √ |
| Retrievability | — | — | √ | √ |
| Verifiability | — | √ | √ | √ |

proposes an improvement to fix our attacks with security and performance analysis. Section 6 concludes this paper.

2. Preliminaries

2.1. Order-Preserving Encryption. Order-preserving encryption (OPE) is deterministic encryption that can preserve numerical ordering on their plaintext space [25].

An order-preserving encryption scheme Π_{OPE} consists of the following algorithms OP-KeyGen and OP-Enc:

- (1) OP-KeyGen is the key generation algorithm, takes as input the security parameter λ , and outputs a secret key opk .
- (2) OP-Enc is the encryption algorithm, takes as input a secret key opk and a plaintext $x \in \{0, 1\}^n$ interpreted as a numerical value $0 \leq x \leq 2^n - 1$, and outputs ciphertext $c \in \{0, 1\}^m$ interpreted as a numerical value $0 \leq c \leq 2^m - 1$.

They satisfy that $OP-Enc(opk, i) < OP-Enc(opk, j)$, for all $opk \leftarrow OP-KeyGen(1^\lambda)$, $0 \leq i < j \leq 2^n - 1$.

For $N, M \subseteq \mathbb{N}$ with $|N| \leq |M|$, a function $f: N \rightarrow M$ is order-preserving if for all $i, j \in N$, $f(i) > f(j)$ iff $i > j$. If an order-preserving encryption scheme Π_{OPE} is secure, then $OP\text{-}Enc(\cdot)$ is a pseudorandom order-preserving function [25].

2.2. Merkle Tree. Merkle tree [26] provides efficient data authentication. A Merkle tree is based on a binary tree and a one-way hash function. The value of its leaf node is the data, and the value of its nonleaf node is the hash of the values of its two child nodes. If the Merkle tree has n leaf nodes, it only needs at most $\lceil \log_{2n} \rceil$ data to authenticate a leaf node, not all n data.

In our improved scheme, we need to calculate the Merkle tree of $node_0^x, node_1^x, node_2^x, \dots, node_{2^N}^x$, and Figure 1 is an illustrative example. In order to authenticate h_3 , we only need to provide h_4 and h_6 , then calculate $h_7' = Hash(h_3 || h_4)$ and $h_8' = Hash(h_6 || h_7')$ in sequence, and finally verify whether $h_8' = h_8$.

3. Review of Ji Et Al.'s BMPLS Scheme

This section reviews the architecture of BMLS and Ji et al.'s scheme.

3.1. Architecture of BMPLS. BMLS can be used as a module of TMIS to manage and share patient location information. The architecture of BMPLS is shown in Figure 2. There are two types of entities in the system: location data owner (LDO) and location data requestor (LDR). LDOs, such as infectious disease or chronic disease patients, record their location information in the blockchain. LDRs request LDOs' location information in different levels according to their trust levels and actual needs. For example, mobile clinics need to know the precise location of patients to provide them with on-site services; the medical center also requires an exact location to deliver the medicine. In contrast, infectious disease investigators only need to know the range of the patient, not the precise location.

3.2. Ji Et Al.'s BMPLS Scheme. Ji et al.'s BMPLS scheme consists of the following three phases [9].

3.2.1. Initialization. The location data owner (LDO) represents his visit region as a coordinate region $S = \{(x, y) | 0 \leq x \leq X, 0 \leq y \leq Y\}$, runs Algorithm 1 to generate the registration record $regRec$, and puts it into the blockchain.

In Algorithm 1, the partition function $Parti(S, N) = \{x_i = i \times X/2^N, 1 \leq i \leq 2^N\} \cup \{y_i = i \times Y/2^N, 1 \leq i \leq 2^N\}$. $OP\text{-}Enc(\cdot, \cdot)$ is the encryption algorithm of an order-preserving encryption scheme Π_{OPE} and $Hash$ is a hash function. $genMT(\cdot)$ denotes the function of using leaf nodes to generate the complete Merkle tree, $Translate^{-1}$ denotes the function of converting the location record into the actual geographic location, and $Sig_{LDO}(\cdot)$ denotes the LDO's signature.

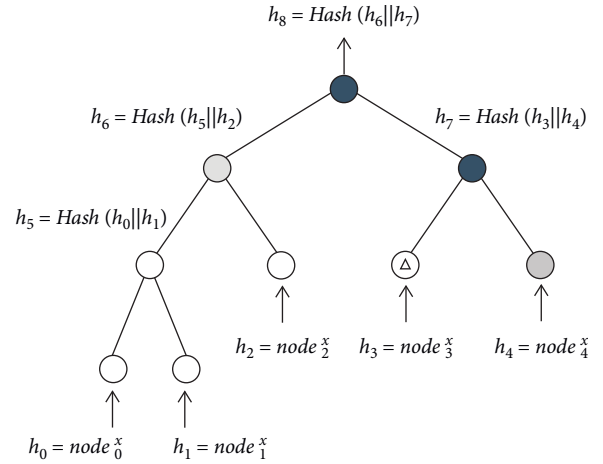


FIGURE 1: Merkle tree with leaf nodes $node_0^x, node_1^x, node_2^x, node_3^x$, and $node_4^x$.

3.2.2. Location Record. The LDO runs Algorithm 2 to generate a location record $record_j^{LDO}$ and puts it into the blockchain.

In Algorithm 2, $Enc(\cdot)$ is the encryption algorithm of the symmetric encryption scheme.

3.2.3. Location Sharing. The location sharing phase consists of the location sharing stage and location verification stage.

(1) *Location Sharing.* When the location data requestor (LDR) wants to obtain the location corresponding to $record_k^{LDO}$, he generates a request $request \leftarrow pub_{LDR} || recoId_k^{LDO} || n || signature_{LDR}$ and sends it to LDO.

The LDO returns location with corresponding granularity according to the trust level of the LDR. (1) If the LDR is fully trusted (level $n = \infty$), the LDO returns an accurate location. (2) If the LDR is semitrusted with level n , the LDO returns a rectangular border with side length 2^n . See Algorithm 3 for details.

In Algorithm 3, $PK\text{-}Enc(\cdot, \cdot)$ is the encryption algorithm of a public-key encryption scheme. id_1, id_2, id_3 , and id_4 are the subscripts of $x_{min}, x_{max}, y_{min}$, and y_{max} , respectively. $nodes^x$ and $nodes^y$ are the Merkle tree data subsets required to authenticate $node_{id_1}^x, node_{id_2}^x$ and $node_{id_3}^y, node_{id_4}^y$, respectively.

(2) *Location Verification.* After receiving the response, LDR runs Algorithm 4 to verify it.

In Algorithm 4, $Dec(\cdot, \cdot)$ is the decryption algorithm corresponding to $Enc(\cdot, \cdot)$. $A.c$ denotes c in $A = b || c$ and $MerkleHash$ denotes the function that uses a Merkle tree data subset to calculate its root value.

4. Cryptanalysis of Ji Et Al.'s BMPLS Scheme

Ji et al. [9] claim that their BMPLS scheme achieves decentralization, untamperability, confidentiality, multilevel privacy-preserving, retrievability, and verifiability. Unfortunately, we find that their scheme has fatal flaws in

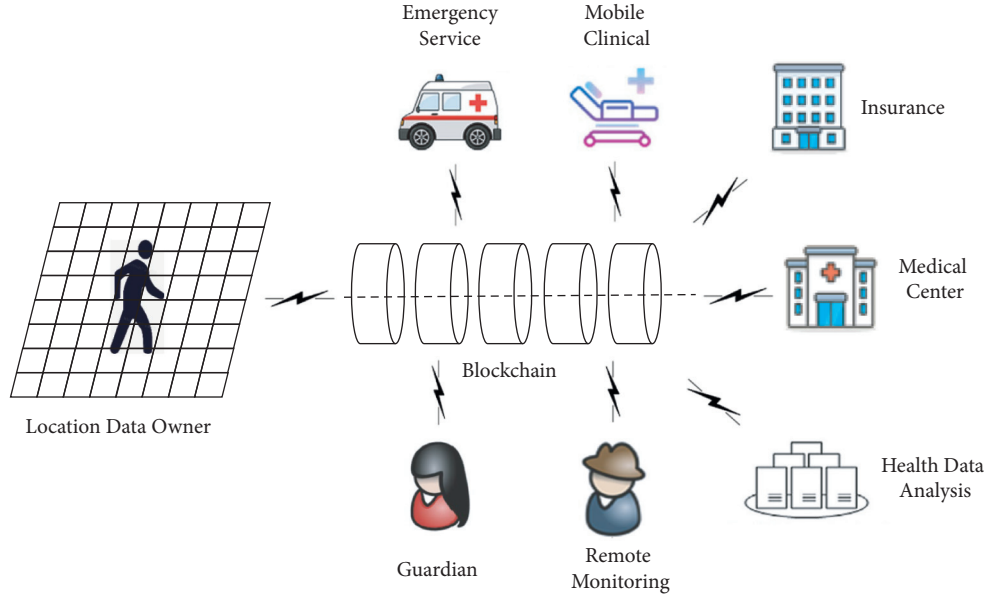


FIGURE 2: Architecture of BMPLS.

Input:

Region $S = \{(x, y) | 0 \leq x \leq X, 0 \leq y \leq Y\}$; Maximum level of location partition N ; LDO's secret key $k_{LDO} = k_{LDO}^x \| k_{LDO}^y$

Output:

Registration record $regRec$

- (1) $\{x_1, x_2, \dots, x_{2^N}\} \cup \{y_1, y_2, \dots, y_{2^N}\} \leftarrow \text{Parti}(S, N)$;
- (2) **for** $i = 1; i \leq 2^N; i++$ **do**
- (3) $ciph_i^x = \text{OP-Enc}(k_{LDO}^x, x_i)$;
- (4) $ciph_i^y = \text{OP-Enc}(k_{LDO}^y, y_i)$;
- (5) $node_i^x = \text{Hash}(i \| x_i \| ciph_i^x)$;
- (6) $node_i^y = \text{Hash}(i \| y_i \| ciph_i^y)$;
- (7) **end for**
- (8) $horTree \leftarrow \text{genMT}(node_1^x, node_2^x, \dots, node_{2^N}^x)$;
- (9) $verTree \leftarrow \text{genMT}(node_1^y, node_2^y, \dots, node_{2^N}^y)$;
- (10) $regRec \leftarrow \text{Sig}_{LDO}(\text{Translate}^{-1} \| horTree_{root} \| verTree_{root})$;
- (11) **return** $regRec$.

ALGORITHM 1: Generation of registration record.

Input:

LDO's j -th location (x_j, y_j) ; LDO's secret keys $k_{LDO} = k_{LDO}^x \| k_{LDO}^y, k_{sym}$; LDO's public-key pub_{LDO}

Output:

Location record $record_j^{LDO}$

- (1) **LDO executes:**
- (2) $ciph_j^x = \text{OP-Enc}(k_{LDO}^x, x_j)$;
- (3) $ciph_j^y = \text{OP-Enc}(k_{LDO}^y, y_j)$;
- (4) $ciph_j \leftarrow ciph_j^x \| ciph_j^y$;
- (5) $\text{OpeHash}_j \leftarrow \text{Hash}(ciph_j)$;
- (6) $\text{LocHash}_j \leftarrow \text{Hash}(x_j \| y_j)$;
- (7) $\text{SymCih}_j \leftarrow \text{Enc}(k_{sym}, x_j \| y_j)$;
- (8) $\text{LocInfo}_j \leftarrow \text{OpeHash}_j \| \text{LocHash}_j \| \text{SymCih}_j \| \text{timestamp}_j$;
- (9) $record_j^{LDO} \leftarrow pub_{LDO} \| \text{LocInfo}_j \| recId_{j-1}^{LDO} \| \text{signature}_{LDO}$;
- (10) **return** $record_j^{LDO}$;

ALGORITHM 2: Generation of location record.

Input:
 Location record ID recoId_k^{LDO} ; Session key between LDO and LDR k_{ses} ; Privacy protection level n ; LDR's public-key pub_{LDR}

Output:
 Shared location information response

- (1) **LDR executes:**
- (2) $\text{request} \leftarrow \text{pub}_{LDR} \parallel \text{recoId}_k^{LDO} \parallel n \parallel \text{signature}_{LDR}$;
- (3) LDR sends request to LDO;
- (4) **LDO executes:**
- (5) **if** $n = \infty$ **then**
- (6) $\text{response} \leftarrow \text{Enc}(k_{ses}, x_k \parallel y_k) \parallel \text{PK-Enc}(pub_{LDR}, k_{ses})$
- (7) **else if** $0 \leq n \leq N$ **then**
- (8) find the border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$ in level n
- (9) $\text{borInfo}_{id_1} \leftarrow id_1 \parallel x_{\min} \parallel \text{ciph}_{id_1}^x$;
- (10) $\text{borInfo}_{id_2} \leftarrow id_2 \parallel x_{\max} \parallel \text{ciph}_{id_2}^x$;
- (11) $\text{borInfo}_{id_3} \leftarrow id_3 \parallel y_{\min} \parallel \text{ciph}_{id_3}^y$;
- (12) $\text{borInfo}_{id_4} \leftarrow id_4 \parallel x_{\max} \parallel \text{ciph}_{id_4}^y$;
- (13) $\text{borInfo} \leftarrow \text{borInfo}_{id_1} \parallel \text{borInfo}_{id_2} \parallel \text{borInfo}_{id_3} \parallel \text{borInfo}_{id_4}$;
- (14) $\text{nodes}^x \leftarrow \{\text{node}_{x_1}^x, \text{node}_{x_2}^x, \dots\}$;
- (15) $\text{nodes}^y \leftarrow \{\text{node}_{y_1}^y, \text{node}_{y_2}^y, \dots\}$;
- (16) $\text{response} \leftarrow \text{Enc}(k_{ses}, \text{ciph}_k \parallel \text{borInfo} \parallel \text{nodes}^x \parallel \text{nodes}^y) \parallel \text{PK-Enc}(pub_{LDR}, k_{ses})$;
- (17) **end if**
- (18) **return** response.

ALGORITHM 3: Location sharing.

confidentiality and multilevel privacy-preserving. In addition, in some cases, their scheme cannot be executed.

4.1. On Confidentiality. The adversary can recover any LDO's location effectively.

Ji et al.'s BMPLS scheme uses the one-way property of the hash function to protect location. It is feasible and secure in the case of infinite (enough) locations. However, it is insecure in current practical applications because the amount of user locations is not enough to resist brute force attacks. The attack is as follows:

- (1) Calculate the coordinate hash table T of all possible locations of the LDO's visit region.
- (2) Obtain a record recoId_j^{LDO} from the blockchain, and extract the hash value LocHash_j .
- (3) Find out the location $x_j \parallel y_j$ corresponding to LocHash_j in table T .

We take the commonly used Global Positioning System (GPS) as an example to evaluate the feasibility of the above attack. In the GPS, the longitude and latitude output formats are $dddmm.mmmm$ and $ddmm.mmmm$, respectively. So there are only 3.60×10^{11} positions in a square area of 1° in longitude and latitude. If estimated by 40° north latitude where New York City is located, 1 degree of longitude and 1 degree of latitude are equivalent to 85 and 111 kilometers, respectively, and a square area with 1 degree of longitude and latitude is 9,435 square kilometers. The land area of New York City is 789 square kilometers, so the number of available GPS locations is about 3.01×10^{10} .

We experiment on a personal computer using Python-3.9.7 and PyCharm Community Edition 2021.2.2 (64 bits).

The configuration is CPU: Intel(R) Core(TM) i9-10900 CPU @ 2.80 GHz~2.81 GHz, RAM: 64 G, and OS: Windows 10 Home (Chinese) 64 bits (10.0.19041). As in [9], we also choose SHA-256 as the hash function. The simulation result shows that the time cost to calculate the hash values of 10^9 position coordinates is 27.01 minutes, and the size of the coordinate hash table is 77.20 GB. Therefore, using such a personal computer can calculate the hash value of all locations of New York City in about 13 hours and 33 minutes.

The above analysis and experiments show that if Ji et al.'s BMPLS scheme is used for New York City, the adversary can completely break the system in about 13.55 hours using a personal computer.

If we use supercomputers or adopt distributed computing or cloud computing technology, we can break the system with very little time overhead.

4.2. On Multilevel Privacy-Preserving. Semitrusted LDR Can Obtain Accurate Location. Ji et al. claim that, in their scheme [9], semitrusted LDR is impossible to reduce the privacy protection region. Unfortunately, we found that semitrusted LDR can recover the LDO's accurate location effectively. This is a more severe attack than reducing the privacy protection region. The attack is as follows:

- (1) Obtain a record recoId_j^{LDO} from the blockchain, and extract the hash value LocHash_j .
- (2) Run Algorithm 3 interactively with LDO to obtain the response response.
- (3) Run Algorithm 4 to verify the response response, and extract the rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$.

Input:
Response from LDO response; Record in the blockchain recold_k^{LDO} ; Session key with $\text{LDO}k_{\text{ses}}$

Output:
Boolean variable b

```

(1) initialize  $b \leftarrow \text{False}$ ;
(2) if  $x'_k \| y'_k \leftarrow \text{Dec}(k_{\text{ses}}, \text{response})$  then
(3)   if  $\text{Hash}(x'_k \| y'_k) \text{ recold}_k^{LDO}.\text{LocInfo}_k \text{ LocHash}_k$  then
(4)      $b \leftarrow \text{True}$ 
(5)   end if
(6) else if  $\text{ciph}_k \| \text{borInfo} \| \text{nodes}^x \| \text{nodes}^y \leftarrow \text{Dec}(k_{\text{ses}}, \text{response})$  then
(7)    $\text{borInfo}_{id_1} \| \text{borInfo}_{id_2} \| \text{borInfo}_{id_3} \| \text{borInfo}_{id_4} \leftarrow \text{borInfo}$ 
(8)    $\text{node}_{id_1}^x \leftarrow \text{Hash}(\text{borInfo}_{id_1})$ ;
(9)    $\text{node}_{id_2}^x \leftarrow \text{Hash}(\text{borInfo}_{id_2})$ ;
(10)   $\text{node}_{id_3}^x \leftarrow \text{Hash}(\text{borInfo}_{id_3})$ ;
(11)   $\text{node}_{id_4}^x \leftarrow \text{Hash}(\text{borInfo}_{id_4})$ ;
(12)   $\text{horTree}_{root} \leftarrow \text{MerkleHash}\{\text{nodes}^x, \text{node}_{id_1}^x, \text{node}_{id_2}^x\}$ ;
(13)   $\text{vorTree}_{root} \leftarrow \text{MerkleHash}\{\text{nodes}^y, \text{node}_{id_3}^y, \text{node}_{id_4}^y\}$ 
(14)  if  $\text{horTree}_{root} = \text{horTree}_{root}$  and  $\text{vorTree}_{root} = \text{vorTree}_{root}$  then
(15)    if  $\text{Hash}(\text{ciph}_k) = \text{recold}_k^{LDO}.\text{LocInfo}_k.\text{OpeHash}_k$  and  $\text{borInfo}_{id_1}.\text{ciph}_{id_1}^x < \text{ciph}_k.\text{ciph}_k^x < \text{borInfo}_{id_2}.\text{ciph}_{id_2}^x$  and
       $\text{borInfo}_{id_3}.\text{ciph}_{id_3}^x < \text{ciph}_k.\text{ciph}_k^x < \text{borInfo}_{id_4}.\text{ciph}_{id_4}^x$  then
(16)       $b \leftarrow \text{True}$ 
(17)    end if
(18)  end if
(19) end if
(20) return  $b$ ;

```

ALGORITHM 4: Location verification.

Input:
Location hash value LocHash_j ; Rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$

Output:
Location information $x_j \| y_j$

```

(1) for  $x$  in  $(x_{\min}, x_{\max})$  do
(2)   for  $y$  in  $(y_{\min}, y_{\max})$  do
(3)     Hash  $\leftarrow \text{Hash}(x \| y)$ ;
(4)     if Hash =  $\text{LocHash}_j$  then
(5)       return  $x \| y$ ;
(6)     end if
(7)   end for
(8) end for

```

ALGORITHM 5: Location information extraction.

- (4) Search for the location where the hash value is equal to $\text{recold}_j^{LDO}.\text{LocInfo}_j.\text{LocHash}_j$ in the rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$; that is, run Algorithm 5 to obtain the accurate location information $x_j \| y_j$.

Take $x_{\min} = 00587654$, $x_{\max} = 01012345$ as an example to illustrate how to make x traverse (x_{\min}, x_{\max}) . Because order-preserving encryption requires the plaintext to be a positive integer, the format of the location needs to be changed from $ddmm.mmmm$ to $ddmmmmmm$. Note that there are no location coordinates like $dd60mmmm$, $dd61mmmm$, ..., $dd99mmmm$, and $(00587654, 01012345)$ must be divided into $(00587654, 00599999)$ and $(01000000, 01012345)$. Then, the former is expressed as " $x = 00587655; x \leq 0059999; x++$ " and the latter as " $x = 01000000; x \leq 01012344; x++$ ".

We evaluated the feasibility of this attack in the above environment. Assume that the rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$ is a square area with a longitude and latitude of 0.5 minutes each. This is a rectangular area with a length of 928 meters and a width of 710 meters. The simulation result shows that the average cost to calculate the accurate location $x_j \| y_j$ in such an area is 20.75 seconds.

If a supercomputer is used or distributed computing or cloud computing technology is adopted, the time for LDR to find the accurate location will be milliseconds.

4.3. Other Weaknesses

- (1) In [9], the coordinates $x_i = i \times X/2^N$ and $y_i = i \times Y/2^N$ may not be integers. But in Definition 1 of [9],

the plaintext space is clearly defined as the integer set $[m] = \{i | 1 \leq i \leq m\}$. Therefore, *it will not be feasible to encrypt x_i and y_i in Algorithm 1 (lines 3 and 4).*

- (2) The LDO's location coordinate x_k or y_k may happen to be a position grid coordinate x_i or y_i . If it happens, *multilevel privacy protection will not be achieved.* For example, if $x_k = x_{2^{N-1}}$, then we have $x_k = x_{\min}$ or $x_k = x_{\max}$ for any trust level n . Then, the inequality in line 15 of Algorithm 4 will not hold, so LDR will not accept any rectangular border provided by LDO. Further, because $\text{borInfo}_{id_1} \cdot \text{ciph}_{id_1}^x = \text{ciph}_k \cdot \text{ciph}_k^x$ or $\text{ciph}_k \cdot \text{ciph}_k^x = \text{borInfo}_{id_2} \cdot \text{ciph}_{id_2}^x$, LDR will determine that the x -coordinate of LDO is x_{\min} or x_{\max} .
- (3) The n -level rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$ must be in the form of $\{x_{i \times 2^n}, x_{(i+1) \times 2^n}, y_{j \times 2^n}, y_{(j+1) \times 2^n}\}$, to ensure that the privacy protection region will not be reduced. If the LDO's location coordinate x_k or y_k satisfies $x_k < x_1$ or $y_k < y_1$, then have $x_{\min} = x_0 = 0$ or $y_{\min} = y_0 = 0$. In this case, both Algorithms 3 and 4 *cannot be executed.*
 - (a) Because 0 is not in the plaintext space of order-preserving encryption, Algorithm 3 cannot generate the ciphertext $\text{ciph}_{id_1}^x$ (or $\text{ciph}_{id_1}^y$) of $x_{\min} = 0$ (or $y_{\min} = 0$) (see Definition 1 of [9]).
 - (b) node_0^x and node_0^y are used when calculating $\text{horTree}_{\text{root}}$ and $\text{vorTree}_{\text{root}}$, but they are not used when calculating horTree and verTree , which will make neither $\text{horTree}_{\text{root}} = \text{horTree}_{\text{root}}$ nor $\text{vorTree}_{\text{root}} = \text{verTree}_{\text{root}}$ holds.

5. Improvement

We present an improvement to fix our attacks in this section and optimize the BMLS scheme to make it correct and executable.

5.1. Ideas for Improvement. The main improvement ideas are as follows:

- (1) Add **Setup** and **Key generation** phases to the scheme to provide the foundation for other phases.
- (2) Replace the **Location verification** phase with the **Location extraction** phase. The original location verification algorithm only returns the verification result but not the location, while our extraction algorithm returns the location with the verification result.
- (3) Use "Salt" to prevent brute force attacks. When calculating the hash value LocHash_j of the location $x_j \| y_j$, select a random number (salt) r_j and let $\text{LocHash}_j = \text{Hash}(x_j \| y_j \| r_j)$. As long as the length of the random number is large enough, the brute force attacks in Sections 4.1 and 4.2 cannot be implemented, even with supercomputers or cloud computing.
- (4) Use integer steps to generate grid coordinates to ensure that the coordinate values are positive integers.

- (5) When the LDO's location coordinate x_k or y_k is equal to the grid coordinate x_i or y_i , we make a minimum positive disturbance to the location coordinates to ensure they are not equal. In GPS, the minimum positive disturbance is 0.0001 minutes. This disturbance is only 18.5 cm and does not affect the actual performance of the system.

5.2. Improved Scheme. The improved BMPLS scheme consists of the following five phases.

5.2.1. Setup

- (1) Choose an order-preserving encryption scheme $\Pi_{\text{OPE}} = (\text{OP-KeyGen}, \text{OP-Enc})$ and set up public parameters [25].
- (2) Choose a public-key encryption scheme $\Pi_{\text{PKE}} = (\text{PK-KeyGen}, \text{PK-Enc}, \text{PK-Dec})$, such as RSA1024 and RSA2048, and set up public parameters.
- (3) Choose a signature scheme $\Pi_{\text{Sig}} = (\text{Sig-KeyGen}, \text{Sig}, \text{Ver})$, such as DSA1024 and DSA2048, and set up public parameters.
- (4) Choose a symmetric encryption scheme $\Pi_{\text{SE}} = (\text{S-Enc}, \text{S-Dec})$, such as AES128 and AES256, and set up public parameters.
- (5) Choose a one-way collision resistance hash function Hash, such as SHA256 and SHA512.
- (6) Set up and deploy a membership service provider (MSP) to maintain the identity of all users in the system and issue certificates for authentication and signature verification [27].
- (7) Set up and deploy a trust level service provider (TLSP) to maintain the trust level of all users in the system and provide trust level query services [28].

5.2.2. Key Generation

- (1) LDO runs Π_{OPE} 's key generation algorithm OP-KeyGen twice and obtains his x -coordinate encryption secret key $\text{opk}_{\text{LDO}}^x$ and y -coordinate encryption secret key $\text{opk}_{\text{LDO}}^y$, respectively.
- (2) LDO runs Π_{Sig} 's key generation algorithm Sig-KeyGen and obtains his public-key pk_{LDO} and signing key sk_{LDO} .
- (3) LDO chooses his secret key k_{LDO} for Π_{SE} .
- (4) LDR runs Π_{PKE} 's key generation algorithm PK-KeyGen and obtains his public-key pk_{LDR} and private key sk_{LDR} .

5.2.3. Initialization. LDO selects the origin (\hat{x}_0, \hat{y}_0) , step lengths x_{sl}, y_{sl} , and partition level N such that $\{(x, y) | \hat{x}_0 \leq x \leq \hat{x}_0 + 2^N \times x_{sl}, \hat{y}_0 \leq y \leq \hat{y}_0 + 2^N \times y_{sl}\}$ can cover his visit region properly. Assume that $\hat{x}_0, \hat{y}_0, x_{sl}$, and y_{sl} have all been converted into positive integers. For example, convert *dddmm.mmmm* and *ddmm.mmmm* into

Input:
 Region $\text{region} = \{\hat{x}_0, \hat{y}_0, x_{sl}, y_{sl}, N\}$; LDO's secret keys opk_{LDO}^x and opk_{LDO}^y ; LDO' signing key sk_{LDO}

Output:
 Registration record regRec

- (1) **for** $i = 0; i \leq 2^N; i + \mathbf{do}$;
- (2) $x_i = i \times x_{sl}$;
- (3) $y_i = i \times y_{sl}$;
- (4) $\text{ciph}_i^x = \text{OP-Enc}(opk_{LDO}^x, x_i)$;
- (5) $\text{ciph}_i^y = \text{OP-Enc}(opk_{LDO}^y, y_i)$;
- (6) $\text{node}_i^x = \text{Hash}(i \| x_i \| \text{ciph}_i^x)$;
- (7) $\text{node}_i^y = \text{Hash}(i \| y_i \| \text{ciph}_i^y)$;
- (8) **end for**
- (9) $\text{horTree} \leftarrow \text{genMT}(\text{node}_0^x, \text{node}_1^x, \text{node}_2^x, \dots, \text{node}_{2^N}^x)$;
- (10) $\text{verTree} \leftarrow \text{genMT}(\text{node}_0^y, \text{node}_1^y, \text{node}_2^y, \dots, \text{node}_{2^N}^y)$;
- (11) $\text{regRec} \leftarrow \text{region} \| \text{horTree}_{\text{root}} \| \text{verTree}_{\text{root}} \| \text{Sig}(sk_{LDO}, \text{region} \| \text{horTree}_{\text{root}} \| \text{verTree}_{\text{root}})$;
- (12) **return** regRec .

ALGORITHM 6: Registration record generation.

$ddmmmmmm$ and $ddmmmmmm$, respectively. We choose the order-preserving encryption scheme Π_{OPE} that allows the plaintext to be 0.

Denote the visit region as $\text{region} = \{\hat{x}_0, \hat{y}_0, x_{sl}, y_{sl}, N\}$. LDO executes Algorithm 6 to generate the registration record regRec and puts it into the blockchain.

5.2.4. Location Record. Suppose that the j -th location of the LDO is $(\hat{x}_0 + x_j, \hat{y}_0 + y_j)$. LDO runs Algorithm 7 to generate a location record record_j^{LDO} and puts it into the blockchain.

5.2.5. Location Sharing. The location sharing phase consists of the location sharing stage and location extraction stage.

(1) *Location Sharing.* When LDR wants to obtain the location corresponding to record_k^{LDO} , he generates a request $\text{request} = \text{recoId}_k^{LDO} \| n \| \text{Sig}(sk_{LDR}, \text{recoId}_k^{LDO} \| n)$ and sends it to LDO.

LDO generates a response response and returns it to LDR as follows:

- (1) It requests the LDR's certificate from the membership service provider (MSP) and then verifies the request. If it is invalid, it returns $\text{response} = \perp$ and aborts.
- (2) It requests the LDR's privacy protection level n_{LDR} from the trust level service provider (TLSP). If $n_{LDR} \neq \infty$ and $n < n_{LDR}$, it returns $\text{response} = \perp$ and aborts.
- (3) It gets the location record record_k^{LDO} from the blockchain.
- (4) It runs Algorithm 8 and returns response to LDR.

(2) *Location Extraction.* Receiving the response response , LDR gets the record record_k^{LDO} from the blockchain and then runs Algorithm 9 to extract the location information.

5.3. Security and Performance Analysis

5.3.1. Security Analysis. We follow the definitions of [9] to analyze the security of the improved BMPLS scheme as follows:

(1) *Decentralization.* One of the main advantages of blockchain is decentralization. The BMLS scheme uses blockchain to manage and store location information, so it inherits decentralization.

(2) *Untamperability.* The unforgeability mentioned in [9] is just untamperability, which is another main property of blockchain. For the same reason as above, the BMLS scheme also inherits untamperability.

(3) *Confidentiality.* In the improved BMPLS scheme, the only data that needs confidentiality protection is LDO's location coordinate (x_j, y_j) . The scheme uses four different cryptographic techniques to protect it: hash function, symmetric encryption, order-preserving encryption, and hybrid encryption. Because the hash function cannot achieve indistinguishable security, the improved BMPLS scheme can only achieve confidentiality in the "all-or-nothing" sense, that is, *prevent the adversary from recovering the location coordinates*. We analyze the probability of a successful attack in different situations as follows:

- (1) *Get (x_j, y_j) from $\text{LocHash}_j = \text{Hash}(x_j \| y_j \| r_j)$.* Let k be the binary bit length of r_j . According to the one-way property of the hash function, the probability that the adversary obtains $x_j \| y_j$ from LocHash_j is less than 2^{-k} , even if he knows a small rectangular border. When k is large enough, such as $k = 160$, the probability is negligible.
- (2) *Get (x_j, y_j) from $\text{SymCih}_j = \text{S-Enc}(k_{LDO}, x_j \| y_j \| r_j)$.* Under the assumption that the symmetric encryption scheme Π_{SE} is secure, the probability that

Input:
 LDO's j -th location (x_j, y_j) ; Region $\text{region} = \{\hat{x}_0, \hat{y}_0, x_{sl}, y_{sl}, N\}$; LDO's secret keys $(opk_{LDO}^x, opk_{LDO}^y)$ and k_{LDO} ;
 LDO's signing key sk_{LDO}

Output:
 Location record record_j^{LDO}

- (1) **for** $i = 0; i \leq 2^N; i + +$ **do**;
- (2) **if** $x_j = i \times x_{sl}$ **then**
- (3) $x_j \leftarrow x_j + 1$;
- (4) **end if**
- (5) **if** $y_j = i \times y_{sl}$ **then**
- (6) $y_j \leftarrow y_j + 1$;
- (7) **end if**
- (8) **end for**
- (9) $\text{ciph}_j^x = \text{OP-Enc}(opk_{LDO}^x, x_j)$;
- (10) $\text{ciph}_j^y = \text{OP-Enc}(opk_{LDO}^y, y_j)$;
- (11) $\text{ciph}_j \leftarrow \text{ciph}_j^x \parallel \text{ciph}_j^y$;
- (12) $\text{OpeHash}_j \leftarrow \text{Hash}(\text{ciph}_j)$;
- (13) Generate a random number r_j ;
- (14) $\text{LocHash}_j \leftarrow \text{Hash}(x_j \parallel y_j \parallel r_j)$;
- (15) $\text{SymCih}_j \leftarrow S - \text{Enc}(k_{LDO}, x_j \parallel y_j \parallel r_j)$;
- (16) $\text{LocInfo}_j \leftarrow \text{OpeHash}_j \parallel \text{LocHash}_j \parallel \text{SymCih}_j \parallel \text{timestamp}_j$;
- (17) $\text{record}_j^{LDO} \leftarrow \text{LocInfo}_j \parallel \text{recId}_{j-1}^{LDO} \parallel \text{Sig}(sk_{LDO}, \text{LocInfo}_j \parallel \text{recId}_{j-1}^{LDO})$;
- (18) **return** record_j^{LDO} ;

ALGORITHM 7: Location record generation.

Input:
 Location record record_k^{LDO} ; LDO's secret keys k_{LDO} and $(opk_{LDO}^x, opk_{LDO}^y)$; Privacy protection level n ; LDR's public-key pk_{LDR}

Output:
 Shared location information response

- (1) $x_k \parallel y_k \parallel r_k \leftarrow S - \text{Dec}(k_{LDO}, \text{record}_k^{LDO}. \text{LocInfo}_k. \text{SymCih}_k)$;
- (2) choose a session key k_{ses}
- (3) **if** $n = \infty$ **then**
- (4) $\text{response} \leftarrow S - \text{Enc}(k_{ses}, x_k \parallel y_k \parallel r_k) \text{PK-Enc}(pk_{LDR}, k_{ses})$
- (5) **else if** $0 \leq n \leq N$ **then**
- (6) find i such that $i \times 2^n \times x_{sl} < x_k < (i + 1) \times 2^n \times x_{sl}$
- (7) $id_1 \leftarrow i \times 2^n$; $id_2 \leftarrow (i + 1) \times 2^n$;
- (8) $x_{\min} \leftarrow i \times 2^n \times x_{sl}$; $x_{\max} \leftarrow (i + 1) \times 2^n \times x_{sl}$;
- (9) find j such that $j \times 2^n \times y_{sl} < y_k < (j + 1) \times 2^n \times y_{sl}$
- (10) $id_3 \leftarrow j \times 2^n$; $id_4 \leftarrow (j + 1) \times 2^n$;
- (11) $y_{\min} \leftarrow j \times 2^n \times y_{sl}$; $y_{\max} \leftarrow (j + 1) \times 2^n \times y_{sl}$;
- (12) $\text{ciph}_{id_1}^x = \text{OP-Enc}(opk_{LDO}^x, x_{\min})$; $\text{ciph}_{id_2}^x = \text{OP-Enc}(opk_{LDO}^x, x_{\max})$;
- (13) $\text{ciph}_{id_3}^y = \text{OP-Enc}(opk_{LDO}^y, y_{\min})$; $\text{ciph}_{id_4}^y = \text{OP-Enc}(opk_{LDO}^y, y_{\max})$;
- (14) $\text{borInfo}_{id_1} \leftarrow id_1 \parallel x_{\min} \parallel \text{ciph}_{id_1}^x$; $\text{borInfo}_{id_2} \leftarrow id_2 \parallel x_{\max} \parallel \text{ciph}_{id_2}^x$;
- (15) $\text{borInfo}_{id_3} \leftarrow id_3 \parallel y_{\min} \parallel \text{ciph}_{id_3}^y$; $\text{borInfo}_{id_4} \leftarrow id_4 \parallel y_{\max} \parallel \text{ciph}_{id_4}^y$;
- (16) $\text{borInfo} \leftarrow \text{borInfo}_{id_1} \parallel \text{borInfo}_{id_2} \parallel \text{borInfo}_{id_3} \parallel \text{borInfo}_{id_4}$;
- (17) $\text{nodes}^x \leftarrow \{\text{node}_{x_1}^x, \text{node}_{x_2}^x, \dots\}$; $\text{nodes}^y \leftarrow \{\text{node}_{y_1}^y, \text{node}_{y_2}^y, \dots\}$;
- (18) $\text{ciph}_k^x = \text{OP-Enc}(opk_{LDO}^x, x_k)$; $\text{ciph}_k^y = \text{OP-Enc}(opk_{LDO}^y, y_k)$;
- (19) $\text{ciph}_k \leftarrow \text{ciph}_k^x \parallel \text{ciph}_k^y$;
- (20) $\text{response} \leftarrow S - \text{Enc}(k_{ses}, \text{ciph}_k \parallel \text{borInfo} \parallel \text{nodes}^x \parallel \text{nodes}^y) \parallel \text{PK-Enc}(pk_{LDR}, k_{ses})$;
- (21) **end if**
- (22) **return** response .

ALGORITHM 8: Location sharing.

```

Input:
Response from LDO response; Record in the blockchain record $_k^{LDO}$ ; LDR's private key  $sk_{LDR}$ 
Output:
Location  $x_k \| y_k$  or rectangular border  $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$ , or False
(1)  $k_{ses} \leftarrow \text{PK-Dec}(sk_{LDR}, \text{response.PK-Enc}(pk_{LDR}, k_{ses}))$ ;
(2) if  $x'_k \| y'_k \| r'_k \leftarrow \text{S-Dec}(k_{ses}, \text{response})$  then
(3)   if  $\text{Hash}(x'_k \| y'_k \| r'_k) = \text{record}_k^{LDO} \cdot \text{LocInfo}_k \cdot \text{LocHash}_k$  then
(4)     return  $x'_k \| y'_k$ ;
(5)   end if
(6) else if  $\text{ciph}_k \| \text{borInfo} \| \text{nodes}^x \| \text{nodes}^y \leftarrow \text{Dec}(k_{ses}, \text{response})$  then
(7)    $\text{borInfo}_{id_1} \| \text{borInfo}_{id_2} \| \text{borInfo}_{id_3} \| \text{borInfo}_{id_4} \leftarrow \text{borInfo}$ 
(8)    $\text{node}_{id_1}^x \leftarrow \text{Hash}(\text{borInfo}_{id_1})$ ;  $\text{node}_{id_2}^x \leftarrow \text{Hash}(\text{borInfo}_{id_2})$ ;
(9)    $\text{node}_{id_3}^y \leftarrow \text{Hash}(\text{borInfo}_{id_3})$ ;  $\text{node}_{id_4}^y \leftarrow \text{Hash}(\text{borInfo}_{id_4})$ ;
(10)   $\text{horTree}_{\text{root}}' \leftarrow \text{MerkleHash}\{\text{nodes}^x, \text{node}_{id_1}^x, \text{node}_{id_2}^x\}$ ;
(11)   $\text{vorTree}_{\text{root}}' \leftarrow \text{MerkleHash}\{\text{nodes}^y, \text{node}_{id_3}^y, \text{node}_{id_4}^y\}$ ;
(12)  if  $\text{horTree}_{\text{root}}' = \text{horTree}_{\text{root}}$  and  $\text{vorTree}_{\text{root}}' = \text{vorTree}_{\text{root}}$  then
(13)    if  $\text{Hash}(\text{ciph}_k) = \text{record}_k^{LDO} \cdot \text{LocInfo}_k \cdot \text{OpeHash}_k$  and
       $\text{borInfo}_{id_1} \cdot \text{ciph}_{id_1}^x < \text{ciph}_k \cdot \text{ciph}_k^x < \text{borInfo}_{id_2} \cdot \text{ciph}_{id_2}^x$  and
       $\text{borInfo}_{id_3} \cdot \text{ciph}_{id_3}^y < \text{ciph}_k \cdot \text{ciph}_k^y < \text{borInfo}_{id_4} \cdot \text{ciph}_{id_4}^y$  then
(14)       $x_{\min} \leftarrow \text{borInfo}_{id_1} \cdot x_{\min}$ ;  $x_{\max} \leftarrow \text{borInfo}_{id_2} \cdot x_{\max}$ ;
(15)       $y_{\min} \leftarrow \text{borInfo}_{id_3} \cdot y_{\min}$ ;  $y_{\max} \leftarrow \text{borInfo}_{id_4} \cdot y_{\max}$ ;
(16)      return  $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$ ;
(17)    end if
(18)  end if
(19) end if
(20) return False.

```

ALGORITHM 9: Location extraction.

the adversary obtains $x_j \| y_j$ from SymCih_j is negligible.

- (3) *Get* (x_k, y_k) from $\text{response} = \text{S-Enc}(k_{ses}, x_j \| y_j \| r_j) \| \text{PK-Enc}(pk_{LDR}, k_{ses})$. If the selected symmetric encryption scheme Π_{SE} and public-key encryption scheme Π_{PKE} are secure, then the probability that the adversary obtains (x_k, y_k) from $\text{S-Enc}(k_{ses}, x_k \| y_k \| r_k) \| \text{PK-Enc}(pk_{LDR}, k_{ses})$ is also negligible.
- (4) *Get* (x_k, y_k) from $\text{ciph}_k^x = \text{OP-Enc}(opk_{LDO}, x_k)$; $\text{ciph}_k^y = \text{OP-Enc}(opk_{LDO}, y_k)$. Similarly, if the selected order-preserving encryption scheme Π_{OPE} is secure, the probability of recovering the plaintext from the ciphertext is also negligible.

Therefore, the probability that the adversary obtains the LDO's location coordinate (x_j, y_j) is negligible, and the improved BMPLS scheme achieves confidentiality.

(4) *Multilevel Location Privacy Protection*. Multilevel location privacy protection aims to ensure that the semitrusted LDR cannot reduce the rectangular border of the LDO's location.

Firstly, confidentiality guarantees that semitrusted LDR cannot obtain the exact location coordinates of LDO.

Secondly, the n -level rectangular border obtained by LDR is $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$, where $x_{\min} = i \times 2^n \times x_{sl}$, $x_{\max} = (i+1) \times 2^n \times x_{sl}$, $y_{\min} = j \times 2^n \times y_{sl}$, and $y_{\max} = (j+1) \times 2^n \times y_{sl}$. Requesting n -level rectangular borders at

different times will get the same results, which will not help reduce the rectangular border. In addition, when $n_1 < n_2$, the n_1 -level rectangular border must be included in the n_2 -level rectangular border. It is impossible to reduce the rectangular border by finding the intersection of the rectangular borders of different levels.

Thirdly, because the order-preserving encryption algorithm OP-Enc is a pseudorandom order-preserving function controlled by a key, as long as the ciphertext space is large enough, such as 2^{160} , the probability of reversing ciph_i^x (or ciph_i^y) from $\text{node}_i^x = \text{Hash}(i \| x_i \| \text{ciph}_i^x)$ (or $\text{node}_i^y = \text{Hash}(i \| y_i \| \text{ciph}_i^y)$) will be negligible. Therefore, the adversary's probability of reducing the rectangular border by finding the coordinate ciphertext is negligible.

Finally, the x -coordinate and y -coordinate are encrypted with different keys, so there is no order-preserving relationship between the ciphertext of the x -coordinate and the ciphertext of the y -coordinate. $\text{ciph}_k^x < \text{ciph}_i^y$ (or $\text{ciph}_k^y > \text{ciph}_i^x$) does not imply $x_k < y_i$ (or $x_k > y_i$). It is impossible for the adversary to use y_{\min} and y_{\max} to narrow the range of x_k . The same is true for y_k .

(5) *Retrievability*. Retrievability means that the LDO's location coordinates can be retrieved effectively [9]. In the improved BMPLS scheme, the location coordinate (x_j, y_j) is stored in the blockchain via ciphertext $\text{SymCih}_j = \text{S-Enc}(k_{LDO}, x_j \| y_j \| r_j)$. LDO knows the decryption key k_{LDO} and can retrieve the location coordinate efficiently. On the other hand, in the location sharing phase, the LDO sends the

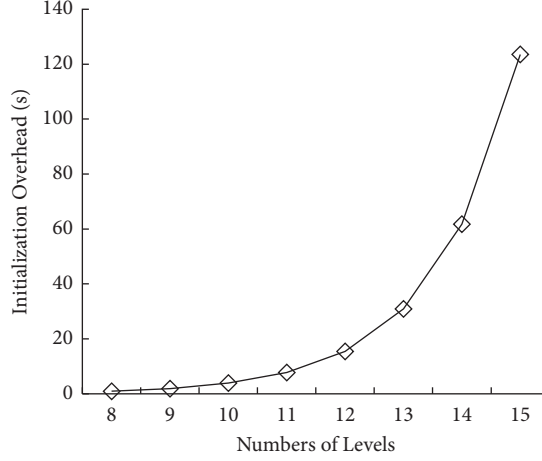


FIGURE 3: Initialization overhead.

location coordinate (x_k, y_k) to the LDR in the form of response = S-Enc($k_{ses}, x_k \| y_k \| r_k$) || PK-Enc(pk_{LDR}, k_{ses}). LDR can use its decryption key sk_{LDR} to decrypt the location coordinate (x_k, y_k) . So, the improved BMPLS scheme achieves retrievability.

(6) *Verifiability*. Verifiability means that LDR can use blockchain to verify whether the location information he gets is correct. The improved BMPLS scheme provides different verification methods in two cases:

- (1) The trusted LDR firstly decrypts the location coordinates and random number $x'_k \| y'_k \| r'_k$ and calculates the hash value $\text{Hash}(x'_k \| y'_k \| r'_k)$. Then, he extracts the hash value LocHash_k corresponding to the coordinate from the blockchain and checks whether $\text{Hash}(x'_k \| y'_k \| r'_k) = \text{LocHash}_k$. The untamperability of the blockchain and the signature of the LDO ensure that LocHash_k will not be tampered with, and the collision resistance of the hash function ensures that $x'_k \| y'_k \| r'_k$ will not be replaced, so (x'_k, y'_k) is correct.
- (2) The semitrusted LDR gets the coordinates of a rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$, with their order-preserving ciphertext $\{\text{ciph}_{id_1}^x, \text{ciph}_{id_2}^x, \text{ciph}_{id_3}^y, \text{ciph}_{id_4}^y\}$, and then verifies it by the Merkle tree and comparing the order of the ciphertext. Similarly, the untamperability and the signature ensure that $\text{horTree}_{\text{root}}$ and $\text{verTree}_{\text{root}}$ will not be tampered with, and the collision resistance of the Merkle tree ensures that $\text{ciph}_{id_1}^x, \text{ciph}_{id_2}^x, \text{ciph}_{id_3}^y, \text{ciph}_{id_4}^y$ are correct. Finally, the property of order-preserving encryption ensures that LDO's location (x_k, y_k) is in the rectangular border $\{x_{\min}, x_{\max}, y_{\min}, y_{\max}\}$.

The security comparison of related schemes is shown in Table 2. It shows that our improved scheme has apparent advantages in security.

5.3.2. *Performance Analysis*. We perform performance simulations under the same environment in Section 4. To ensure security, we set the ciphertext space of order-preserving encryption to be 2^{160} , denoted as OPE (2^{160}). Choose SHA, AES, RSA, and DSA as hash, symmetric encryption, public-key encryption, and digital signature algorithms. We simulate on two data scales, AES128+RSA1024+DSA1024 and AES256+RSA2048+DSA2048. We also use two hash functions, SHA256 and SHA512. Since the time cost of the hash function is very small, it has no significant impact on the simulation. The simulation results of each phase are as follows:

(1) *Initialization*. For initialization, we simulated the partition level N from 8 to 15. The result is shown in Figure 3. When $N = 10$, the LDO's initialization time is only 3.87 seconds. Estimated with a step length of 10 meters, the visit region region is a rectangle with a side length of 10.24 kilometers in this case. When $N = 13$, the LDO's initialization time is 30.89 seconds. In this case, the visit region's side length is 81.92 kilometers, which can meet the needs of most scenarios. When $N = 15$, the initialization of the LDO takes 123.53 seconds. At this time, the side length of the visit region has been as long as 327.68 kilometers.

The simulation shows that the initialization time is generally tens of seconds. Because each LDO only needs to be initialized once, this overhead is acceptable.

(2) *Location Record*. We conducted eight simulations of location record generation, and the results are shown in Figure 4. In the case of OPE(2^{160}), AES128, and DSA1024, the minimum overhead is 7.81 ms, the maximum is 8.17 ms, and the average is 8.00 ms. When the case of OPE(2^{160}), AES256, and DSA2048, the minimum is 8.53 ms, the maximum is 9.05 ms, and the average is 8.78 ms. In both cases, the overheads of location record generation are very small.

(3) *Location Sharing*. There are two cases of location sharing: $n = \infty$ (trusted LDR) and $0 \leq n \leq N$ (semitrusted LDR). The simulation results are shown in Figures 5(a) and 5(b).

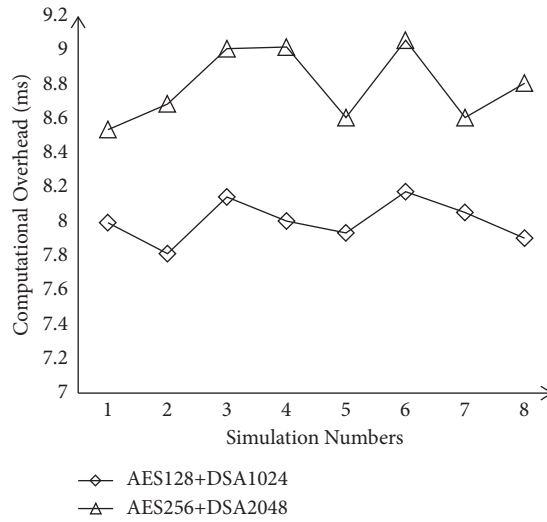


FIGURE 4: Record generation overhead.

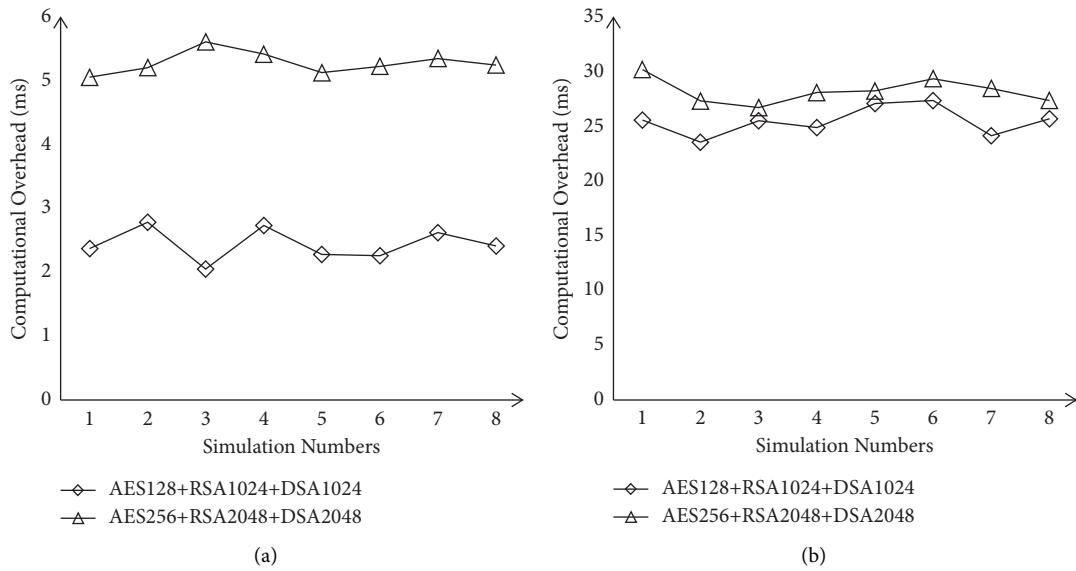


FIGURE 5: Location sharing overhead, (a) $n = \infty$, (b) $n = 3$.

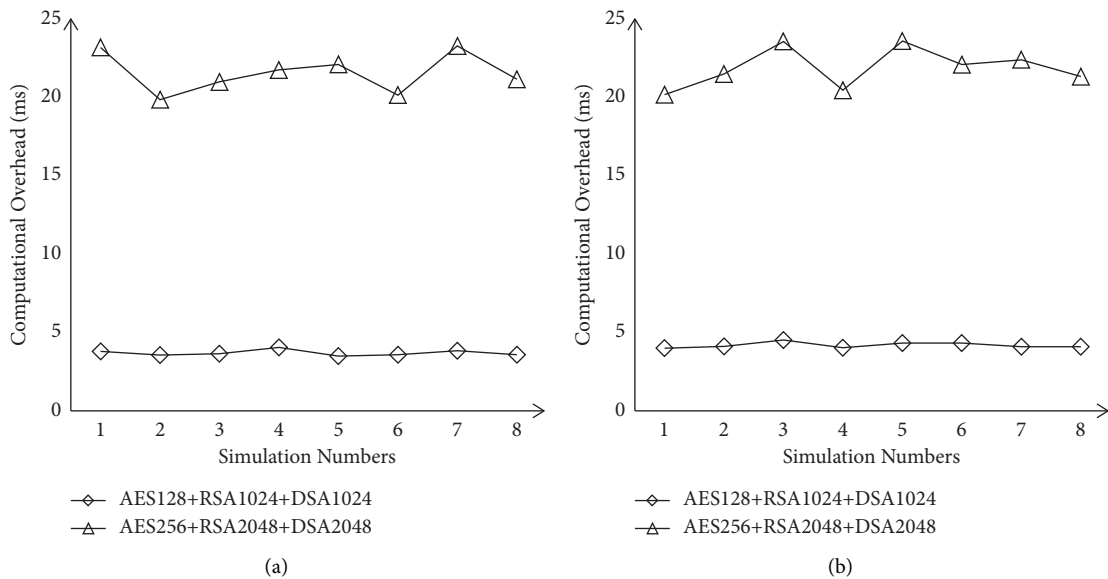


FIGURE 6: Location extraction overhead, (a) $n = \infty$, (b) $n = 3$.

Case of $n = \infty$. When using AES128, RSA1024, and DSA1024, the average overhead is 2.43 ms. While using AES256, RSA2048, and DSA2048, it is 5.25 ms.

Case of $0 \leq n \leq N$. The computational overheads increase to 25.35 ms and 28.07 ms, respectively. Most of the overhead is used to calculate six order-preserving encryption, which is about 22.70 ms. In this case, the computational overhead of our scheme is greater than that of [9] since we use order-preserving encryption with large ciphertext space, and [9] uses small ciphertext space. This is the time cost to improve security.

(4) *Location Extraction*. There are also two cases for location extraction, and the simulation results are shown in Figures 6(a) and 6(b). Unlike location sharing, there is little difference in overheads between the two cases. When using AES128, RSA1024, and DSA1024, the average overheads are 3.67 ms and 4.16 ms, respectively. While using AES256, RSA2048, and DSA2048, they are 21.41 ms and 21.741 ms, respectively.

The above simulation shows that the computational overhead of the location record phase is within 10 ms, and the computational overheads of the location sharing and location extraction phases are both within 30 ms. Therefore, our improved BMPLS scheme is practical.

6. Conclusion

In the telecare medical information systems, patient location information is sensitive data that needs to be protected. Blockchain technology provides a new method for patient location privacy protection and secure sharing. Recently, Ji et al. [9] proposed a blockchain-based multilevel privacy-preserving location sharing (BMPLS) scheme for telecare medical information systems. In this paper, we show that Ji et al.'s BMPLS scheme does not achieve confidentiality and multilevel privacy-preserving. Using salting technology, we propose an improved BMPLS scheme to fix our attacks. We also optimized the BMLS scheme to make it correct and executable. Analysis shows that the improved BMPLS scheme achieves all security while maintaining its high performance. Our improved BMPLS scheme is currently based on the general blockchain, and we may further study the use of edge-of-thing blockchain to enhance system performance [2, 3]. For different scenarios, designing an applicable trust level management scheme is another future research direction.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would also like to thank Ms. Hongying Chen for her help in performance simulation. This work was supported by the National Social Science Fund of China (no.

21XTQ015), the Natural Science Foundation of Fujian Province of China (nos. 2019J01750, 2019J01752, and 2020J01814), the Education and Scientific Research Project for Young and Middle-Aged Teachers of Fujian Province (no. JAT201387), and the Natural Science Foundation of Zhangzhou (no. ZZ2019J27).

References

- [1] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," pp. 1-9, 2009, <https://bitcoin.org/bitcoin.pdf>.
- [2] P. B. N. Deepa, Q.-V. Pham et al., "Toward blockchain for edge-of-things: a new paradigm, opportunities, and future directions," *IEEE Internet of Things Magazine*, vol. 4, no. 2, pp. 102-108, 2021.
- [3] T. R. Gadekallu, Q.-V. Pham, D. C. Nguyen et al., "Blockchain for edge of things: applications, opportunities, and challenges," *IEEE Internet of Things Journal*, p. 1, 2021.
- [4] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and PUF-based lightweight Authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, p. 1, 2021.
- [5] W. Wang, H. Huang, L. Zhang, and C. Su, "Secure and efficient mutual authentication protocol for smart grid under blockchain," *Peer-to-Peer Networking and Applications*, vol. 14, no. 5, pp. 2681-2693, 2021.
- [6] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for IIoT devices," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.
- [7] G. Zyskind, O. Nathan, and A. S. Pentland, "Decentralizing privacy: using blockchain to protect personal data," in *2015 Proceedings of the 2015 IEEE Security and Privacy Workshops*, pp. 180-184, San Jose, CA, USA, May 2015.
- [8] M. Amoretti, G. Brambilla, F. Medioli, and F. Zanichelli, "Blockchain-based proof of location," in *Proceedings of the 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 146-153, Lisbon, Portugal, June 2018.
- [9] Y. Ji, J. Zhang, J. Ma, C. Yang, and X. Yao, "BMPLS: blockchain-based multi-level privacy-preserving location sharing scheme for telecare medical information systems," *Journal of Medical Systems*, vol. 42, no. 8, 2018.
- [10] T.-F. Lee, H.-Z. Li, and Y.-P. Hsieh, "A blockchain-based medical data preservation scheme for telecare medical information systems," *International Journal of Information Security*, vol. 20, no. 4, pp. 589-601, 2021.
- [11] H. Li, L. Zhu, M. Shen, F. Gao, X. Tao, and S. Liu, "Blockchain-based data preservation system for medical data," *Journal of Medical Systems*, vol. 42, no. 8, 2018.
- [12] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control," *Journal of Medical Systems*, vol. 40, no. 10, 218 pages, 2016.
- [13] S. Cao, G. Zhang, P. Liu, X. Zhang, and F. Neri, "Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain," *Information Sciences*, vol. 485, pp. 427-440, 2019.
- [14] J. Wang, K. Han, A. Alexandridis et al., "A blockchain-based eHealthcare system interoperating with WBANs," *Future Generation Computer Systems*, vol. 110, pp. 675-685, 2020.
- [15] H. Huang, P. Zhu, F. Xiao, X. Sun, and Q. Huang, "A blockchain-based scheme for privacy-preserving and secure sharing of medical data," *Computers & Security*, vol. 99, Article ID 102010, 2020.

- [16] Y. Zhuang, L. R. Sheets, Y.-W. Chen, Z.-Y. Shae, J. J. P. Tsai, and C.-R. Shyu, "A patient-centric health information exchange framework using blockchain technology," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2169–2176, 2020.
- [17] S. Shamshad, K. Minahil, K. Mahmood, S. Kumari, and C.-M. Chen, "A secure blockchain-based e-health records storage and sharing scheme," *Journal of Information Security and Applications*, vol. 55, Article ID 102590, 2020.
- [18] H. Huang, X. Sun, F. Xiao, P. Zhu, and W. Wang, "Blockchain-based eHealth system for auditable EHRs manipulation in cloud environments," *Journal of Parallel and Distributed Computing*, vol. 148, pp. 46–57, 2021.
- [19] H. Zhu, Y. Guo, and L. Zhang, "An improved convolution Merkle tree-based blockchain electronic medical record secure storage scheme," *Journal of Information Security and Applications*, vol. 61, Article ID 102952, 2021.
- [20] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "Blockchain leveraged decentralized IoT eHealth framework," *Internet of Things*, vol. 9, Article ID 100159, 2020.
- [21] S. Tanwar, K. Parekh, and R. Evans, "Blockchain-based electronic healthcare record system for healthcare 4.0 applications," *Journal of Information Security and Applications*, vol. 50, Article ID 102407, 2020.
- [22] K. Miyachi and T. K. Mackey, "hOCBS: a privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design," *Information Processing & Management*, vol. 58, no. 3, Article ID 102535, 2021.
- [23] Z. Chen, W. Xu, B. Wang, and H. Yu, "A blockchain-based preserving and sharing system for medical data privacy," *Future Generation Computer Systems*, vol. 124, pp. 338–350, 2021.
- [24] K. Mohammad Hossein, M. E. Esmaili, T. Dargahi, A. Khonsari, and M. Conti, "BCHealth: a novel blockchain-based privacy-preserving architecture for IoT healthcare applications," *Computer Communications*, vol. 180, pp. 31–47, 2021.
- [25] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, "Order-preserving symmetric encryption," in *Advances in Cryptology - EUROCRYPT 2009*, A. Joux, Ed., vol. 5479, pp. 224–241, Springer, Cham, 2009.
- [26] R. C. Merkle, "A certified digital signature," in *Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, pp. 218–238, Santa Barbara, CA, USA, July 1989.
- [27] E. Androulaki, A. Barger, C. Cachin et al., "Hyperledger fabric," *Proceedings of the Thirteenth EuroSys Conference*, vol. 30, pp. 1–15, 2018.
- [28] Z. Yang, K. Yang, L. Lei, K. Zheng, and V. C. M. Leung, "Blockchain-based decentralized trust management in vehicular networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1495–1505, 2019.

Research Article

Differential Evolution and Multiclass Support Vector Machine for Alzheimer's Classification

Jhansi Rani Kaka ¹ and K. Satya Prasad²

¹Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Kakinada, India

²Department of Electronics and Communication Engineering,
Vignana's Foundation for Science, Technology and Research (VFSTR), Guntur, India

Correspondence should be addressed to Jhansi Rani Kaka; jhsankaka@outlook.com

Received 30 October 2021; Revised 9 December 2021; Accepted 15 December 2021; Published 13 January 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Jhansi Rani Kaka and K. Satya Prasad. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Early diagnosis of Alzheimer's helps a doctor to decide the treatment for the patient based on the stages. The existing methods involve applying the deep learning methods for Alzheimer's classification and have the limitations of overfitting problems. Some researchers were involved in applying the feature selection based on the optimization method, having limitations of easily trapping into local optima and poor convergence. In this research, Differential Evolution-Multiclass Support Vector Machine (DE-MSVM) is proposed to increase the performance of Alzheimer's classification. The image normalization method is applied to enhance the quality of the image and represent the features effectively. The AlexNet model is applied to the normalized images to extract the features and also applied for feature selection. The Differential Evolution method applies Pareto Optimal Front for nondominated feature selection. This helps to select the feature that represents the characteristics of the input images. The selected features are applied in the MSVM method to represent in high dimension and classify Alzheimer's. The DE-MSVM method has accuracy of 98.13% in the axial slice, and the existing whale optimization with MSVM has 95.23% accuracy.

1. Introduction

Alzheimer's disease (AD) is a cognitive degenerative disorder leading to dementia and is considered a mental and physical disability. AD has stages like very mild, mild, and moderate dementia class, and patients' treatment is dependent on the stage of AD [1]. AD is one of the fastest-growing and challenging diseases to treat and affects the livelihood of not just patients but also close family members, nurses, caregivers of the patients. Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans are common imaging techniques to analyze Alzheimer's [2]. Mild Cognitive Impairment (MCI) is AD at its transition state, and this is necessary to classify the stages for therapeutic measures to delay the disease progression [3]. Clinical neuroimaging techniques such as MRI and PET scans are suitable for analyzing brain changes with AD progression and MCI [4]. The structural MRI scans provide

detailed information of the brain anatomical structures that can detect and measure AD of brain atrophy patterns [5].

Machine learning methods help in analyzing high-dimensional data and automated classification to learn the complex structural changes of complex patterns in different imaging modalities. Feature extraction helps in training the features for classification algorithms to build predictive models that are useful for clinical processes [6]. Various methods have been applied to develop the early classification of AD on an individual basis like deep learning networks and machine learning methods [7]. Predefined features like voxel and regional measures were obtained from image preprocessing pipelines for the combination of various algorithms with classifiers such as random forests or Support Vector Machines (SVM) [8]. Various existing methods have been applied for Alzheimer's classification and have the limitations of overfitting and imbalance data problems [9, 10]. Existing methods have the limitation of being easily

trapped into local optima that select some of the irrelevant features. Solving the local optima problem in the feature selection further improves the relevant feature selection to improve learning of feature difference and improves the classification performance. The proposed DE-MSVM method applies the threshold value to escape from local optima, and a set of samples is selected to fine-tune the model for classwise training. The contribution of the proposed DE-MSVM method is discussed in the following:

- (1) Multiobjective optimization method is proposed to select the features based on the data instances and classes of the training data. This helps to fine-tune the model to learn the difference between the features and improves the classification performance.
- (2) The feature learning and feature selection method provides higher performance in three-slice classification. The proposed method provides higher performance in MRI and PET images.
- (3) The proposed DE-MSVM method has accuracy of 98.3%, and the existing RFE-GA-SVM method has 95.79% accuracy.

Alzheimer's disease diagnosis is an integral part of the clinical assessment and is usually carried out in MRI and PET images. Various models have been proposed for Alzheimer's disease classification to improve its reliability. Some of the notable researches in Alzheimer's classification were surveyed in this section for a better understanding of existing methods.

Basaia et al. [11] applied Convolutional Neural Network (CNN) for the classification of Alzheimer's into 3 classes on MRI images. The ADNI dataset was used to test the performance of the proposed CNN model in Alzheimer's classification. The CNN has higher accuracy in Alzheimer's classification without feature engineering. Ramzan et al. [12] applied the deep learning method of ResNet-18 architecture to improve the efficiency of Alzheimer's classification. The model was fully trained from scratch and performed transfer learning, and an extended network architecture was applied to fine-tune the model. The ADNI benchmark dataset was used to test the model performance in Alzheimer's classification. The ResNet-18 model has a higher performance than existing methods in Alzheimer's classification. Naz et al. [13] applied freeze features from ImageNet for binary and ternary classification for Alzheimer's classification. Various CNN models such as VGG, InceptionResNet, Inception v3, DenseNet, ResNet, GoogLeNet, and AlexNet were applied with freeze features to test the performance. The ADNI dataset was applied to test the performance of the proposed method in Alzheimer's classification. The result shows that the VGG model with freeze features has a higher performance in Alzheimer's classification. Janghel and Rathore [14] applied VGG-16 architecture of the CNN model with conversion and resizing of images for feature extraction. The SVM, k means clustering, and decision tree were applied for the classification of Alzheimer's. The developed method was tested on MRI and PET images of the ADNI dataset for Alzheimer's

classification. The developed method has higher performance on MRI and PET images compared to the existing CNN model. González et al. [15] applied the framework of preprocessing, feature extraction, and classification for Alzheimer's classification in three datasets. The preprocessing involves unified segmentation of the input images. The Neuromorphometrics, Hammers, and atlas-based features were extracted for the feature extraction. The linear SVM, logistic regression, and random forest were used for Alzheimer's classification based on extracted features. The result shows that the proposed method has higher performance in Alzheimer's classification existing methods.

AbdulAzeem et al. [16] applied an end-to-end framework of the CNN model for Alzheimer's classification in MRI images. The ADNI dataset was used in this study to test the developed model in binary and multiclass classification. The Glorot Uniform Weight Initializer was applied to enable the weight in the activation function to prevent the network from starting from the saturated region. The Adam optimizer was applied to fine-tune the model to improve the performance of Alzheimer's classification. The developed model has a higher performance in Alzheimer's classification compared to existing methods. Eitel and Ritter [17] applied four attributes such as occlusion, layerwise relevance propagation, guided backpropagation, and gradient input with the CNN model for Alzheimer's classification. The developed method was applied in the MRI ADNI dataset to test the performance of Alzheimer's classification. Buvaeswari and Gayathri [18] applied a deep learning-based SegNet method for segmentation, and ResNet-101 was applied for Alzheimer's classification. The seven morphological characteristics were extracted for feature extraction. The developed ResNet-based model has higher performance in Alzheimer's classification compared to the existing method. Alam et al. [19] applied the Twin SVM model with Dual-Tree Complex Wavelet Transform (DTCWT) for Alzheimer's classification. The developed method has higher efficiency on the ADNI and OASIS datasets for Alzheimer's classification. Deep learning-based models in Alzheimer's classification suffer from the limitations of the overfitting problem and generate more irrelevant features for the classification. Some of the researchers were involved in applying the feature selection and SVM-based model for Alzheimer's classification to overcome the overfitting problem.

Asim et al. [20] applied Principal Component Analysis (PCA) for the feature reduction and SVM method for the classification of Alzheimer's classification. The developed method evaluated on the ADNI dataset shows that the developed method has higher efficiency in Alzheimer's classification than the existing method. Shakarami et al. [21] proposed the AlexNet-SVM model for Alzheimer's classification to improve the accuracy of the classification. The developed method was tested on the PET images from the ADNI dataset. The developed model has higher performance than the existing methods in Alzheimer's classification. Zeng et al. [22] applied the switching delay PSO method to optimize the SVM parameters for Alzheimer's classification. The PCA method was applied for the feature reduction, and

SVM was applied for the classification. The developed method was tested on the ADNI dataset and showed higher performance in Alzheimer's classification. Neffati et al. [23] applied downsized kernel Principal Component Analysis and Multiclass SVM model for Alzheimer's classifier. The developed method was tested on the OASIS MRI dataset and showed higher performance. Divya et al. [24] applied Recursive Feature Elimination (RFE) and Genetic Algorithm (GA) for the feature selection and SVM for the classification of Alzheimer's classification. The developed method has higher performance on the ADNI dataset than the existing methods in Alzheimer's classification. Nanni et al. [25] applied various texture descriptions and SVM for the classification of Alzheimer's classification. The MRI ADNI dataset was used to test the performance of Alzheimer's classification. The texture features and SVM method improve the performance of Alzheimer's classification than the existing methods. The SVM-based models suffer from imbalance data problem that affects the performance of classification.

Reddy et al. [26] applied an adaptive Genetic Algorithm with fuzzy logic for the prediction of heart disease at early stage. The developed model consists of a rough set-based heart disease feature selection method and a fuzzy rule-based classification model. Rough set theory selects the important features for heart disease classification, and the selected features were applied for the heart disease classification. Gadekallu et al. [27] applied Principal Component Analysis (PCA), Grey Wolf Optimization (GWO), and Deep Neural Network (DNN) for the classification of diabetic retinopathy. The standard scaler normalization method is applied to normalize the input dataset. The PCA method performs feature reduction in the input dataset, and GWO selects the optimal parameter for the DNN model.

Summarization of the related papers in Alzheimer's classification in taxonomywise is given in Table 1.

This paper is formulated as follows: an explanation about the DE-MSVM method, normalization, and AlexNet feature extraction is given in Section 2. The simulation setup of the proposed method is given in Section 3, and the results of the proposed model in Alzheimer's classification are given in Section 4. The conclusion of this proposed research is presented in Section 5.

2. Proposed Method

In this research, the DE-MSVM method is proposed to increase the performance of Alzheimer's classification. The ADNI datasets were used to test the performance of the proposed DE-MSVM and existing methods for Alzheimer's classification. The normalization is applied to enhance the quality of the images and applied for feature extraction. The AlexNet model is applied to extract the deep features from the input images. The Differential Evolution method is applied to select the relevant features from the extracted features. The MSVM model is applied for the classification based on the selected features. The block diagram of the proposed DE-MSVM model is shown in Figure 1.

2.1. Normalization. The original data differences affect the classification performance, and usually, image data have different intensities. The min-max normalization technique is applied to enhance the image intensity to provide clear information and improve classifier performance. The min-max intensity normalization is shown in the following equation:

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

where x_{\min} denotes the image minimum intensity, x_{\max} denotes the image maximum intensity, x_N denotes the normalized image, and x denotes the input image.

2.2. Laplacian Redecomposition for Multimodal Medical Image Fusion. The LRD scheme is used to decompose the source images to obtain LSI images and nonoverlapping and overlapping domain images. The LEM fusion rule [28–30] is used to fuse LSI, and two fusion rules such as OD and NOD are used to fuse overlapping and nonoverlapping domain images, respectively. The IRS fusion rule is applied to reconstruct HIS fusion image, and inverse Laplacian transform is applied to reconstruct the fusion image.

2.3. LEM Fusion Rule. Texture details are information of interest in anatomical images, and low-frequency information is present in some functional images. The LEM is defined as the square of the sum of local window pixels [28]. The square operation leads to unstable energy acquisition due to differences in functional and anatomical images. The square operation in normalized pixels' range is smaller if the sum of pixels is less than 1 and larger if the sum of pixels is greater than 1. This affects the accurate acquisition of decision results. So, the direct addition operation is applied instead of the square operation in the following equation:

$$Q_{\mu}(i, j) = \lambda \sum_{x=-1}^1 \sum_{y=-1}^1 G_{\mu}^T(i+x, j+y), \quad (2)$$

where filtering template is denoted as $\lambda = [111; 111; 111]$ and local window sizes are represented as x and y . Then, calculate the maximum value of Q_{μ} as E_{μ} and this defines LEM, as given in the following equation:

$$E_{\mu}(i, j) = \max \left(\sum_{x=-1}^3 \sum_{y=-1}^3 Q_{\mu}(i+x, j+y) \right), \quad (3)$$

where m_7 is denoted as binary decision graph and G_F^3 is used to construct an LSI fusion image, as given in equations (4) and (5):

$$m_7 = \begin{cases} 1, & E_A(i, j) > E_B(i, j), \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

$$G_F^3(i, j) = m_7 \times G_A^T + \sim m_7 \times G_B^T(i, j), \quad (5)$$

where reverse operator in range of 0 to 1 is denoted as \sim .

TABLE 1: The proposed DE-MSVM method performance on axial slice images.

| Method | Type | Advantages | Limitations |
|-----------------------------------|------------------------|---|--|
| CNN models [11–14, 16–18, 20, 21] | Deep learning method | Feature extraction is efficient, and feature learning process helps to provide reliable classification. | Overfitting problem in the training and validation. |
| SVM [15, 19, 22–25] | Machine learning model | SVM model has the capacity to handle the high-dimensional dataset. | SVM model has lower efficiency in learning the feature differences, which affects the sensitivity and specificity. |
| Fuzzy logic [26] | Fuzzy model | Fuzzy model updates the rules to improve the classification process. | Fuzzy model is highly sensitive to outliers and has lower efficiency in feature learning. |
| GWO-DNN [27] | Deep learning | GWO method selects the parameter for the DNN to improve the classification performance. | GWO method is easily trapped into local optima, and the DNN model suffers from an overfitting problem. |

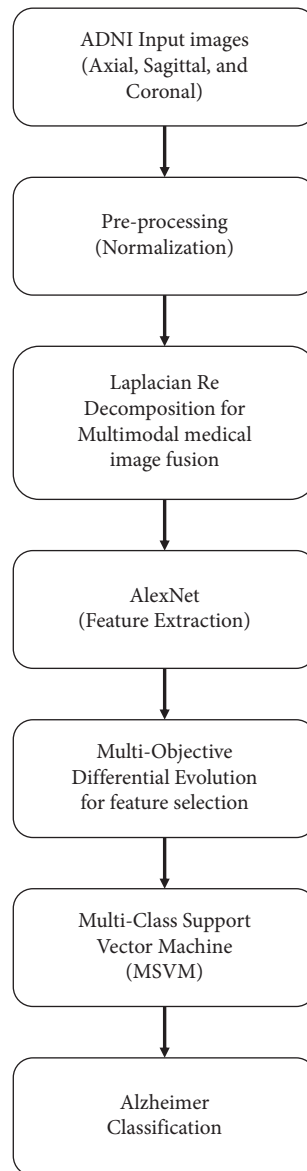


FIGURE 1: The block diagram of the proposed DE-MSVM in Alzheimer's classification.

2.4. OD Fusion Rule. Domain fusion images have more useful information to make overlapping and have three major efforts: (1) a Local Decision Maximums (LDM) based on MLD and LEM is applied to mark edges and anatomical images details; (2) another LDM marker is applied based on LEM and MLD decision scheme for functional images of abnormal areas; (3) binary decision graph is developed based on two LDMs through comparing the sizes, and fusion images of the overlapping domain are obtained. The OD has more advantages of infusion of anatomical feature information and functional images. The algorithm description is given as follows.

Anatomical and functional images of LDM are calculated using LEM and MLD, as given in the following equation:

$$M_A(i, j) = D_A(i, j) + E_A(i, j), \quad (6)$$

where the anatomical image of LDM is denoted as M_A and MLD obtains $D_A(i, j)$, while LEM obtains $E_A(i, j)$.

Equations (7) and (8) measure M_B .

$$m_8 = \begin{cases} 1, & \text{if } D_B(i, j) > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

$$M_B(i, j) = D_B(i, j) + \sim m_8 \times E_B(i, j), \quad (8)$$

where LDM value of the functional image is denoted as M_B and binary decision graph is denoted as m_8 .

The fusion rule is given in equations (9) and (10).

$$m_9 = \begin{cases} 1, & \text{if } M_A(i, j) > M_B(i, j), \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

$$O_F^y(i, j) = m_9 \times O_A^y(i, j) + \sim m_9 \times O_B^y(i, j), \quad (10)$$

where binary decision graph is denoted as m_9 and anatomical and functional overlapping domain images are denoted as O_A^y and O_B^y . The overlapping domain fusion image is denoted as O_F^y .

2.5. NOD Fusion Rule. According to DGR nonoverlapping domain definition, the nonoverlapping domain of fusion image is obtained using the following fusion algorithm, as given in the following equation:

$$N_F^y(i, j) = N_A^y(i, j) + N_B^y(i, j), \quad (11)$$

where N_A and N_B denote anatomical and functional non-overlapping domain images. The fusion image of the non-overlapping domain is denoted as N_F .

2.6. IRS Fusion Rule. The subband fusion image with high frequency is reconstructed using the IRS fusion rule to eliminate artifacts in images since complementary and redundant information fuse in HIS separately, which leads to image artifacts in subband fusion image reconstruction with high frequency. Overlapping domain surrounding pixels are replaced, and two global decision graphs are constructed to solve this problem. The reconstruction task is completed by

the first decision graph, and the local mean algorithm is combined with the second decision graph to eliminate artifacts.

2.7. Reconstructed Fused Image. Fused medical images are obtained by reconstruction of Laplacian multiscale. The traditional inverse Laplacian transform is applied to reconstruct fused images, and the decomposition process of inverse operation obtains the fusion image. The fusion image F is measured in the following equation:

$$F(i, j) = \sum_{\gamma=1}^{\tau-1} L_F^\gamma(i, j) + G_F^\tau(i, j). \quad (12)$$

2.8. AlexNet Feature Extraction. The AlexNet is a deep learning technique applied for feature extraction. A fully connected layer of AlexNet CNN is applied for the feature extraction from the fused image. The AlexNet CNN consists of 22 layers of feature extractor based on transfer learning technique, plus a fully connected (FC) layer with $1 \times 1 \times 64$ dimensions [21]. Extracted features were applied to the Differential Evolution method for feature selection.

2.9. Multiobjective Differential Evolution for Feature Selection. Differential Evolution (DE) has a population of solutions $x_i = \{x_{i,j}\}$ for $i = 1, \dots, N_p$ and $j = 1, \dots, n$, where decision variables are represented by $x_{i,j} \in IR$, the number of vectors is denoted as N_p , and the number of vector elements is denoted as n . The MOOP concept is explained before the DEMO algorithm discussion in detail [31, 32].

The MOOP has a number of objective functions that are either minimized or maximized. Several constraints are needed to be satisfied during optimization, and MOOP is given in equations (13) and (14).

$$\frac{\min}{\max} f_m(x_i), \quad (13)$$

where $m = 1, \dots, M \wedge$ and $i = 1, \dots, N_p$.

Subject to $g_k(x_i) \geq 0$,

$$h_l(x_i) = 0, \quad (14)$$

where $k = 1, \dots, K$, $l = 1, \dots, L$, $x_j^{\text{Lower}} \leq x_{i,j} \leq x_j^{\text{Upper}}$, and $j = 1, \dots, n$.

The number of objective functions is denoted as M , the number of inequality constraints g_k is denoted as K , the number of equality constraints h_l is denoted as L , and lower and upper boundaries of search space are denoted as $x^{\text{Lower}} = \{x_j^{\text{Lower}}, j = 1, \dots, n\}$ and $x^{\text{Upper}} = \{x_j^{\text{Upper}}, j = 1, \dots, n\}$.

Decision space D is divided by constraints into infeasible and feasible regions, where objective space is represented by feasible region $S \subseteq D$. The objective functions values are determined in the multidimensional space of objective space. This finds the point in the objective space $f(x_i) = (f_1(x_i), \dots, f_m(x_i))$ for each feasible solution $x_i \in S$.

The objective space of solution quality is measured based on its dominance. x_i dominates x_j solution when x_i is better than x_j in all objectives and strictly, x_i should be better than x_j at least in one objective.

Pareto Optimal Front (POF) is the dominant solution applied in this method. The POF is determined by MOOP in the decision space. Methods for nondominant selection are used for this task, and after mutation and crossover to original population size N_p , a grown population is truncated.

The MOOP of feature selection uses two objective functions $f(x) = (f_1(x), f_2(x))$. The function f_1 is given in

$$\min f_1(x) = \sum_{j=1}^n I(x_j), \quad (15)$$

where a maximum number of features in vector x is denoted as n and the function $I(x_j)$ for $j = 1, \dots, n$ is given in

$$I(x_j) = \begin{cases} 1, & \text{if } x_j \geq 0.5, \\ 0, & \text{if } x_j < 0.5. \end{cases} \quad (16)$$

The number of used features is defined as the objective function f_1 and the f_2 function is sophisticated. The classes are defined as $E = \{e_1, \dots, e_{NE}\}$, where a maximum number of labels are defined as NE. A set of valid samples E_k^{valid} is belonging (\mapsto) to k^{th} , as given in the following equation:

$$E_k^{\text{valid}} = \{x = \{x_j\} \vee \forall j: I(x_j) = 1 \wedge x \mapsto e_k\}. \quad (17)$$

A set of samples E_k^{SVM} in k^{th} class and classification SVM(x) is defined as in the following equation:

$$E_k^{\text{SVM}} = \{x = \{x_j\} \vee \exists j: I(x_j) = 1 \wedge \text{SVM}(x) \mapsto e_k\}, \quad (18)$$

where classification of vector x to class E is denoted using a symbol \mapsto . The ratio between the size of both introduced sets is function f_2 , as given in the following equation:

$$\max f_2(x) = \sum_{k=1}^{\text{NE}} \frac{E_k^{\text{SVM}}}{E_k^{\text{valid}}}, \quad (19)$$

where equation (19) helps to fine-tune the model to achieve higher accuracy for SVM in the training and validation set. Equation (15) helps the model to find a similar instance to fine-tune the model to reduce the error rate of the model. Furthermore, equation (15) helps to fine-tune the model based on the number of instances in the dataset, and equation (19) helps to fine-tune the model related to the labels. Equations (15) and (19) help the model to learn the difference in the features, which makes it easy for the hyperplane of SVM to classify the data.

The classification accuracy of function f_2 is based on a selected feature subset.

DEMO algorithm is one of the successful DE realizations for solving MOOPs. In this study, the DEMO method is applied to improve the strength of the Pareto evolutionary algorithm-SPEA2 for nondominant selection.

3. Multiclass Support Vector Machine

Binary classifiers f_1, f_2, \dots, f_N are constructed for $1 \dots N$ classes, each trained to be different from one class to the other [33–37]. A multiclass category is obtained based on the maximal output before applying the SGN function: $\text{argmax } g^k(x)$, where $g^k(x) = \sum_{i=1}^n y_i \alpha_i^k k(x, x_i) + b^k$, in which $k = 1, \dots, N$.

Hyperplane distance to the point x of a signed real value is denoted as $g^k(x)$ which is referred as the confidence value. The higher value increases the confidence, where x belongs to the positive class. The highest confidence value is assigned with x .

The input data is denoted as $X = \{x_1, x_2, \dots, x_m\} \in R^d$, the hypersphere radius is denoted as r , and the center is denoted as $c \in R^d$. The minimum hypersphere which encloses the optimization problem is given in equation (20).

Minimize r^2 , subject to $\|\Phi(x_j) - c\|^2 \leq r^2$, $j = 1, \dots, m$

$$L(c, r, \alpha) = r^2 + \sum_{j=1}^m \alpha_j \{ \|\Phi(x_j) - c\|^2 - r^2 \}. \quad (20)$$

Derive. $\partial L(c, r, \alpha) / \partial c = 2 \sum_{j=1}^m \alpha_j (\Phi(x_j) - c) = 0$.

The following equation is obtained:

$$\begin{aligned} \sum_{j=1}^m \alpha_j &= 1, \\ \sum_{j=1}^m \alpha_j \Phi(x_j) &= 0. \end{aligned} \quad (21)$$

Hence, equation (10) becomes

$$L(c, \gamma, \alpha) = \sum_{j=1}^m \alpha_j k(x_j, x_j) - \sum_{i,j=1}^m \alpha_i \alpha_j k(x_i, x_j). \quad (22)$$

The optimization problem is solved based on the dual form of α , as given in the following:

Maximize

$$W(\alpha) = \sum_{i=1}^m \alpha_i k(x_i, x_i) - \sum_{i,j=1}^m \alpha_i \alpha_j k(x_i, x_j), \quad (23)$$

subject to $\sum_{i=1}^m \alpha_i = 1$ and $\alpha_i \geq 0$, $i = 1$ to m .

Lagrange multiplier's possibilities are nonzero if the inequality constraints are equality solution.

Optimal solution complementarity conditions for α , (c, γ) are given in

$$\alpha_i \{ \|\Phi(x_i) - c\|^2 - r^2 \}, \quad i = 1, \dots, m. \quad (24)$$

Training samples x_i lie on the surface of the optimal hypersphere related to $\alpha_i > 0$.

The following equation provides the decision function solution:

$$f(x) = \text{sgn}(r^2 - \|\Phi(x) - c\|^2). \quad (25)$$

Equations (26) and (27) are provided:

$$f(x) = \text{sgn} \left(r^2 - \left\{ \Phi(x) \cdot \Phi(x) - 2 \sum_{i=1}^m \alpha_i \Phi(x) \cdot \Phi(x_i) + \sum_{i,j=1}^m \alpha_i \alpha_j (\Phi(x_i) \cdot \Phi(x_j)) \right\} \right), \quad (26)$$

$$f(x) = \text{sgn} \left(r^2 - \left\{ k(x, x) - 2 \sum_{i=1}^m \alpha_i k(x, x_i) + \sum_{i,j=1}^m \alpha_i \alpha_j k(x_i, x_j) \right\} \right). \quad (27)$$

The method aims to obtain the minimum enclosing hypersphere consisting of satisfying all training samples.

4. Simulation Setup

The proposed Differential Evolution-Multiclass Support Vector Machine (DE-MSVM) model is implemented on the ADNI dataset and compared with existing methods. This section provides the implementation details of the proposed DE-MSVM model and dataset.

Dataset: ADNI fMRI and PET datasets were used to evaluate the performance of the proposed DE-MSVM method [38, 39]. The ADNI fMRI dataset consists of 3692 images, which contains 1775 normal images and 1917 Alzheimer's disease images. The ADNI PET dataset consists of 1775 normal images and 900 diseased images. The images of axial, coronal, and sagittal planes are present in the dataset. The sample images of the MRI and PET dataset for three slices are shown in Figure 2.

System requirement: the proposed DE-MCSVM method is implemented in the system consisting of an Intel i7 processor, 16 GB RAM, 6 GB graphics card, and Windows 10 64-bit OS. The MATLAB R2018b tool was used to implement and measure the performance metrics for classification.

Metrics: the performance metrics include Accuracy, Sensitivity, Specificity, False Omission Rate (FOR), False Discovery Rate (FDR), and MCC. The formula for metrics is given as follows:

$$\begin{aligned} \text{accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100, \\ \text{sensitivity} &= \frac{TP}{TP + FN} \times 100, \\ \text{specificity} &= \frac{TN}{TN + FP} \times 100, \\ \text{FDR} &= \frac{FP}{FP + TP} \times 100, \\ \text{FOR} &= \frac{FN}{FN + TN} \times 100, \end{aligned} \quad (28)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

5. Experimental Results

In this study, the DE-MSVM model is proposed to increase the performance of Alzheimer's classification. The ADNI fMRI and PET images were used to test the performance of Alzheimer's classification. The normalization method is applied to enhance the quality of the images. AlexNet feature extraction method and DE feature selection are applied to select the relevant features for the classification. The MSVM model is applied with selected features and classifies Alzheimer's images. This section provides detailed information on the results of the proposed DE-MSVM method.

Extracted features size is 4096, and the proposed feature selection method selected 2078 features for the classification. Equation (16) discards most of the features based on the threshold of more than 0.5-feature important score.

The proposed DE-MSVM method is applied on the ADNI axial slice for Alzheimer's classification and compared with existing methods, as shown in Table 2. The proposed DE-MSVM model has higher performance compared to the existing method in Alzheimer's classification. Pareto Optimal Front in differential entropy feature selection selects the relevant features to represent the characteristics of the input in a nondominated manner. The differential entropy feature selection method provides clear separation of feature characteristics based on multiobjective optimization. The MSVM model performs well in the classification in case of clear separation of margin and is more efficient in high-dimensional space. The AdaBoost classifiers are sensitive to the outlier in the feature, and the autoencoder classifier has a limitation of lower efficiency in many features. The proposed DE-MSVM model has accuracy of 98.13%, and AdaBoost has 96.67% accuracy.

The proposed DE-MSVM method is tested on the axial slice of the ADNI dataset and compared with existing methods, as shown in Figure 3. This shows that the DE-MSVM method has higher performance compared to existing feature selection and classifier models. Pareto Optimal Front in differential entropy applies nondominated feature selection to effectively represent the characteristics. The existing feature selection methods such as whale optimization, grey wolf, and bat methods have limitations of being easily trapped into local optima and having poor convergence. The proposed DE-MSVM method has accuracy of 98.31%, and the whale-MSVM method has 95.23% accuracy.

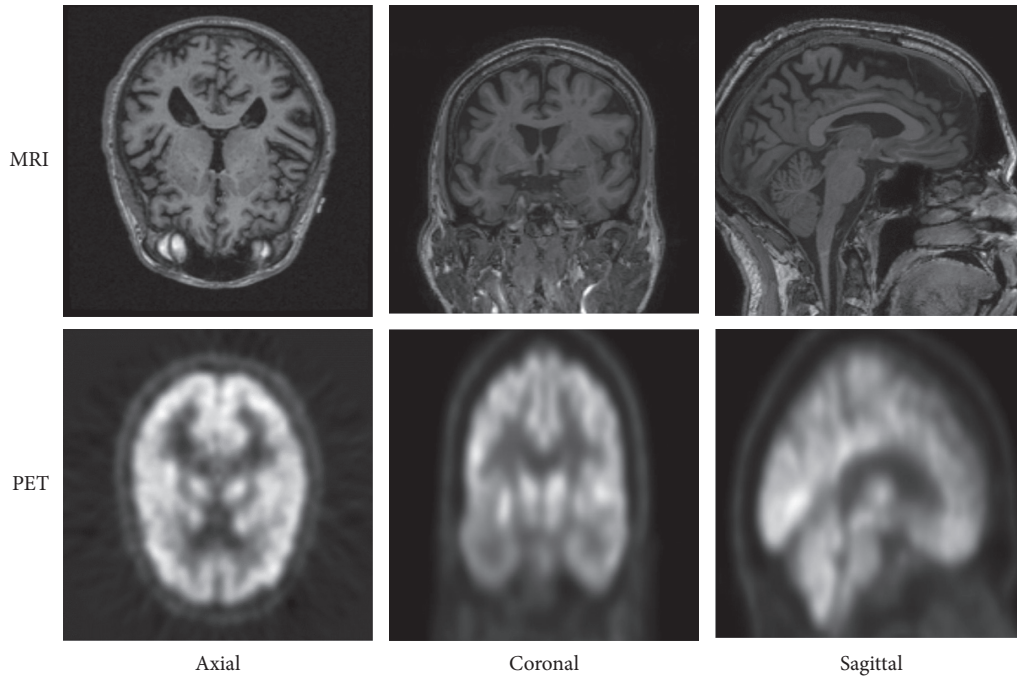


FIGURE 2: Sample images of MRI and PET ADNI dataset on three slices.

TABLE 2: The proposed DE-MSVM method performance on axial slice images.

| | | ADNI dataset (axial slice) | | | | | |
|--|---------------|----------------------------|-----------------|-----------------|---------|---------|---------|
| Feature selection | Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | FOR (%) | FDR (%) | MCC (%) |
| Without feature selection | Autoencoder | 83.4 | 84.69 | 84.53 | 80.51 | 85.54 | 70.36 |
| | AdaBoost | 82.71 | 84.99 | 86.32 | 75.97 | 92.76 | 80.52 |
| | MSVM (linear) | 86.25 | 88.26 | 72.51 | 78.45 | 86.69 | 74.28 |
| | MSVM | 87.7 | 89.51 | 87.36 | 88.93 | 89.28 | 87.17 |
| Bat feature selection algorithm | Autoencoder | 85.35 | 85.35 | 84.56 | 80.37 | 86.41 | 72.06 |
| | AdaBoost | 83.15 | 81.3 | 82.82 | 76.04 | 83.43 | 80.94 |
| | MSVM (linear) | 84.31 | 83.07 | 72.11 | 79.42 | 82.86 | 74.63 |
| | MSVM | 89.72 | 90.98 | 89.8 | 90.28 | 89.9 | 91.63 |
| Grey wolf algorithm | Autoencoder | 88.53 | 87.44 | 88.09 | 82.23 | 88.86 | 74.46 |
| | AdaBoost | 86.33 | 87.54 | 88.66 | 78.58 | 86.85 | 81.37 |
| | MSVM (linear) | 85.98 | 85.33 | 75.05 | 82.26 | 89.76 | 75.99 |
| | MSVM | 90 | 91.85 | 90.32 | 91.13 | 90.59 | 91.21 |
| Whale optimization algorithm | Autoencoder | 87.89 | 89.75 | 87.43 | 84.23 | 90.34 | 76.3 |
| | AdaBoost | 87.48 | 86.8 | 87.47 | 81.82 | 91.98 | 86.78 |
| | MSVM (linear) | 89.55 | 93.26 | 75.23 | 82.25 | 89.17 | 80.58 |
| | MSVM | 95.23 | 92.76 | 94.26 | 95.48 | 94.6 | 93.02 |
| Multiobjective differential evolutionary algorithm | Autoencoder | 90.98 | 85 | 89.86 | 89.43 | 90.01 | 85.61 |
| | AdaBoost | 96.67 | 95.96 | 95.07 | 92.82 | 94.09 | 91.87 |
| | MSVM (linear) | 97.33 | 96.92 | 97.41 | 97.65 | 96.76 | 92.35 |
| | MSVM | 98.13 | 98.96 | 98.2 | 98.36 | 97.03 | 96.06 |

The proposed DE-MSVM method is tested on the sagittal slice images and compared with existing methods, as shown in Table 3. The proposed DE-MSVM method has a higher performance in Alzheimer's classification than existing methods. The Pareto Optimal Front is applied in the DE method to select the features in a nondominated manner. The selected features are applied in the MSVM for the classification, and the MSVM method performs well on high-dimensional data. The autoencoder classifiers have the

limitations of lower performance in many features and imbalanced class problems. The AdaBoost classifier is sensitive to an outlier in the features and has lower performance. The linear MSVM model does not consider the nonlinear relationship between the features and target. The proposed DE-MSVM method has 98.65% accuracy, and the autoencoder has 97.89% accuracy.

The proposed DE-MSVM method is tested on the sagittal slice and compared with the existing feature selection

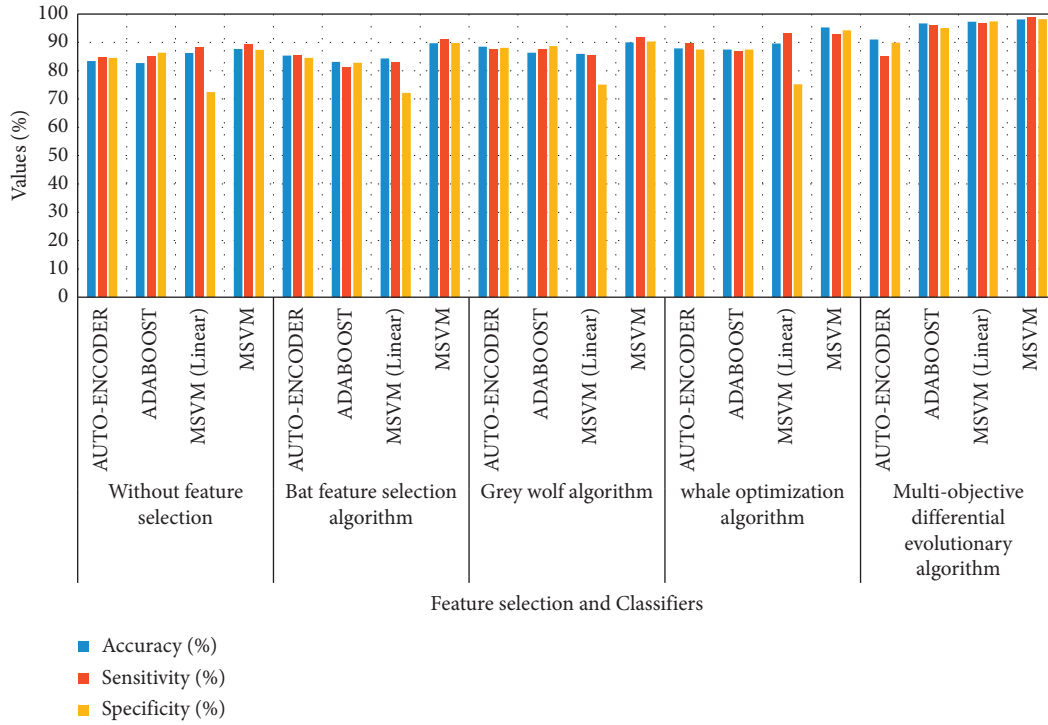


FIGURE 3: Performance analysis on the axial slice of ADNI dataset.

TABLE 3: The proposed DE-MSVM method performance analysis on the sagittal slice.

| ADNI dataset (sagittal slice) | | | | | | | |
|--|---------------|--------------|-----------------|-----------------|---------|---------|---------|
| Feature selection | Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | FOR (%) | FDR (%) | MCC (%) |
| Without feature selection | Autoencoder | 82.7 | 84.65 | 82.31 | 79.87 | 86.05 | 79.32 |
| | AdaBoost | 82.6 | 84.96 | 85.8 | 75.33 | 92.19 | 79.55 |
| | MSVM (linear) | 85.47 | 87.38 | 83.9 | 80.89 | 85.23 | 79.74 |
| | MSVM | 85.73 | 91.34 | 89.91 | 83.14 | 86.05 | 84.04 |
| Bat feature selection algorithm | Autoencoder | 84.16 | 88.94 | 71.49 | 79.4 | 86.03 | 74.09 |
| | AdaBoost | 83.82 | 84.45 | 86.43 | 75.15 | 83.19 | 80.36 |
| | MSVM (linear) | 85.57 | 85.14 | 83.94 | 80.17 | 85.58 | 81.44 |
| | MSVM | 87.31 | 89.19 | 90.93 | 89.29 | 89.45 | 87.87 |
| Grey wolf algorithm | Autoencoder | 85.03 | 90.57 | 84.6 | 81.44 | 88.9 | 85.6 |
| | AdaBoost | 85.41 | 87.03 | 88.23 | 78.21 | 93.24 | 80.8 |
| | MSVM (linear) | 88.52 | 86.72 | 87.36 | 81.86 | 88.74 | 74.46 |
| | MSVM | 88.09 | 91.6 | 89.8 | 90.74 | 90.27 | 89.84 |
| Whale optimization algorithm | Autoencoder | 87.24 | 89.71 | 82.98 | 83.99 | 84.02 | 75.5 |
| | AdaBoost | 87.26 | 86.19 | 82.84 | 84.38 | 86.87 | 86.58 |
| | MSVM (linear) | 89.1 | 90.2 | 84.39 | 86.46 | 88.49 | 80.54 |
| | MSVM | 95.18 | 94.21 | 93.27 | 96.76 | 89.72 | 92.19 |
| Multiobjective differential evolutionary algorithm | Autoencoder | 97.89 | 96.78 | 96.86 | 97.45 | 97.05 | 72.33 |
| | AdaBoost | 95.08 | 95.03 | 94.6 | 94.4 | 94.53 | 91.42 |
| | MSVM (linear) | 90.76 | 89.07 | 90.12 | 90.26 | 89.2 | 85.04 |
| | MSVM | 98.65 | 98.32 | 97.81 | 98.69 | 98.78 | 96.06 |

method, as shown in Figure 4. The proposed DE-MSVM model has higher performance than other existing feature selections. The Pareto Optimal Front helps to select the relevant features from AlexNet feature extraction for classification. The existing feature selection methods such as whale, grey wolf, and bat have the limitations of being easily trapped into local optima and having poor convergence.

The proposed DE-MSVM method is tested on the coronal slice and compared with existing methods, as shown in Table 4. The proposed DE-MSVM method has higher performance than existing methods in Alzheimer’s classification. The classwise learning of the proposed method helps the model to learn the feature difference that improves the sensitivity and specificity of the model. The Pareto Optimal Front is applied in

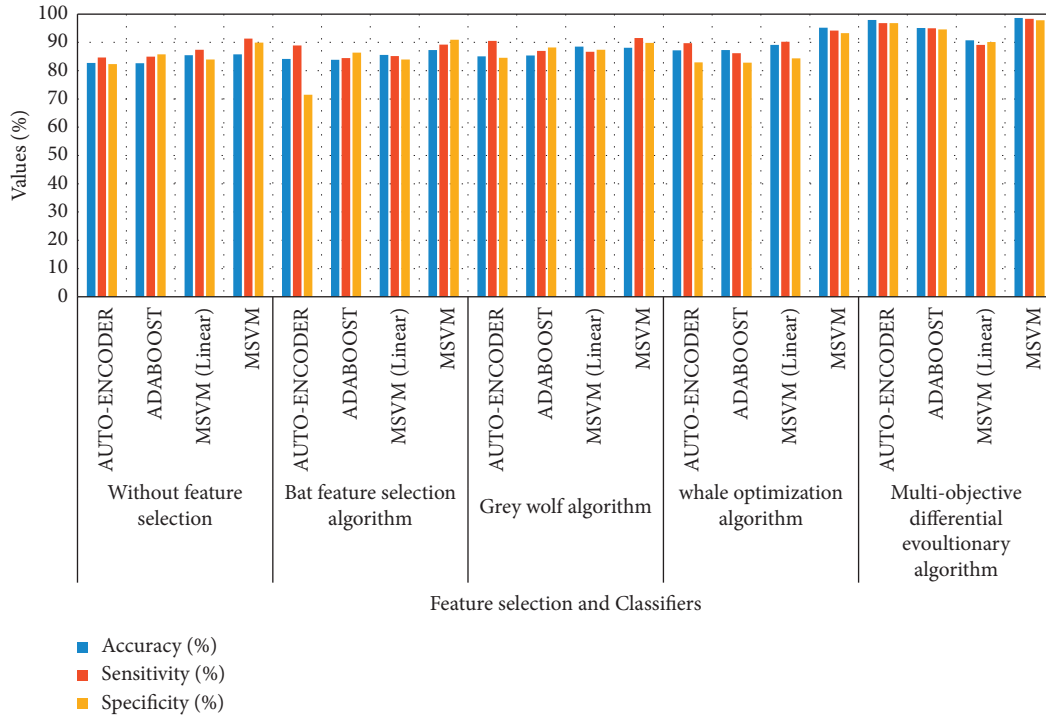


FIGURE 4: The proposed DE-MSVM method on the sagittal slice.

TABLE 4: The proposed DE-MSVM method on the coronal slice.

| | | ADNI dataset (coronal slice) | | | | | |
|--|---------------|------------------------------|-----------------|-----------------|---------|---------|---------|
| Feature selection | Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | FOR (%) | FDR (%) | MCC (%) |
| Without feature selection | Autoencoder | 77.49 | 84.79 | 83.09 | 75.6 | 84.42 | 77.97 |
| | AdaBoost | 75.11 | 82.58 | 82.41 | 73.18 | 84.7 | 73.39 |
| | MSVM (linear) | 79.72 | 81.07 | 65.98 | 73.24 | 82.65 | 76.07 |
| | MSVM | 85.31 | 89.77 | 85.3 | 86.31 | 86.2 | 79.87 |
| Bat feature selection algorithm | Autoencoder | 80.07 | 78.99 | 77.06 | 78.33 | 83.82 | 78.33 |
| | AdaBoost | 77.03 | 83.98 | 84.8 | 70.14 | 81.35 | 78.52 |
| | MSVM (linear) | 79.75 | 83.5 | 79.96 | 76.71 | 77.44 | 70.39 |
| | MSVM | 84.1 | 84.39 | 85.07 | 84.8 | 83.87 | 80.03 |
| Grey wolf algorithm | Autoencoder | 83.29 | 82.43 | 81.25 | 84.96 | 84.9 | 79.41 |
| | AdaBoost | 86.77 | 87.32 | 87.51 | 80.45 | 86.66 | 87.35 |
| | MSVM (linear) | 87.09 | 86.66 | 93.23 | 86.98 | 89.35 | 84.05 |
| | MSVM | 95.07 | 98.28 | 90.39 | 92.28 | 96.08 | 89.33 |
| Whale optimization algorithm | Autoencoder | 85.41 | 87.85 | 84.14 | 82.68 | 83.5 | 80.41 |
| | AdaBoost | 90.31 | 88.03 | 87.76 | 89.06 | 89.82 | 89.25 |
| | MSVM (linear) | 89.62 | 87.99 | 86.46 | 86.59 | 89.01 | 88.1 |
| | MSVM | 96.21 | 96.68 | 95.28 | 94.71 | 94.26 | 95.34 |
| Multiobjective differential evolutionary algorithm | Autoencoder | 91.03 | 90.8 | 91.34 | 90.87 | 90.21 | 76.59 |
| | AdaBoost | 95.9 | 94.08 | 94.9 | 95.49 | 95.64 | 81.26 |
| | MSVM (linear) | 97.53 | 96.06 | 97.65 | 97.34 | 96.79 | 92.7 |
| | MSVM | 98.12 | 97.78 | 98.7 | 98.06 | 97.89 | 95.81 |

Differential Evolution to select features in a nondominated manner. The selected feature is applied in the MSVM method for Alzheimer's classification. The autoencoder method has the limitation of overfitting problem, AdaBoost classifier is sensitive to the outlier of features, and linear MSVM method fails to analyze the nonlinear relationship between the feature and target. The proposed DE-MSVM method has 98.12% accuracy, and the existing DE-AdaBoost has 95.9% accuracy.

The proposed DE-MSVM method is evaluated on the coronal slice and compared with existing methods, as shown in Figure 5. The proposed DE-MSVM method has higher performance than existing methods in Alzheimer's classification. The result shows that other fine-tuned models are less sensitive to the feature difference, and this affects the sensitivity and specificity of the model. The Pareto Optimal Front in Differential Evolution selects the

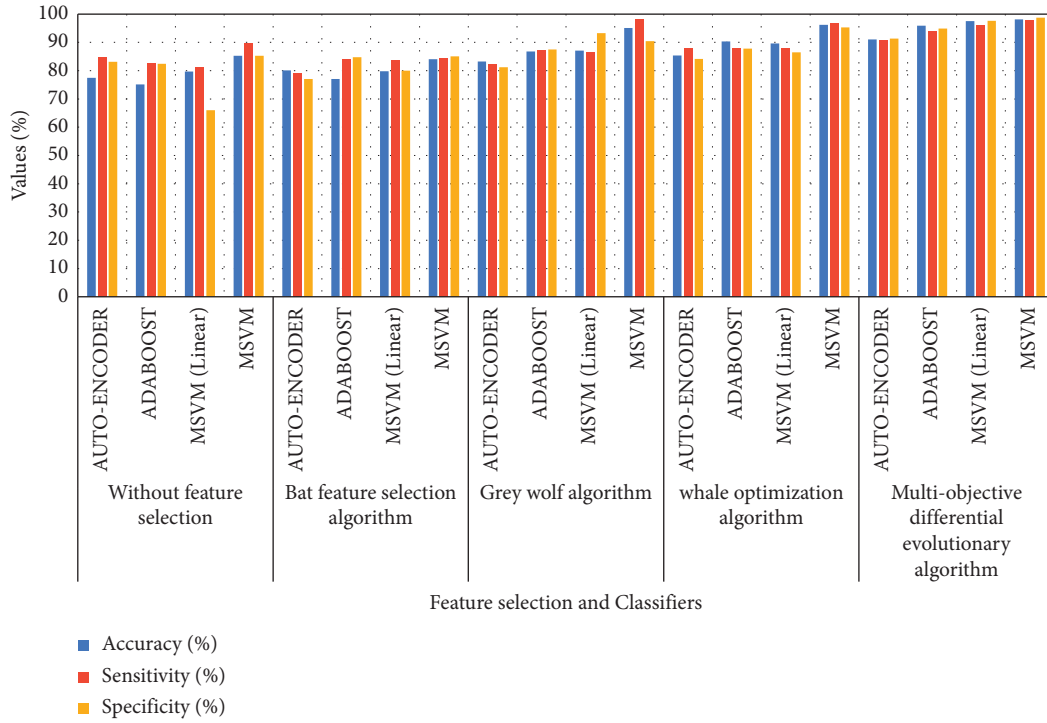


FIGURE 5: The proposed DE-MSVM method on the coronal slice.

TABLE 5: Comparative analysis of ADNI dataset.

| Methods | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|------------------------|--------------|-----------------|-----------------|
| SegNet-ResNet-101 [18] | 96.3 | 96.7 | 93.9 |
| Twin SVM [19] | 90 | 94 | 71 |
| PCA-SVM [20] | 94 | 95 | 93 |
| AlexNet-SVM [21] | 96.39 | 95 | 97.78 |
| RFE-GA-SVM [24] | 95.79 | 89.44 | 98.92 |
| DE-MSVM | 98.3 | 98.35 | 98.23 |

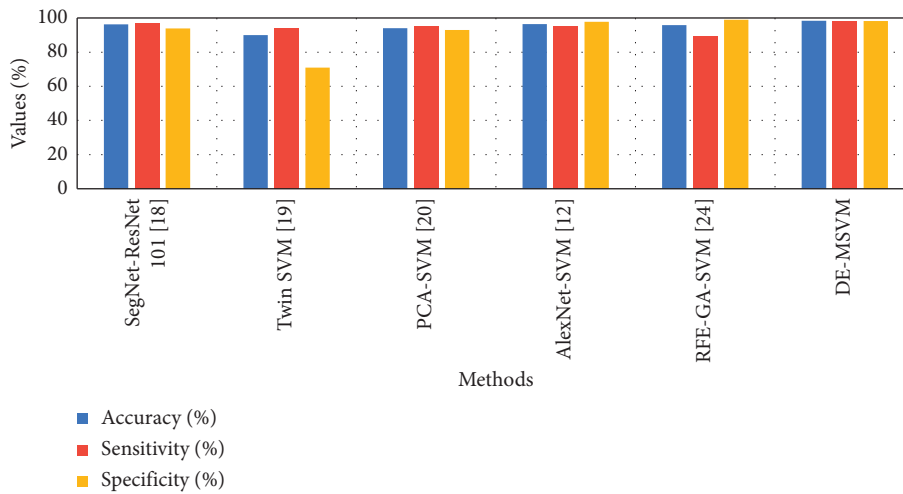


FIGURE 6: Comparative analysis of developed method in ADNI dataset.

features in a nondominated manner and applies them for classification. The MSVM method provides the classification of Alzheimer’s based on the selected features. The

existing whale, grey wolf, and bat methods have limitations of being easily trapped into local optima and having poor convergence.

5.1. Comparative Analysis. The proposed DE-MSVM method is compared with the existing methods in ADNI dataset to analyze the performance.

The proposed DE-MSVM is compared with existing methods in the ADNI dataset, as shown in Table 5 and Figure 6. The proposed DE-MSVM method has higher performance than existing deep learning methods and SVM-based methods. The proposed DE-MSVM method has applied Parent Optima Front to select the features in a non-dominated manner. The proposed DE-MSVM model selects the features based on the data instances and classwise learning of features. This helps to learn the feature difference in the model that improves the classification efficiency. The sensitivity and specificity of the proposed method have achieved 98.35% and 98.23%, respectively. This shows that the classwise learning process in the proposed method improves the efficiency of the model. The deep learning methods like SegNet-ResNet-101 [18] and AlexNet-SVM [21] have limitations of overfitting problems. The RFE-GA-SVM [24] method has the limitation of being easily trapped into local optima and having poor convergence in the feature selection. The proposed DE-MSVM method has accuracy of 98.3%, and AlexNet-SVM [21] has 96.39% accuracy.

6. Conclusion

Alzheimer's is a neurodegenerative disorder, and the early classification of Alzheimer's helps in providing better treatment. The existing models in Alzheimer's classification have the limitations of overfitting problems and local optima in the feature selection. In this study, the DE-MSVM method is proposed to improve the performance of Alzheimer's classification. The Pareto Optimal Front in Differential Evolution selects the relevant features in a nondominated manner. The AlexNet model extracts the features from the input images and apply for the feature selection. The Differential Evolution method selects the features to represent the characteristics of the images. The selected features were applied to MSVM for Alzheimer's classification on the ADNI dataset. The proposed DE-MSVM method has accuracy of 98.13% in the axial slice, and the existing whale-MSVM method has 95.23% accuracy. The future work of this proposed method is applied with a deep learning model for the classification.

Data Availability

The datasets analyzed during the current study are available in Alzheimer's Disease Neuroimaging Initiative (ADNI) repository, <https://adni.loni.usc.edu/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.



References

- [1] H. Nawaz, M. Maqsood, S. Afzal, F. Aadil, I. Mehmood, and S. Rho, "A deep feature-based real-time system for Alzheimer disease stage detection," *Multimedia Tools and Applications*, vol. 80, no. 28-29, pp. 35789-35807, 2020.
- [2] I. Saied, T. Arslan, and S. Chandran, "Classification of Alzheimers disease using RF signals and machine learning," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 1, 2021.
- [3] R. Sharma, T. Goel, M. Tanveer, S. Dwivedi, and R. Murugan, "FAF-DRVFL: fuzzy activation function based deep random vector functional links network for early diagnosis of Alzheimer disease," *Applied Soft Computing*, vol. 106, Article ID 107371, 2021.
- [4] H. J. Son, J. S. Oh, M. Oh, J.-H. Lee, J. H. Roh, and J. S. Kim, "The clinical feasibility of deep learning-based classification of amyloid PET images in visually equivocal cases," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 2, pp. 332-341, 2020.
- [5] M. Liu, F. Li, H. Yan et al., "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, Article ID 116459, 2020.
- [6] V. P. S. Rallabandi, K. Tulpule, M. Gattu, and Alzheimer's Disease Neuroimaging Initiative, "Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis," *Informatics in Medicine Unlocked*, vol. 18, Article ID 100305, 2020.
- [7] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, "A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification," *Magnetic Resonance Imaging*, vol. 78, pp. 119-126, 2021.
- [8] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo et al., "Convolutional neural networks for classification of Alzheimer's disease: o," *Medical Image Analysis*, vol. 63, Article ID 101694, 2020.
- [9] S. Basheera and M. Satya Sai Ram, "A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI," *Computerized Medical Imaging and Graphics*, vol. 81, Article ID 101713, 2020.
- [10] J. Liu, Y. Pan, F.-X. Wu, and J. Wang, "Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification," *Neurocomputing*, vol. 400, pp. 322-332, 2020.
- [11] S. Basaia, F. Agosta, L. Wagner et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinica*, vol. 21, Article ID 101645, 2019.
- [12] F. Ramzan, M. U. G. Khan, A. Rehmat et al., "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," *Journal of Medical Systems*, vol. 44, no. 2, p. 37, 2020.
- [13] A. Naz, A. Ashraf, and A. Zaib, "Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset," *Multimedia Systems*, pp. 1-10, 2021.
- [14] R. R. Janghel and Y. K. Rathore, "Deep convolution neural network based system for early diagnosis of Alzheimer's disease," *IRBM*, vol. 42, no. 4, pp. 258-267, 2021.
- [15] J. Samper-González, N. Burgos, S. Bottani et al., "Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data," *NeuroImage*, vol. 183, pp. 504-521, 2018.
- [16] Y. AbdulAzeem, W. M. Bahgat, and M. Badawy, "A CNN based framework for classification of Alzheimer's disease," *Neural Computing and Applications*, vol. 33, no. 16, pp. 10415-10428, 2021.
- [17] F. Eitel, K. Ritter, and K. Ritter, "Testing the robustness of attribution methods for convolutional neural networks in

- MRI-based Alzheimer's disease classification," in *Proceedings of the Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support*, In *Second International Workshop, iMIMIC 2019, and 9th International Workshop, ML-CDS*, pp. 3–11, Springer, Cham, Switzerland, October 2019.
- [18] P. R. Buvanewari and R. Gayathri, "Deep learning-based segmentation in classification of Alzheimer's disease," *Arabian Journal for Science and Engineering*, vol. 46, no. 6, pp. 5373–5383, 2021.
- [19] S. Alam, G. R. Kwon, J. I. Kim, and C. S. Park, "Twin SVM-based classification of Alzheimer's disease using complex dual-tree wavelet principal coefficients and LDA," *Journal of Healthcare Engineering*, vol. 2017, Article ID 8750506, 2017.
- [20] Y. Asim, B. Raza, A. K. Malik, S. Rathore, L. Hussain, and M. A. Iftikhar, "A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning," *International Journal of Imaging Systems and Technology*, vol. 28, no. 2, pp. 113–123, 2018.
- [21] A. Shakarami, H. Tarrah, and A. Mahdavi-Hormat, "A CAD system for diagnosing Alzheimer's disease using 2D slices and an improved AlexNet-SVM method," *Optik*, vol. 212, Article ID 164237, 2020.
- [22] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, "A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 320, pp. 195–202, 2018.
- [23] S. Neffati, K. Ben Abdellafou, I. Taouali, and K. Bouzrara, "An improved machine learning technique based on downsized KPCA for Alzheimer's disease classification," *International Journal of Imaging Systems and Technology*, vol. 29, no. 2, pp. 121–131, 2019.
- [24] R. Divya, R. S. S. Kumari, and Alzheimer's Disease Neuroimaging Initiative, "Genetic algorithm with logistic regression feature selection for Alzheimer's disease classification," *Neural Computing and Applications*, vol. 33, pp. 8435–8444, 2021.
- [25] L. Nanni, S. Brahmam, C. Salvatore, and I. Castiglioni, "Texture descriptors and voxels for the early diagnosis of Alzheimer's disease," *Artificial Intelligence in Medicine*, vol. 97, pp. 19–26, 2019.
- [26] G. T. Reddy, M. P. K. Reddy, K. Lakshmana, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evolutionary Intelligence*, vol. 13, no. 2, pp. 185–196, 2020.
- [27] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, and G. Srivastava, "Deep neural networks to predict diabetic retinopathy," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.
- [28] X. Li, X. Guo, P. Han, X. Wang, H. Li, and T. Luo, "Laplacian redecomposition for multimodal medical image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6880–6890, 2020.
- [29] J. Jose, N. Gautam, M. Tiwari et al., "An image quality enhancement scheme employing adolescent identity search algorithm in the NSSST domain for multimodal medical image fusion," *Biomedical Signal Processing and Control*, vol. 66, Article ID 102480, 2021.
- [30] H. Li, X.-J. Wu, and T. Durrani, "NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9645–9656, 2020.
- [31] U. Mlakar, I. Fister, J. Brest, and B. Potočnik, "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Expert Systems with Applications*, vol. 89, pp. 129–137, 2017.
- [32] H. Faris, M. M. Mafarja, A. A. Heidari et al., "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems," *Knowledge-Based Systems*, vol. 154, pp. 43–67, 2018.
- [33] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification," *Informatics in Medicine Unlocked*, vol. 16, Article ID 100151, 2019.
- [34] K. Thirumala, S. Pal, T. Jain, and A. C. Umarikar, "A classification method for multiple power quality disturbances using EWT based adaptive filtering and multiclass SVM," *Neurocomputing*, vol. 334, pp. 265–274, 2019.
- [35] I. Aswani Kumar and A. K. Cherukuri, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 4, pp. 462–472, 2017.
- [36] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "New imbalanced bearing fault diagnosis method based on Sample-characteristic Oversampling Technique (SCOTE) and multi-class LS-SVM," *Applied Soft Computing*, vol. 101, Article ID 107043, 2021.
- [37] Y. Liu, J.-W. Bi, and Z.-P. Fan, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Information Sciences*, vol. 394–395, pp. 38–52, 2017.
- [38] L. Sørensen, C. Igel, N. Liv Hansen et al., "Early detection of Alzheimer's disease using MRI hippocampal texture," *Human Brain Mapping*, vol. 37, no. 3, pp. 1148–1161, 2016.
- [39] D. B. Akhila, S. Shobhana, A. L. Fred, and S. N. Kumar, "Robust Alzheimer's disease classification based on multi-modal neuroimaging," in *Proceedings of the 2016 IEEE International Conference on Engineering and Technology (ICETECH)*, pp. 17–18, IEEE, Coimbatore, India, March 2016.

Research Article

Research on Privacy Protection Technology of Mobile Social Network Based on Data Mining under Big Data

Jiawen Du ¹ and Yong Pi ²

¹Law School, Wuhan University, Wuhan 430072, Hubei Province, China

²Shanghai International College of Intellectual Property, Tongji University, Shanghai 200092, China

Correspondence should be addressed to Jiawen Du; djwkira@whu.edu.cn

Received 24 November 2021; Accepted 18 December 2021; Published 13 January 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Jiawen Du and Yong Pi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advent of the era of big data, people's lives have undergone earth-shaking changes, not only getting rid of the cumbersome traditional data collection but also collecting and sorting information directly from people's footprints on social networks. This paper explores and analyzes the privacy issues in current social networks and puts forward the protection strategies of users' privacy data based on data mining algorithms so as to truly ensure that users' privacy in social networks will not be illegally infringed in the era of big data. The data mining algorithm proposed in this paper can protect the user's identity from being identified and the user's private information from being leaked. Using differential privacy protection methods in social networks can effectively protect users' privacy information in data publishing and data mining. Therefore, it is of great significance to study data publishing, data mining methods based on differential privacy protection, and their application in social networks.

1. Introduction

In recent years, with the development of communication technology, social software has brought convenience to user communication, ensured real-time contact between users, and accelerated the dissemination of information and news. Therefore, more and more users are attracted to register and use them. Their social circles have moved to social platforms, and their various activities and behaviors on social platforms have accumulated a lot of data [1]. With the rapid development of database technology and the continuous improvement of hardware level, as well as the increasing demand for information dissemination and sharing, a large amount of useful data can be saved [2]. Faced with such massive data storage, data mining and data publishing have become two important research directions for database applications [3]. Data mining is intended to extract meaningful rules and models from data, and data publishing is to present the data in an appropriate form [4]. Data release and data mining in social networks are likely to cause the personal sensitive information included in the social

network and the relationship between users to be destroyed or information leaked, which greatly affects the security of the use of social networks. There is a great risk of privacy leakage [5]. Therefore, how to better publish and mine the massive information in social networks without destroying its private information has become an important research topic in social networks [6].

Continuously enhancing the security of social networks and continuously improving privacy protection capabilities will help people use social networks more safely and securely [7]. At present, many privacy protection technologies have been proposed for user privacy and security issues in social networks. The easiest way to implement the technology is only to hide user identity information and not to process other information [8]. Although this technology protects the user's personal privacy within a certain range, malicious actors can still identify the individual's identity through the background knowledge of the target user's social network relationship, leading to the disclosure of user privacy [9]. Therefore, how to ensure the privacy and security of users when performing data mining on social network data is of

great significance [10]. The social network recommendation system not only helps users find valuable information for themselves but also allows the information to be displayed to interested users so as to achieve a win-win situation for information producers and information consumers [11]. Privacy protection mainly includes two aspects: the protection of sensitive knowledge and the protection of sensitive data. Sensitive knowledge mainly refers to sensitive knowledge such as association rules and classification rules extracted from the database; sensitive data refers to the private data that can correspond to an individual, thereby causing the individual to be exposed [12]. The paper explores and analyzes the privacy issues in current social networks and proposes user privacy data protection strategies based on data mining algorithms so as to hope that in the era of big data, the privacy of users in social networks will no longer be illegally violated [13].

Mobile social networking has become a rapidly growing application among domestic and foreign mobile users. It is urgent to protect user privacy [14]. The existing simple data processing methods cannot meet the needs of privacy protection, and the existing laws and regulations have restricted the application and development of data mining technology [15]. If certain protection measures are not taken for the information, the private information of a specific individual will be exposed, which will cause harm to the owner of the data. Similarly, if the protection measures taken are improper or too simple, then reasonable data mining methods will be used to obtain the private information of a specific individual, resulting in privacy leakage [16]. The privacy protection of social network data is to perform some artificial operations on the original network data, such as adding, deleting, or modifying parts, so that the attacker cannot obtain the user's sensitive information and avoid information leakage [17]. The data mining algorithm proposed in this paper can well protect the user's identity from being identified and the user's private information from being leaked. The algorithm can decompose the data, reconstruct the features, and store the data vertically, which can effectively prevent the data from being threatened by security and will not cause the loss of mining accuracy. Only the processed data can be released to the public. Of course, while protecting the user's sensitive information, making the processed information still have certain usability is also an important factor in measuring data anonymity.

2. Related Work

Literature [18] classifies privacy protection technologies into three categories according to different specific applications. They are privacy protection based on data perturbation technology, data encryption, and data anonymization.

Literature [19] put forward the concept of database anonymization and used the generalization method to hide sensitive attributes in groups of scale.

Literature [20] proposed a k -degree model for privacy protection of node degrees in social networks, which made it impossible for attackers to identify target nodes by collecting node degrees as background knowledge.

Literature [21] proposed to minimize information loss while generating a k -degree anonymous model.

Literature [22] proposed to construct the k -degree anonymous graph by using the idea of dynamic programming to protect the privacy of social network structure.

Literature [23] proposed that the parameter k of many existing K -anonymity models is predefined, and K represents the privacy protection of nodes in social networks. The idea of personalized privacy protection is formally introduced, and a K -anonymity model based on personalized privacy protection requirements is proposed.

Literature [24] divided the original network into k isomorphic subgraphs, which effectively prevented the node reidentification attack.

Literature [25] constructs K anonymity model for path privacy, and the construction method is to modify different types of edges based on greedy ideas.

Literature [26] combines L diversity on the basis of the k -degree model to protect the sensitive attributes of nodes or edge relations in social networks.

In literature [27], through clustering technology, the nodes in the original network are clustered to obtain an anonymous network composed of super nodes, and the super nodes are generalized to achieve the purpose of privacy protection.

Due to the development of technology sharing, big data are widely used in every aspect of life, and unreasonable use also brings great troubles and even terrible threats to users. However, at present, there is still no mature technology and relevant perfect laws and regulations for the protection of users' privacy. The lack of this aspect makes it impossible for relevant industry standards to have clear boundaries and implementation criteria and to implement effective measures to overcome this shortcoming. In order to ensure users' privacy, this paper carries out effective data mining and analysis on social networks. Combined with the KD tree optimization algorithm, a social network model based on data mining is built to protect the privacy of social networks, and experimental verification and algorithm analysis are carried out on data sets.

3. Methodology

Big data is like a huge spider web, weaving the network information of today's society. It is a large-scale and quite complex project, with the collection and processing function irreplaceable by other modern technologies. Thus, complexity, diversity, scale, and convenience are the outstanding characteristics of big data. It is such a combination of characteristics that big data technology has incomparable advantages over other technologies. The main problem of attribute and relationship-oriented data privacy protection is how to hide data in a relational database. The three common directions are data anonymization, secure multiparty computing, and data distortion. The comprehensive application of the three directions can effectively reduce the risk of personal data leakage.

The goal of the anonymous triangle protection principle is to protect those anonymous triangles in the process of

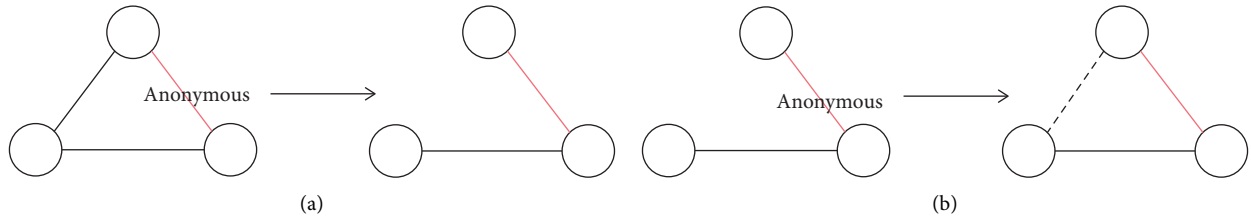


FIGURE 1: Anonymization triangle protection principle: (a) delete edge and (b) add edge.

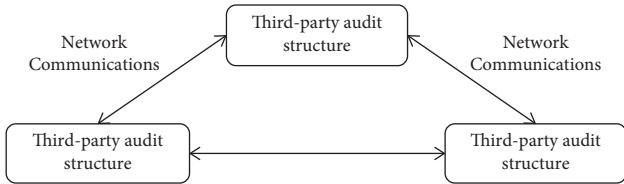


FIGURE 2: Data storage security architecture based on cloud computing.

graph anonymization. If multiple edges are generated, the original triangle that has been anonymized will not be included, and the triangle will not be anonymized gradually, so as to protect the original relationship as shown in Figure 1.

The common friends in the social network are all distributed on a scale, so only a small part of the connected edges have a higher relationship value. In order to participate in convenient social activities and enjoy all-round customized services in the era of big data, users cannot have absolute privacy rights [28]. But this does not mean that social network users can relax the protection of personal privacy but should pay more attention to the awareness of personal privacy protection. Only in this way can we ensure that we can enjoy normal services and social activities in the torrent of the big data era and can protect our privacy from being violated. At present, the release of dynamic social network data divides the privacy protection needs in social networks into different levels and at the same time provides privacy protection for users' sensitive attributes and sensitive edges in social networks. The data storage security system architecture based on cloud computing is shown in Figure 2.

K -anonymity technology has been widely used in anonymous relational data. In the privacy protection of graph data, many researchers still use k -anonymity technology to expand its application to graph data. K -nearest neighbor anonymity extracts all nodes with similar neighbors, encodes them, and divides them into the same group until each group is composed of at least k nodes. Then each group is anonymized so that any node in the same group has at least $k - 1$ isomorphic neighbor nodes. This method can effectively resist neighborhood attacks. For social networks, social networks have the characteristics of a "small world," and nodes with the same background are more likely to generate connections and aggregate in a small group. Therefore, the anonymous data after clustering privacy protection still retains the macro characteristics of the original network. Social network analysts can carry out data mining on social networks on the premise of ensuring users'

privacy and security so as to ensure the effectiveness of anonymous data.

Traditional data mining refers to the process of discovering new knowledge based on the original data and using corresponding mining algorithms. Traditional data mining algorithms cannot effectively protect private data, and security is affected. The KD3 framework is based on traditional data mining technology to process the privacy data that needs to be protected to form a new release database D' . Then reconstruct its features to form a new data feature F . And use the algorithm on it to adjust to get a new data mining algorithm M' . Finally, get a new mining result X' , make X' and X as close as possible. In this way, privacy data is effectively protected, and almost consistent mining results are obtained. The frame is shown in Figure 3.

At present, the development and utilization of various social software include the signing of privacy treaties, but most of these treaties are mandatory terms, and users can only be forced to accept them. Social network users cannot check individual options in the privacy terms according to their actual situation, so they check "agree" in order to have the right to use the software. In the process of software development, each merchant should take more initiative to consider the initiative of user authorization, rather than blindly forcing users to accept terms. Clustering-based privacy protection is also one of the mainstream protection technologies of graph data. The idea of aggregation is to aggregate the points or edges in the social network into a super point or super edge according to the similarity and perform the same anonymous operation on the members in the super point or super edge. Figure 4 shows the structure of the intrusion detection system.

Describe computer intrusion data by ω and v . Among them, ω represents the horizontal domain vector of computer network intrusion data, and v represents the vertical domain vector of computer network intrusion data. α represents the initial filtering result of the intrusion feature data; then α is expressed as follows:

$$\alpha = \sqrt{W}\omega \cdot s(\omega) + s(v) + m, \quad (1)$$

where W represents the norm vector of the intrusion signal, $s(\omega)$ represents the norm coefficient of the horizontal domain vector, $s(v)$ represents the norm coefficient of the vertical domain vector, and m represents the initial filtering constant. The signal processing result of the intrusion feature data can be expressed as follows:

$$R = W + 2\alpha\zeta(n) \cdot \omega v, \quad (2)$$

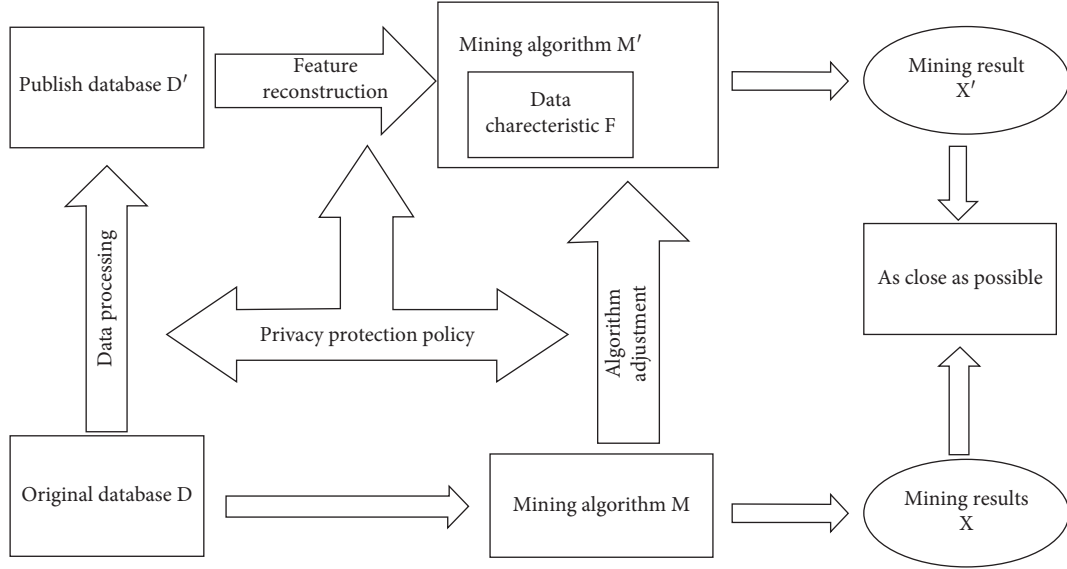


FIGURE 3: Data mining method framework for privacy protection.

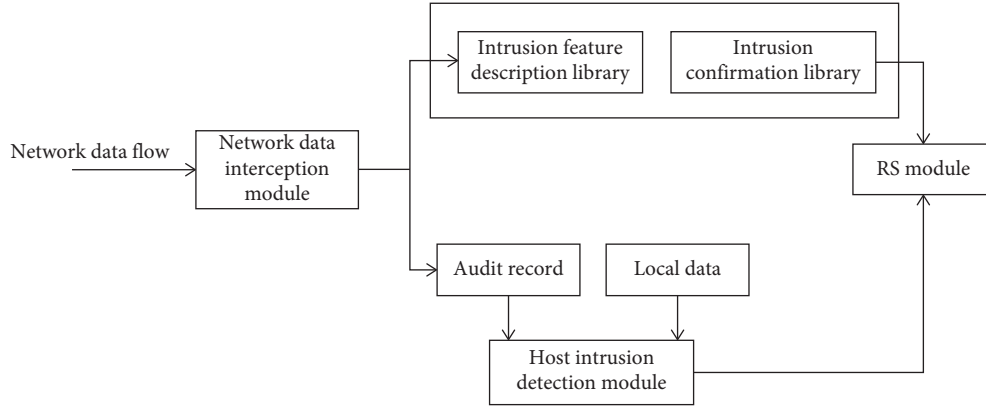


FIGURE 4: Schematic diagram of the structure of the intrusion detection system.

where $c(n)$ represents the superimposed signal processing result of computer intrusion communication data.

The agglomeration coefficient is generally used in social networks to describe the degree of interconnection between a point and its neighboring points, that is to say, the agglomeration coefficient can reflect the degree of mutual understanding between a user's friends. The local agglomeration coefficient is used to describe the properties of a specific vertex, and the average agglomeration coefficient is used to describe the average of the local agglomeration coefficients of all vertices in the entire social network. In the social network $G = (V, E, L)$, G is an undirected graph. The local agglomeration coefficient C_i of a vertex V_i in G is shown in the following formula:

$$C_i = \frac{2|e_{jk}|}{k_i(k_i - 1)}v_j, \quad v_k \in N_i, e_{jk} \in E, \quad (3)$$

where e_{jk} is the edge between vertices i and j , and N_i is $N_i = \{v_j: e_{ij} \in E \cap e_{ji} \in E\}$, which is the set of adjacent vertices of vertex v_i . k_i is the number of adjacent vertices in

v_i ; then in the social network $G = (V, E, L)$, the average agglomeration coefficient is shown in the following formula:

$$CC = \frac{1}{n} \sum_{i=1}^n C(i), \quad (4)$$

where n is the number of vertices in social network G .

Social network researchers can still use the clustered graph features to investigate the macro characteristics of the original graph. The main idea of the algorithm is: cluster the nodes of the social network according to the comprehensive distance between the nodes, cluster them into several super points, and the specific details in the super points are hidden. As long as the nodes in the two super points have one edge connected, there is only one edge connected between the two super points.

In the social network $G = (V, E, L)$, the average path length APL is the average of the shortest distance between all vertices, as shown in the formula:

$$APL = \frac{2}{n(n-1)} \sum_{v_i, v_j \in G} d(v_i, v_j), \quad (5)$$

where $d(v_i, v_j)$ is the shortest distance between the vertices v_i and v_j and n is the number of vertices in the social network G .

With the rapid development of the Internet and information technology, all kinds of data in social networks are constantly accumulating. With the progress of the times and the passage of time, big data has spread all over various fields and platforms. It also ushered in the generation of massive data. In the social network-oriented application, it is particularly important to protect users' privacy. By adopting certain protection strategies, users' data cannot be leaked, and their security can be guaranteed. KD tree is a kind of data structure, which can be used to divide data nodes into K -dimensional space. KD tree is a binary tree in which each node represents a spatial range. In order to further study the KD tree optimization center point selection method, define the following formulas. First, set the sample data set $A\{a_1, a_2, \dots, a_n\}$.

The number of data elements contained in a single rectangular cell Num

$$\text{Num} = \frac{n}{m \times k}, \quad (6)$$

where n represents the number of elements in the sample data set, k represents the number of clusters, and m represents the number of sub-blocks contained in a cluster. The data can be adjusted timely according to the size of the given data set. Usually, when there is little difference in the number of data set samples, m can be taken as 10. A complete KD tree can be constructed by knowing the three parameters of N , M , and K , while the parameters of K and M represent the depth of KD tree and the number of contained leaf nodes, respectively.

Rectangular unit center C_i

$$c_i = \frac{S_i}{W_i}, \quad (7)$$

where S_i represents the linear sum of all elements in the rectangular unit, W_i represents the weight of the rectangular unit, and its value mainly represents the number of sample elements contained in the rectangular unit.

The density Den of the rectangular unit is mainly used to indicate the density between the data elements contained in the rectangular unit.

$$\begin{aligned} \text{Den}_i &= \frac{W_i}{V_i} \\ &= \frac{W_i}{(\max(d_{\max} - d_{\min}))^2}, \end{aligned} \quad (8)$$

where W_i represents the number of sample elements contained in the rectangular cell, V_i represents the area of the rectangular cell, and d_{\max} and d_{\min} represent the maximum and minimum data elements in the corresponding rectangular cell, respectively.

With the improvement of data sharing and the development of data mining technology, people are getting more information, and the leakage of personal privacy data is

getting more and more attention. The hierarchical information security organization is shown in Figure 5.

The existence of vertices is one of the most basic privacy information in social networks. Everyone may be on many different social networks. The same user may disclose different privacy in different social networks. Vertex is a necessary condition for the existence of a social network, and the attribute of the vertex is easy to obtain information in the social network graph. Although differential privacy protection can effectively protect users' social relations, it is mainly based on that the attacker has mastered some information about the attack object. Therefore, the ability of the attacker should be reasonably evaluated before designing the privacy protection algorithm.

4. Result Analysis and Discussion

Data mining is a process of extracting hidden patterns from data. It is an important way to transform data into information and knowledge, and it is one of the effective means to analyze and process large amounts of data. At present, data mining technology has been widely used in biology, natural language processing, information retrieval, and other fields. Applying data mining methods to the research of social networks has become a new branch in the field of data mining.

The above mainly introduces some basic theories in social networks and the background knowledge that an attacker may have to launch an attack. And the algorithm of KD tree optimization to select the center point is analyzed experimentally. Brief introduction and summary of structured privacy protection technology and privacy protection technology with label attribute data. Although there are endless methods to protect user privacy in social networks, with the vigorous development of social applications and the large-scale increase in the number of people using social networks, social network data will become more and more complex, and privacy protection technologies need to be more perfect. The efficiency and usability of the algorithm for selecting the initial center point based on the KD tree optimization are analyzed. All the experimental results are simulated in MATLAB. The data set used in the experiment comes from UCI Machine Learning Repository, and the five data sets used in UCI are Iris, Ecoli, AcuteInflammations, Breastcancer, and Thyroid for related research. Table 1 is a description of these five data sets.

The accuracy of the KD tree optimization center point selection algorithm and the traditional K -medoids clustering algorithm when performing the same clustering are compared, as shown in Table 2.

From Table 2, we can see that compared with the traditional K -medoids algorithm, the accuracy of the KD-tree optimized center point selection algorithm proposed in this paper has been significantly improved, which shows that the KD-tree optimized center point selection algorithm is very effective. However, in the experiment, due to the KD tree optimization algorithm, it is necessary to build a KD tree and calculate the center and density of rectangular elements, so the time consumption is relatively large, which is inevitable.

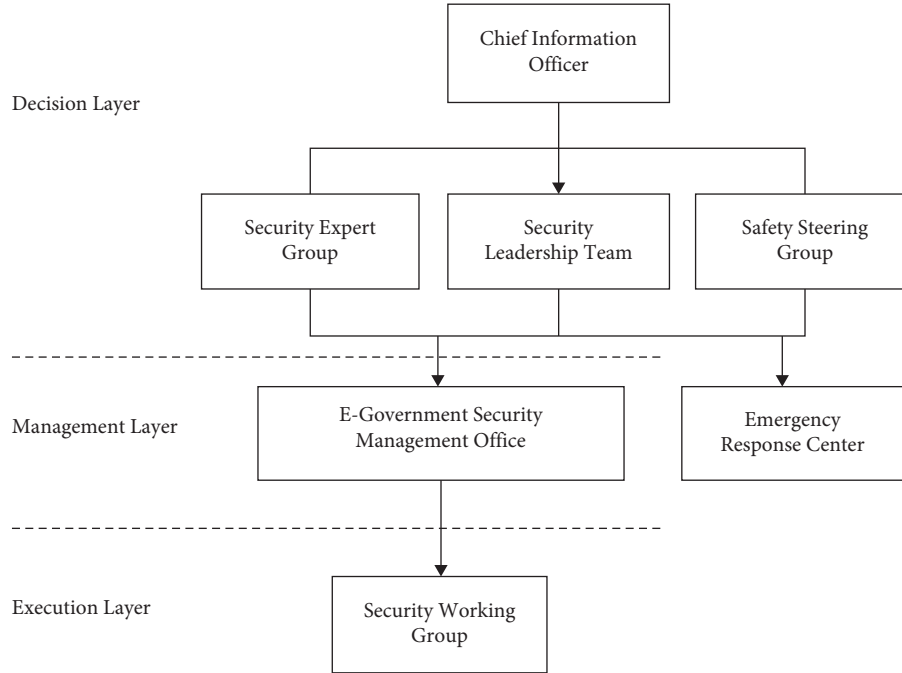


FIGURE 5: Hierarchical information security organization.

TABLE 1: Composition description of data set.

| Data set name | Type of data | Number of records in the data set | Number of attributes | Number of clusters |
|---------------|--------------|-----------------------------------|----------------------|--------------------|
| Iris | Multivariate | 152 | 5 | 4 |
| Ecoli | Multivariate | 334 | 7 | 8 |
| Breastcancer | Multivariate | 697 | 9 | 3 |
| Thyroid | Multivariate | 222 | 7 | 5 |

TABLE 2: Accuracy analysis of experimental results of KD tree optimization selection algorithm and traditional K -medoids algorithm.

| | K -medoids algorithm | | KD tree optimization algorithm | |
|--------------|------------------------|--------------|--------------------------------|--------------|
| | Running time (ms) | Accuracy (%) | Running time (ms) | Accuracy (%) |
| Iris | 36 | 76.42 | 45 | 87.12 |
| Ecoli | 79 | 73.66 | 108 | 85.54 |
| Breastcancer | 84 | 93.24 | 99 | 96.21 |
| Thyroid | 74 | 79.36 | 86 | 84.53 |

Therefore, the KD tree optimization algorithm proposed in this paper has a high accuracy for the processing of data with low dimensions.

Next, we further verify the effectiveness of the algorithm for higher dimension data. Table 3 shows the attribute description of related data sets.

The above data are applied to the KD tree optimization selection algorithm proposed in this paper and the traditional K -medoids algorithm, and five independent experiments are performed on each group of data to analyze the accuracy rate in detail and select each group of data. The average of the results of the five experiments was recorded, and the results of the experimental analysis are shown in Figure 6.

As can be seen from Figure 6, the KD tree optimization algorithm proposed in this paper is also suitable for high-dimensional data, and the accuracy is also high, but the

TABLE 3: Data set description.

| Number of data sets | Data dimension | Number of clusters |
|---------------------|----------------|--------------------|
| D_1 | 10 | 14 |
| D_2 | 20 | 14 |
| D_3 | 30 | 14 |
| D_4 | 40 | 14 |
| D_5 | 50 | 14 |

accuracy of the traditional algorithm is decreasing. Of course, when the data dimension is high, the time cost of the algorithm will increase accordingly.

The performance of the algorithm is analyzed from two aspects: data validity and algorithm running time. The evaluation of data validity focuses on the information loss caused by the algorithm after anonymity to the original social network. As shown in Figure 7, under the same degree

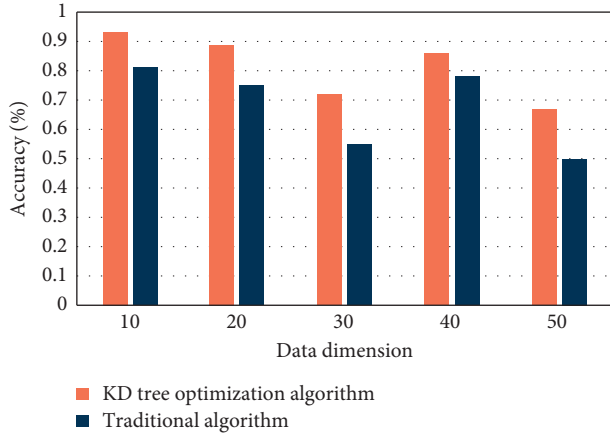


FIGURE 6: Accuracy analysis of experimental results in different dimensions.

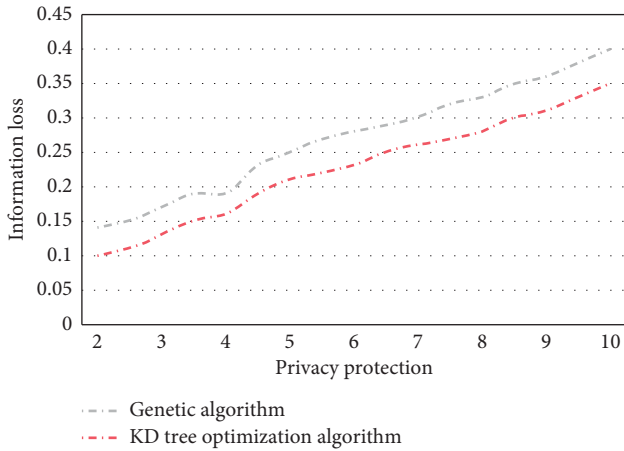


FIGURE 7: Comparison chart of algorithm information loss.

of privacy protection, the algorithm proposed in this paper has higher data availability.

As shown in Figure 8, under the same privacy protection strength, the KD tree algorithm takes less time, is more efficient, and has a higher time efficiency.

The algorithm in this paper first uses KD tree optimization to select k clustering centers. When there are new data, the nearest neighbor search method is used to cluster the new data reasonably so as to cluster the dynamic data quickly and efficiently. The algorithm only needs to process the incremental data so as to avoid reclustering all the data when the incremental data appears, thus improving the efficiency of clustering the incremental data to a certain extent. The increase in weight varies with k as shown in Figure 9, which basically changes linearly. The number of node splits varies with k as shown in Figure 10, which is positively correlated with the overall, but also related to the size of the data set. Because it needs to be affected by node grouping, if there are too many remaining nodes less than the value, too many nodes need to be split.

There are many ways to protect users' privacy in social networks, but what we cannot ignore is how to ensure the practicability and availability of anonymized data.

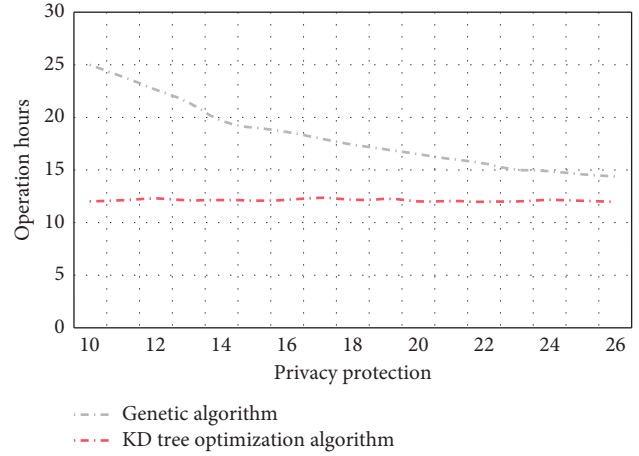


FIGURE 8: Comparison chart of the algorithm running time.

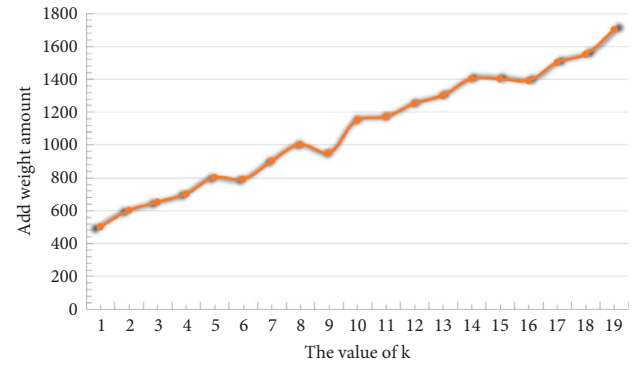


FIGURE 9: Schematic diagram of the change of the weight addition amount.

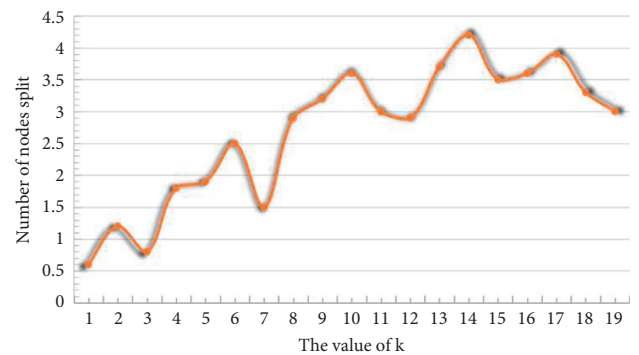


FIGURE 10: Schematic diagram of changes in the amount of node splitting.

Anonymous social network graph should ensure that the user's identity is not identified and the user's sensitive information is not leaked. Although different applications may have different anonymous methods to process data, they should ensure the authenticity of the processed data, which has its due research and mining value when it is released. On the one hand, data mining of privacy data protection should take certain protection measures for privacy data; on the other hand, in data mining, the protected data is mined, and

the algorithm adopted needs to eliminate and reconstruct the data, but the protection of privacy data that is damaged by decomposition will increase the storage capacity in decomposition, which wastes the storage space to a certain extent; at the same time, the damaged decomposition also effectively prevents data leakage and plays a very good security role. In the current era of big data, it is of great significance to explore the privacy protection of social networks.

5. Conclusions

With the continuous development of social network software and platform, a large number of data with social value and research significance have been accumulated. Data mining and analysis may lead to the disclosure of users' privacy. Therefore, how to ensure users' privacy security when effectively mining social networks is particularly important. Future research will mainly focus on the optimization of the algorithm so as to make it better applied to massive data. This paper proposes a KD tree optimal selection center point algorithm. Because it is easy to be attacked by external data in the process of dynamic clustering, the algorithm also introduces noise to disturb the data so as to achieve the effect of privacy protection. The data mining algorithm proposed in this paper can well protect the user's identity from being identified and the user's private information from being leaked. The algorithm can decompose the data and reconstruct the features and store the data vertically, which can effectively prevent the data from being threatened by security and will not cause the loss of mining accuracy. Only the processed data can be released to the public. Of course, while protecting the user's sensitive information, making the processed information still have certain usability is also an important factor in measuring data anonymity. The proposed anonymization algorithm is used in social networks to retain the structure of the original network and the effectiveness of the original data. While solving user identity authentication, data privacy disclosure, and information loss, it also protects the better application of social networks and creates more value. It is an important carrier for the development of the information age. Huge data is like a mine. The game between data mining and privacy protection is also a contest of technological development. Research on privacy protection technology of social network data still faces many new challenges, and there are still many problems to be further studied.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This paper was supported by the Research on Criminal Law Guarantee System of Network Security in China from the

perspective of new development concept and overall national security concept (no. 21AZD082) and Research on Criminal Compliance Issues of Cross-Border e-Commerce Enterprises in Shaanxi Province under the background of "One Belt and One Road" (no.2021E003).

References

- [1] P. Su, D. Yuan, and Martin, "Research on big data mining technology based on privacy protection," *Modern Computer: Professional Edition*, vol. 20, pp. 26–29, 2017.
- [2] Z. Ali, M. Imran, S. McClean, N. Khan, and M. Shoaib, "Protection of records and data authentication based on secret shares and watermarking," *Future Generation Computer Systems*, vol. 98, pp. 331–341, 2019.
- [3] F. Zhang and B. Shang, "The ethical dilemma and countermeasures of privacy protection in the era of big data," *People's Tribune Frontiers*, vol. 20, pp. 76–87, 2021.
- [4] Y. Zhou, H. Chai, and Y. Zhao, "Analysis of the status quo and trends of big data research in the field of international library information," *Library Journal*, vol. 38, no. 12, pp. 18–29+46, 2019.
- [5] W. Youke, W. Haiyang, W. Ningyun, and W. Yue, "An incentive-based protection and recovery strategy for secure big data in social networks," *Information Sciences*, vol. 508, pp. 79–91, 2020.
- [6] N. A. Khan, S. Zhang, W. Zhou, A. Almogren, I. Ud Din, and M. Asif, "Inferring ties in social IoT using location-based networks and identification of hidden suspicious ties," *Scientific Programming*, vol. 2020, no. 1, pp. 1–16, 2020.
- [7] B. Hu, "Implementation of data mining algorithms in big data security defense," *Industry and Technology Forum*, vol. 019, no. 7, pp. 48–49, 2020.
- [8] Atiquzzaman, N. Yen, and Z. Xu, "Big data analytics for cyber-physical system in smart city," in *Proceedings of the BDCPS: International conference on Big Data Analytics for Cyber-Physical-Systems*, BDCPS, Shengyang, China, December 2019.
- [9] W. Yan, "Research on the protection of women's rights and the countermeasures of social support based on big data network background," *International Journal for Engineering Modelling*, vol. 31, no. 1, pp. 252–257, 2018.
- [10] M. Zhou, Z. Duan, and C. Shang, "Research on the three dimensions of education privacy protection in the big data era," *Guangxi Radio and TV University*, vol. 3, pp. 25–28, 2021.
- [11] L. Zhang and B. Ashuri, "BIM log mining: discovering social networks," *Automation in Construction*, vol. 91, no. 7, pp. 31–43, 2018.
- [12] G. Zuo, "Research on distributed privacy protection clustering mining algorithm based on big data," *Intelligent Computers and Applications*, vol. 8, no. 6, pp. 63–66, 2018.
- [13] X. Zhang and Y. Wu, "Research progress of empirical asset pricing based on network big data mining," *Economic Trends*, vol. 000, no. 6, pp. 129–140, 2018.
- [14] R. Yang, "Research on empirical asset pricing based on network big data mining," *National Circulation Economy*, vol. 2224, no. 28, pp. 69–70, 2019.
- [15] X. Wang, "Artificial intelligence and privacy protection in the era of big data medical care," *Electronic Product World*, vol. 26, no. 6, pp. 84–86, 2019.
- [16] R. Toujani and J. Akaichi, "Event news detection and citizens community structure for disaster management in social

- networks,” *Online Information Review*, vol. 43, no. 1, pp. 113–132, 2019.
- [17] W. Yamin, Z. Fuanguo, W. Huaxiong, G. Zheng, M. Yinbin, and D. Yuqiao, “A new secret handshake scheme with multi-symptom intersection for mobile healthcare social networks-ScienceDirect,” *Information Sciences*, vol. 520, pp. 142–154, 2020.
- [18] P. Zhao, K. Bian, T. Zhao et al., “Understanding smartphone sensor and app data for enhancing the security of secret questions,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 552–565, 2017.
- [19] J. Zhang, Y. Ma, and W. Xie, “Research on differential privacy protection for location big data,” *Software Guide*, vol. 017, no. 11, pp. 206–208, 2018.
- [20] X. Chen, “Research on social network data mining technology based on naive bayes algorithm,” *Computer Measurement & Control*, vol. 25, no. 6, pp. 199–202, 2017.
- [21] M. Hou, R. Wei, X. Lan, L. Xing, T. Na, and L. Lu, “Application research of medical big data privacy protection model based on differential privacy,” *China Digital Medicine*, vol. 014, no. 12, pp. 86–88, 2019.
- [22] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho, and S. Baik, “Image steganography using uncorrelated color space and its application for security of visual contents in online social networks,” *Future Generation Computer Systems*, vol. 86, no. 9, pp. 951–960, 2016.
- [23] H. Chen, G. Wang, and P. Zhang, “Key nodes mining algorithm in Sina Weibo social network based on Hadoop cloud platform,” *Dongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Southeast University (Natural Science Edition)*, vol. 48, no. 4, pp. 590–595, 2018.
- [24] P. Pinto, I. Theodoro, M. Arrais, and J. Oliveira, “Data mining and social web semantics: a case study on the use of hashtags and memes in Online Social Networks,” *IEEE Latin America Transactions*, vol. 15, no. 12, pp. 2276–2281, 2017.
- [25] A. Farasat, G. Gross, R. Nagi, and A. G. Nikolaev, “Social network analysis with data fusion,” *IEEE Transactions on Computational Social Systems*, vol. 3, no. 2, pp. 1–12, 2016.
- [26] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, and P. Viswanath, “Hiding the rumor source,” *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6679–6713, 2017.
- [27] H. Zhu, Y. Zhang, and Z. Yan, “A provably password authenticated key exchange scheme based on chaotic maps in different realm,” *International Journal on Network Security*, vol. 18, no. 4, pp. 688–698, 2016.
- [28] B. Martin, “\Let’s protest\”: surprises in communicating against repression,” *IEEE potentials*, vol. 35, no. 5, pp. 16–18, 2016.

Research Article

Cloud-Assisted Privacy-Preserving Method for Healthcare Using Adaptive Fractional Brain Storm Integrated Whale Optimization Algorithm

S. Thanga Revathi ¹, A. Gayathri,² J. Kalaivani,¹ Mary Subaja Christo ¹, Danilo Pelusi,³ and M. Azees ⁴

¹Department of Networking and Communications (School of Computing), SRM Institute of Science and Technology, Kattankulathur 603203, Chennai, India

²Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS, Chennai, India

³Faculty of Communication Sciences, University of Teramo, Via Balzarini, Teramo 64100, Italy

⁴Department of ECE, GMR Institute of Technology, Rajam, Andhra Pradesh 532127, India

Correspondence should be addressed to Mary Subaja Christo; marysubaja@gmail.com and M. Azees; azees.m@gmrit.edu.in

Received 22 October 2021; Revised 25 November 2021; Accepted 1 December 2021; Published 23 December 2021

Academic Editor: Thippa Reddy G

Copyright © 2021 S. Thanga Revathi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The security of medical data in the cloud is the key consideration of cloud customers. While publishing the medical data, the cloud distributor may suffer from data leakages and attacks such that the data may leak. In order to resolve this, this article devises the developed Adaptive Fractional Brain Storm Integrated Whale Optimization Algorithm (AFBS_WOA), which is the hybridization of Adaptive Fractional Brain Storm Optimization (AFBSO) and Whale Optimization algorithm (WOA). The developed AFBS_WOA algorithm generates the key matrix coefficient for retrieving the perturbed database in order to preserve the privacy of healthcare data in the cloud. The developed AFBS-WOA scheme utilized the fitness function involving utility and privacy measures for calculating the secret key. Here, the privacy-preserved database is obtained by multiplying the input database with a key matrix based on developed AFBS-WOA using the Tracy–Singh product. For data retrieval, the secret key is shared with the service provider in order to retrieve the database, and then the data are accessed. Moreover, the experimental result demonstrates that the developed AFBS_WOA model attained the maximum utility and privacy measure of 0.1872 and 0.8755 using the Hungarian dataset.

1. Introduction

Healthcare involves various complex processes, such as treatment, diagnosis, prevention, and injury. Due to the advancement of technology, the healthcare industry has been introduced to reduce the services provided by hospitals. Nowadays, the healthcare industry is one of the tremendous emerging economies of the country. The healthcare industry performs various processes, such as medical data storage, medical data sharing, and providing medical advice to patients. Healthcare in a cloud environment is used to store the medical information of patients, including medical records, medical reports, scanning reports, and patient history. This

information should be kept confidential and preserved in the cloud environment to maintain secrecy. The major challenge of healthcare in cloud computing is the privacy preservation of sensitive data. Since the healthcare model shares the sensitive information of patients with the service provider and third party through the cloud environment, this information may be hacked by the attackers [1]. Hence, various privacy preservation techniques have been introduced to conserve medical information. The important parameters considered for maintaining the security of medical information is privacy measure and utility measure.

Privacy preservation is the process of conserving the sensitive information of an individual before publishing.

Some of the common privacy preservation techniques are perturbation approaches, generalization approaches, and synthetic data generators [2]. The perturbation approach produces some changes to input data, whereas the generalization approach replaces the original elements with less accurate elements, and synthetic data generators generate the synthetic data similar to the original data [3, 4]. Moreover, other protection methods employed to ensure the secrecy of information are data sanitation, blocking, cryptography, and anonymization. Data publishing with privacy conservation needs controlling the distribution while utilizing the personal details of a person. Data publishing involves two stages: data gathering and data publishing. In the initial phase, data is gathered from data owners through the data publisher, whereas the data publisher shares gathered information public or data recipient. In cloud storage, the stored information is offered by the Third-Party Auditor (TPA). Three service schemes, Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS), are widely utilized for the computation and configuration of applications through Internet [2].

This article devises a novel optimization technique, namely, AFBS-WOA, to generate the retrievable data perturbation model in order to secure the data in the cloud. Here, the developed AFBS-WOA model is designed by incorporating AFBSO and WOA. Initially, the input database is combined with a secret key using the Tracy–Singh product, where the secret key is obtained by the developed AFBS-WOA scheme. The developed AFBS-WOA scheme utilized the fitness function, such as the utility and privacy measure for calculating the secret key. Moreover, the generated secret key is employed for the privacy-preserved healthcare data publishing in cloud computing.

1.1. Major Contribution of Developed AFBS-WOA Technique. The developed AFBS-WOA technique is devised to recognize the optimal key coefficient generation in order to preserve the privacy of healthcare data in the cloud. The algorithm utilizes the fitness function to calculate the secret key. The generated key will be used for the perturbation of the data in order to secure the patients' data in the cloud. The stored data are retrieved at the required place after performing the retrieval process using the generated key.

The remaining section of this article is formed in the mentioned manner. Section 2 describes the literature survey based on privacy preservation in cloud computing; Section 3 describes the developed AFBS-WOA technique; Section 4 demonstrates the discussion of results; Section 5 provides the conclusion of this research.

2. Motivation

In the medical field, the medical records are conveyed to the research panel for deciding the kind, defect, severity, and effects of diseases. While publishing the information, the publishers do not leak any patients' information with others for sustaining the data privacy. This motivates the researchers to do research in this domain.

2.1. Literature Survey. This section describes the literature survey of various existing techniques based on the privacy preservation of data in cloud computing.

Benifa and Mini [5] developed the Genetic Grey Wolf Optimization Algorithm (GGWO) to preserve the secrecy of information. Although the GGWO method obscures the sensitive information effectively, the information loss attained by this method was high. George and Sumathi [2] devised the Crow search-based Lion algorithm for generating the key matrix coefficient in order to preserve the information in the cloud. This method attained the maximum privacy measure and utility measure. However, the computational complexity of this method was high. Majeed [6] modeled the secure anonymization scheme for conserving the secrecy of medical information saved in the cloud. Although the privacy and utility attained by this method were high, this method has failed with a diverse environment. Yousra and Mazleena [7] developed the Privacy-Preserving Data Mining (PPDM) scheme for preserving the privacy of datasets. Although the processing speed of this method was high, the computation cost of this method was high.

Vijayakumar et al. [8, 9] proposed an alert system for helping patients with heart diseases during an emergency. The system sends a private and confidential message from the heart patient to the healthcare entities, including hospital, ambulance service, and personal doctor. The system has ensured a comparatively high level of security with low computational overhead and communication overhead. Zhou et al. [10, 11] have proposed an identity-based distributed decryption scheme for a personal health record sharing system. In this method, the data can be shared with multiple parties without reconstructing the decryption key. Moreover, it is proposed that it is secure against chosen ciphertext attack (CCA). The dynamic searchable symmetric encryption (DSSE) technique allows the user to search the dynamic information from the IIOTH system. Liu et al. [12] proposed a privacy-preserving DSSE scheme for IIOTH for the database with forward security. A secure index is developed based on the hash chain to overcome the file injection attack. Furthermore, the fine-grained operations are executed over the encrypted files, which return only the matched attribute instead of the whole file. This article also proposed a scheme to achieve attribute-based access control. Qian et al. [13] proposed a private set intersection scheme for fine-grained profile matching. The medical data are secured by reencryption techniques and the patient's data are divided using multitag to perform the fine-grained operations. This proposed system has demonstrated that this scheme has improved efficiency by reducing the bilinear pairs.

Wang et al. [14] proposed a lightweight and reliable authentication protocol to handle the physical layer security problem and overcentralized server problem using cutting-edge blockchain technology and physically unclonable functions. In addition to this, a future extractor scheme was also proposed to handle the biometric information. The reliability of the system is proved using security evaluation methods, which illustrate that the authentication protocol

requires the least computational and communication cost. Wang et al. [15] proposed a system to address the common security weaknesses, such as the man-in-the-middle attack, key generation center, and denial-of-service attacks, by a novel pairing-free certificateless scheme. The system is based on the blockchain technique and smart contract to construct a reliable and efficient lightweight certificateless signature (CLS) scheme. The system is evaluated and proved to be reliable with less computational cost and communication cost.

2.2. Challenges. The challenges faced by the various privacy preservation techniques in cloud computing are listed as follows:

- (i) The performance of the GGWO method can be enhanced by extending the technique with the group of optimization models using numerous datasets [1].
- (ii) In [2], the security of the developed scheme is enhanced by including dyadic products; however, the performance of the dyadic product is not effective for all possibilities. Hence, the dyadic product can be replaced with some other advanced concepts for further improvements.
- (iii) In [3], the selection of applicable trusted infrastructure, service provider, and algorithms is still inadequate to satisfy user confidentiality requirements.
- (iv) The security method in [4] can be extended by including some effective sensitive attributes from the anonymous data in order to improve security.

3. Proposed AFBS-WOA for Privacy-Preserved Healthcare Data Publishing

This section describes the developed AFBS-WOA model for resolving the security issues in the cloud. Figure 1 shows the structural design of the privacy preservation scheme in the cloud using the developed AFBS-WOA. Here, the original database is multiplied with the optimal key matrix produced by the developed AFBS-WOA model using the Tracy–Singh product [16, 17]. The size of the matrix generated from the Tracy–Singh product is large; hence, the matrix size is reduced to be the same as the input data size for further effective processing. From the reduced matrix, an optimal key coefficient is selected, which acts as a key for the retrieval of the perturbation database. For data retrieval, the secret key is shared with the service provider, and then the data are accessed.

3.1. Privacy-Protected Data Publishing. In order to attain privacy-protected data publishing, there is a need to generate a retrievable perturbation database. For that, let us assume the database to be R and its matrix size $M * N$; then, the input data matrix is represented as follows:

$$R_{M \times N} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1v} \\ r_{21} & r_{22} & \cdots & r_{2v} \\ \vdots & \vdots & & \\ r_{u1} & r_{u2} & \cdots & r_{uv} \end{bmatrix}, \quad (1)$$

where r_{uv} represents the coefficients of data and the values of u and v range from M to N , respectively. After that, the matrix multiplication is performed for the input matrix with optimal key matrix produced by developed AFBS-WOA model using the Tracy–Singh product. The expression for Tracy–Singh product is signified as follows:

$$T_{MO \times NP} = R_{M \times N} \circ W_{O \times P}, \quad (2)$$

where the term $W_{O \times P}$ denotes the optimal key matrix acquired from AFBS-WOA, which is signified as follows:

$$W_{O \times P} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1q} \\ h_{21} & h_{22} & \cdots & h_{2q} \\ \vdots & \vdots & & \\ h_{p1} & h_{p2} & \cdots & h_{pq} \end{bmatrix}, \quad (3)$$

where h_{pq} specifies the optimal key matrix coefficient. The mathematical expression for the Tracy–Singh product among optimal key matrix with input data matrix is signified as follows:

$$V_{SR \times UW} = (R_{uv} \circ W)_{uv} = \left((R_{uv} \otimes W_{pq})_{pq} \right)_{uv}. \quad (4)$$

For the Tracy–Singh product, the product of the input partial matrix and the key coefficients are represented as follows:

$$V_{SR \times UW} = \begin{bmatrix} R_{11} \circ W & R_{12} \circ W \\ R_{21} \circ W & R_{22} \circ W \end{bmatrix}. \quad (5)$$

Here, the terms R_{11} , R_{12} , R_{21} , and R_{22} indicate the matrix of input data and $V_{SR \times UW}$ represents the matrix produced by the Tracy–Singh product. The obtained matrix size of the Tracy–Singh product is large, which is diminished to a size similar to that of the input database matrix for secure processing. Then, the mathematical notation for the reduced matrix is represented as follows:

$$G_{M \times N} = \begin{bmatrix} G_1 & G_2 \\ G_3 & G_4 \end{bmatrix}, \quad (6)$$

where G_1 , G_2 , G_3 , and G_4 denote the data matrix elements. Then, the secret key $J_{1 \times 1} = r_{11}$ is assessed from the reduced optimal key matrix, which is the primary element of the optimal key matrix. Moreover, the perturbed database R^* is recognized by performing EX-OR operation for both reduced matrix $G_{M \times N}$ and the secret key $J_{1 \times 1}$, and it is given in the following:

$$R_{M \times N}^* = G_{M \times N} \oplus J_{1 \times 1}, \quad (7)$$

where $G_{M \times N}$ depicts the reduced matrix.

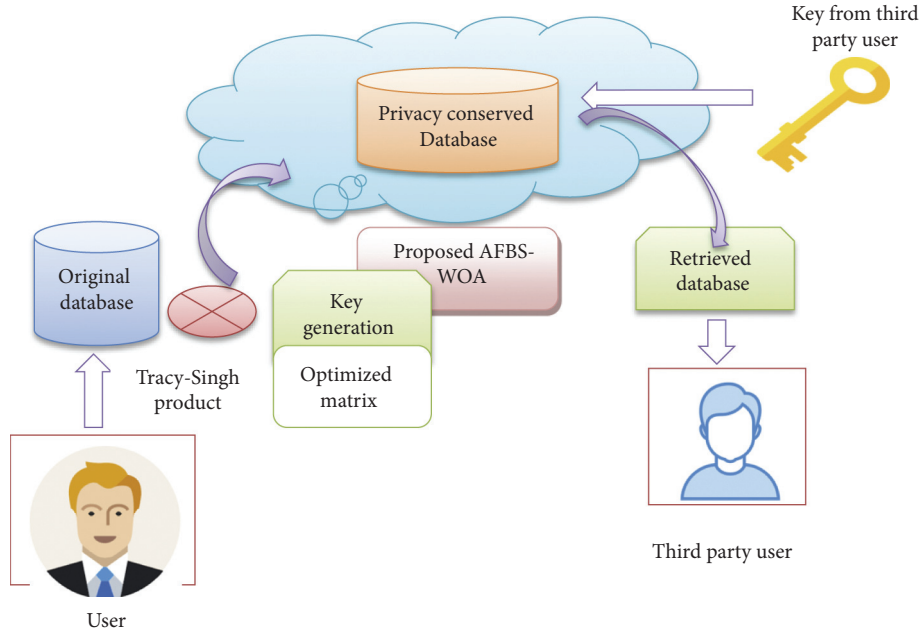


FIGURE 1: Retrievable data perturbation model using the proposed AFBS-WOA.

3.2. Optimal Key Coefficient Generation Using Developed AFBS-WOA. This section describes the novel optimization method, namely, AFBS-WOA, for the optimal key coefficient generation. The developed AFBS-WOA algorithm is designed by integrating AFBSO [18] and WOA. The developed AFBS-WOA algorithm selects the key matrix coefficient without changing the characteristics of the original database. WOA is a metaheuristic method, which mimics the hunting activities of whales, and the AFBSO algorithm is designed based on the activities of the human brain. Although the searching behavior of WOA is high, the detection performance of this method is poor. Thus, the WOA algorithm does not attain the local optimum. In order to overcome this, the AFBS-WOA is developed, which improves the optimization issues. Thus, the developed AFBS-WOA algorithm selects the optimal key matrix coefficient. The processing steps for the developed AFBS-WOA algorithm are given as follows.

3.2.1. Initialization of Whale Population. The initialization function describes the size of the solution vector that relies on the size of the input matrix, which is demonstrated as follows:

$$S = \{S_u; 1 \leq u \leq O\}, \quad (8)$$

where O represents the size of solution space and S_u represents the initialization of whales.

3.2.2. Fitness Measure. The fitness with maximum values of privacy and utility measure is considered as an optimal solution. The fitness function is used to predict the optimal solution. Here, the privacy measure is based on the modification degree and utility is based on the numerical

characteristics, which are required to be preserved in order to attain the privacy-preserved database. The fitness function is described as follows:

$$\text{Fitness} = \left[\frac{H + I}{2} \right], \quad (9)$$

where H and I denote the privacy and utility, which are expressed as follows:

$$H = \frac{1}{M \times N} \sum_{u=1}^M \sum_{v=1}^N \frac{(R_{uv} - R_{uv}^*)}{\text{Max}(R_{uv}, R_{uv}^*)}, \quad (10)$$

$$I = \frac{a + b}{2}, \quad (11)$$

where R_{uv}^* depicts the data elements of retrieved data, a depicts the mean, and b depicts the covariance. Moreover, the numerical aspects of the database rely on both mean and covariance acquired from the original and retrieved databases. The expression for mean and covariance is stated as follows:

$$\begin{aligned} \text{Mean } a &= 1 - \left[\frac{1}{M \times N} \sum_{u=1}^M \sum_{v=1}^N \frac{R_{uv}}{\text{Max}(R_{uv})} - \frac{1}{M \times N} \sum_{u=1}^M \sum_{v=1}^N \frac{R_{uv}^*}{\text{Max}(R_{uv}^*)} \right], \\ \text{Covariance } b &= 1 - \left[\frac{1}{M \times N} \sum_{u=1}^M \sum_{v=1}^N Q_{uv} - \frac{1}{M \times N} \sum_{u=1}^M \sum_{v=1}^N Q_{uv}^* \right], \end{aligned} \quad (12)$$

where Q_{uv} and Q_{uv}^* defines the covariance of both original and retrieved database, correspondingly.

3.2.3. Solution Update Phase for Encircling Prey. In this phase, the solution update is done on the searching probability condition $j < 0.5$ and the value of $|Y|$ is less than 1.

The distance measure is notified as \vec{I} , and its solution update is stated as follows:

$$\begin{aligned}\vec{I} &= \left| \vec{U} \vec{S}^*(f) - \vec{S}(f) \right|, \\ \vec{S}(f+1) &= \vec{S}^*(f) - \vec{Y} \cdot \vec{I},\end{aligned}\quad (13)$$

where \vec{Y} and \vec{U} indicate two coefficients such that \vec{Y} relies on two constraints, like \vec{m} and \vec{n} . The value of \vec{U} ranges between 0 and 1, whereas the value of \vec{Y} differs between 2 and 0.

$$\begin{aligned}\vec{Y} &= \vec{m} \cdot \vec{n} - \vec{m}, \\ \vec{U} &= 2 \cdot \vec{n}.\end{aligned}\quad (14)$$

3.2.4. Solution Update Phase for the Exploitation of WOA. After the solution update is done based on searching probability condition $j > 0.5$ and the value of $|Y|$ is fewer than 1, the exploitation process initiates, and it is stated as follows:

$$\begin{aligned}\vec{S}(f+1) &= \vec{I}^b \cdot e^{bf} \cdot \cos(2\pi\alpha) + \vec{S}^*(f), \\ \vec{I}^b &= \left| \vec{S}^*(f) - \vec{S}(f) \right|,\end{aligned}\quad (15)$$

where b and α specify the constant for search space, and the value of α differs between $[-1, 1]$.

3.2.5. Solution Update for the Exploration of WOA. When the condition of searching probability $j < 0.5$ and $|Y|$ is bigger than 1, then update solution becomes

$$\begin{aligned}\vec{S}(f+1) &= \vec{S}_{\text{rand}} - \vec{Y} \cdot \vec{I}, \\ \vec{I} &= \left| \vec{U} \vec{S}_{\text{rand}} - \vec{S} \right|,\end{aligned}\quad (16)$$

where \vec{S}_{rand} states the random solution from optimization.

3.2.6. Solution Update Using AFBSO. The purpose of the AFBSO algorithm [18] is to select the optimal matrix coefficient. The optimal matrix coefficient is selected by updating the following equation:

$$\begin{aligned}S(f+1) &= \mu S(f) + \frac{1}{2} \mu S(f-1) + \frac{1}{6} \mu (1-\mu) S(f-2) \\ &+ \frac{1}{24} \mu (1-\mu) (2-\mu) S(f-3) + \chi L(a, b),\end{aligned}\quad (17)$$

where $\mu = ((f_{ct} - f_{\min}) / (f_{\max} - f_{\min}))$, which represents the adaptive factor, $S(f)$ represents the idea chosen from the previous iteration, $S(f-1)$ represents the idea chosen from $(f-1)$ th iteration, $S(f-2)$ represents the idea chosen from $(f-2)$ th iteration, $S(f-3)$ represents the idea chosen from $(f-3)$ th iteration, and $L(a, b)$ represents the Gaussian random value with mean a and variance b .

$$\chi = d \log \text{sig} \left(\frac{L_{c\max}/2 - L_c}{K} \right); \quad d = [0, 1], \quad (18)$$

where $L_{c\max}$ represents the maximum iteration and L_c represents the correct iteration.

3.2.7. Reevaluation of Fitness Criterion. The solution acquired from the entire process is reevaluated using the fitness function, from which the maximum fitness value is considered an optimal solution.

3.2.8. Termination. At the final stage of iteration Z , a key matrix is obtained, which is considered as an optimal solution, and the obtained key matrix is employed to generate a retrievable perturbation database.

$$W_{O \times P} = S_{\text{optimized}}. \quad (19)$$

3.3. Retrieval from Perturbed Database. For the retrieval phase, the perturbed database is EXOR^{ed} with a secret key, which provides the reduced matrix $G_{M \times N}^*$. Furthermore, the reduced matrix $G_{M \times N}^*$ is partitioned using secret key J , such that the original database is retrieved at the receiver side.

$$\begin{aligned}G_{M \times N}^* &= R_{M \times N}^* \oplus J_{1 \times 1}, \\ R_{M \times N}^* &= \frac{G_{M \times N}^*}{J_{1 \times 1}},\end{aligned}\quad (20)$$

where $G_{M \times N}^*$ represents the retrievable reduced matrix and $R_{M \times N}^*$ shows the retrievable perturbation matrix. Table 1 describes the pseudocode of the developed AFBS-WOA for the retrievable data perturbation model.

3.4. Security Analysis. The developed AFBS-WOA model is developed for resolving the security issues in the data stored in the cloud. The original data are multiplied with the generated optimal key matrix before being stored in the cloud. An optimal key matrix is generated by the proposed AFBS-WOA model using the Tracy–Singh product. The generated matrix size is reduced to the size of the input for further processing. The stored data can be retrieved using an identified optimal key coefficient from the reduced matrix. This secret key shall be shared with the service providers for the data to be accessed. The following subsection discusses threats, security issues, challenges, and solutions for different kinds of attacks in the proposed system.

3.4.1. Password Guessing Attack. Password guessing attack is commonly known as Brute-Force Attack. The attacker tries to guess the password with certain combinations of the user credentials. To avoid this attack, the password should be set strongly. In our proposed system, the key to providing data security is generated separately for every data stored in the cloud. This makes it more secure against the Password Guessing Attack.

TABLE 1: Algorithmic procedure of the developed AFBS-WOA for privacy-preserved healthcare data publishing.

| | |
|----|---|
| 1 | Perturbed database retrieval using proposed AFBS-WOA |
| 2 | Input: input database |
| 3 | Output: perturbed data |
| 4 | Parameters: Search agent S , highest iteration Z , optimized search agent S^* |
| 5 | Initiate the algorithmic parameters |
| 6 | For all element in database |
| 7 | Calculate the secret key (apply AFBS-WOA) |
| 8 | Calculate the Tracy Singh product based on input database as well as optimal key matrix |
| 9 | Calculate secret key J from optimal key matrix |
| 10 | Perturbed database $R_{M \times N}^* = G_{M \times N} \oplus J_{1 \times 1}$ |
| 11 | End for |
| 12 | End |
| 13 | //Developed AFBS-WOA algorithm |
| 14 | Initiate |
| 15 | Arbitrarily initialize whale population $S = \{S_u; 1 \leq u \leq O\}$ |
| 16 | Estimate the fitness measure using maximum privacy as well as utility |
| 17 | While ($f < Z$) |
| 18 | For every S |
| 19 | If ($j < 0.5$) |
| 20 | If ($ Y < 1$) |
| 21 | $\vec{S}(f+1) = S^*(f) - \vec{Y} \cdot \vec{I}$ |
| 22 | Else if ($ Y \geq 1$) |
| 23 | $\vec{S}(f+1) = \vec{S}_{rand} - \vec{Y} \cdot \vec{I}$ |
| 24 | End if |
| 25 | Else if ($j < 0.5$) |
| 26 | If ($ Y < 1$) |
| 27 | $\vec{S}(f+1) = \vec{I}^j \cdot e^{bf} \cdot \cos(2\pi\alpha) + \vec{S}^*(f)$ |
| 28 | Else if ($ Y \geq 1$) |
| 29 | Update the solution using AFBSO |
| 30 | End if |
| 31 | End if |
| 32 | End for |
| 33 | Evaluate Y if exceeds the search space |
| 34 | Renew S^* for optimal solution |
| 35 | $f = f + 1$ |
| 36 | End while |
| 37 | Return optimal solution S^* |
| 38 | End |

3.4.2. *Data Breaches.* Data breach refers to the leakage of data to the unauthorized user [19]. This attack can have a huge impact on the organization, including the leakage of sensitive data. This may occur due to problems in application designing, operational issues, or access by unauthorized users. In our proposed system, the data are stored in a perturbed format and can be retrieved only with the help of the unique secret key used for the perturbation. Thus, this system holds better for the data breach.

3.4.3. *Man-in-the-Middle Attack.* The man-in-the-middle attack is when an attacker positions himself between the two endpoints and alters the communication between the two parties. The proposed system perturbs the original data with an optimal key matrix $V_{SR \times UW} = (R_{uv} \circ W)_{uv} = ((R_{uv} \otimes W_{pq})_{pq})_{uv}$ before being storing in the cloud. Moreover, the data can be retrieved only using the secret key $G_{M \times N}^* = R_{M \times N}^* \oplus J_{1 \times 1}$, which is known only to the authorized persons. If the attacker is present between the cloud and the user, the attacker could not read the message as it is in the perturbed format and it can be retrieved only using the secret key.

4. Results and Discussion

This section describes the experimental outcomes recorded using the developed AFBS-WOA for privacy-conserved healthcare data publishing in cloud computing. Moreover, the simulation tool, database description, and comparative techniques are also described.

4.1. *Experimental Setup.* The experimentation of the developed AFBS-WOA-based privacy preservation model is implemented in Java with CloudSim tool, and the simulation requires PC, Intel I3 processor, 4 GB RAM with Windows 10 OS.

4.1.1. *Database Description.* The database utilized by the experimentation of the developed AFBS-WOA model is the heart disease dataset [20]. A total of three datasets are taken from the heart disease dataset, namely the Cleveland dataset, Hungarian dataset, and Switzerland dataset. The data size of the Cleveland dataset is 303 instances (rows) \times 14 attributes (columns), Hungarian dataset is 294

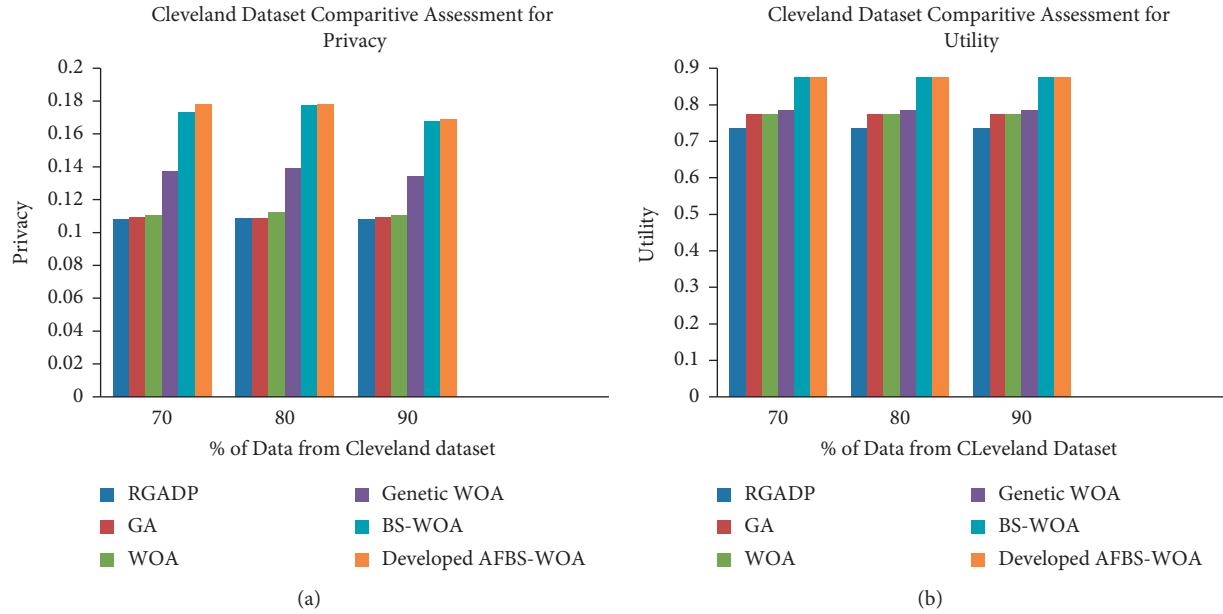


FIGURE 2: Comparative assessment using the Cleveland dataset using (a) privacy and (b) utility.

instances (rows) \times 14 attributes (columns), and Switzerland data size is 123 instances (rows) \times 14 attributes (columns).

4.1.2. Metrics for Evaluation. The evaluation metrics employed for the developed AFBS-WOA algorithm are privacy and utility measures. The privacy measure depends on the modification degree, while the utility measure depends on the numerical characteristics; hence, the value of the privacy and the utility should be obtained as high as possible. The explanation for privacy and utility is already given in equations (10) and (11), correspondingly.

4.1.3. Techniques for Comparison. The performance improvement of the developed AFBS-WOA algorithm is calculated by comparing the recorded outcome with the outcome of existing techniques, such as Retrievable General Additive Strategy Database (RGADB) [21], Genetic Algorithm (GA) [22], WOA [23], and the genetic-WOA and BS-WOA.

4.2. Comparative Assessment of Developed AFBS-WOA. The performance improvement of the developed AFBS-WOA model is assessed by varying the percentage of data using three datasets, Cleveland dataset, Hungarian dataset, and Switzerland dataset, based on privacy and utility measures.

4.2.1. Comparative Assessment Based on the Cleveland Dataset. This section describes the assessment of the developed AFBS-WOA model based on evaluation metrics using the Cleveland dataset. Figure 2(a) describes the comparative assessment of the developed AFBS-WOA model by varying the percentage of data based on the privacy

measure. When the percentage of data = 70, the privacy measure recorded by the existing methods, RGADP is 0.1092, GA is 0.1099, WOA is 0.1122, Genetic-WOA is 0.1231, BS-WOA is 0.1484, whereas the privacy of the developed model measured is 0.1971. Figure 2(b) shows the comparative assessment of the developed model based on utility by changing the data percentage. For the data percentage = 80, the developed model recorded the utility measure of 0.8741, whereas the existing techniques, like RGADB, GA, WOA, Genetic-WOA, and BS-WOA, recorded the utility measure of 0.734, 0.7741, 0.7742, 0.7842, and 0.8739, correspondingly.

4.2.2. Comparative Assessment Based on the Hungarian Dataset. This section deliberates the comparative assessment of the developed AFBS-WOA model based on evaluation metrics using Hungarian dataset. Figure 3(a) shows the graphical outcome of existing comparative methods with the developed model based on privacy by changing the data. When the data percentage = 90, the developed model obtained the privacy of 0.1872, and the existing techniques, such as RGADB, GA, WOA, Genetic-WOA, and BS-WOA, measured the privacy values of 0.1055, 0.1076, 0.1077, 0.1527, and 0.1715, respectively. Figure 3(b) depicts the graphical representation of comparative results in terms of utility by adjusting the data percentage. When the data percentage = 80, then the utility measured by the developed AFBS-WOA model is 0.8751, RGADP is 0.7353, GA is 0.7749, WOA is 0.7752, Genetic-WOA is 0.7853, and BS-WOA is 0.8751.

4.2.3. Comparative Assessment Based on the Switzerland Dataset. This section described the comparative discussion of the developed AFBS-WOA model based on evaluation metrics using the Switzerland dataset. Figure 3(a) shows the comparative assessment of the developed model based on



FIGURE 3: Comparative assessment using the Hungarian dataset based on (a) privacy and (b) utility.

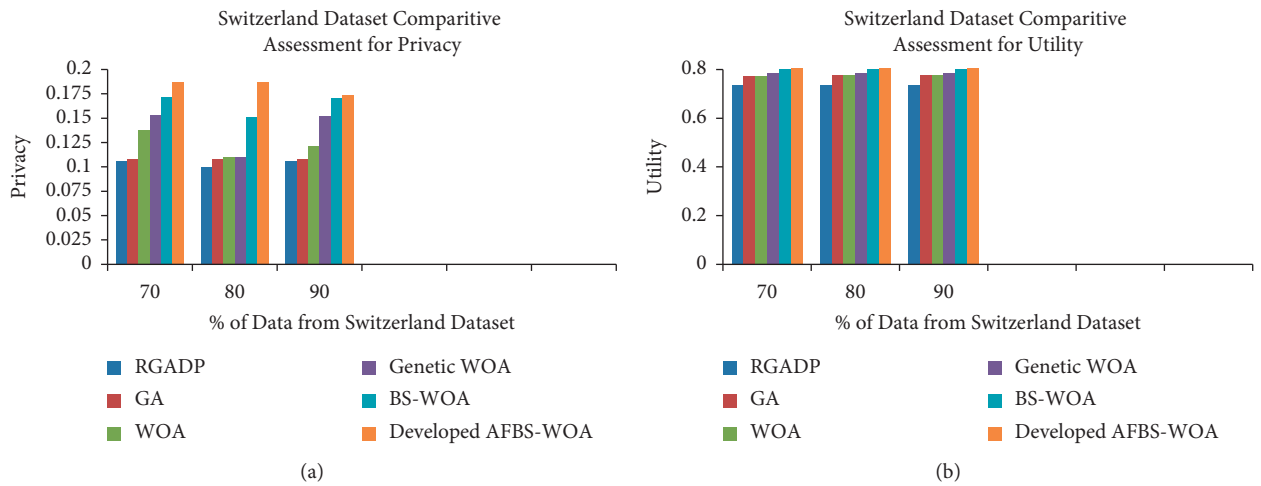


FIGURE 4: Comparative assessment using the Switzerland dataset based on (a) privacy and (b) utility.

privacy by changing the data percentage. For the data percentage = 80, the developed model recorded the privacy measure of 0.1836, whereas the existing techniques, like RGADP, GA, WOA, Genetic-WOA and BS-WOA, recorded the privacy measure of 0.1007, 0.1091, 0.1114, 0.1121, and 0.1462, correspondingly. Figure 4(b) describes the comparative assessment of the developed AFBS-WOA model by varying the percentage of data based on the utility measure. When the percentage of data = 70, the utility measure recorded by the existing models, like RGADP is 0.7235, GA is 0.7634, WOA is 0.7637, Genetic-WOA is 0.7731, and BS-WOA is 0.8628, whereas the developed model measured the utility of 0.8633.

4.3. Comparative Discussion. Table 2 describes the comparative discussion of the developed AFBS-WOA model for constructing the perturbation database. From the table, it is

clearly declared that the developed model attained the maximum privacy of 0.1872 and maximum utility of 0.8755, correspondingly. The existing methods, like RGADP, GA, WOA, Genetic-WOA and BS-WOA, attained the privacy of 0.1055, 0.1076, 0.1077, 0.1527, and 0.1715 and the utility of 0.7355, 0.7753, 0.7755, 0.7853, and 0.8752. The comparative analysis clearly shows that the proposed AFBS-WOA achieves the maximum privacy and utility parameters compared with the other discussed existing methods. Medical data are shared all over the world for research purposes to enhance the healthy environment. Moreover, this proposed method can be applied in healthcare organizations to protect the medical data shared through the cloud. The patient's data are stored and retrieved from the cloud in a secured way with more utility factor. The performance of the proposed system is implemented and compared using three different datasets with the fitness function.

TABLE 2: Comparative discussion.

| Database | Evaluation metrics | RGADP | GA | WOA | Genetic-WOA | BS-WOA | Developed AFBS-WOA |
|-------------|--------------------|--------|--------|---------|-------------|--------|--------------------|
| Cleveland | Privacy | 0.1081 | 0.1093 | 0.1101 | 0.1375 | 0.1735 | 0.1781 |
| | Utility | 0.7349 | 0.7746 | 0.7747 | 0.7849 | 0.8747 | 0.8749 |
| Hungarian | Privacy | 0.1055 | 0.1076 | 0.1077 | 0.1527 | 0.1715 | 0.1872 |
| | Utility | 0.7355 | 0.7753 | 0.7755 | 0.7853 | 0.8752 | 0.8755 |
| Switzerland | Privacy | 0.1090 | 0.1107 | 0.1178 | 0.1480 | 0.1662 | 0.1691 |
| | Utility | 0.7258 | 0.7650 | 0.76519 | 0.7750 | 0.8648 | 0.8657 |

The GGWO algorithm was proposed to enhance the security of the cloud data by employing the k-anonymization method to enhance the privacy policies of the stored data. The system evaluated has reduced fitness value with an increase in the number of iterations, which enhances the privacy of the system. Even though the system performs comparatively better than some of the other existing systems, the percentage of the information loss is comparatively high in this method. From the observations of our proposed system, we propose that it is likely to overcome the drawback of the GGWO with the maximum fitness value, privacy, and utility by enhancing the function using the proposed developed AFBS-WOA.

5. Conclusion

This article presents the developed AFBS-WOA model for generating the optimal key coefficient matrix. The developed AFBS-WOA method is formed by the combination of AFBSO and WOA methods in order to generate the optimal key coefficient matrix for privacy-preserved healthcare data publishing. The optimal key matrix selected by the developed AFBS-WOA method used utility and privacy measures for constructing the perturbation database in order to attain the privacy-conserved healthcare data publishing in cloud computing. Moreover, the secret key is shared with the service provider for retrieving the original perturbation database. The database from the data owner is multiplied with optimal key matrix produced by the developed AFBS-WOA model using the Tracy-Singh product to obtain the privacy-preserved healthcare database. The database retrieval is done by the key generated using the developed AFBS-WOA model. Moreover, the experimental result demonstrates that the developed AFBS-WOA model attained the maximum utility and privacy measure of 0.1872 and 0.8755 using the Hungarian dataset. In addition, the future enhancement of this research can be done by including some other effective optimization techniques for further improving the performance. In the future, the proposed method can be extended with more optimization algorithms on different datasets to enhance the GGWO system performance with various datasets.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J.-J. Yang, J.-Q. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," *Future Generation Computer Systems*, vol. 43-44, pp. 74–86, 2015.
- [2] A. George and A. Sumathi, "Dyadic product and crow lion algorithm based coefficient generation for privacy protection on cloud," *Cluster Computing*, vol. 22, no. 1, pp. 1277–1288, 2019.
- [3] J. Herranz, J. Nin, P. Rodríguez, and T. Tassa, "Revisiting distance-based record linkage for privacy-preserving release of statistical datasets," *Data & Knowledge Engineering*, vol. 100, pp. 78–93, 2015.
- [4] L. M. Kaufman, "Data security in the world of cloud computing," *IEEE Security & Privacy Magazine*, vol. 7, no. 4, pp. 61–64, 2009.
- [5] J. V. B. Benifa and G. V. Mini, "Privacy based data publishing model for cloud computing environment," *Wireless Personal Communications*, vol. 113, no. 4, pp. 2215–2241, 2020.
- [6] A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data," *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 4, pp. 426–435, 2019.
- [7] S. A. Yousra and S. Mazleena, "A new heuristic anonymization technique for privacy preserved datasets publication on cloud computing," *Journal of Physics: Conference Series*, vol. 1003, no. 1, Article ID 012030, 2018.
- [8] P. VijayaKumar, P. Pandiaraja, M. Karuppiah, and L. J. Deborah, "An efficient secure communication for healthcare system using wearable devices," *Computers & Electrical Engineering*, vol. 63, pp. 232–245, 2017.
- [9] P. Vijayakumar, S. M. Ganesh, L. J. Deborah, and B. S. Rawal, "A new SmartSMS protocol for secure SMS communication in m-health environment," *Computers & Electrical Engineering*, vol. 65, pp. 265–281, 2018.
- [10] T. Zhou, J. Shen, D. He, P. Vijayakumar, and N. Kumar, "Human-in-the-Loop-Aided privacy-preserving scheme for smart healthcare," *IEEE Transaction on Emerging Topics in Computational Intelligence*, pp. 1–10, 2020.
- [11] Y. Zhang, D. He, M. S. Obaidat, P. Vijayakumar, and K.-F. Hsiao, "Efficient identity-based distributed decryption scheme for electronic personal health record sharing system," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 384–395, 2021.
- [12] Y. Liu, J. Yu, J. Fan, P. Vijayakumar, and V. Chang, "Achieving privacy-preserving DSSE for intelligent IoT healthcare system," *IEEE Transactions on Industrial Informatics*, 2021.
- [13] Y. Qian, J. Shen, P. Vijayakumar, and P. K. Sharma, "Profile matching for IoMT: a verifiable private," *Set Intersection Scheme*, vol. 25, no. 10, pp. 3794–3803, 2021.
- [14] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and PUF-based lightweight Authentication protocol for wireless medical

- sensor networks,” *IEEE Internet of Things Journal*, vol. 14, no. 8, pp. 2327–4662, 2015.
- [15] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, “Blockchain-based reliable and efficient certificateless signature for IIoT devices,” *IEEE Transactions on Industrial Informatics*, pp. 1551–3203, 2021.
- [16] S. Liu and T. Li, “A new hypernetwork model based on matrix operation,” in *Proceedings of the 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 176–182, Taipei, Taiwan, China, November 2015.
- [17] A. N. Langville and W. J. Stewart, “The Kronecker product and stochastic automata networks,” *Journal of Computational and Applied Mathematics*, vol. 167, no. 2, pp. 429–447, 2004.
- [18] P. Yadav, “Case retrieval algorithm using similarity measure and fractional brain Storm optimization for health informaticians,” *The International Arab Journal of Information Technology*, vol. 16, no. 2, 2019.
- [19] M. Kazim and S. Y. Zhu, “A survey on top security threats in cloud computing,” (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 3, 2015.
- [20] Archive.ics.uci.edu, “The heart disease dataset taken from,” 2021, <https://archive.ics.uci.edu/ml/datasets/heart+disease>.
- [21] P. Yang, X. Gui, J. An, J. Yao, J. Lin, and F. Tian, “A retrievable data perturbation method used in privacy-preserving in cloud computing,” *China Communications*, vol. 11, no. 8, pp. 73–84, 2014.
- [22] M. Qiu, Z. Ming, J. Li, K. Gai, and Z. Zong, “Phase-change memory optimization for green cloud with genetic algorithm,” *IEEE Transactions on Computers*, vol. 64, no. 12, pp. 3528–3540, 2015.
- [23] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

Research Article

A Personalized Eccentric Cyber-Physical System Architecture for Smart Healthcare

Amutha Balakrishnan,¹ Ramana Kadiyala ,² Gaurav Dhiman ,³ Gokul Ashok,¹ Sandeep Kautish ,⁴ Kusum Yadav,⁵ and J. Maruthi Nagendra Prasad ⁶

¹*School of Computing, SRM Institute of Science and Technology, Chennai, India*

²*Department of Artificial Intelligence & Data Science, Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh, India*

³*Department of Computer Science, Government Bikram College of Commerce, Patiala, India*

⁴*Dean-Academics with LBEF Campus, Kathmandu, Nepal*

⁵*College of Computer Science and Engineering, University of Ha'il, Ha'il, Saudi Arabia*

⁶*Department of Computer Science and Engineering, Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh, India*

Correspondence should be addressed to Sandeep Kautish; sandeep.kautish@lbef.edu.np

Received 27 September 2021; Revised 28 October 2021; Accepted 2 November 2021; Published 17 December 2021

Academic Editor: Thippa Reddy G

Copyright © 2021 Amutha Balakrishnan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The development and technological advancement of wireless sensor networks in different fields has been a revolution for mankind. To meet the high-end requirements, the support of the cloud that provides the resources for the application is very much essential. This paper presents an architecture called cloud sense to connect cyber and physical spaces for wireless body area networks with varying high-end workflow at different perspectives. The scalability issue in collecting patient data and processing the data is established using ganglia that is a scalable, distributed monitoring system to support high-performance computing in clusters for the set of input events such as electrocardiogram (ECG), blood pressure (BP), saturation of peripheral oxygen (SPO₂), temperature, and skin conductance of the kind of human body parameters. Various parameter metrics have been analyzed based on the equivalent creation of instances. The connectivity mechanism behind the proposed cyber-physical system is unique of its kind; it is exhibited through wireless Internet on a small scale of three remote locations; the system works well with specific network parameter metrics; and the results proved that availability and scalability issues were addressed with numerical analysis.

1. Introduction

A cyber-physical system (CPS) is required to interconnect the physical devices in the hospital for healthcare monitoring and to analyze the data stored in the cloud. Besides, an analytics platform through the internet of things (IoT) is the need of the hour for efficient healthcare delivery in the world. The proposed CPS will act as an interface between the physical and cyber worlds. Physical world comprises the body sensors and the electronic devices that can be interconnected together to form the physical space. Cyberspace consists of the data where it can be transferred to the doctors

and the researchers to analyze and make decisions for further need. The CPS that is proposed using the smart health application was operational under the following categories: (1) patient-centric, (2) network-centric, (3) hardware-centric, and (4) data-centric. CPS will function as an intelligent monitoring platform for timely diagnostic decisions in the critical care unit of hospitals and home care patients. This research lies in blending recent cutting-edge technologies that are wireless body area network (WBAN), IoT, and streaming big data analytics to handle an enormous amount of data. The concept of fog computing has been introduced to save energy and time to provide timely and

needy services at the doorstep of the patient. Also, an algorithm for disseminating the patients' health parameters using a priority queuing mechanism is proposed. By introducing fog, data analytics is possible in the terminals themselves, which again improves adeptness in the proper functioning of the system with localized decision-making, the geographical distribution of data within less time, and optimized usage of resources. Fog computing enables people to collect data from various devices and has a larger capacity to process more data than edge computing, whereas edge computing performs much of the processing on embedded computing platforms kept with the patients in a WBAN system as it is directly interfaced with the sensors and controllers.

The proposed system will reduce latency, improve operational efficiency, and will provide effective service to save human life using built-in decision-making policies by the introduction of fog controllers, which are used for effective data dissemination locally with reduced time complexity.

1.1. Problem Definition. The research tries to interconnect the recent cutting-edge technologies that are wireless body area network, internet of things (IoT), and streaming big data analytics. The concept of fog computing has been introduced to save energy and time to provide timely and needy services nearer to the doorstep of the patient regardless of the location of the patient. Hospitals, doctors, and patients are interconnected through local and remote servers, through the fog controllers into the cloud. Also, an algorithm for disseminating the patients' health parameters using a priority queuing mechanism is proposed. By introducing fog, data analytics is possible in the end terminal itself, which again improves adeptness in the proper functioning of the system with localized decision-making, the geographical distribution of data in lesser time with optimized usage of resources. When a massive amount of data needs decision-making, scalability and reliability issues have been solved by the concept of availability.

2. Literature Review

Future global deployment of WSNs could provide data in petabytes or exabytes every year for, for example, environmental monitoring. Whether the related cloud environment model is appropriate for the processing of sensor information is, however, far from clear [1]. The next-generation network sensor platforms should aim for a multi-application model of popular infrastructure with a strong separation of concerns between infrastructure providers and application developers. The WBAN ecosystem can be applied to the cloud, and infrastructure as a service (IaaS) providers such as Amazon EC2 or Eucalyptus can provide an infrastructure for healthcare [2]. The WBAN networks provide a way to capture physiological data for use in several distributed applications. To provide end-to-end physiologic monitoring and diagnosis [3], Amazon EC2 can respond to the complex needs of medical services and can be incorporated with wireless body sensor network technologies.

However, the large amount of data stored in the cloud is easily measured for performance in real time. MapReduce is the most prevalent cloud computing programming model [2]. It is the programming model for the processing and generation of large data sets. In the MapReduce model, several real-world activities are expressible. Functional-style programmes are automatically paralleled and run on a wide variety of commodity machines. The run-time framework provides descriptions of the partitioning of input data, coordinating the execution of the programme across several machines, handling machine failures, and maintaining contact between the machines [3]. The WBAN data can be fit into a model by updating the model parameters that enables two or more databases to appear as one, whether on premises or in the cloud. Teradata QueryGrid, IBM PureData Systems with Fluid Query, and SAP HANA that work with smart data services are offering data federation capabilities. The model itself is sufficient for the physicians to go for a decision without affecting the underlying WBAN data.

Previous research in the health sector centered on creating the prototype of the body area network with wireless sensors for the use of routing protocols. Nodes in the body sensor were used to detect critical human parameters such as ECG, blood pressure, level of oxygen, heart rate, and body temperature. The paper "Alerts for mobile healthcare: criteria and pilot studies" provides efficient routing and tracking of alerts to quality and cost-effective health facilities [4]. In their paper, Lee et al. proposed that high blood pressure and arrhythmia can be effectively avoided and regulated by continuous physiological surveillance [5]. Previously, a smart, mobile care system focused on roles with an alerting mechanism was proposed and implemented [2]. For further study, variations of human body parameter values in various patients are reported when standard human body parameters are retained as median values. An algorithm has been developed to classify human values anomalies leading to the diagnosis of disease and medicines [6]. Several values of human body parameters were collected and translated into unique data packets for doctoral evaluation and wirelessly forwarded to a hospital server. When the observed human body parameters were greater than the threshold values, an alert message was suggested for the caregiver assigned to the patient with encrypted contact [7]. The ZigBee system for fall tracking, incorporated through drop detection, indoor positioning, and ECG monitoring, offered insights into a secure transmission protocol based on anycast routing for the wireless patient surveillance process [8]. WBAN writers performed a report on medical and nonmedical uses. It offers a great deal of insight into the applications. The medium access protocol has been revised to collect patient data for context awareness purposes [9]. The authors in [10] proposed a knowledge interview mechanism for globally accessing patient data that was clarified in the cloud-based wireless body sensor network. In [11], the authors made a profound survey study of wireless body area network architecture problems.

In [12], the authors suggested a fuzzy logic application to diagnosis of anemia for expert fuzzy system presentation. It

has an expert system based on an inflammatory system that was only considered for the diagnosis of amnesia and other body parameters. The authors discussed a 1 kg rise in body weight, correlated with the systemic increase in blood pressure between 3 and 6 mm Hg in the healthy and ingravescens classes [13]. Soon, telehealth initiatives are also underway. Researchers also discussed the role of an intelligent mobile care system with a warning mechanism in the chronic care climate. The device tracks patients' medical records and sends rapid avoidance measures to avoid unexpected promises [14]. A structure for collaborative software agents has been presented by separate software agents with three main components: information management, the reasoning for confusion, and software agents [15]. The framework for the structured handling of warning messages was created as an alert monitor that complies with the needs of the medical staff or its mobile devices for receiving warnings within a specific deadline [16].

The authors' findings suggested that the cloud-integrated sensor data provides a special hybrid platform for remote health surveillance [5]. The goal was very well defined to provide valuable insight for the designers of WBANs and highlight key problems concerning the efficiency of the collection of healthcare data [17]. An overview of the computer environment has been studied for the collection of personal data from remote mobile patients [18]. The authors proposed a smart health solution through the use of a clustering mechanism for wireless sensor networks [19]. Mobile ZigBee and Bluetooth health gateways have been briefly analyzed [20]. An integrated gateway for various PHDs was implemented to collect measurements from different PHDs. It functions in two modes, namely, immediate transmission and integrated transmission. The overhead transmission can be minimized by the gateway consisting of an activity monitor, a drug dispenser, and a pulse oximeter [21]. For remote patient monitoring applications as a trial, an intelligent smart health portal with fog was initialized [22].

As IoT leads to an exponential proliferation of endpoint systems, fog computing is known to expand the hierarchically distributed architecture from the edge of the network to the heart. In addition to big data and analytics, IoT introduces a new dimension to its wide distribution of sources [23]. In the field of healthcare applications, the basic computing materials of fog were treated where data can be moved without delay [24]. The technology acceptance model (TAM) has been developed and has shown the difference in health conditions between adoption factors because of the advancement of medical technologies and their perceived ease of use [25]. A research by Megalingam et al. [18] suggested a portable system to give warnings to the caretakers. The researchers addressed the development of virtual group enablers (VGE) between patient, nurse, and doctor devices to allow the remote analysis of WBAN data. The study involves GMS, the medical data recording server (MDRS), the policy engine (PE) and medical officers' equipment, WBANs for patients, and environmental sensors. Group preparation and adjustment have been undertaken depending on the circumstances and needs of

patients and medical officers, which can be easily modified by high-level policies. Medical officers provide input on the consistency of the obtained WBAN data by using quality of health monitoring [26]. Authors have suggested a secure transmission protocol based on all cast routing for wireless patient monitoring, which automatically selects the nearest recipient data in anycast category to minimize millisecond latency and control overhead [27]. The above literature did not concentrate on dispersed needy service without latency at the appropriate geographical locations, whereas the proposed project focuses mainly on managing loads by utilizing resources properly, providing geographically distributed customer needs with minimal latency and maximizing resource usage from nearest points of interest, during the sequence.

So there is a need for a scalable architecture for healthcare as a case study with many numbers of instances created using virtual machines in a cloud computing environment to interconnect patients, doctors, and hospitals geographically. The paper is organized as follows. Section 3 represents the five-tier methodology and subsection details. Section 4 reveals the cyber-physical system components. Section 5 presents the mathematical modeling of five vital parameters of WBAN. Section 7 presents the concluding remarks with a case study after Section 6 with performance metrics.

3. Proposed System

The proposed system will reduce latency, improve operational efficiency, and will provide effective service to save human life using built-in decision-making policies in the observed WBAN data. Remote healthcare through fog computing is one of the new approaches that can handle some of the challenges of smart healthcare in terms of localized decision-making, geographical distribution of data, and smart load balancing with security, sharing, integration, and management. In the proposed work, we implemented the significance and opportunities of fog essentials to reduce the tasks offered by cloud computing is pervasive in healthcare's future challenges it faces as of today. The proposed architecture consists of five tiers as shown in Figure 1. The methodology in the context of tier-wise software, hardware, and the proposed algorithms is described in the consecutive flow charts. Figure 1 indicates the proposed architecture to fuse cyber and physical phases in the domain of wireless healthcare in terms of five tiers. When the computational needs of individual tasks are high, the workflow is categorized as calculation-intensive. Similarly, when data specifications are fantastic (e.g., size of and data file, number of files, data storage, etc.), the workflow is categorized as data-intensive. Data-intensive workflows may use the architecture of environments such as data clouds. Data clouds offer services such as low-latency transportation protocols and reproduction mechanisms for data delivery, for which massive data sets stored in distributed repositories need to be accessed, processed, and transmitted. The period depends on the time you spend dealing with the input and output files and the time you compute them.

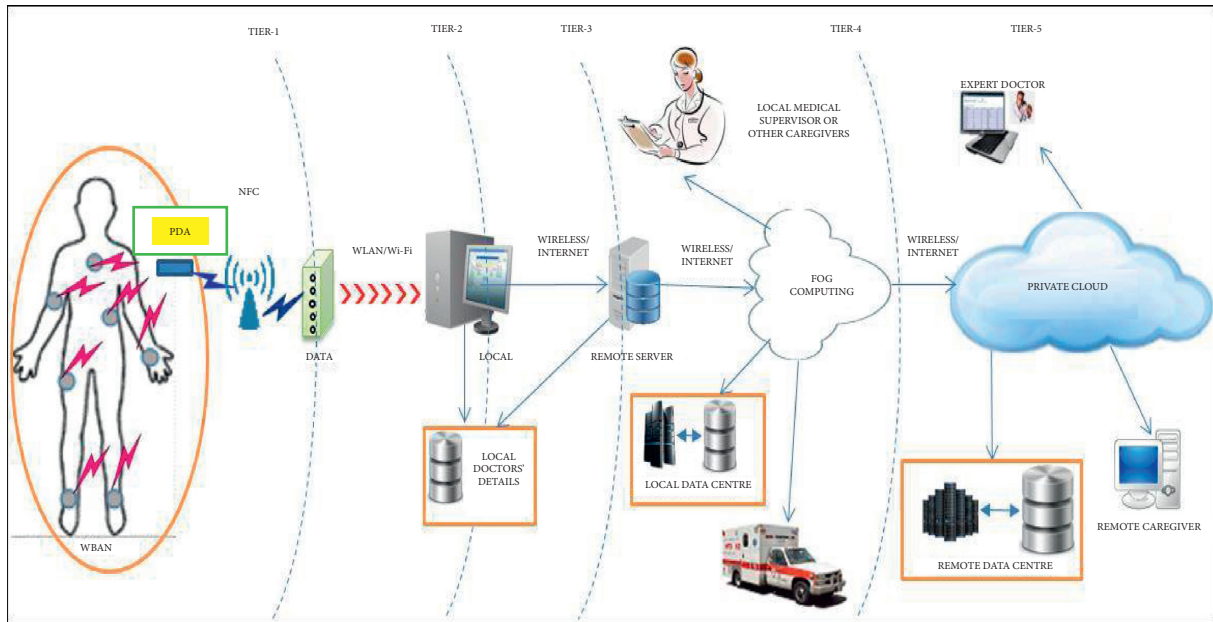


FIGURE 1: Proposed five-tier architecture for remote medical care fusing cyber and physical spaces.

In a traditional healthcare monitoring system, we rely on human intervention along with physical devices for observing patient information periodically and consistently. The internet of things (IoT) is a connecting system, such as electronic devices, buildings, and even more medical centers and hospitals, in which full access to all patient information is communicated at needy times with assured data through the Internet. To assist in the healthcare modernization process, WBAN along with IoT devices connected to the Internet plays a major role. The human healthcare framework is a patient-resource-based monitoring system that includes informational, audiovisual coordination, and the retrieval of health data through the sensors. The integration of IoT in healthcare monitoring systems is cumbersome because of the large amounts of data and the need for secured data transfer to protect patient's personal medical information from being seeped [28]. Any malicious person's intervention or stealing of data and altering or modifying the patient data for any unwanted and unethical purposes lead to data violation, and manipulation of patient's vital data in any way may have serious implications; even it may lead to the death of the patient [29]. Therefore, building a definite architecture for a wireless body area network with varying workflow is considered with location information as an application domain. The scalability problem is addressed by suitable algorithms when data flow becomes enormous. The virtualization concept is applied to address when a large handling issue occurs in the cloud environment.

3.1. Tier I: Data Collection from WBAN to PDA. The medical data collected from the body sensor network comprises ECG, SPO_2 , pulse rate, temperature, skin conductance, and blood pressure. A data collector that is built within the microcontroller unit collects all the six-sensor data and processes the data based on prioritization. The body sensor

data is prioritized as normal, abnormal, and critical. A personal digital assistant (PDA) is responsible for collecting the data from the sensors using near-field communication (NFC). NFC, Bluetooth, and ZigBee are provided with information on chip vendors and application product vendors' deployment in smart healthcare services. Based on the availability, we can go for the communicating device. The algorithm resides in the controller to check the incoming data with a normal human body data set. If there is no data from the sensors for a specified period of an interval, the loop re-executes to access the data from the body sensors with a significant waiting time as represented in Figure 2.

3.2. Tier II: Routing Data from PDA to Local Server. The PDA checks the integrity of the incoming data with the previous data. The data collected from the respective PDAs was routed to the local server. The PDA ID and the doctor ID are mapped according to the specialization, using a local doctor's database available in the local server. The streaming analytics engine continuously monitors the parameters for critical and abnormal patients. The summary of parameters was sent to the local doctor and the emergency response team in the hospital. The mapping is stored and could be retrieved through a web page, using the categorized status of normal, abnormal, and critical as represented in Figure 3.

3.3. Tier III: Routing Data from Local Server to Remote Server. The mapped information was routed from the local servers at a different location to the specified remote server. The remote server has a database that consists of a cluster of expert physicians to be referred across different hospitals and geographical locations for needy patients. It works on a cluster binding algorithm, which binds the patient information to the appropriate doctor from the cluster while

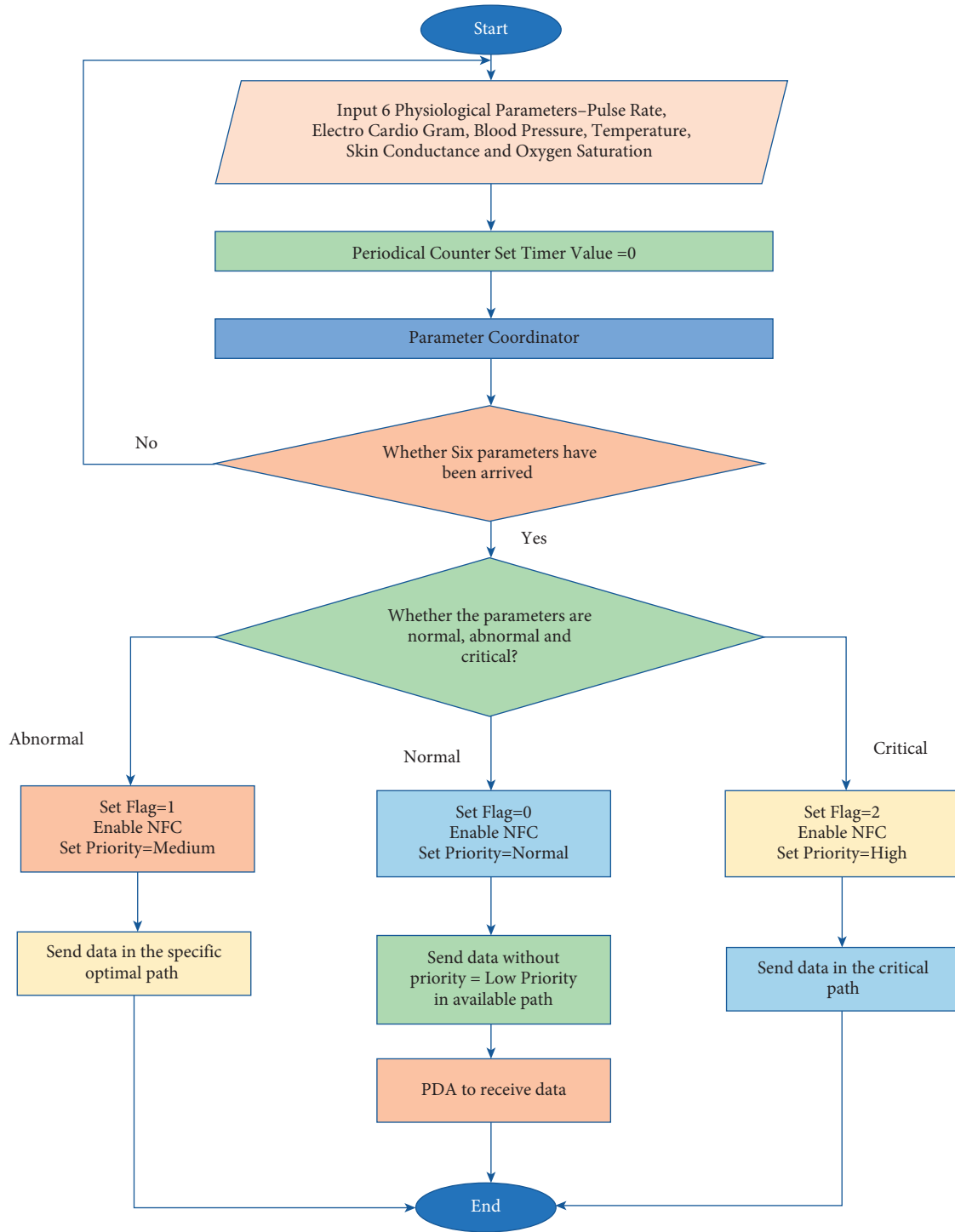


FIGURE 2: TIER I: from medical kit to PDA; mode of communication: NFC.

operating and gathering the data. This is an improved system where clustering has been centralized for collecting and managing data depending on the network parameters and the availability of appropriate doctors. This acts as an aid in the remote server to view the mapped information. The previous mapping of PDA with a doctor in Tier II is done for doctors in the same location whereas the binding in the remote server does the mapping with an additional doctor in the nearest neighboring locations as represented in Figure 4.

Figure 2 emphasizes collecting the six physiological parameters and verifies whether all the six vital parameters of the patient have been received. An abnormality table is constructed in such a way that, whether they received six vital parameters fall within the threshold, say a safer health status, assuming normal and if the parameters go behind the threshold slightly lesser or higher, it can be treated as abnormal and if the parameters shoot up and show turbulence and are also lesser or higher beyond the threshold, it is

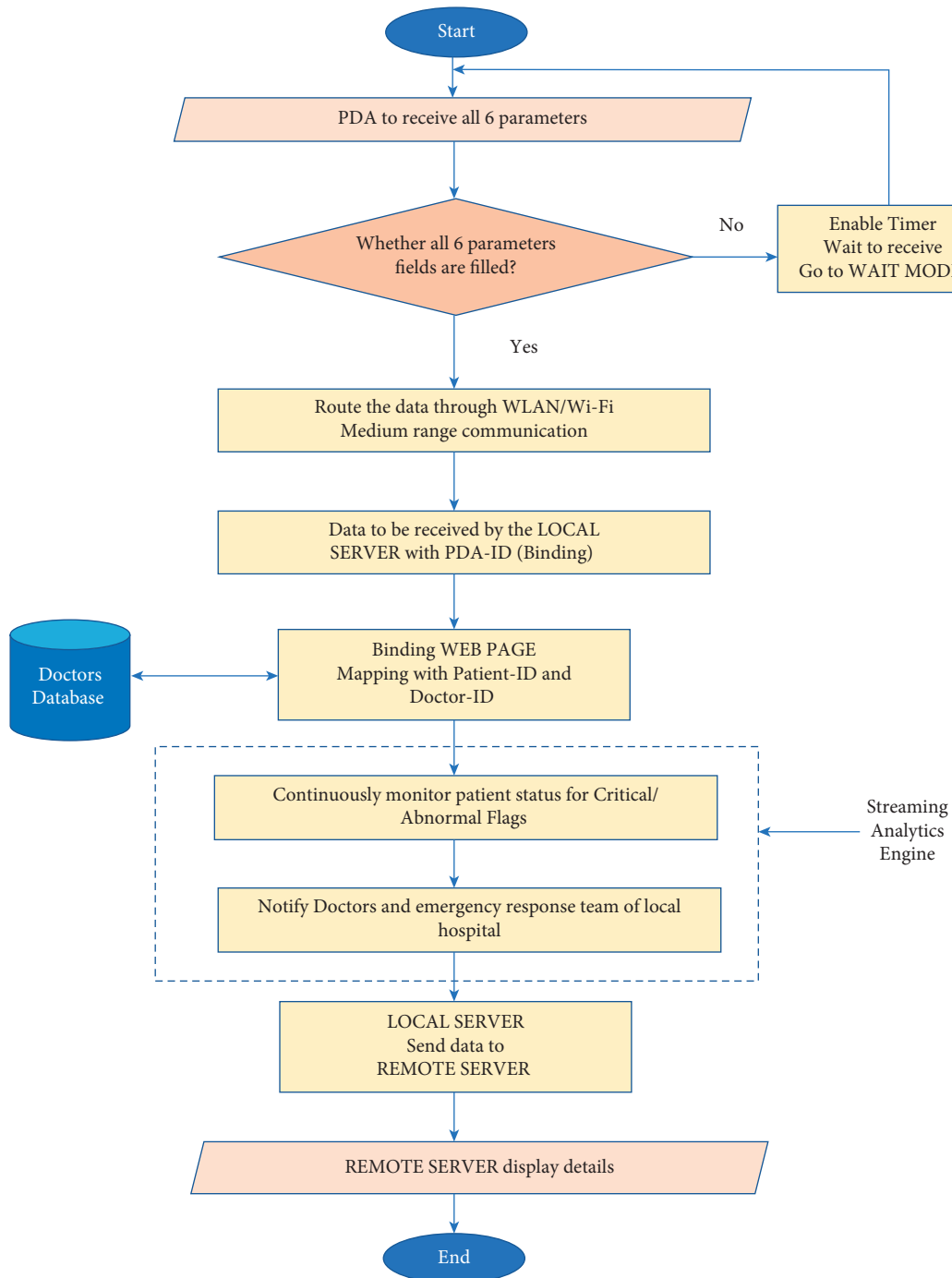


FIGURE 3: TIER II: from PDA to local server; mode of communication: wireless internet.

subsumed as critical after the comparison of the benchmarked data set.

Figure 3 prioritizes the routing of data after receiving from the PDA, and a mapping is done with respect to patient and physician. A streaming engine starts monitoring and flowing the status of the patient data continuously.

3.4. Tier IV: Routing Data from Remote Server to Fog Controller. The PDA-doctor mapped information from the

remote server is categorized according to the status (normal, abnormal, or critical). Normal data is directly stored in the database. The fog controller has a decision time interval configured for abnormal and critical states. The abnormal and critical data from the corresponding PDA is continuously sent to the appropriate hospitals and expert doctors. For the critical data, the patient status and the location information were sent to the caregivers. Fog controller does three jobs that are the geographical distribution of patient data with less latency, localized decision mapping, and

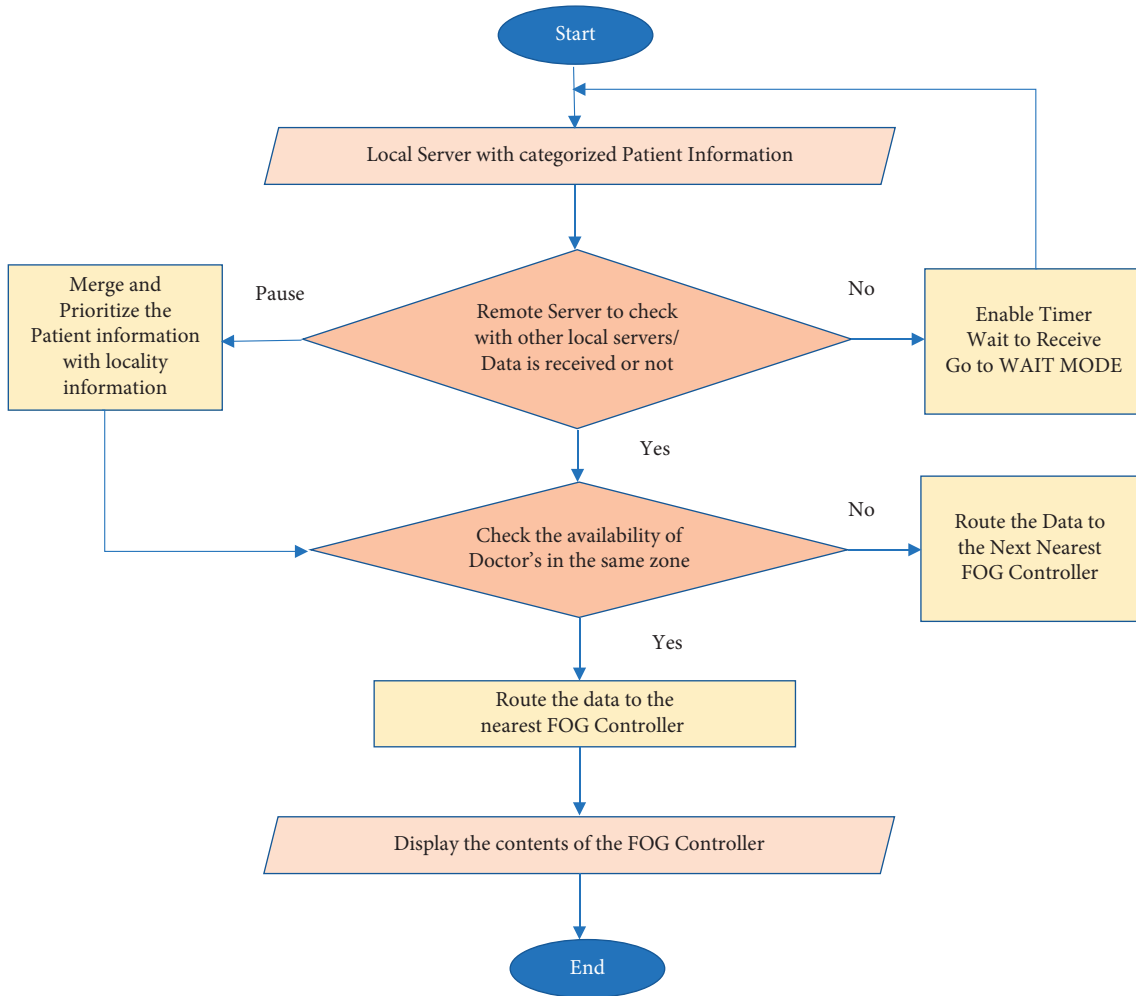


FIGURE 4: TIER III: from local server to remote server; mode of communication: wireless internet.

appropriate load balancing based on the available time of the doctor in the nearest location. If the expertise physician is engaged with some critical tasks, the next available physician in the nearest location has to be referred by the fog controller in an optimized way with reduced latency as the time required for providing the healthcare services is less compared to the cloud services as represented in Figure 5.

3.5. Tier V: Routing Data from Fog Controller to Private Cloud. The data received by the fog is routed towards the private cloud as represented in Figure 6 after the specific tasks have been completed.

- (1) Categorize patient data rendering to the locality of the server
- (2) Categorize patient data conferring to the criticality of data
- (3) Categorize patient data bestowing to the availability of doctor
- (4) Correlate the patient parameters per flag values
- (5) Notify the respective hospital or doctor based on criticality and availability.

- (6) Immediate action to both yes or no cases as per flag values so as the patient can get the medical attention without time lag

Figure 7 presents the timely remote medical diagnosis system proposed with fog essentials, fog reduces the tasks of the cloud and takes local decisions then and there and provides the required data to the needy patients regardless of the geographic boundaries globally.

An enhanced hospital management system swearing WBAN with the patient in an autonomous and also in an emergency situation is represented in Figure 7. The fog controller mechanism has been introduced in the architecture of CPS. The WBAN circuit that is worn by the patient is received by the personal data collector. All the heterogeneous data collected from different sensors in different units were digitized. The digitized data is made as a data packet with additional security code, header, and other priority relevant data that is used to emphasize the urgency of the patient to be given preference over the other continuously monitoring patients. This data sequence was collected by the local server in the same hospital. Simultaneously, the same data was sent to the remote servers in the adjacent localities based on the location of the current and

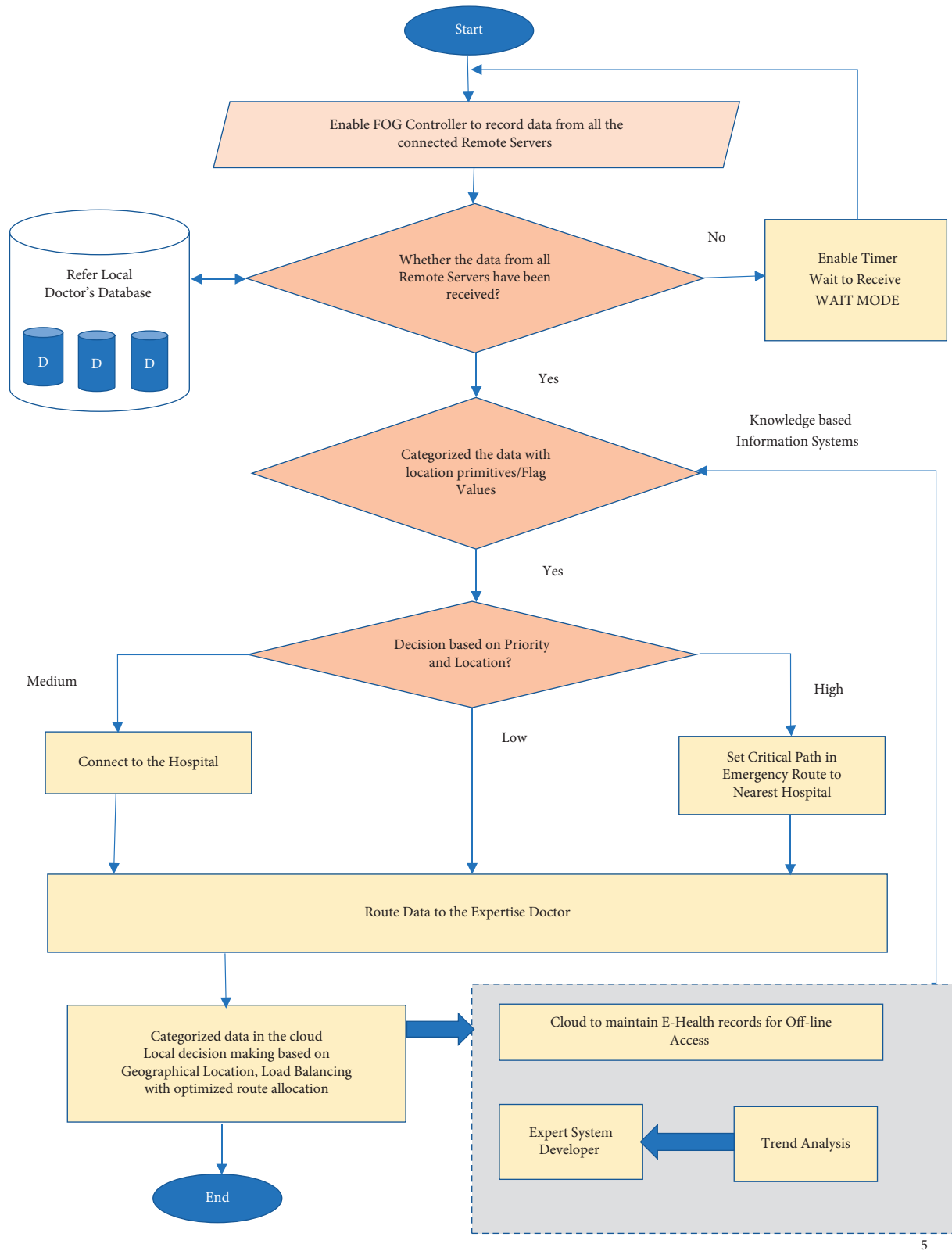


FIGURE 5: TIER IV: from remote server to fog controller; mode of communication: wireless internet.

adjacent hospitals and the readiness of doctors to attend to the patient. The remote servers were available in a sufficient number of hospitals and not in all the hospitals where all the

local servers operate. Then the data from the remote server travels via the fog controller where the decision-making is done based on geographical distances and the emergency of

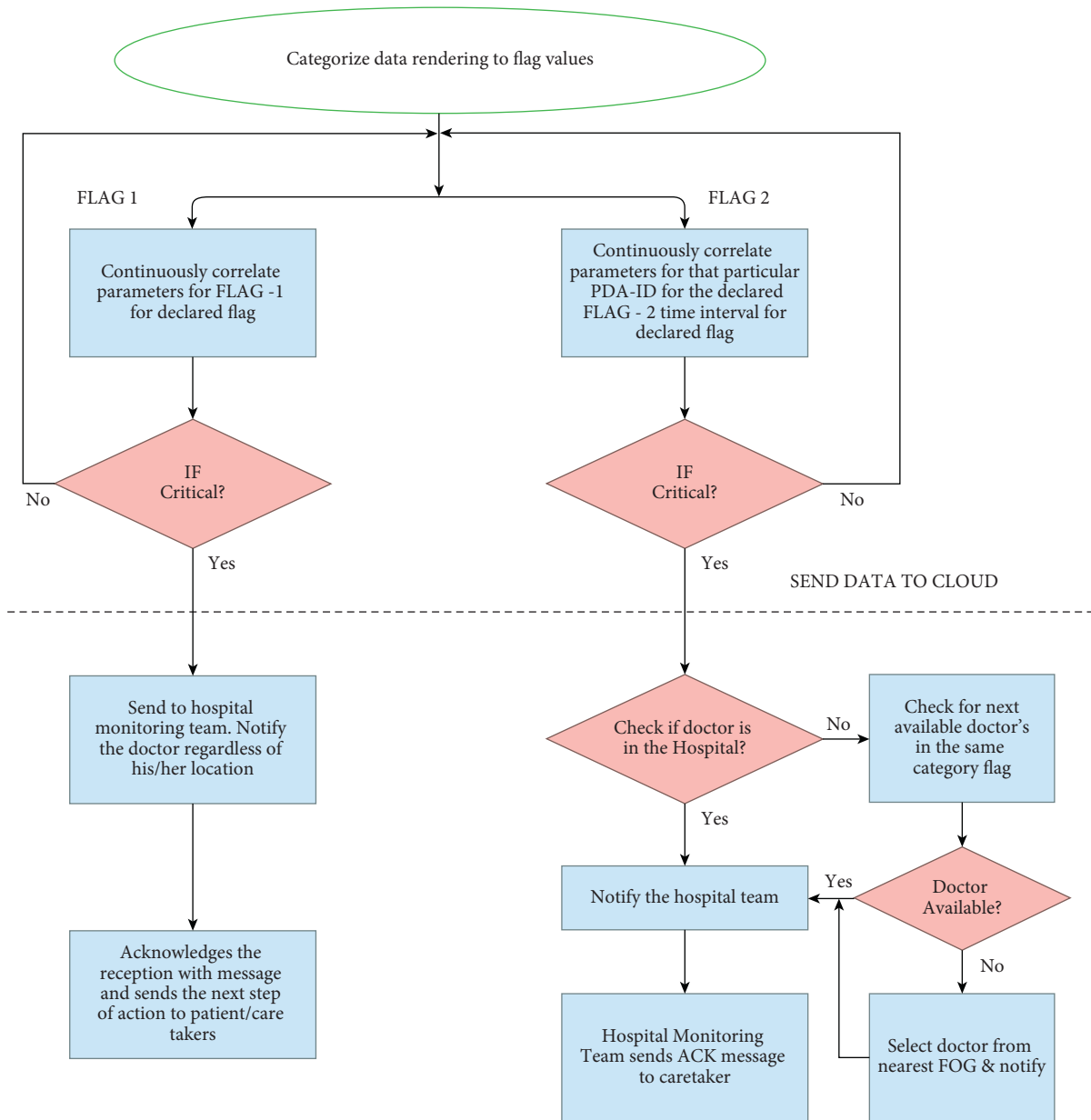


FIGURE 6: TIER V: from remote server via fog controller to cloud; mode of communication: wireless internet.

the patient data along with the availability of the physicians in the nearby hospitals. Finally, the data reaches the cloud.

The characteristics of fog have been defined as low latency and location awareness, widespread geographical distribution, network mobility with a large number of interoperable nodes, and predominant wireless access with real-time streaming capability with heterogeneous applications [30, 31]. The requirement of a smart hospital for providing accurate and timely healthcare architecture using CPS was proposed by the authors of [32, 33]. The concept of the smart city involving multiple disciplines like smart community, smart transportation, smart healthcare, and smart parking was proposed with the proposed architecture [34].

Some reliable transport layer protocols that provide end-to-end reliability of data transmitted in healthcare wireless sensor networks, and the advantages and disadvantages of

MAC, routing, and transport layer protocols have been proposed by the authors of [35, 36]. The paper reviews the existing schemes on security solutions in wireless healthcare scenarios with a summary of open security research issues that need to be explored for future healthcare applications using WMSNs [37, 38].

The authors have proposed software architecture for information sharing and collaboration that dealt with analysis, modeling, and development. It also elucidates the interoperability in sharing the medical records with the semantic details [39]. The authors proposed an IoT-based new semantic interoperability model (IoT-SIM) to provide semantic interoperability among heterogeneous IoT devices in the domain of healthcare. The doctors communicate to their patients with IoT devices to monitor the patients' current health conditions, and the information between the

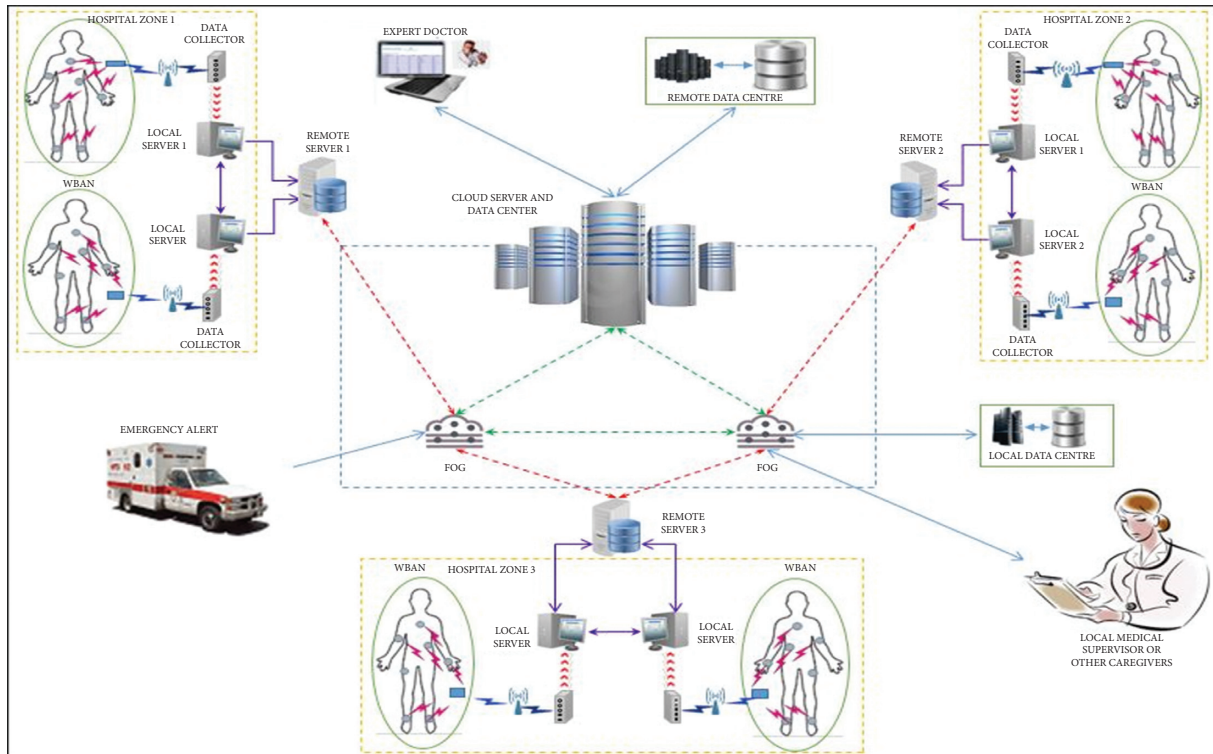


FIGURE 7: WBAN with fog essentials swarming internet of things.

doctors and the patient is semantically annotated and communicated. The semantic web technologies provide the tools that allow to process data more effectively and accurately, create the framework for interoperability between HS, and also integrate data from various sources with their semantic meaning [40]. The doctor allocation process from Tiers II–V is illustrated in Figure 8.

4. An Application Framework for Cyber-Physical System with Ganglia

To investigate system load and connectivity issues, we use the Ganglia monitoring tool. Every instance to be monitored should run the ganglia monitoring daemon (gmond). Aggregated data instances should contain additional packages. There is only one configuration file per cluster, and the configuration file does not change as long as the server is active. New instances are discovered immediately and terminated instances are forgotten within 90 seconds.

Figure 9 indicates the architecture for the system, which fuses the physical and cyberspaces. Whenever a new instance comes up, the monitoring tool will discover all instances in its cluster and send metrics to them. A newly launched instance is discovered and added to a cluster immediately. It will also do a rediscovery every 90 seconds so that instances that have been terminated are removed from its list of destinations as shown in Figure 9.

4.1. Cyber-Physical System. Cyber-physical system is a driving force for interconnecting the physical and cyber

world. It is the basis for connecting wireless sensors to the cloud even. There is plenty of scope for WSN acting as a wireless body area network. The term cyberspace is used for managing physical spaces. But it is applied to the virtual space that is created within the Internet. Cyberspace is more than a symbolic and figurative space that exists on the Internet. As cyber-physical systems are developed with the products, sensors, equipment, systems, hardware, software, and API, they can bring ubiquity everywhere in the world.

A Sensor-Cloud can be one of the pervasive applications using sensors as an interface between the physical and cyber world.

Sensor-Cloud is a part of cloud computing that uses the physical sensors for different sensing modalities to accumulate its data and transmit all the sensor data into a cloud computing infrastructure. Sensor-Cloud handles the sensor data efficiently, which is used for many monitoring applications. We use Sensor-Cloud for the WBAN application. The following points specify the necessity of a Sensor-Cloud:

- (1) To acquire the data from sensors, sensor modeling language can be used. The metadata is very important, as sensor data without location is meaningless in a sensor-centric world.
- (2) The Sensor-Cloud architecture provides instances as created from the event-driven mechanism that is devised as virtual sensors, which are going to travel in the network based on the requirement and the emergency of the patient's data.
- (3) These virtual sensors can be put into the cluster CPU, cluster memory, cluster network, cluster process,

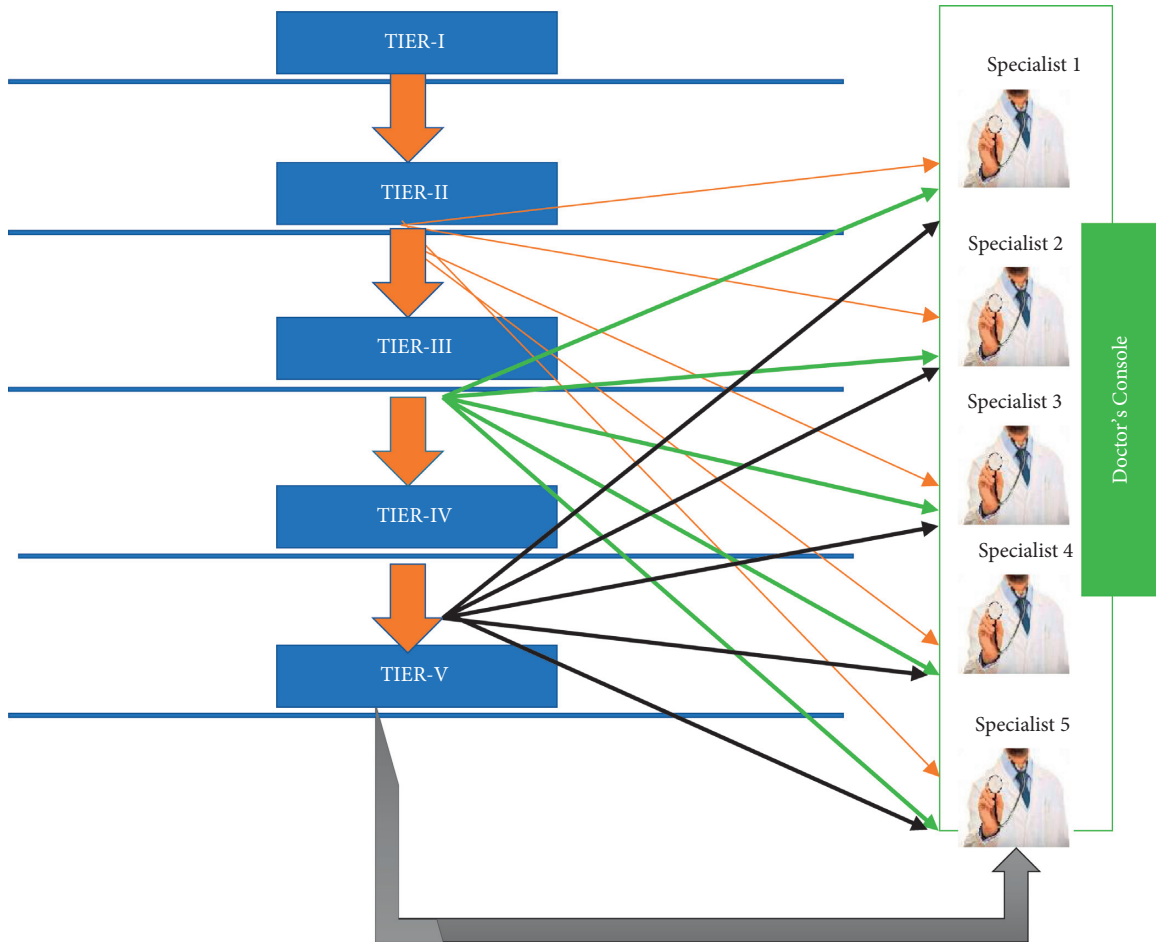


FIGURE 8: Doctor allocation process.

cluster storage, and so on. These service instances can be used through the graphical user interface.

- (4) Once the service instances become useless, they can be deleted by the users to avoid changes.
- (5) Automation of service played a crucial role in providing cloud computing sessions.
- (6) The performance metrics of the cloud can be categorized as efficiency, flexibility to avail cloud services, the delivery time of the service with tacit, and explicit knowledge.
- (7) Using the instances properly, the allocation of resources was done incorporating a policy-based resource provisioning control mechanism through software.
- (8) High importance is given to healthcare as per IOM (Institute of Medicine).
- (9) Fidelity of the system is considered as the most important criterion because of the amount of data collected from the sensor field and delivered to the intended person within the stipulated time and as it is also dependent on the application.
- (10) As fidelity is a critical factor, it means receiving the required amount of data packets to detect features of interest in a given time instant.

- (11) The intensity variations in sensor data can be used to model the trend analysis.
- (12) To route the data packets, the greedy perimeter stateful routing technique can be applied as it forwards the packets to nodes that are close to their destination. Also, it needs knowledge about the geographical coordinates of the sensor net.
- (13) In regions where such a greedy path does not exist, the node can deliver the data to the perimeter node where the packet travels successfully through the planar region of the network until it reaches a node closer to the destination.
- (14) When there is any network problem, it must be resilient to withstand and recover quickly from different complicated conditions.

4.2. Objectives of Cyber-Physical System Establishment. The following five objectives have been conceived to implement the cyber-physical system to address the scalability and reliability issues in remote medical diagnosis:

- (1) Design and development of a data acquisition system with threshold detection policies

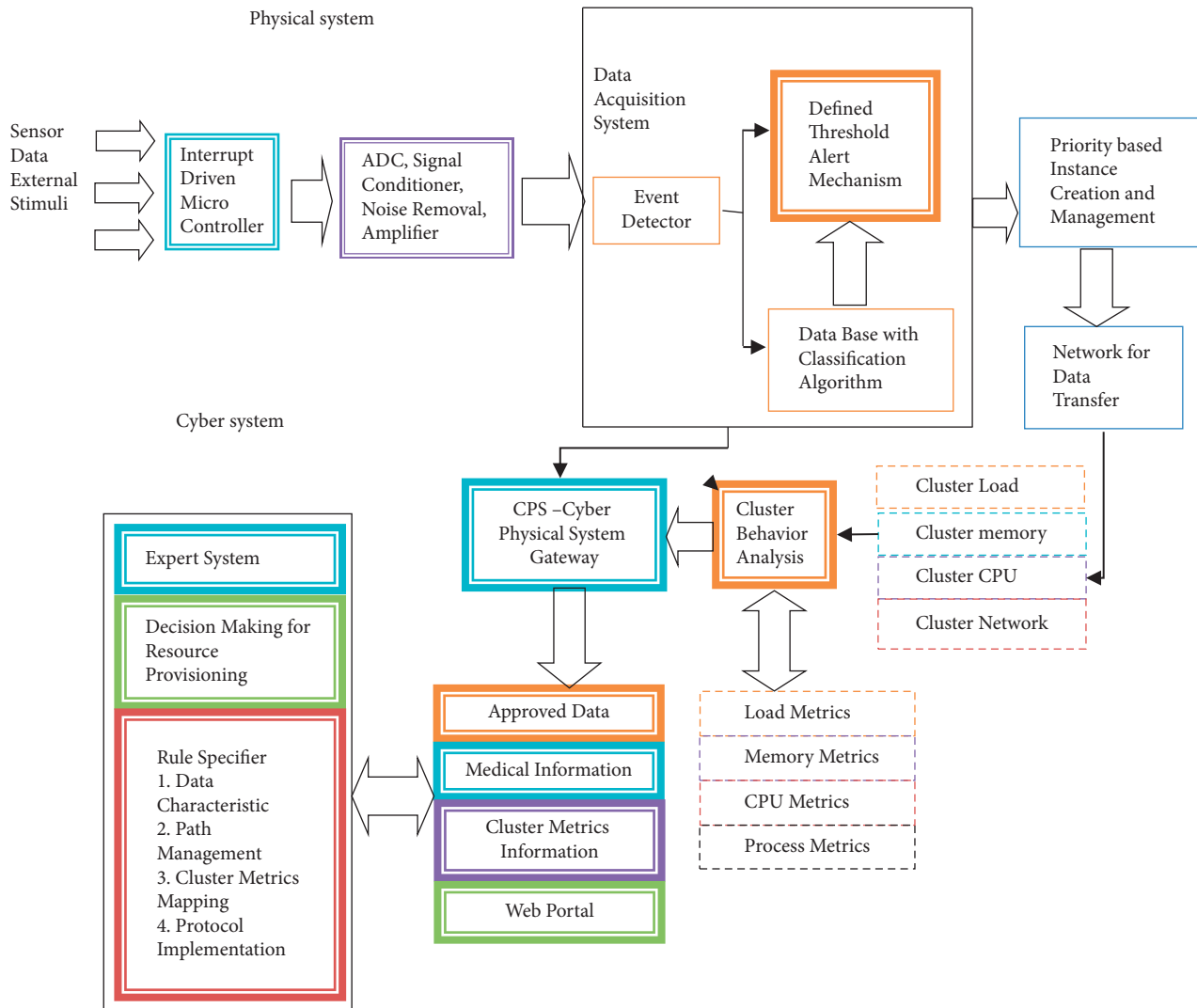


FIGURE 9: Architecture for cyber-physical system (CPS), CloudSense.

- (2) Design and development of a software-defined controller to analyze the network characteristics
- (3) Development of a path planning strategy for time-critical data
- (4) Development of a knowledge base for patient data and network data correlation aspects
- (5) Design and development of a mapping table to map the input events with network instances based on priority queuing

4.3. *Description of CPS Architecture.* The building blocks of the architecture are as follows.

4.3.1. *Physical System.* The physical system comprises the physical wireless body area sensors connected to the appropriate microcontroller units with required signal conditioning mechanisms. The noise removal and amplification aspects are carried out so that the output from one unit of the WBAN is given as one event from the physical device with

the timestamp. The sensors in the WBAN unit were of different modalities such as temperature, blood pressure, skin conductance, oxygen saturation (SPO_2), and ECG. These parameters were in different units and different formats. The data packet formation for the medical data has to incorporate all the sensor data in the digital format with source ID, destination ID, sequence no., and length payload with security codes. The sample format is shown in Figure 10.

4.3.2. *DATA Acquisition Systems (DASs).* The DAS comprises an event detector, defined threshold alert mechanism, and a database with a classification algorithm built-in. An event detector is a mechanism through which if any input signal has to be attended to and when the physiological body parameters have been accessed. Defined threshold alert mechanism consists of a two-way threshold-based alerting where the input signal that goes beyond the critical level was identified. Identification of the threshold value is based on the comparison between the normal body parameter values

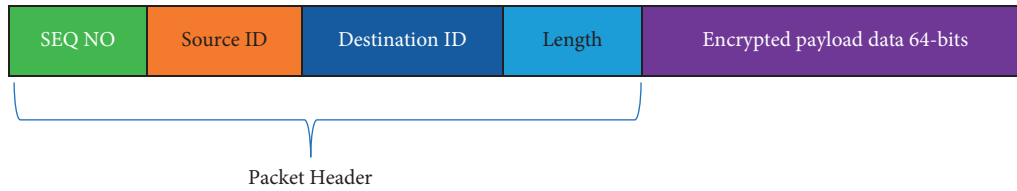


FIGURE 10: Data packet format.

and the measured body parameter values. The threshold monitoring tool has to sit all day long watching and waiting for the event to occur based on the comparison. This is a monitoring system that will alert when the body parameter values become critical. If the disease is identified properly, proper physicians were connected without delay.

(1) *Database with Built-in Classification Algorithm: (Stochastic Modeling)*. Stochastic modeling is concerned with the application of probability to real-world situations characterized by uncertainty. Due to the pervasive nature of uncertainty, the tools have the potential to demonstrate their measurement quality in almost every aspect of the medical diagnosis system.

4.3.3. Priority-Based Instance Creation and Management.

The interlinking of the physical and cyber systems occurs at the priority-based instance creation and management system. This is again a mechanism that provides priority queuing based on the criticality of instances that occur. During the situation, when a critical issue occurs, based on the emergency of the human's health condition, the situations call for a queuing scheme that allows having priority over all other conditions, priority queuing (PQ) is being considered for the scenario. PQ has been allocated four queues in our study, each with a different priority: high priority, medium priority, regular priority, and low priority. After the highest priority queue has been emptied that must be served first, the data packets from each queue are transmitted. Within each queue, packets are transmitted first in, first out based on a crucial event in the calculation of human body parameters. The queue size does not affect the amount of time the packets obtain in that queue. PQ queue size is optimized for data packets. In order of priority, each queue is served. The high-priority queue is often first served because it concentrates on emergency data packets; if the high-priority queue is empty, then the medium queue is emptied and also serves next to the patient's critical condition.

Whenever a high queue packet, which is an emergency data packet, is received, the doctor must attend immediately. The queue is filled before any other queues are processed. When the medium-priority queue is emptied, the usual queue that holds normal body parameter values is emptied if there are no packets in the high-priority queue. Finally, the low-priority queue is emptied if the high, medium, and regular queues are emptied. Therefore, when PQ is used, packets in lower priority queues cannot be forwarded in the required time, causing network apps to run out of time for

applications with packets using lower priority protocols. If a packet does not fit any of the configured queues, the packet goes to the usual queue. Since PQ is not dynamic, it does not respond to network trends. When used, it is a good idea to conduct network baselines regularly and to review traffic to ensure that queue size and protocol distributions are correctly configured to manage the traffic at peak times shown in Figure 11.

(1) *High-Priority Queue*. Packets arriving at the high-priority queue shall immediately be served. The medium-, regular-, and low-priority queues are serviced after the high-priority queue has been emptied. If packets arrive for the high-priority queue at any time, they were transmitted before the high-priority queue has been emptied before any other queue receives operation. The high-priority queue's default size is 20 packets.

(2) *Medium-Priority Queue*. The medium queue is serviced after the high-priority queue has been emptied. When a high-priority queue arrives, the packets in the high-priority queue are forwarded first, until the queue is empty, and then the medium queue receives attention again. The medium-priority queue's default size is 40 packets.

(3) *Normal Priority Queue*. If the high or medium queues have no packets, the usual queue is serviced. When packets arrive at the high or medium queues, these are forwarded to medium in order, and packets are sent to the usual queue after those queues have been cleared. The normal priority queue size is 60 packets. By default, all unspecified traffic is allocated to the usual priority queue, but by using the default argument, you can change this behavior.

(4) *Low-Priority Queue*. Low-priority queue packets were forwarded if all other queues are empty. When a packet comes in any of the other queues, the queues were cleared first, until they are empty, and then the low-priority queue is serviced back. The low-priority queue default size is 80 packets.

A priority list is created to configure PQ. To create a priority list, 16 priority lists can be created. There are 4 queues in each list: high, medium, regular, and low. Packets are allocated to one of the four queues depending on their features: the protocol, the input interface, the size of the packet, the criticality factor, and patient position. Traffic not specified in one of the four queues is sent to the standard queue, the normal queue. The priority-list command, its arguments, keywords, and descriptions are included in Table 1.

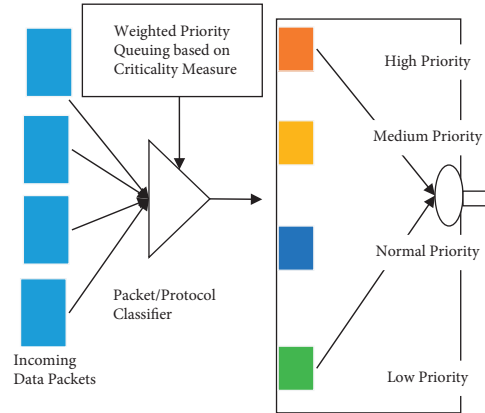


FIGURE 11: Priority-based weighted queuing mechanism for data packet transfer.

TABLE 1: Sample priority table.

| Priority list command (listing number) | Arguments (parameters) | Keywords (percentage of severity) (%) | Criticality factor (critical value) | Timestamp (time of occurrence) | Location ID of the patient (country, state, and IP address) | Assignment of queue (name of queue) |
|--|------------------------|---------------------------------------|-------------------------------------|--------------------------------|---|-------------------------------------|
| 1001 | ECG, BP | 100 | 0 | 9:00 a.m. | Inchlocid01 | High |
| 1002 | BP, TMP | 75 | 1 | 12:00 p.m. | Pkpelocid45 | Medium |
| 1003 | SPO ₂ , TMP | 50 | 2 | 2:45 a.m. | Uscalocid56 | Normal |
| 1004 | SC, TMP | 25 | 3 | 4:54 p.m. | Rumalocid98 | Low |

A sample priority table consists of the following information:

- (1) Priority listing is the event number, based on the incoming traffic received from ingress
- (2) Arguments consist of parameter listing, based on the combination of parameters for the cause of disease and the attendance requirement of the physician
- (3) The keywords column indicates the percentage of the severity of the disease and the need for attention
- (4) Criticality factor presents any one of the four values 0, 1, 2, and 3 depending upon the queue in which the data has to fit in for a diagnosis
- (5) The timestamp indicates the incoming data time, based on the time of occurrence of the event, as it is possible to predict the time from the wireless body sensor data unit (from the data collector within the body sensor unit)
- (6) Location ID consists of the country code, state code, and the IP address of the system from which the patient's information is received
- (7) Assignment of queue indicates the name of the queue through which the patient's information was processed for diagnosis

Table 1 presents the fields such as priority-list command like the patient ID, the parameters that are observed from the patient, the severity of the disease from the observed parameter, the critical value that indicates whether the patient has to be admitted immediately by the doctor, timestamp of the data, location data that includes the country name, state name, and the IP address of the system from where the data

is being routed, and the importance of the queue from where the data have to be fetched.

To assign a public IP address to the EC2 instance, the following procedure is adopted:

- (1) Open the Amazon EC2 console.
- (2) In the navigation pane, choose the Instances icon.
- (3) Select the instance of your own, and choose the Actions icon, Networking, and Manage IP addresses.
- (4) Expand the network interface. Under IPv6 addresses, choose Assign new IP address.
- (5) Then choose Save.
- (6) To assign an ephemeral external IP address that does not persist beyond the life of the resource, we create an instance or forwarding rule without specifying an IP address, the resource is automatically assigned an ephemeral external IP address.
- (7) Spot instances can also be created.

5. Mathematical Modeling

The list of physiological parameters that can be observed through the body sensor network is categorized as follows:

$$\sum_{i=1}^n \text{Phy.Pi} \quad (1)$$

Phy.Pi consists of the five vital body parameter values considered for this research. ECG is modeled as A ; body temperature is modeled as B ; and oxygen saturation is modeled as C .

- (1) ECG can be measured from the human in the range of $+o$ to $-o$, where all the variations observed from the human's ECG values comprise the two extremities of death conditions and all the living conditions before death in both extremes. When we normalize the range of ECG, a human can withstand; it has the maximum ECG value when death can occur as well as the minimum ECG value at where death can occur. The maximum ECG value is normalized as $+1$, and the minimum ECG value is normalized as -1 . The normal ECG value, a human can withstand, is termed as 0 . The acronym ALD indicates all ranges of ECG values including the two extremities of death conditions along with the normal life condition and the intermediate life conditions between normal and $+1$ and between normal and -1 . These between life conditions have been viewed as disease conditions and ready for diagnosis.

$$A_{LD} = \int_{-o}^{+o} d(\text{NECG}). \quad (2)$$

Equation (2) indicates the total number of ECG values that can be measured from a human at any time instant.

- (2) Body temperature can be measured from the human in the range of $+p$ to $-p$, where all the possible temperature values that can be observed from the human can include both the extreme values of temperature meeting death in maximum and minimum level, which are $+1$ and -1 . Along with the extreme values, the normal human body temperature's normalized value is 0 . Between 0 and $+1$ and between 0 and -1 , both intermediate values will lead to diseases. These values need immediate diagnosis based on their level of severity.

$$B_{LD} = \int_{-p}^{+p} d(\text{NTMP}). \quad (3)$$

Equation (3) indicates the total number of temperature values that can be measured from a human at any time instant.

- (3) Oxygen content- (SPO_2) can be measured from the human in the range of $+q$ to $-q$, where all the possible SPO_2 values that can be observed from the human can include both the extreme values of oxygen saturation meeting death in maximum and minimum level, which are $+1$ and -1 . Along with the extreme values, the normal human's oxygen saturation's normalized value is 0 . Between 0 and $+1$ and between 0 and -1 , both intermediate values will lead to diseases. These values need immediate diagnosis based on their level of severity.

$$C_{LD} = \int_{-q}^{+q} d(\text{NSPO2}). \quad (4)$$

Equation (4) indicates the total number of Oxygen saturation values that can be measured from a human at any time instant.

$$D_{LD} = \int_{-r}^{+r} d(\text{NBP}). \quad (5)$$

- (4) Blood pressure can be systolic and diastolic and be measured from the human that indicates the pumping mechanism of the heart.

Equation (5) indicates the total number of blood pressure values containing a separate column of low blood pressure and high blood pressure. The extreme values for the human's death conditions about maximum and minimum values along with the in-between values from the normal value in both directions have to be observed as 0 to $+1$ and 0 to 1 .

$$E_{LD} = \int_{-s}^{+s} d(\text{NSC}). \quad (6)$$

Pulse oximetry measures the amount of oxygen being carried in our blood, as a percentage that can be measured at the finger using a pulse oximeter. This measurement is known as the SPO_2 and the saturation of peripheral oxygen is SAO_2 , which is the saturation of arterial oxygen. A decrease in oxygen saturation and increases in pulse rate and heart rate variability have been found to be associated.

- (5) Skin conductance: equation (6) presents the total number of skin conductance values about the normal skin conductance value as zero and the extreme values in both directions indicating death as $+1$ and -1 . Excluding the normal value, the entire set of values has to be diagnosed with prioritization parameters.

Excluding the death conditions from equations (2)–(6), the following can be the equations for life and diagnosis:

$$A_L = \int_{-o}^{+o} d(\text{NECG}) - [\text{NECG}_{+o}, \text{NECG}_{-o}], \quad (7)$$

$$B_L = \int_{-p}^{+p} d(\text{NTMP}) - [\text{NTMP}_{+p}, \text{NTMP}_{-p}], \quad (8)$$

$$C_L = \int_{-q}^{+q} d(\text{NSPO2}) - [\text{NSPO2}_{+q}, \text{NSPO2}_{-q}], \quad (9)$$

$$D_L = \int_{-r}^{+r} d(\text{NBP}) - [\text{NBP}_{+r}, \text{NBP}_{-r}], \quad (10)$$

$$E_L = \int_{-s}^{+s} d(\text{NSC}) - [\text{NSC}_{+s}, \text{NSC}_{-s}]. \quad (11)$$

5.1. Case I: Instance Creation for Parameter 1 – ECG. As mentioned, $\text{ins}_{t_0} - \text{ins}_{t_t}$ indicates the life and death instances about ECG. Instances between $+o$ and $-o$ can be termed as t .

All the ECG instances that we can expect become, ins_{ot_0} , ins_{ot_1} , ins_{ot_2} , \dots , ins_{ot_t} from the time instant t_0 to t_t .

Time instants of occurrences are represented in

$$ECG = \int_{ins_{ot0}}^{ins_{ott}} d(A_{LD}), \quad (12)$$

where A_{LD} is represented in equation (2).

A random sample of any of the ECG instances is exhibited in the equation. Equation 12 shows a random sampling of any of the ECG instances. Any subsample of ECG instances that requires critical diagnosis is created from the random sample.

Immediate diagnosis of ECG is represented in

$$ECG_{IM} = \int_{ins_{ot0}}^{ins_{ott}} d\{A_{LD}\}_{CF}. \quad (13)$$

5.2. Case II: Instance Creation for Parameter 2 – Temperature. As mentioned, $ins_{to}-ins_{tt}$ indicates the life and death instances about human body temperature. Instances between $+p$ and $-p$ can be termed as t .

From the time instant t_0 to t_t , all the temperature instances that we can predict become $ins_{pto}, ins_{pt}, ins_{pt2}, \dots, ins_{tt}$. Temperature instances are represented in

$$TMP = \int_{ins_{pt0}}^{ins_{ptt}} d(B_{LD}), \quad (14)$$

where B_{LD} is represented in equation (3).

A random sample of any of the temperature instances is exhibited in the equation. Equation 14 shows a random sampling of any of the temperature instances. Any subsample of temperature instances that requires critical diagnosis is created from the random sample. Immediate diagnosis of TMP is represented in

$$TMP_{IM} = \int_{ins_{pt0}}^{ins_{ptt}} d\{B_{LD}\}_{CF}. \quad (15)$$

5.3. Case III: Instance Creation for Parameter 3 – SPO₂. As mentioned, $ins_{to}-ins_{tt}$ indicates the life and death instances about human body oxygen saturation. Instances between $+q$ and $-q$ can be termed as t .

From the time instant t_0 to t_t , all the SPO2 instances that we can predict become $ins_{qto}, ins_{qt1}, ins_{qt2}, \dots, ins_{qtt}$.

The oxygen saturation instances are represented in

$$SPO2 = \int_{ins_{qt0}}^{ins_{qtt}} d(C_{LD}), \quad (16)$$

where C_{LD} is represented in equation (4).

A random sample of any of the oxygen saturation instances is exhibited in the equation. Equation 16 shows a random sampling of any of the SPO2 instances. Any subsample of SPO2 instances that requires critical diagnosis is created from the random sample. Immediate diagnosis of SPO₂ is represented in

$$SPO2_{IM} = \int_{ins_{qt0}}^{ins_{qtt}} d\{C_{LD}\}_{CF}. \quad (17)$$

5.4. Case IV: Instance Creation for Parameter 4 – Blood Pressure. As mentioned, $ins_{to}-ins_{tt}$ indicates the life and

death instances about low or high blood pressure values. Instances between $+r$ and $-r$ can be termed as t .

From the time instant t_0 to t_t , all the blood pressure instances that we can predict become $ins_{rto}, ins_{rt1}, ins_{rt2}, \dots, ins_{rtt}$.

The blood pressure instances are represented in

$$BP = \int_{ins_{rto}}^{ins_{rtt}} d(D_{LD}), \quad (18)$$

where D_{LD} is represented in equation (5).

A random sample of any of the blood pressure instances is exhibited in the equation. Equation 18 shows a random sampling of any of the blood pressure instances. Any subsample of blood pressure instances that requires critical diagnosis is created from the random sample. Immediate diagnosis of blood pressure is represented in

$$BP_{IM} = \int_{ins_{rto}}^{ins_{rtt}} d\{D_{LD}\}_{CF}. \quad (19)$$

5.5. Case V: Instance Creation for Parameter 5 – Skin Conductance. As mentioned, $ins_{to}-ins_{tt}$ indicates the life and death instances about low or high skin conductance values. Instances between $+s$ and $-s$ can be termed as t . From the time instant t_0 to t_t , all the skin conductance instances that we can predict become $ins_{st0}, ins_{st1}, ins_{st2}, \dots, ins_{stt}$.

The skin conductance instances are represented in

$$SC = \int_{ins_{st0}}^{ins_{stt}} d(E_{LD}), \quad (20)$$

where E_{LD} is represented in equation (6).

A random sample of any of the skin conductance instances is exhibited in the equation. Equation 20 shows a random sampling of any of the skin conductance instances. Any subsample of skin conductance instances that requires critical diagnosis is created from the random sample. Immediate diagnosis of skin conductance is represented in

$$SC_{IM} = \int_{ins_{st0}}^{ins_{stt}} d\{E_{LD}\}_{CF}. \quad (21)$$

Combining all the critical instances for the mentioned five parameters become

$$\begin{aligned} Z = & \int_{ins_{ot0}}^{ins_{ott}} d\{A_{LD}\}_{CF} + \int_{ins_{pt0}}^{ins_{ptt}} d\{B_{LD}\}_{CF} \\ & + \int_{ins_{qt0}}^{ins_{qtt}} d\{C_{LD}\}_{CF} + \int_{ins_{rto}}^{ins_{rtt}} d\{D_{LD}\}_{CF} \\ & + \int_{ins_{st0}}^{ins_{stt}} d\{E_{LD}\}_{CF} + . \end{aligned} \quad (22)$$

The questions arise like, how to include the entire set of critical instances that have to be given high prioritization. How to handle if the number of virtual instances has come more in terms of CPU, memory, network, cluster usage, and load/process time, which is shown in Figure 12.

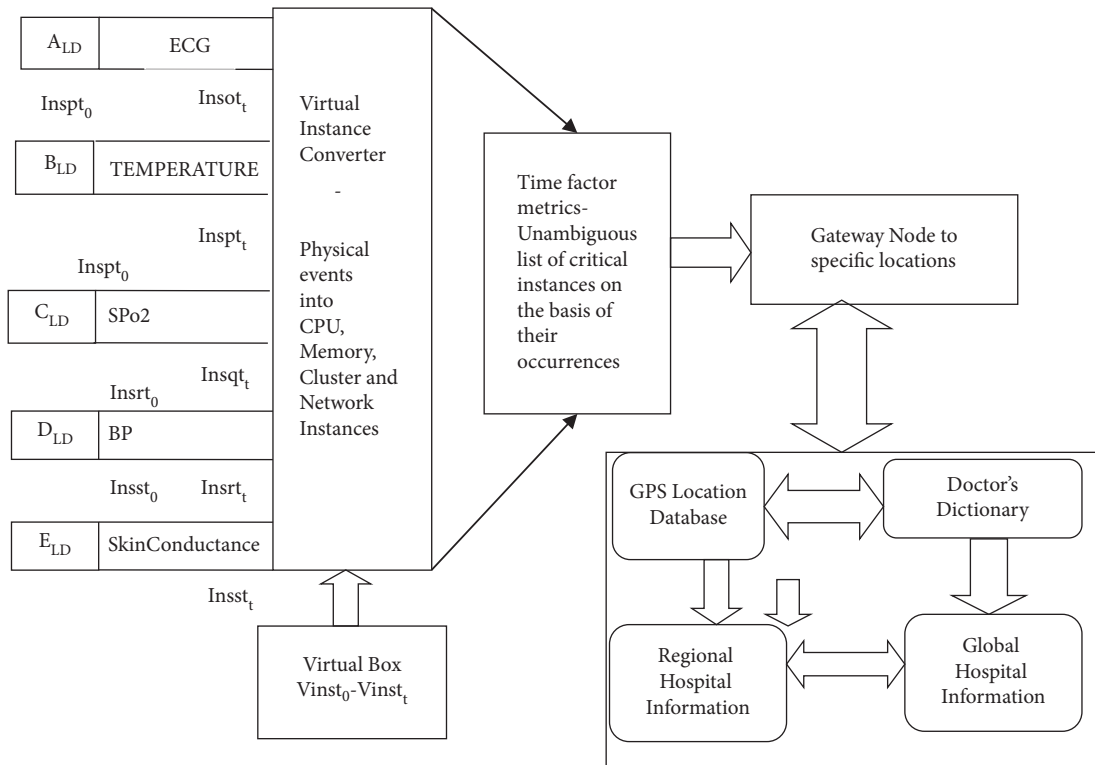


FIGURE 12: Instance creation and management for smart health data.

6. Addressing the WBAN: Data Scalability Issue

6.1. *Cyber System Components.* Eucalyptus enables pooling compute, storage, and network resources that can be dynamically scaled up or down as application workloads change. It is compatible with Amazon’s EC2.

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. An Elastic Compute Cloud instance is a virtual server that can run applications in Amazon Web Services (AWS). When setting up an EC2 instance, we can custom configure CPU, storage, memory, and networking resources, and when an instance is created, we can create it with an Amazon Machine Image (AMI).

Ganglia gets installed on the cluster at the bootstrap stage that provides insight into how each box in the cluster is performing. If the CPU was running high, it might be wise to choose EC2 instances that had larger cores, or if the memory was overutilized also, then EC2 instances can be chosen with a larger memory capacity.

6.2. *Image Management in CloudSense.* At this juncture, the cloud is ready to operate. The node at “192.168.20.2” was registered, and the resource the cloud has to offer is within the availability zone of the cloud. The type m1 small indicates that there are four machines available and can be created each with one processing unit, RAM 192 MB, and a disk with 2 GB space. The VM images can be downloaded from the Eucalyptus store. Eucalyptus has provided links to

prepackaged virtual machines that are ready to run in the private Eucalyptus cloud.

6.3. *Instance Management.* The images have been registered. They are now ready to be uploaded and operated upon. An instance can be sprung up after the assignment of a public IP by the DHCP daemon. Instances can be terminated as and when they are not required.

6.4. *Cloud Monitoring.* There are myriad ways to view the performance of the cloud. Many tools have been developed. Of them, the tool “Ganglia” is a very efficient and effective tool that can be used to view the performance through various metrics.

Ganglia is a scalable distributed monitoring system [41]. It scales well with very large numbers of servers and is useful for viewing performance metrics in real time. The applications of ganglia are represented in Figure 13.

On the back end, Ganglia is made up of the subsequent components:

- (1) Gmond (Ganglia monitoring daemon): it is a small service that collects information about a node. This is installed on every server that is to be monitored.
- (2) Gmetad (Ganglia meta daemon): it is a daemon on the master node that collects data from all the Gmond daemons (and also from the other Gmetad daemons, on condition).

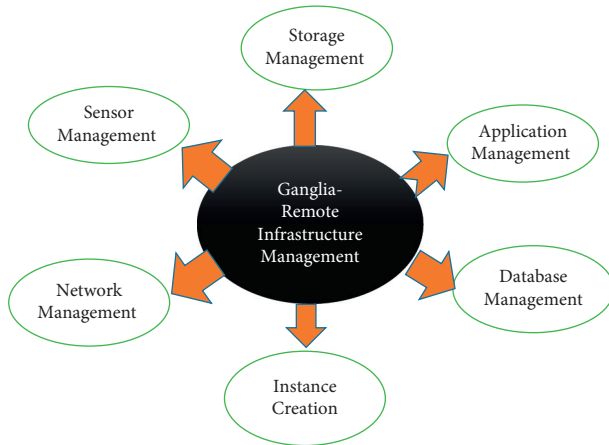


FIGURE 13: Ganglia applications.

- (3) RRD (round-robin database) tool: it is a tool on the master node used to store data and visualizations can be possible from Ganglia in the time series mode of operation.
- (4) PHP web front end: it is a web interface on the master node that displays graphs and metrics from data in the RRD tool.

Every node (server) that we want to monitor has Gmond to be installed. Every node uses Gmond to send data to the single master node running Gmetad, which should collect all the node data and send it to the RRD tool to be stored. The data in the web browser can be viewed with the help of PHP scripts and Apache.

Ganglia grid with the master node is shown as the Ganglia server running the Gmetad daemon, and the other nodes have been shown as connecting servers running the Gmond daemon:

To monitor the data through the web interface, the data is organized on several levels as shown in Figure 14. Ganglia organizes the nodes, which are individually monitored machines, into clusters, nothing but groups of similar nodes. On a higher level, collections of clusters can also be organized into grids.

Important limitations for wider acceptance of the existing WBAN systems for continuous monitoring are as follows:

- (a) Unwieldy wires between sensors and a processing unit
- (b) Lack of system integration of individual sensors
- (c) Interference on a wireless communication channel shared by multiple devices
- (d) Nonexistent support for massive data collection and knowledge discovery

Depending on the application requirements, the WBAN coordinator is further connected to telemedicine and medical servers for relevant recommendations, like connecting to the edge and fog terminals as everything can be visualized by the Ganglia monitoring system as it supports scalability in an efficient way.

Ganglia is a free and open-source monitoring tool for high-performance computing systems such as clusters and grids. It is made up of three components: gmond, gmetad, and a web front end. Gmond is a multithreaded daemon that runs on each node and communicates with gmetad. Gmetad collects and archives metrics and generates reports. Internally, it makes use of the rrdtool for data logging and graphing. The web front end included with Ganglia is used to visualize the data. Ganglia is a comprehensive monitoring solution that visualizes and archives system metrics. It comes preconfigured to monitor over 30 metrics without requiring any additional coding. The installation procedure consists of downloading ganglia packages and their dependency packages [37]. Figure 15 was taken from the web interface provided by ganglia to monitor the eucalyptus cloud.

It can be clearly understood about the moment that no heavy processes were going on, and hence, the utilization rates are quite low. The following figure indicates the cluster load as a graph, which is measured as load/process versus time. The number of node participation is 1 as minimum and 2 as maximum. The number of CPUs is 2 as minimum and 6 as maximum. The process minimum is termed as zero, and the maximum process is termed as 3. As it is not specified, it is the cluster load at the hour when it was measured in Figure 16.

As per the diagrammatic representation of cluster load last hour in Figure 16, the node considers the instance as it indicates the patient data loading. One patient, one instance is formed when the number of nodes participating is 1, and two patient nodes, two instances are created when the number of nodes participating is 2.

Figure 17 is a graph specifying the cluster memory at a given time in analyzing the cluster memory in terms of use, share, cache, buffer, swap, and total. Memory utilization can be done using the free command in Linux. The command displays the total amount of free, used physical, and swap memory in the system, as well as the buffers used by the kernel. We can also estimate the total physical memory; the total used memory, the amount of free physical memory, the size of the file cache, the total size of the swap space, the amount of swap space used, and the amount of swap space free.

The representation in Figure 18 implies that the cluster memory storage corresponds to the hospital information of patient data in the memory pertaining to the collective patient information at a particular time.

Figure 18 shows the cluster CPU usage at some time during the operation of the cluster. To monitor per-process CPU utilization, the parameters to be monitored are the CPU, memory, page faults, stack, disk I/O, context switching on a per-process for a specified time interval. The addition of the parent and child process can also be monitored using special CPU and network monitoring devices. The parameters can be indicated as threads, processes, users, and groups that are resource hogs. The amount of time that the system spent in user mode is commonly measured in units of user-HZ. User mode with low priority is observed as system mode, the idle task, and the waiting time for an I/O to finish are measured.

Ganglia Monitoring System

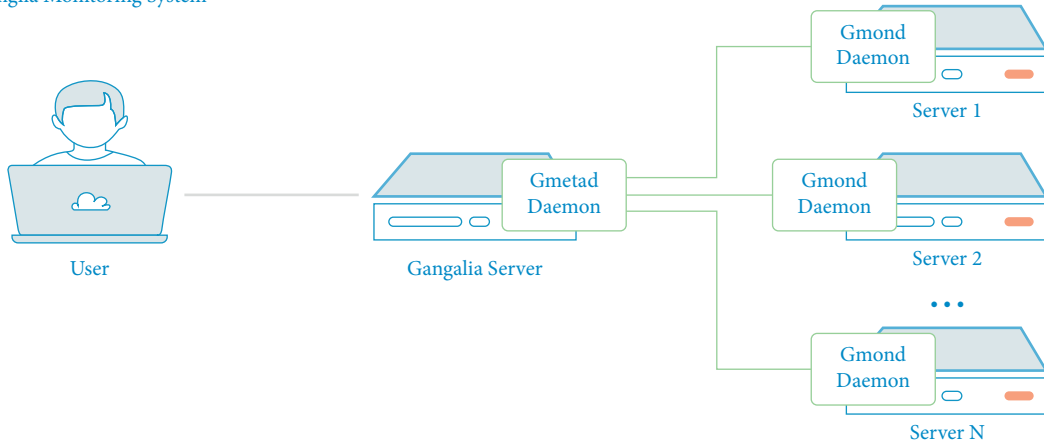


FIGURE 14: Ganglia monitoring system.

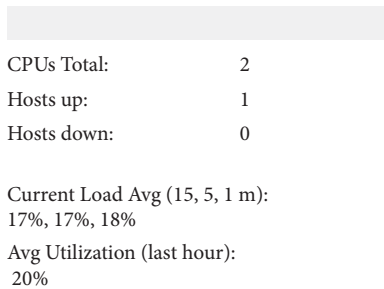


FIGURE 15: Cluster information.

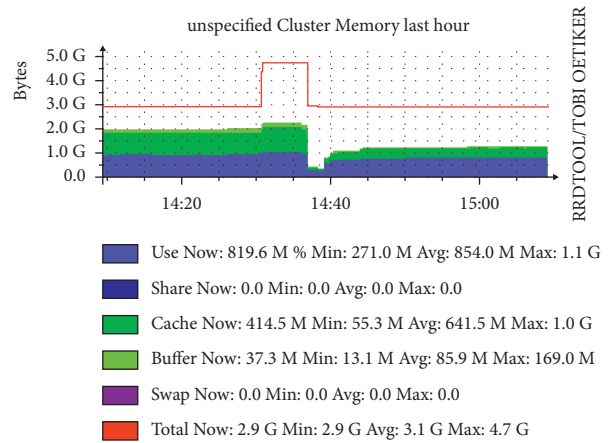


FIGURE 17: Cluster memory at a particular time.

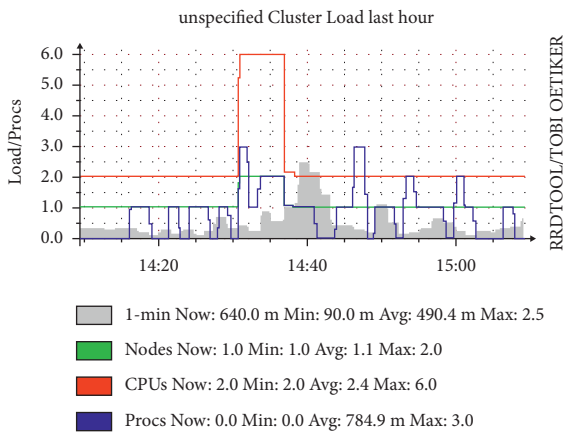


FIGURE 16: Cluster load at a particular time.

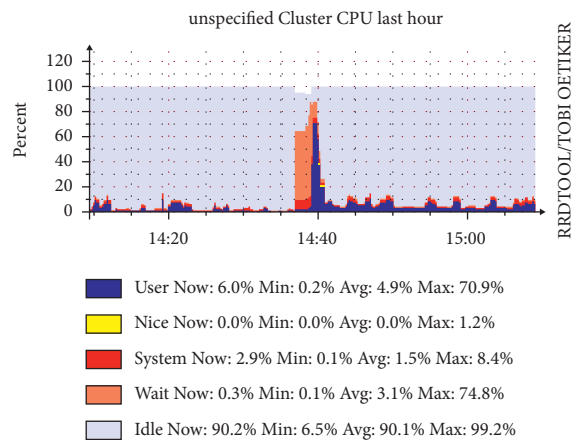


FIGURE 18: Cluster CPU usage at a particular time.

The prioritization of data pertaining to the patient’s abnormality condition is shown in Figure 18. Context switching indicates that the data can be forwarded to the port immediately as per the severity condition of the patient. If the same type of prioritized information is received, it could be mapped with parent and child processes as they fall in the same category, due to medical emergency.

Figure 19 shows the network usage over a particular hour during the operation of the cloud. The bandwidth usage by connection can be viewed as whether it is incoming traffic or outgoing traffic and which application is utilizing how much

amount of bandwidth. It can also be viewed as how many hosts are connected at a particular instant of time and what type of traffic is going on in the network.

The indication in Figure 19 corresponds to how many patients are connected to the fog at a particular time and

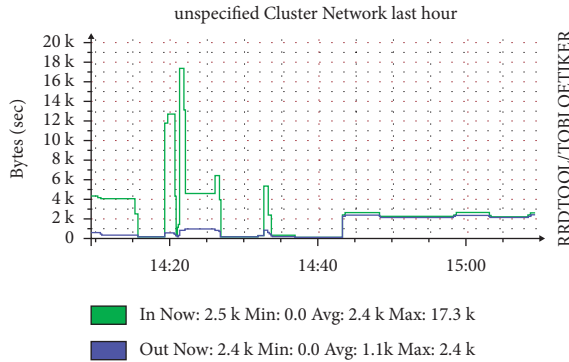


FIGURE 19: Cluster network at a particular time.

whether the patient traffic is within the bandwidth limitation or not, by comparing the incoming and outgoing traffic.

Figure 20 is the system load as it varies over time in the cluster. Clustering makes one instance of an application server into a master controller. This master controller will process all the requests and distribute them in too many numbers of instances. The algorithms used for this purpose are the normal operating system's scheduling algorithms such as round-robin and so on. Clustering is done for load balancing to enable scalability as the purpose of this research is to increase scalability when the number of inputs becomes enormous, which is shown in Figure 20. As scalability is the ability to add more instances of an application server, it increases the capacity of the system through the reduction of the response time of the application.

Figure 20 corresponds to one instance as it relates to one host like so many instances and group those instances of the healthcare application and process the data to do the data analysis and rank the data. It is highly essential to rank, in order to give importance to high-prioritized data rather than normal data, which does not require critical care.

6.5. WBAN Data. WBAN is used for people who require emergency and continuous medical care. They have made the lives of mankind a lot easier. At present, the data produced by the LS-WBAN is processed on local systems. The data is brought to the local machine via gateways and then stored on relational database systems. But the main problem is the issue of scalability. A network may increase its size over a huge geographical area. Such spatial data is massive and huge data sets are generated, which need to be crunched in real time for real-time analysis.

6.6. The Solution. Data can be streamed to the eucalyptus cloud and instantaneously stored using blocks or buckets. These data can be analyzed, and the analysis has been done by springing up instances assuming the data scales up the storage blocks in the cluster. The analysis has been done by bringing up the cluster to its limit and making a study of how the cluster behaves accordingly.

6.7. Scaling the Cluster. The cluster resources are based on the front end for storage and the creation and functioning of

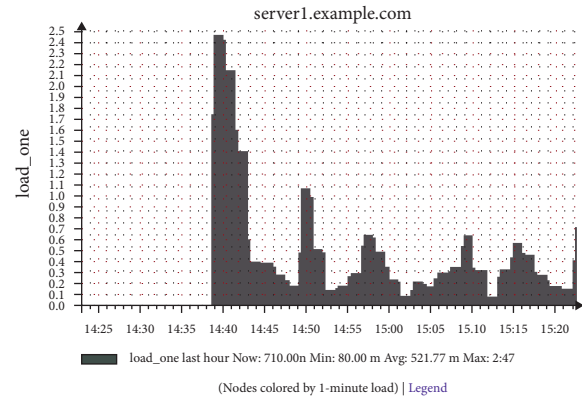


FIGURE 20: Cluster load versus time.

the VMs in the node. The nodes are brought up one by one, and the cluster behavior is studied based on the metrics provided by the ganglia interface. An instance is started using the instance command. The resources are now reduced as shown in Figure 21.

Figure 21(a): resources after starting one instance. The figure shows the load/process metric. One patient is monitored.

Figure 21(b): load/process after spawning one instance. As the graph indicates, the number of nodes is indicated as 1; the number of CPUs is indicated as 2; and the number of processes is increased from 2.2 to 11 between the timings 16–16:20. The figure shows the cluster memory metric. After one instance is spawned, cache, buffer, and total memory size are shown in Figure 16. The patient information after a time period gets accumulated, and the number of processes to save and store temporarily is being decided and represented.

Figure 21(c): cluster memory after one instance spawn. This figure shows the CPU usage in terms of user, nice, system, wait, and idle. After one instance is spawned, the number of users' cache, buffer, and total memory size. The group of patient data nodes is collected by the cluster after instance creation.

Figure 21(d): cluster CPU usage after one instance spawn. This figure shows network usage. The entire health network usage is being depicted.

Figure 21(e): cluster network after 1 instance spawn. This figure shows the legend and the cluster network variation after the data received are depicted.

Figure 21(f): load versus time after one instance spawn. It is a sample load pertaining to the hospital server.

Figure 22 represents the time at which another instance is brought up and the resource availability.

Figure 22(a): availability after 2 instances spawn. The figure shows the load/process. The data path availability is shown for data transfer.

Figure 22(b): cluster load after 2 instances spawn. This figure shows the cluster memory usage. After monitoring 2 patient nodes, the memory occupied is depicted.


```

bonny@myueccluster:~$ euca-describe-availability-zones verbose
AVAILABILITYZONE myueccluster 192.168.10.121
AVAILABILITYZONE | - vm types free / max cpu ram disk
AVAILABILITYZONE | - m1.small 0003 / 0004 1 192 2
AVAILABILITYZONE | - c1.medium 0003 / 0004 1 256 5
AVAILABILITYZONE | - m1.large 0001 / 0002 2 512 10
AVAILABILITYZONE | - m1.xlarge 0001 / 0001 2 1024 20
AVAILABILITYZONE | - c1.xlarge 0000 / 0000 4 2048 20
bonny@myueccluster:~$
    
```

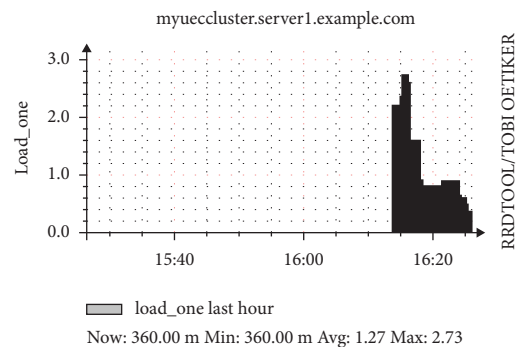
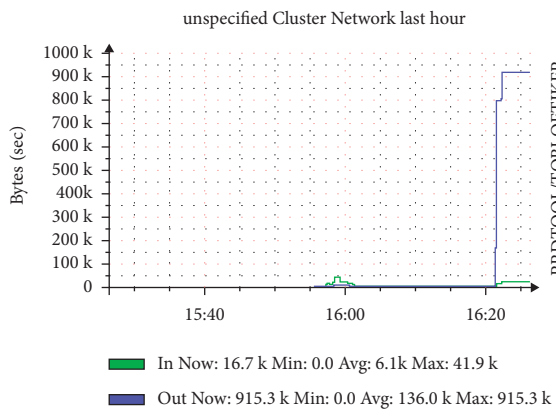
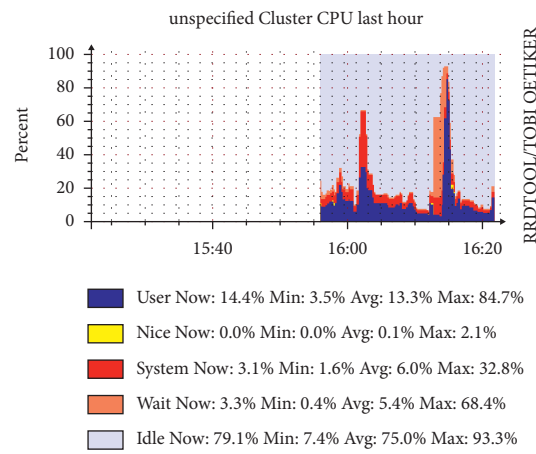
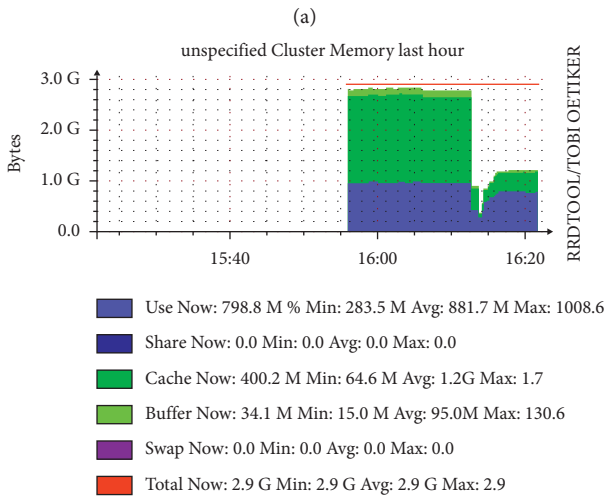
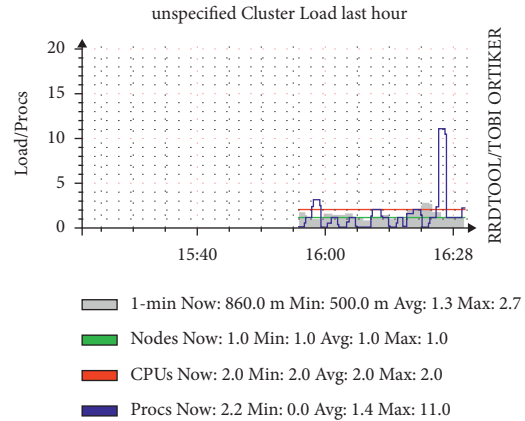


FIGURE 21: Various stages of network resources at a specific time instance.

Figure 22(c): cluster memory after 2 instances spawn. This figure shows CPU usage. After the monitoring of 2 nodes, the CPU processing is depicted.

Figure 22(d): cluster CPU after 2 instance spawn. This figure shows network usage. The cluster CPU, which resides and does the operation over the node data, after acquiring 2 nodes' data is shown in this figure.

Figure 22(e): cluster CPU after 2 instances spawn. This figure shows network usage. The network corresponds to the hospital is depicted with regard to its workload.

Figure 22(f): cluster network after 2 instances spawn. This figure shows the legend. The combination of clusters in a cluster network is depicted.

Figure 23 shows the resources available after the spawn of the third instance. The resources for computing, memory, and storage are exhibited.


```

bonny@myueccluster:~$ euca-describe-availability-zones verbose
AVAILABILITYZONE myueccluster 192.168.10.121
AVAILABILITYZONE | - vm types free / max cpu ram disk
AVAILABILITYZONE | - m1.small 0002 / 0004 1 192 2
AVAILABILITYZONE | - c1.medium 0002 / 0004 1 256 5
AVAILABILITYZONE | - m1.large 0000 / 0002 2 512 10
AVAILABILITYZONE | - m1.xlarge 0000 / 0001 2 1024 20
AVAILABILITYZONE | - c1.xlarge 0000 / 0000 4 2048 20
bonny@myueccluster:~$
    
```

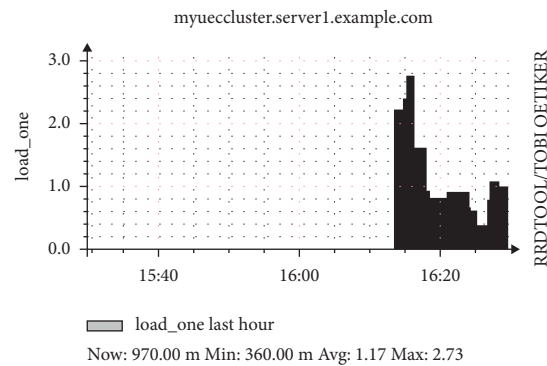
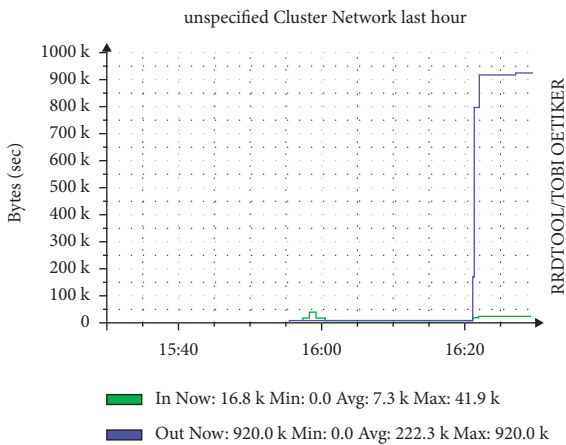
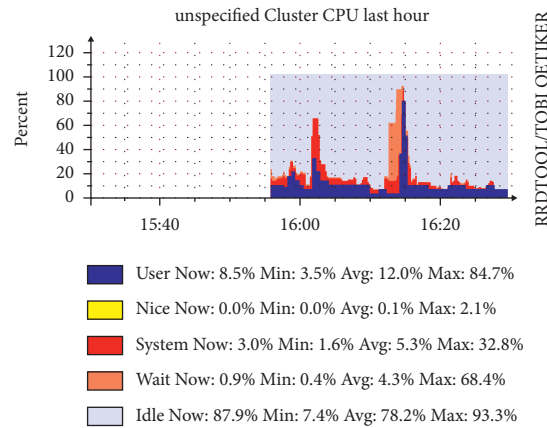
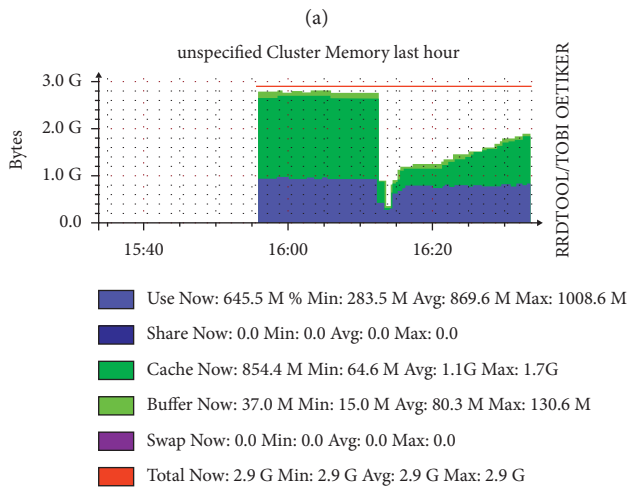
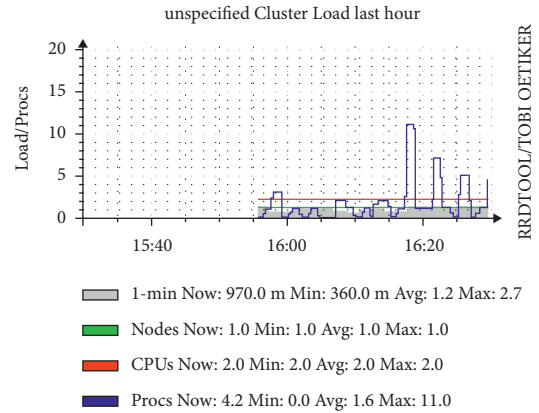


FIGURE 22: Time at which another instance is brought up and the resource availability.

Figure 23(a): the resource usage. This figure shows the load/process. The usage of loading of a patient data per process is depicted.

Figure 23(b): cluster load after 3 instances spawn. This figure shows the cluster memory usage. The cluster is a collection of patient nodes; the cluster load is exhibited.

Figure 23(c): cluster memory after 3 instances spawn. This figure shows CPU usage. The processes acquiring each patient information after the 3 instances of patient node are being depicted.

Figure 23(d): cluster CPU after 3 instances spawn. This figure shows network usage. The cluster CPU resides

```

bonny@myueccluster:~$ euca-describe-availability-zones verbose
AVAILABILITYZONE myueccluster 192.168.10.121
AVAILABILITYZONE | - vm types free / max cpu ram disk
AVAILABILITYZONE | - m1.small 0001 / 0004 1 192 2
AVAILABILITYZONE | - c1.medium 0001 / 0004 1 256 5
AVAILABILITYZONE | - m1.large 0000 / 0002 2 512 10
AVAILABILITYZONE | - m1.xlarge 0000 / 0001 2 1024 20
AVAILABILITYZONE | - c1.xlarge 0000 / 0000 4 2048 20
bonny@myueccluster:~$
    
```

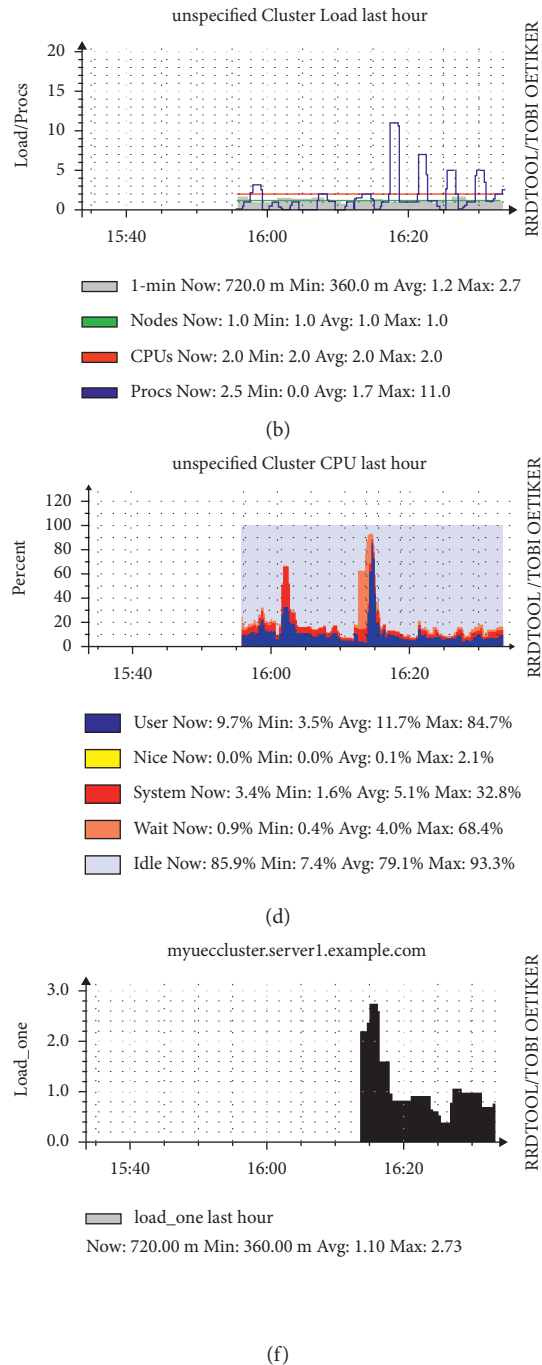
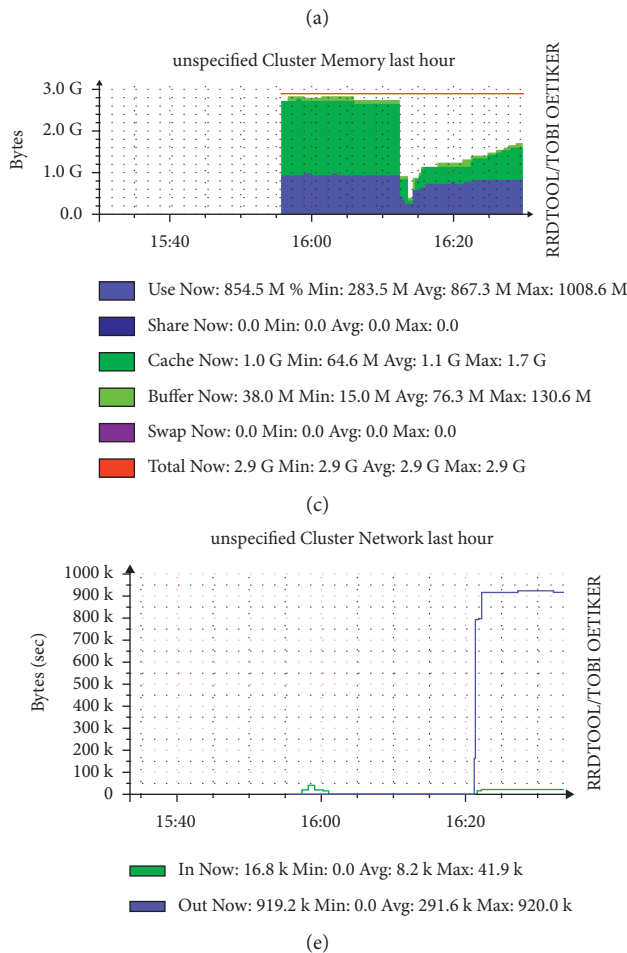


FIGURE 23: Availability zone after 3 instances spawn.

and does the operation over the node data, after acquiring 3 nodes' data, which is shown in the figure.

Figure 23(e): cluster network after 3 instances spawn. This figure shows the legend. In the collection of clusters, the entire health network after the 3 instances is depicted.

Figure 23(f): load after 3 instances spawn. This figure shows the load on that time instant. Cluster load on a particular time instant is depicted.

Figure 24 represents the last available instance, spawned up with the availability zone that shows the resources.

Figure 24(a): cluster load after 4 instances spawn. This figure shows the load/process. When the number of clusters has been increased to 4, the cluster load is depicted.

Figure 24(b): cluster memory after 4 instances spawn. This figure shows the cluster memory. The memory after 4 instances of nodes is depicted.

```

bonny@myueccluster:~$ euca-describe-availability-zones verbose
AVAILABILITYZONE      myueccluster  192.168.10.121
AVAILABILITYZONE      |- vm types   free / max  cpu  ram  disk
AVAILABILITYZONE      |- m1.small   0000 / 0004  1   192  2
AVAILABILITYZONE      |- c1.medium  0000 / 0004  1   256  5
AVAILABILITYZONE      |- m1.large   0000 / 0002  2   512  10
AVAILABILITYZONE      |- m1.xlarge  0000 / 0001  2  1024  20
AVAILABILITYZONE      |- c1.xlarge  0000 / 0000  4  2048  20
bonny@myueccluster:~$
    
```

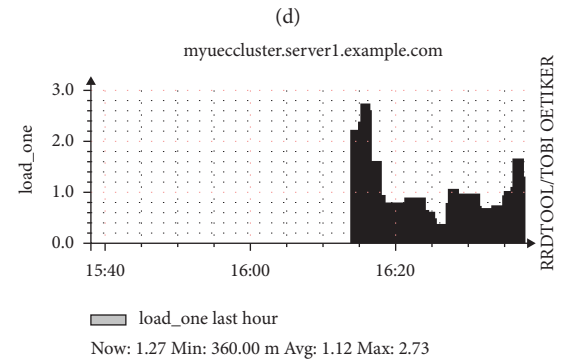
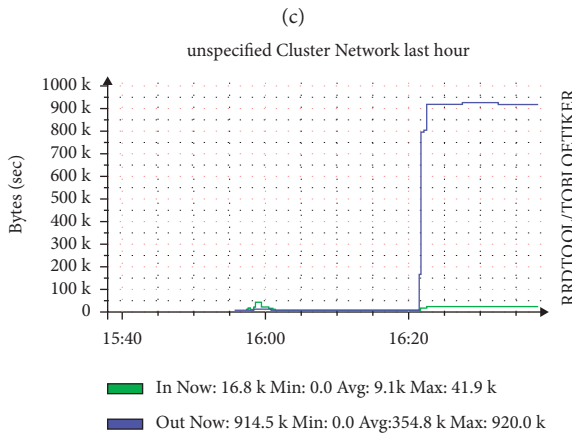
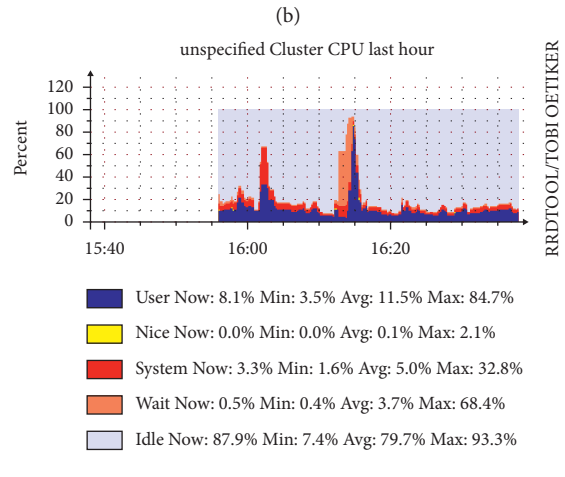
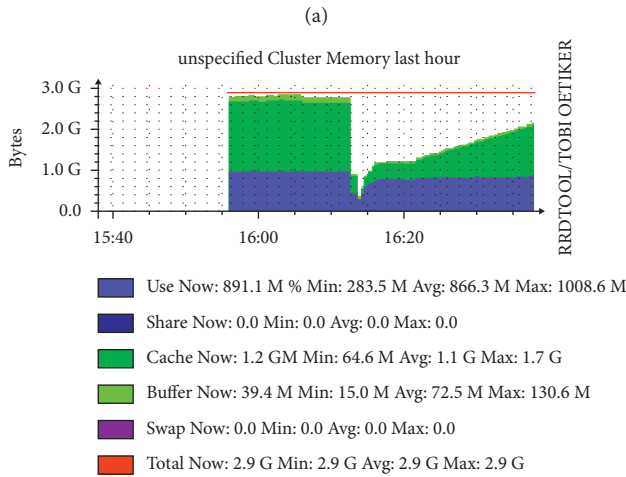
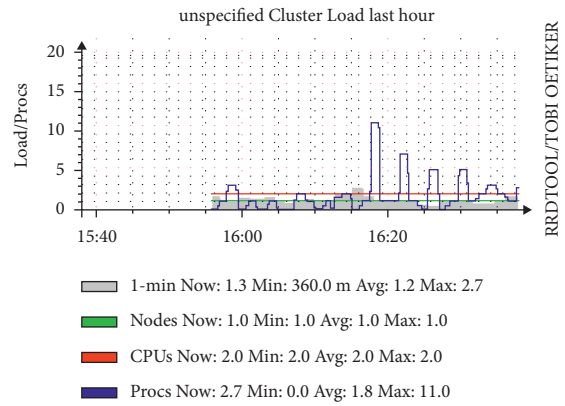


FIGURE 24: The resources are represented by the last available instance, which was generated with the availability zone.

Figure 24(c): cluster CPU after 4 instances start. This figure shows CPU usage. After 4 instances, the CPU utilization is being depicted.

Figure 24(d): cluster network after 4 instances spawn. It shows the cluster network usage and the efficiency in scalability.

Figure 24(e) shows legend after 4 instances spawn. This figure shows the load one metric after all instances are spawn.

Figure 25 shows the network traffic cover the cloud set up. If another instance is to be spawned up, the error that occurs is shown with no availability of resources.

The system information about the cluster is shown in Figure 26.

Figure 27 shows the metrics for load, CPU, memory, and network.

Figure 28 shows the memory metrics. It is based on memory buffers, cached memory, free memory, shared memory, and free swap space.

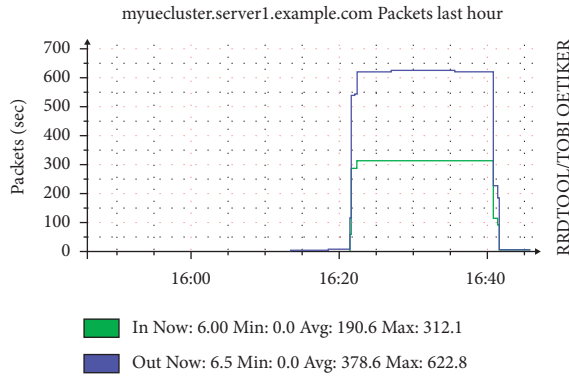


FIGURE 25: Packets/second in the network.

CPU's Total: 2
 Hosts up: 1
 Hosts down: 0

Current Load Avg (15, 5, 1m):
 38%, 52%, 60%

Avg Utilization (last hour):
 58%



Cluster Load Percentages
 50-75 (100.00%)

FIGURE 26: Cluster load percentages.

The network metrics are based on bytes received, bytes sent, packets received, and packets sent, which is shown in Figure 29.

Figure 30 shows the CPU metrics based on CPU idle, CPU idle, CPU Nice, CPU system, CPU user, and CPU wio.

The process metrics are based on total running processes, and the number of total processes is shown in Figure 31.

7. Interpretation of Results

The study of the above graphs makes it all very clear. The cloud is an option for data scalability for handling and analyzing it efficiently and can be used for the same till the resources hit the threshold value of the resources in the cloud. It can be seen from the graphs that there is a steep rise in load as the instances are brought up. The memory usage also rises linearly at an observable high slope value. The network usage steps up at the spawning up of the first instance remains fairly constant thereafter. Bottlenecks can be observed mostly in cluster memory. The CPU bottleneck problem may start if too many complex operations are performed in the cluster. The load on the cloud increases linearly at a fairly low slope value.

Hadoop MapReduce is yet to be implemented on the Eucalyptus cluster, and a study is made on the performance. The same scenario shall be tested on the OpenStack cloud platform, and a comparison shall be made based on the

results with Eucalyptus. Security has not been enforced at the enterprise level. There are myriad works to be done using the cloud and shall be published soon. Considering data availability, when one instance has come many instances, then many servers may have the information at many locations. The locations of the servers were the physician's locations correlating the prioritized instance generation and management mechanism.

7.1. Performance Metrics. Performance of the proposed system can be measured in terms of throughput, bandwidth, delay/latency, packet delivery ratio, signal-to-noise-interference ratio, link quality, nice, buffer size, cache hits, heterogeneity of packets, relationship between group-related instances that can be applied interchangeably with the number of instances and number of packets transmitted generated, relationship between packet loss and SINR and link failure, and scaling due to buffer size increment and path optimization as shown in Table 2.

If $0\% < \text{CPU usage} > 80\%$, redirect the virtual instance, else assign a new CPU—when to provide what type of resource with what capacity?

If $0\% < \text{Memory usage} > 80\%$, increase buffer size, else assign external memory—path optimization and flush out cache memory using a specific set of algorithms.

If $0\% < \text{Network usage} > 80\%$, redirect the virtual instances in an unused path, else assign a new path closer to the destination—load balancing and prioritized packet transfer.

If $0\% < \text{Cluster usage} > 80\%$, redirect the virtual instance towards destination, else assign a new cluster—apply network partitioning and specify a new topology to handle critical instances.

If $0\% < \text{load/process time} > 80\%$, reduce the load by reassigning any of the four mentioned categories—apply utilization factor estimations and time-out mechanisms.

7.2. Software-Defined Policies for Smart Health

7.2.1. Software-Defined Policy for Resource Allocation. Resources offered in a global data network can be as follows:

- (1) CPU
- (2) Memory
- (3) Cluster
- (4) Network
- (5) Grid

The process parameters required can be as follows:

- (1) Process/load time
- (2) Throughput/packet delivery ratio (packets transmitted, received, and dropped)
- (3) Latency/delay/jitter
- (4) Congestion/network traffic
- (5) Link quality/SNR
- (6) Utilization factor/buffer size/BER

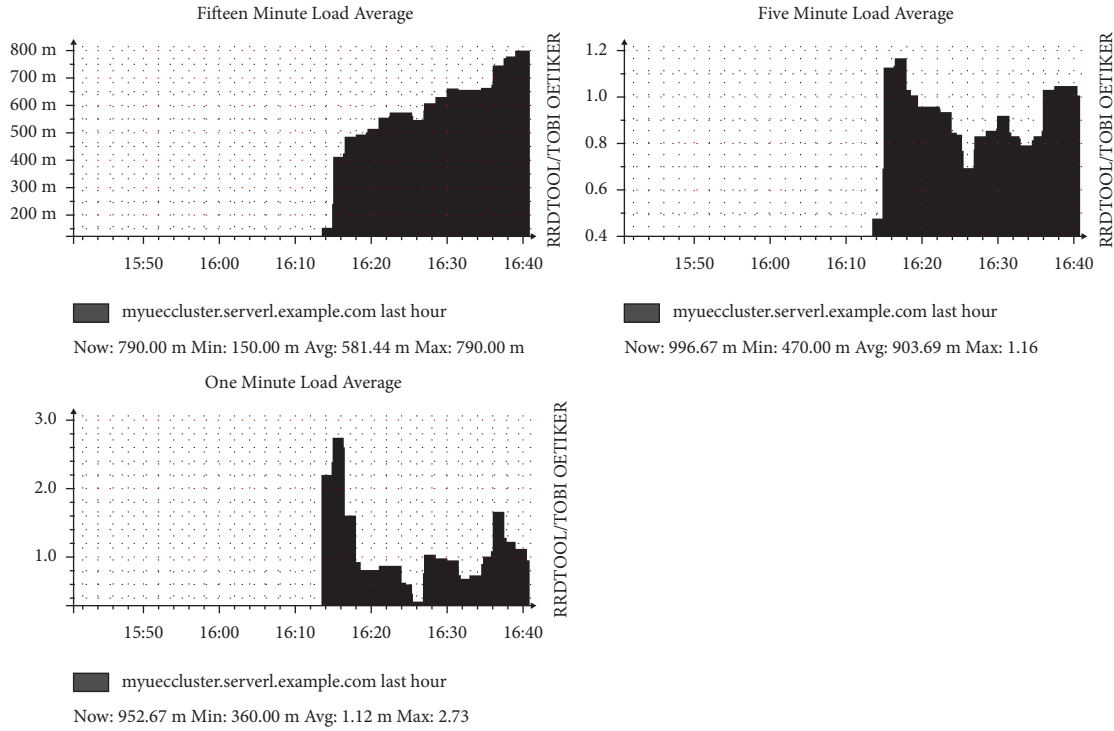


FIGURE 27: Load metrics over time.

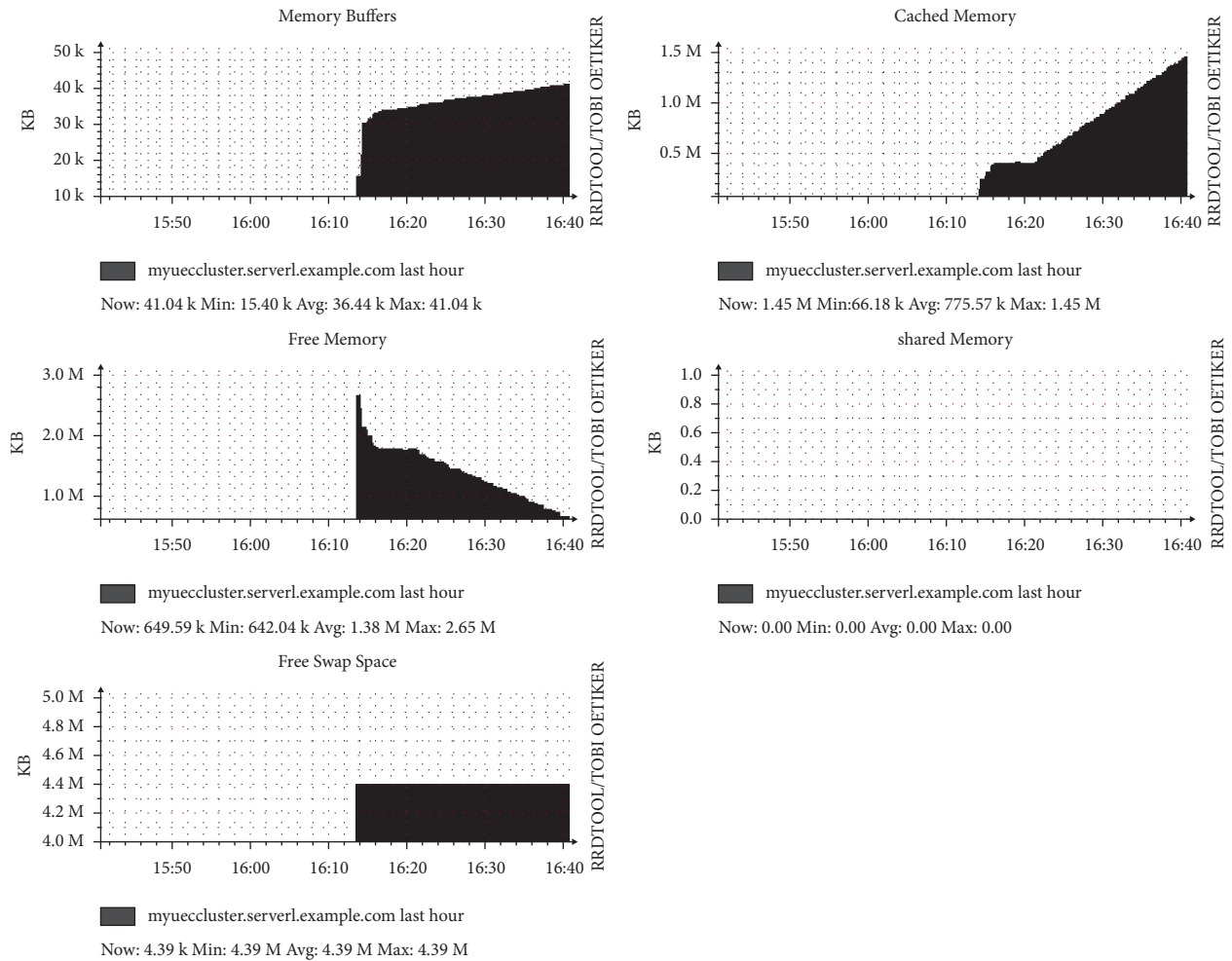


FIGURE 28: Memory metrics.

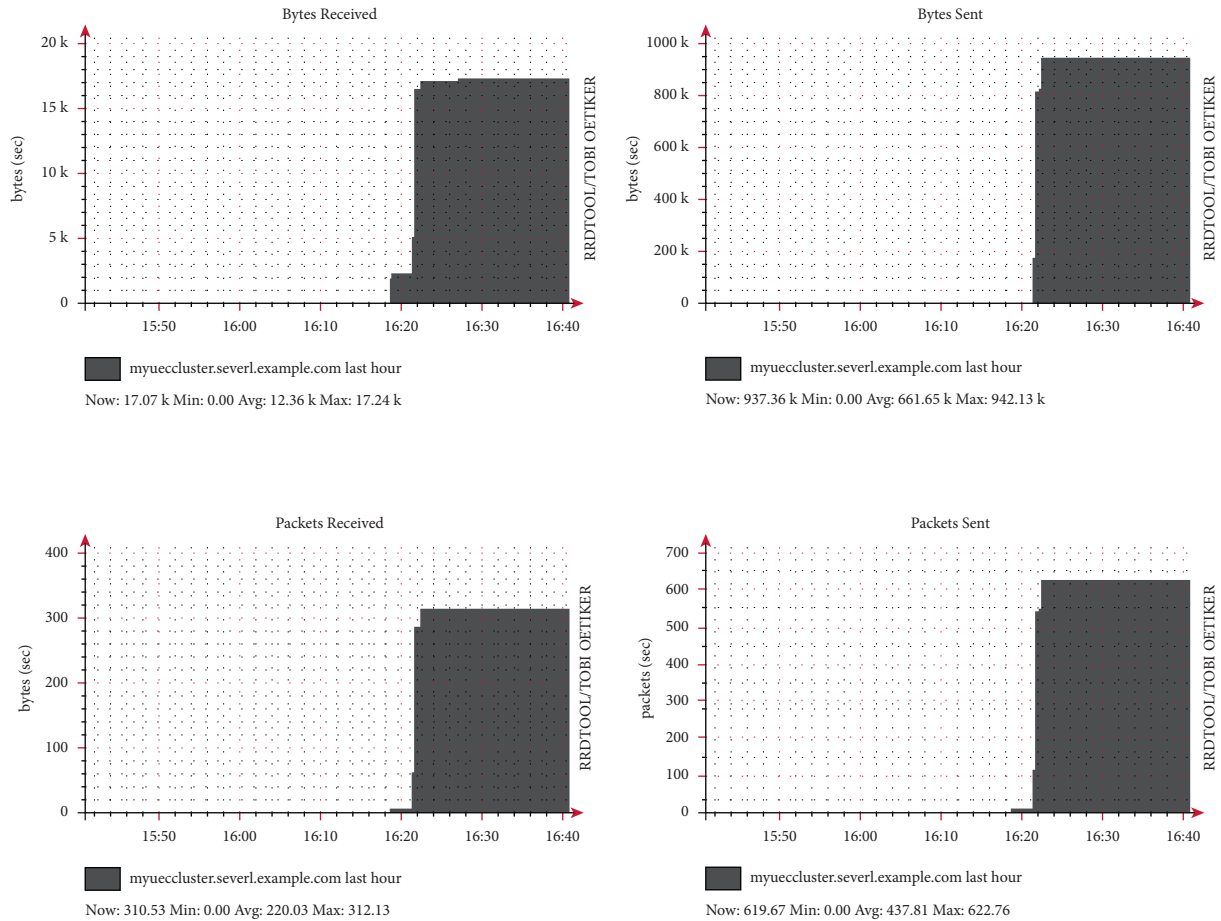


FIGURE 29: Network metrics.

- (7) Software-defined policy for the smart healthcare application was based on the three mentioned categories
- (8) Parameter-/data-centric—high, medium, normal, and low
- (9) Network-centric—topology, flow, and routing protocol
- (10) Node-centric/IP address—location based and distance-based (hop distance)

(1) *Case I: Parameter-/Data-Centric.* If the healthcare data relating to criticality is of high value, it is mentioned as CF_{IM} (critical factor – immediate), which lies on or nearer to both the extremities of death condition values. For this kind of scenario, emergency path allocation was done.

If the healthcare data relating to criticality is of medium value, it is declared as CF_{med} (critical factor – medium), which lies in the medium range of the disease spectrum, where the diagnosis can be done at a moderate level. For this kind of scenario, a congestion-free path can be allocated through the network.

If the healthcare data linking to criticality is of normal value, it is stated as CF_{nor} (critical factor – default), where continuous patient monitoring is required. For this kind of

scenario, the data was allowed to flow through the paths with the least congestion/congestion-free paths to reach the destination.

If the healthcare data relating to criticality is of low value, it is revealed as CF_{low} (critical factor – low), which means these types of patients can wait for a while to meet the doctor for a diagnosis. For this kind of scenario, the data was allowed to flow through the available paths/less-congested paths with an average waiting time of patients in the hospital for the diagnosis.

(2) *Case II: Network-Centric.* Analyzing the network characteristic, the data have to reach the destination in time based upon the following three factors:

- (1) Topology-based: response time
- (2) Flow-based: ingress/egress
- (3) Routing-protocol-based: proactive/reactive

Topology-based: response time. The healthcare data that is fed as an input to a system connected through any topology, say star, mesh, ring, and so on, has to be forwarded to the destination that is the nearest doctor or hospital based on availability. A dynamic software mechanism can be built in the forwarding/routing/gateway node, so the data can reach the

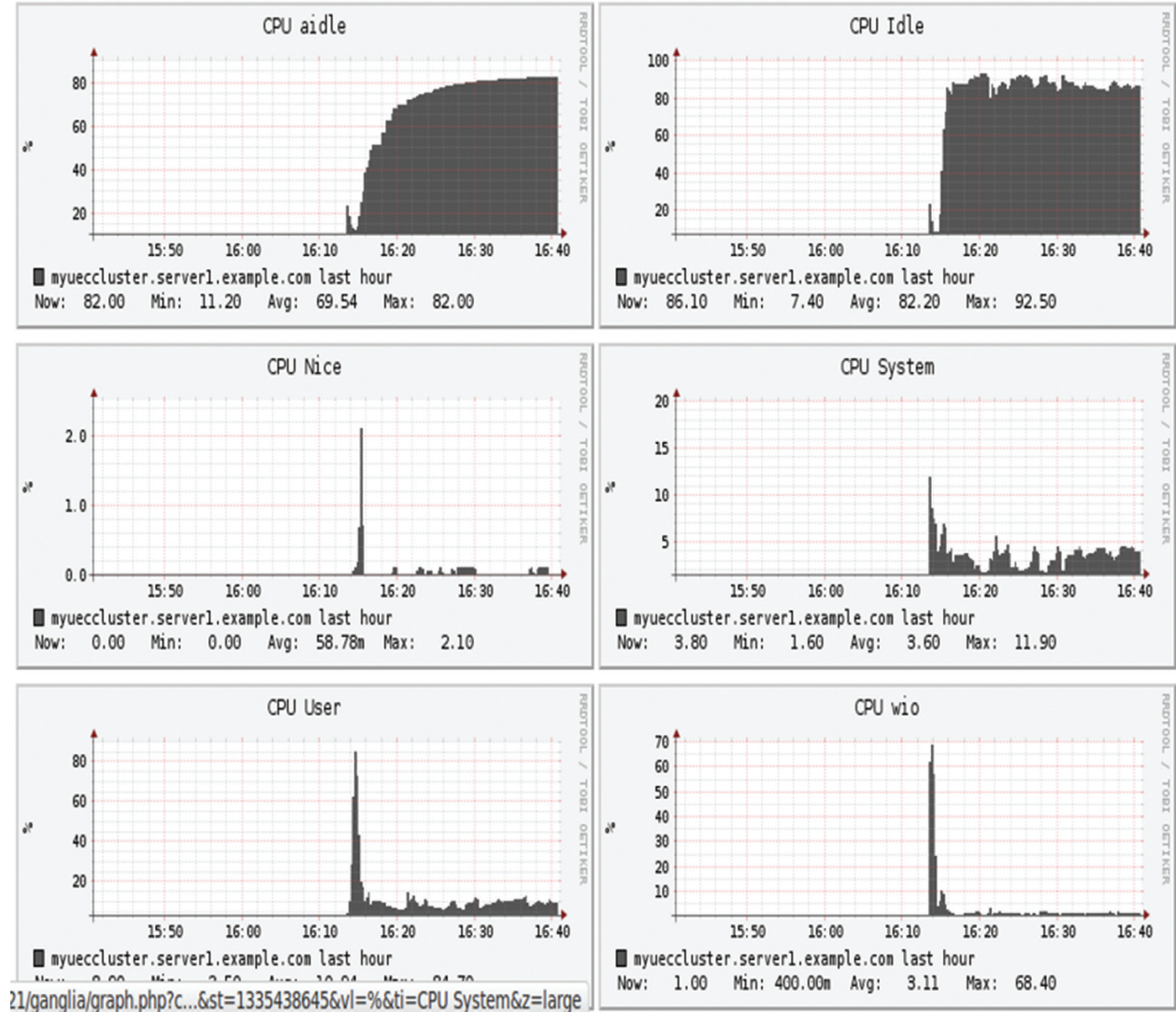


FIGURE 30: CPU metrics.

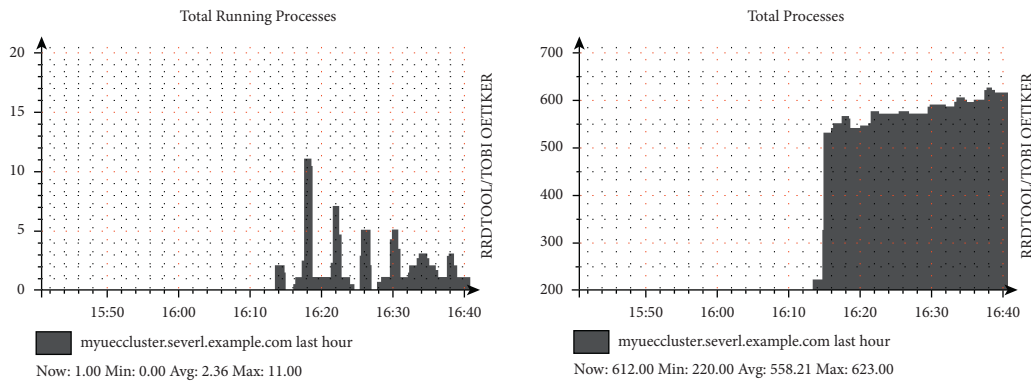


FIGURE 31: Process metrics.

TABLE 2: Sample data requirements for resource allocation.

| Number of instances | Packets transmitted | Packets received | Packet loss | Packet delivery ratio/throughput | Signal-to-noise interference ratio (dB) | Network delay (ns) | BER | Resource needed |
|---------------------|---------------------|------------------|-------------|----------------------------------|---|--------------------|------|-----------------|
| 1,20,000 | 120 | 118 | 02 | 98.33 | 98 | 0.002 | 0.67 | CPU |

destination, based on response time requirements. When the topology change occurs, if the data from the forwarding node can reach the destination node at a lesser time than the previously connected topology, the new topology change has to be accomplished.

Flow-based: ingress/egress. Depending upon the inflows and outflows of a particular node or a system, the waiting time of a healthcare data packet is made nil. For the flow to be processed, according to the flow table entry, the actual number of servers in the up condition was identified using the link quality factor, and then the flow was forwarded towards the responsive server.

Routing-table-based: the routing protocol. It can be used for transferring the patient data is categorized typically as proactive and reactive.

Under proactive protocol-based routing, the patient data that are under critical conditions can be transferred, and the patients under continuous monitoring can also be transferred for doctor's diagnosis at any time.

Under the reactive protocol, if a patient's cumulative healthcare data for a specific time is required for analysis by the expertise, the router will forward the historical data for diagnosis.

(3) *Case III: Node-Centric (IP Address)*. TCP/IP provides end-to-end connectivity. Node-centric data avoid the communication overhead by combining the features of TCP/IP specification of how data can be packetized, addressed, transmitted, routed, and received at the destination. A communication network should allow a user to focus on the data he or she needs, wherein smart healthcare application, the ultimate aim is to provide the required healthcare to any patient at any time regardless of the geographical location, the patient is situated or met an accident or in trauma. Therefore, to refer to a specific, physical location where that data is to be retrieved is required. Node-centric data networking comes along with the concept of TCP/IP such as data caching to reduce congestion and improve delivery speed, simpler configuration of network devices, and building security into the network at the data level shown in Figure 32.

Communication in node-centric networking is driven by receivers at the hospital side through the exchange of two types of information packets: IP address and patient data. Both types of packets carry a location name that can be transmitted in a single data packet.

IP address (source): the patient data is kept in a data packet along with the IP address of the node that is the creator and transfers the data. This data is embedded into an IP packet that is then pushed to flow in the network towards the destination using a congestion-free path based on the response time requirement of the packet.

Destination: once the embedded packet reaches the destination node in the nearby hospital, based on

proactive or reactive mode, the node will return an acknowledgment that contains both the IP address and the patient data packet along with a signature by the destination's IP address, which binds the two. This packet follows in reverse the path taken by the IP addressed data interest to get back to the source.

7.2.2. Software-Defined Policy on Process Parameter Requirement

- (1) Process/load time: based on the number of instances, considering data availability and scalability issues,

$$\sum_{i=1}^n \frac{e_i}{p_i} = \sum_{i=1}^n u_i \leq 1. \quad (23)$$

A task set is schedulable under critical instances; e_i is the execution time of the task in the processor based on its specification. p_i is the period of task, and u_i is the utilization factor. The earliest deadline first algorithm is modified with the criticality and the emergency of the patient, which will suitably take up the process can be preferable.

- (2) Bandwidth demands: a real-time signal processing application like healthcare data processing computes in each sampling period one or more outputs. Each output $X(k)$ is a weighted sum of n inputs $Y(i)$, expressed as follows:

$$X(k) = \sum_{i=1}^n a(k, i)Y(i). \quad (24)$$

The weights $a(k, i)$ are known as per prioritization of the data and fixed. The processor time-bound is also measured based on the response time requirement of producing a specific number of outputs in each sampling period.

- (3) Execution time: execution time is the amount of time required for the processor to complete the execution of the task J_i when it executes alone and has all the resources it requires.
- (4) Precedence constraints: AND/OR precedence constraints can be evaluated to combine the simultaneous critical events at the same location, where the number of emergency cases is more.
- (5) Usefulness function: the usefulness function is used to describe qualitatively the real-time performance objectives of the smart healthcare system. It emphasizes how a single patient can feel concerned about his/her medical diagnosis. The usefulness function can guide the choice and implementation of the scheduling strategies.
- (6) Lateness: lateness is defined as the difference between its completion time and its deadline mentioned to complete that is the response time of the patient data to be attended.
- (7) Priority driven approach for instance execution: the patient data is continuously perceived through the

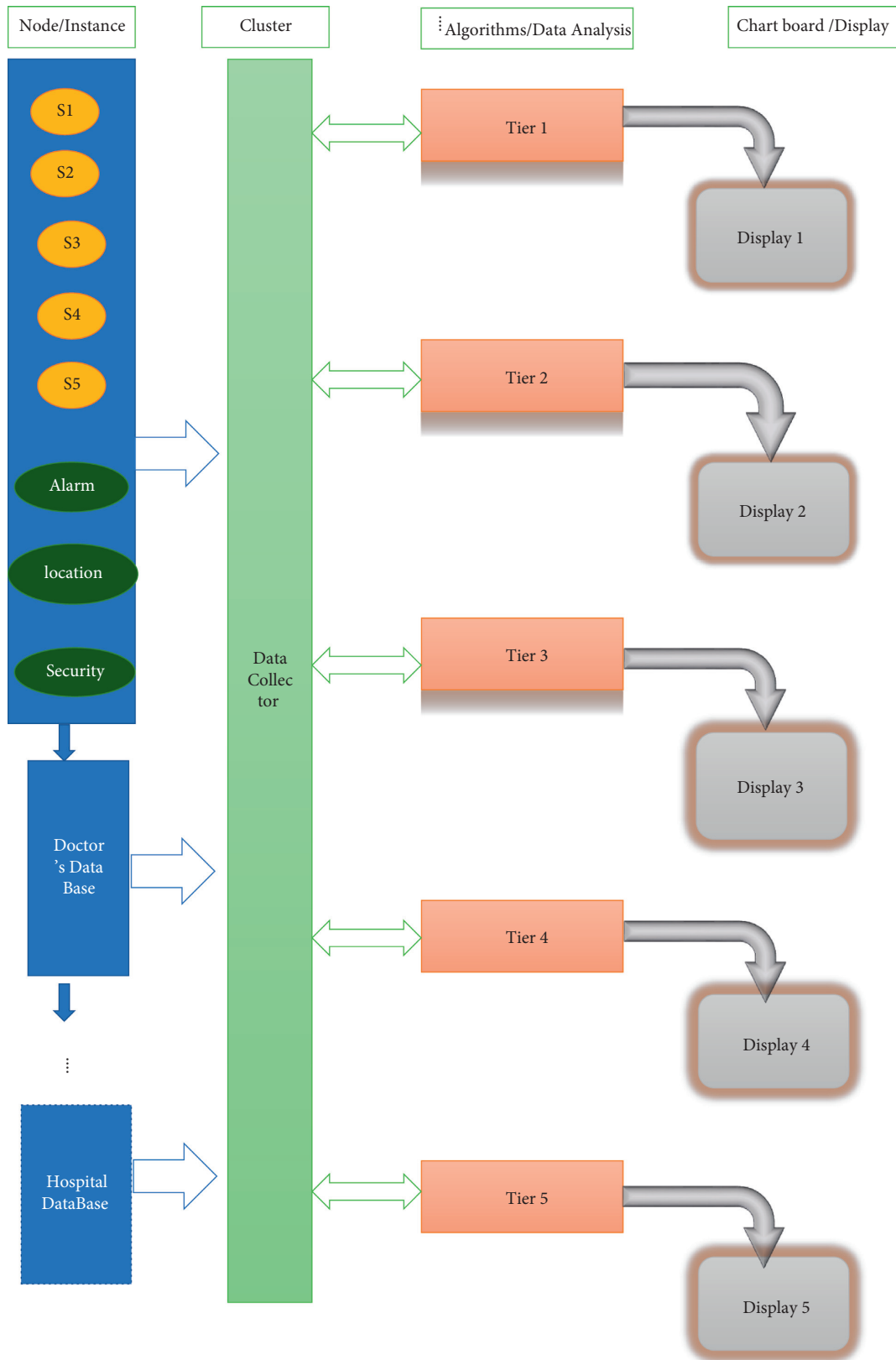


FIGURE 32: Node-centric data.

sensors and the streaming platform streams the patient data. As the patient data has been prioritized and categorized as normal, abnormal, and critical, the simulation environment must create instances appropriately. The patient's data collector has been considered as a single node. The Amazon EC2 console launches an instance for each node, which will be a part of the cluster. Also, the workload for the single cluster is required, based on the number of nodes, as the nodes specify the number of instances, against the cluster, a separate instance for that workload is also created. To know the survivability of the clusters, minimum of three nodes have been run under trial. The instances rely on Amazon Time Sync Service for the clock synchronization among the nodes. When choosing an AMI, some of the machines are preconfigured to use Amazon Time Sync Service, whereas the other machines are not preconfigured.

There are three instance types used such as m for general purpose, c for compute-optimized, and i for storage-optimized, with instance store volumes. We can have, for example, 16 vCPUs and 32 GB of RAM per instance, for internal testing depending upon our workload. When creating an instance for the node, a private key file must be downloaded to be used to securely connect to the instance. The location of the file must be decided and the file path also.

To configure our network, the custom TCP inbound rules to the security group must be enabled to allow TCP communication on two ports for inter- and client-node communication. This task enables the nodes to work as a cluster and make the load balancer route the traffic to the nodes and the healthcare application to connect to the load balancer. A procedure to timer activation, instance creation, network instigation, security mechanisms, and triggering the entire operation through simulation is as follows.

7.2.3. Procedure

- (1) Activate timer
- (2) Instances remaining = 1000; //Observe the number of instances at a specific time period
- (3) Next release time = processor clock + 100 units; // Current processor time at source + required response time //Arrival time at the destination
- (4) While (instances remaining > 0)
- (5) Now = processor clock
- (6) If (now < next release time), do
- (7) Timer sleep until (next release time)
- (8) Instances waiting in the specific type of task; //based on the criticality of the data such as high, medium, normal, and low
- (9) Next release time = next release time + 100 units
- (10) Else

- (11) Instances in the program of the aperiodic task; //Any patient data can be injected into the network at any time
- (12) Next release time = Now + 100 units
- (13) Instances remaining = instances Remaining - 1
- (14) End while
- (15) Thread-destroy (thread-ID)

7.3. Advantages of Ganglia in Performance Monitoring. A clustering coefficient quantifies the degree to which nodes in a graph cluster. To create tightly knit groups with a relatively high density of ties within the WBAN network. The probability of ties is greater than the average probability of establishing a tie randomly between two nodes [38, 39]. There are two types of tie connectivity between nodes: global and local.

The global version was created to provide an overview of the network's clustering, whereas the local version provides information about the embeddedness of individual nodes within WBAN networks. Both are useful in terms of performance dimensions.

The global clustering coefficient is computed using node triplets. Three nodes are connected by either two or three undirected ties to form a triplet. The global clustering coefficient is defined as the ratio of closed triplets to total triplets. This metric indicates network clustering and can be applied to both undirected and directed networks, both of which are referred to as transitivity [42, 43].

The global clustering coefficient is defined as follows:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}} \quad (25)$$

$$= \frac{\text{number of closed triplets}}{\text{number of connected triplets of vertices}}.$$

Each triangle forms three connected triplets. A generalization to weighted networks was proposed [44], and a redefinition to two-mode networks was expressed after that [45].

The local clustering coefficient of a node in a network shows looseness towards its neighborhood to form a clique. Such a graph can be treated as a small-world network.

A graph $G = (V, E)$ formally consists of a set of vertices V and a set of edges E between them. An edge represented as e_{ij} connects vertex v_i with vertex v_j .

The neighborhood N_i for a vertex v_i can be represented by its adjacently connected neighbors as follows:

$$N_i = \{v_j, e_{ij} \in E \cap e_{ji} \in E\}. \quad (26)$$

As e_{ij} and e_{ji} indicates an undirected graph, K_i can be defined as the number of vertices, $|N_i|$, in the neighborhood, and N_i , of a vertex.

The local clustering coefficient C_i for a vertex v_i is given by the proportion of links between the vertices within its neighborhood divided by the number of links that could exist between them [40].

An undirected graph has the property that e_{ij} and e_{ji} are considered identical. Therefore, if a vertex v_i has K_i neighbors, $(k_i(k_i - 1)/2)$ edges could exist among the vertices within the neighborhood. Thus, the local clustering coefficient for undirected graphs can be defined as follows:

$$C_i = \frac{2|\{e_{jk}: v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}. \quad (27)$$

7.3.1. Network Average Clustering Coefficient. In addition to the global clustering coefficient, Watts and Strogatz define the overall level of clustering in a network as the average of the local clustering coefficients of all the vertices n .

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i. \quad (28)$$

This metric favors low degree nodes, whereas the transitivity ratio favors high degree nodes. A weighted average is identical to the global clustering coefficient when each local clustering score is weighted by

$$k_i(k_i - 1). \quad (29)$$

7.4. Case Study: A Sample Smart Health Network. Figure 33 presents the sample network that exists between the patients and the hospitals. S1, S2, . . . , S5 indicates the available patients at different locations, and they are not interconnected with each other. D1, D2, . . . , D5 indicates the availability of hospitals/doctors in predefined locations spreading geographically. The sample network connects both the patients and the hospitals through the existing paths. The available paths are mentioned as follows without considering the network congestion.

Based on the number of existing paths and available paths, patient data have to reach the destination within the mentioned response time. The patient data have to be collected from the patient using a data collector, and from the data collector, it reaches the fog node. From the fog node, it enters into the Internet to the cloud for further processing. The data transfer time between any two nodes depends upon the distance and the speed of data transfer between the nodes in the hospital network. In the simulation, the following procedure has been adopted.

- Step 1. Create instances
- Step 2. Configure the hospital network—assigning the nodes
- Step 3. Synchronize clocks of the nodes in the network
- Step 4. Set up load balancing as the data can flow in any available path
- Step 5. Generate certificates for the nodes
- Step 6. Start nodes to function
- Step 7. Initialize the cluster, which comprises a set of nodes
- Step 8. Testing and configuring the cluster with all the nodes

- Step 9. Run a sample workload using the instances
- Step 10. Monitor the cluster for its performance
- Step 11. Scale the cluster for more number of nodes to participate
- Step 12. Use the database to store and for further compute operations
- Step 13. Estimate the bandwidth for the service and fix the data transfer speed
- Step 14. Consider the data size and ensure the reachability of data

To analyze the performance of the system, the following table exhibits the constraints behind it. The type of topology and congestion in the path are not now considered in the table shown in Table 3.

In Table 4, there is only one path available for the sources and destinations, regardless of the response time requirement, the only existing path that has to be chosen for data transfer, whereas if more than one path exists between the source and the destination, based on the response time requirement, whichever path is having less travel time was chosen for data transfer. If more than one path is having similar travel times, any path that is having less congestion was chosen for data transfer. These constraints were made as policies, which were fed into the software-defined controller to route the data packets based on the criticality of the event at the source. The number of virtual instances indicates that the same data packet can be delivered over many paths, with the request being served by whatever path is active at the time (Figure 34). The number of virtual instances to be created was estimated, based on the actual number of events and the weighting factor of the event. Weighting factor values are related to the emergency of the patient's health condition.

8. Concluding Remarks: Salient Features of the Proposed Cyber-Physical System

The WBAN architecture has been devised in terms of the five-tier architecture, and the scalability of extending the hospital network between interhospital service paradigms over the networking environment has been simulated with required diagrams. The solution we proposed provided the opportunity to interconnect not only doctors to the patients regardless of their locations but also interconnect the hospitals regardless of their geographic locations. There are specific algorithms devised to operate at every tier, and the simulation has been executed using Amazon creating the required instances of EC2 so as to know the performance and the scalability measures. The results show evident information about the data transfer mechanisms at normal and abnormal conditions, based on priority and ranking based on emergency conditions.

The performance factors have been estimated in a promised inexpensive, unobtrusive, and unsupervised monitoring method in normal and abnormal patient conditions. The technology fuses the cyber and physical phase of the system that is ubiquitous and affordable, and the

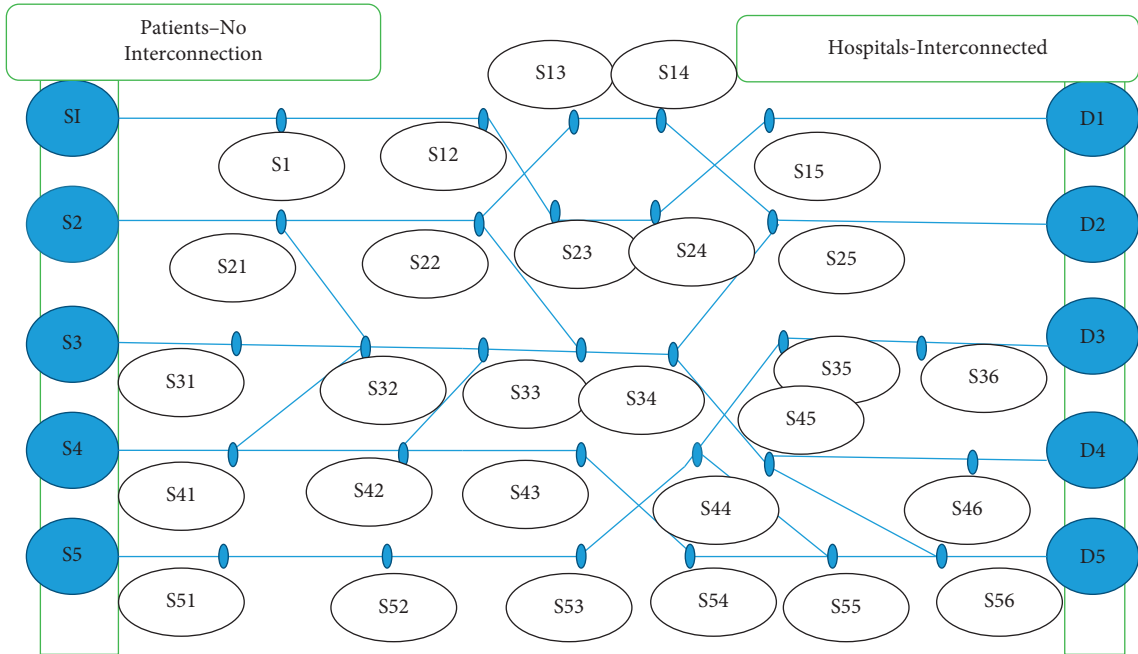


FIGURE 33: Sample network for smart healthcare application.

TABLE 3: Existing path in the sample network between patients and hospitals.

| From patient to hospital | Existing paths | Travel time in units |
|--------------------------|--|----------------------|
| S1-D1 | S1—{S11—S12—S23—S24—S15}—D1 | 0.06 |
| S1-D2 | No path | — |
| S1-D3 | No path | — |
| S1-D4 | No path | — |
| S1-D5 | No path | — |
| S2-D1 | No path | — |
| S2-D2 | S2—{S21—S22—S13—S14—S25}—D2 S2—{S21—S32—S33—S34—S25}—D2 S2—{S21—S22—S34—S25}—D2 | 0.06 0.06 0.06 |
| S2-D3 | No path | — |
| S2-D4 | S2—{S21—S32—S33—S34—S45—S56}—D5 S2—{S21—S22—S34—S33—S32—S41—S42—S43—S54—S55—S56} D5 | 0.07 0.12 |
| S2-D5 | S2—{S21—S32—S33—S34—S45—S56}—D5 S2—{S21—S32—S41—S42—S43—S54—S55—S56}—D5 | 0.07 0.09 |
| S3-D1 | No path | — |
| S3-D2 | S3—{S31—S32—S33—S35—S25}—D2 S3—{S31—S32—S21—S22—S13—S14—S25}—D2 | 0.06 0.08 |
| S3-D3 | No path | — |
| S3-D4 | S3—{S31—S32—S33—S34—S45—S46}—D4 | 0.07 |
| S3-D5 | S3—{S31—S32—S33—S34—S45—S56}—D5 | 0.07 |
| S4-D1 | No path | — |
| S4-D2 | S4—{S41—S32—S21—S22—S13—S14—S25}—D2 S4—{S41—S32—S33—S34—S25}—D2 | 0.08 0.08 |
| S4-D3 | No path | — |
| S4-D4 | S4—{S41—S42—S33—S34—S45—S46}—D4 | 0.07 |
| S4-D5 | S4—{S41—S42—S43—S54—S55—S56}—D5 | 0.07 |
| S5-D1 | No path | — |
| S5-D2 | No path | — |
| S5-D3 | S5—{S51—S52—S53—S44—S35—S36}—D3 | 0.07 |
| S5-D4 | No path | — |
| S5-D5 | S5—{S51—S52—S53—S44—S55—S56}—D5 | 0.07 |

TABLE 4: Chosen path as per response time requirement.

| Type of event (TE) | Number of events (E_i) | Weighting factor (W_i) | Number of actual instances = $\sum_{i=0}^n (A_i) = E_i W_i$ | Number of existing paths from source to destination ($S_p - D_p$) | Number of virtual instances (V_i) | Number of available paths (A_p) | Source-destination (S-D) | Required response time (R_i units) | Chosen path (S-D) |
|--------------------|----------------------------|----------------------------|---|---|---------------------------------------|-------------------------------------|--------------------------|---------------------------------------|----------------------|
| High | 10 | 1 | 10 | 1 | 10 | 1 | S1-D1 S5-D5 | 0.06 0.08 | S1-D1 S5-D5 |
| Medium | 100 | 0.75 | 75 | 3 | 225 | 3 | S2-D2 S2-D4 | 0.06 0.07 | S2-D2 (any) S2-D4 |
| Normal | 200 | 0.5 | 100 | 0 | 0 | 0 | S3-D3 S3-D4 | 0 0.08 | NIL S3-D4 |
| Low | 50 | 0.25 | 12.5 | 1 | 125 | 1 | S4-D4 S5-D5 | 0.07 0.07 | S4-D4 S5-D5 |

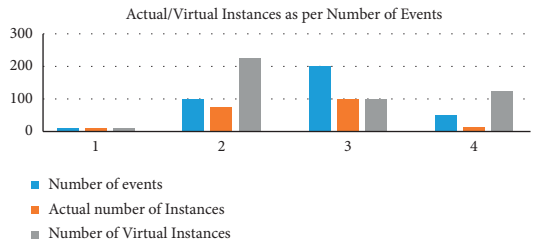


FIGURE 34: Performance analysis as per the number of events.

methodology is a success with the scalability factor also. To make it realizable, the system design, configuration and customization, seamless integration, security and privacy of patient data, and other social issues have to be taken care of.

Fusing physical and cyberspaces through an application is implemented in this research. It can be applied to any domain, where the data usage becomes global. Secure data forwarding is done through the cloud storage system, that is, messages can be forwarded directly between end users over and done with cloud services. Data availability and scalability are maintained using optimization techniques through software control mechanisms. Any patient at any location in the globe can get a timely and proper medical diagnosis through the cyber-physical system architecture. Extensive security and performance analysis are possible from data modification attacks due to the use of encryption using BoxCryptor v2.0 and cloud database storage using Eucalyptus v3.4. User identity management is supported within Eucalyptus with capabilities to control virtual resource pools using fine-grained role-based access control mechanisms for each resource pool. Ganglia supports scalability and automatic resource provisioning techniques with a threshold alert mechanism. This concept can be extended further with the clustering coefficient mechanism. If the rules for physicians are made flexible, to serve any patient anywhere in the world, Ganglia can be used to replicate data across sites [46].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request (sandeep.kautish@lbf.edu.np).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pp. 88–97, Association for Computing Machinery, Atlanta Georgia USA, 28 September 2002.
- [2] K. Lee and D. Hughes, "System architecture directions for tangible cloud computing," in *Proceedings of the 2010 First ACIS International Symposium on Cryptography, and Network Security, Data Mining and Knowledge Discovery, E-Commerce and Its Applications, and Embedded Systems*, pp. 258–262, IEEE, Suzhou, China, 23 October 2010.
- [3] K. Lee, D. Murray, D. Hughes, and W. Joosen, "Extending sensor networks into the cloud using amazon web services," in *Proceedings of the 2010 IEEE International Conference on Networked Embedded Systems for Enterprise Applications*, pp. 1–7, IEEE, Suzhou, China, 25 November 2010.
- [4] D. Johnson, K. Murari, M. Raju, R. B. Suseendran, and Y. Girikumar, "Eucalyptus beginner's guide-uec edition," Ubuntu Server, 2010 May, <https://fdocuments.in/reader/full/book-eucalyptus-beginners-guide-uec-edition1>.
- [5] R.-G. Lee, K.-C. Chen, C.-C. Hsiao, and C.-L. Tseng, "A mobile care system with alert mechanism," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 5, pp. 507–517, 2007 Sep 10.
- [6] S. Ivanov, C. Foley, S. Balasubramaniam, and D. Botvich, "Virtual groups for patient WBAN monitoring in medical environments," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 11, pp. 3238–3246, 2012 Jul 11.
- [7] A. Singh, V. Naik, S. Lal et al., "Improving the efficiency of healthcare delivery system in underdeveloped rural areas," in *Proceedings of the 2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*, pp. 1–6, IEEE, Bangalore, India, 4 January 2011.
- [8] I. D. Chakeres and E. M. Belding-Royer, "AODV routing protocol implementation design," in *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops*, pp. 698–703, IEEE, Tokyo, Japan, 23 March 2004.
- [9] Z. Iqbal, M. Gidlund, and J. Åkerberg, "Deterministic and event triggered MAC protocol for industrial wireless networks," in *Proceedings of the 2013 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1252–1259, IEEE, Cape Town, South Africa, 25 February 2013.
- [10] O. Diallo, J. J. P. C. Rodrigues, M. Sene, and J. Niu, "Real-time query processing optimization for cloud-based wireless body area networks," *Information Sciences*, vol. 284, pp. 84–94, 2014 Nov 10.
- [11] S. Movassaghi, M. Abolhasan, and J. Lipman, "A review of routing protocols in wireless body area networks," *Journal of Networks*, vol. 8, 2013 Apr 8.
- [12] H. Zhang, R. Zhang, Q. Li et al., "A GPRS-based wearable electronic thermometric alarm system," in *Proceedings of the 2012 5th International Conference on BioMedical Engineering and Informatics*, pp. 804–807, IEEE, Chongqing, China, 16 October 2012.
- [13] D. P. Tobón, T. H. Falk, and M. Maier, "Context awareness in WBANs: a survey on medical and non-medical applications," *IEEE Wireless Communications*, vol. 20, no. 4, pp. 30–37, 2013 Sep 12.
- [14] B. Bin Liu, Z. Zhisheng Yan, and C. W. Chang Wen Chen, "MAC protocol in wireless body area networks for E-health: challenges and a context-aware design," *IEEE Wireless Communications*, vol. 20, no. 4, pp. 64–72, 2013 Sep 12.
- [15] A. Pantelopoulou and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 1–2, 2009 Oct 30.
- [16] C. M. Chen, "Web-based remote human pulse monitoring system with intelligent data analysis for home health care," *Expert Systems with Applications*, vol. 38, no. 3, 2011 Mar 1.
- [17] E. Kafeza, D. K. W. Chiu, S. C. Cheung, and M. Kafeza, "Alerts in mobile healthcare applications: requirements and pilot study," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 2, pp. 173–181, 2004 Jun 7.

- [18] R. K. Megalingam, V. Radhakrishnan, D. C. Jacob, D. K. Unnikrishnan, and A. K. Sudhakaran, "Assistive technology for elders: wireless intelligent healthcare gadget," in *Proceedings of the 2011 IEEE Global Humanitarian Technology Conference*, pp. 296–300, IEEE, Seattle, WA, USA, 30 October 2011.
- [19] R. Cavallari, F. Martelli, R. Rosini, C. Buratti, and R. Verdone, "A survey on wireless body area networks: technologies and design challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1635–1657, 2014 Feb 13.
- [20] G. Bhuvaneshwari and A. Rama, "Reliable wireless patient monitoring system using hybrid transmission protocol," *Middle-East Journal of Scientific Research*, vol. 20, no. 4, pp. 468–474, 2014.
- [21] K. Malhi, S. C. Mukhopadhyay, J. Schnepfer, M. Haefke, and H. Ewald, "A zigbee-based wearable physiological parameters monitoring system," *IEEE Sensors Journal*, vol. 12, no. 3, pp. 423–430, 2010 Nov 11.
- [22] J. Aramideh and H. Jelodar, "Application of fuzzy logic for presentation of an expert fuzzy system to diagnose anemia," *Indian Journal of Science and Technology*, vol. 7, no. 7, pp. 933–938, 2014 Jul 1.
- [23] Kim, "A study on the acceptance factor for telehealth service according to health status by group," *Indian Journal of Science and Technology*, vol. 8, no. S1, pp. 542–550, 2015 Jan.
- [24] T. Tamura, I. Mizukura, M. Sekine, and Y. Kimura, "Monitoring and evaluation of blood pressure changes with a home healthcare system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 4, pp. 602–607, 2011 May 19.
- [25] R. S. Dilmaghani, Bobarshad, Ghavami, S. Choobkar, and Wolfe, "Wireless sensor networks for monitoring physiological signals of multiple patients," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 4, pp. 347–356, 2011 Mar 24.
- [26] Arsene, I. Dumitrache, and I. Mihu, "Expert system for medicine diagnosis using software agents," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1825–1834, 2015 Mar 1.
- [27] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, no. 2, pp. 155–163, 2009 May 1.
- [28] M. Hamim, S. Paul, S. I. Hoque, M. N. Rahman, and I. A. Baqee, "IoT based remote health monitoring system for patients and elderly people," in *Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques, (ICREST)*, pp. 533–538, IEEE, Dhaka, Bangladesh, 10–12 January 2019.
- [29] F. A. Almalki and B. O. Soufiene, "EPPDA: An efficient and privacy-preserving data aggregation scheme with authentication and authorization for IoT-based healthcare applications," *Hindawi WCMC*, vol. 2021, Article ID 5594159, 18 pages, 2021.
- [30] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pp. 13–16, Association for Computing Machinery, Helsinki Finland, 17 August 2012.
- [31] S. A. Haque, S. M. Aziz, and M. Rahman, "Review of cyber-physical system in healthcare," *International Journal of Distributed Sensor Networks*, vol. 10, no. 4, pp. 217–415, 2014 Apr 27.
- [32] B. Nathali Silva, M. Khan, and K. Han, "Big data analytics embedded smart city architecture for performance enhancement through real-time data processing and decision-making," *Wireless Communications and Mobile Computing*, 2017.
- [33] E. E. Egbogah and A. O. Fapojuwo, "A survey of system architecture requirements for health care-based wireless sensor networks," *Sensors*, vol. 11, no. 5, pp. 4875–4898, 2011 May.
- [34] P. Kumar and H. J. Lee, "Security issues in healthcare applications using wireless medical sensor networks: a survey," *Sensors*, vol. 12, no. 1, pp. 55–91, 2012 Jan.
- [35] S. Jabbar, F. Ullah, S. Khalid, M. Khan, and K. Han, "Semantic interoperability in heterogeneous IoT infrastructure for healthcare," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9731806, 10 pages, 2017.
- [36] D. D. Koutsouris and A. A. Lazakidou, Eds., Springer, 2014 Sep 25, <https://link.springer.com/book/10.1007/978-3-319-06844-2>.
- [37] A. Datt, A. Goel, and S. C. Gupta, "Monitoring list for compute infrastructure in Eucalyptus cloud," in *Proceedings of the 2015 IEEE 24th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 69–71, IEEE, Larnaca, Cyprus, 15 June 2015.
- [38] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, pp. 3747–3752, 2004 Mar 16.
- [39] M. Latapy, C. Magnien, and N. D. Vecchio, "Basic notions for the analysis of large two-mode networks," *Social Networks*, vol. 30, no. 1, pp. 31–48, 2008 Jan 1.
- [40] M. Kaiser, "Mean clustering coefficients: the role of isolated nodes and leaves on clustering measures for small-world networks," *New Journal of Physics*, vol. 10, no. 8, Article ID 083042, 2008 Aug 29.
- [41] Sacerdoti, Katz, Massie, and Culler, "Wide area cluster monitoring with Ganglia," in *Proceedings of the 2003 Proceedings IEEE International Conference on Cluster Computing*, pp. 289–298, IEEE, Hong Kong, China, 1 December 2003.
- [42] D. Barmpoutis and R. M. Murray, "Networks with the smallest average distance and the largest average clustering," <https://arxiv.org/abs/1007.4031>.
- [43] N. Deepa, B. Prabadevi, P. K. Maddikunta et al., "An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier," *The Journal of Supercomputing*, vol. 77, no. 2, pp. 1998–2017, 2021.
- [44] C. Dhanamjayulu, U. N. Nizhal, P. K. R. Maddikunta et al., "Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning," *IET Image Processing*, 2021.
- [45] G. Thippa Reddy, A. Srivatsava, K. Lakshmana, R. Kaluri, S. Karnam, and G. Nagaraja, "Risk prediction to examine health status with real and synthetic datasets," *Biomedical and Pharmacology Journal*, vol. 10, no. 4, pp. 1897–1903, 2017.
- [46] L. Matthew, Massie, N. Brent, and b Chun, "the ganglia distributed monitoring system: design, implementation, and experience " Elsevier," *Parallel Computing*, vol. 30, pp. 817–840, 2004.

Research Article

A Novel Random Error Approximate Adder-Based Lightweight Medical Image Encryption Scheme for Secure Remote Monitoring of Health Data

Nagarajan Manikandan,¹ Rajappa Muthaiah,¹ Yuvaraja Teekaraman ,²
Ramyia Kuppusamy,³ and Arun Radhakrishnan ⁴

¹School of Computing, SASTRA Deemed University, Thanjavur, Tamil Nadu 613 401, India

²Mobility, Logistics, and Automotive Technology Research Centre, Faculty of Engineering, Vrije Universiteit Brussel, Brussel 1050, Belgium

³Department of Electrical and Electronics Engineering, Sri Sairam College of Engineering, Bangalore City 562 106, India

⁴Department of Electrical & Computer Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia

Correspondence should be addressed to Yuvaraja Teekaraman; yuvarajastr@ieee.org and Arun Radhakrishnan; arun.radhakrishnan@ju.edu.et

Received 20 September 2021; Revised 14 October 2021; Accepted 27 October 2021; Published 23 November 2021

Academic Editor: Thippa Reddy G

Copyright © 2021 Nagarajan Manikandan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the present global scenario, social distancing is an inevitable one. The need for social distancing and advancements of technology to facilitate the patients and doctors around the world mandated the telemedicine and remote monitoring of patient details as the pivotal way to diagnose the disease. In this, it is essential to transmit the patient's information such as X-ray and scan images of them to the doctor in the remote location. Preventing the medical data from the technological adversaries is the need of the hour. Infinitesimal attacks in medical images may cost human lives. This work proposes a lightweight, secure medical image encryption scheme for the remote monitoring of health data. The proposed encryption scheme uses computationally less complex weighted shift approximate adder (WSAA)-based encryption logic. The scheme uses a 256-bit key for the encryption process that strengthens the encryption and robust against various attacks. The proposed encryption scheme deploys the WSAA for diffusing the pixel values. A unique way of key distribution for pixel-wise encryption within the image is proposed that avoids the need for separate logic for the pixel-wise confusion. The proposed Encryption scheme is evaluated for its entropy and horizontal, vertical, diagonal correlation, histogram, key space, and sensitivity. Experimental results affirm that the proposed scheme significantly good with less computational complexity. The peak signal-to-noise ratio (PSNR) value of the decrypted image is infinity, and this matches the ideal requirement of the medical encryption scheme.

1. Introduction

The recent technological development and advancement in the Internet of Things (IoT) facilitate the people to make use of resources available around the world. This rapid development of technology trends and enables the patients to get the support of doctors globally. The patient's information, such as clinical test reports, X-ray, and scan reports, are sent through the Internet medium. Most of the clinical data are in the form of an image [1]. The information over the Internet

is more prone to vulnerabilities and attacks. Any corruption or misinterpretation of medical data due to the intermediate attack is nontolerable, and it may endanger to human lives. This necessitates the scheme for the secure remote monitoring of health data. There are many image encryption schemes presented in the literature for the general images, but when it comes to medical image, those methods may not be secure enough to preserve the information of the patients. This is due to the strong correlative nature of pixels with its adjacent pixels in the medical images and various modalities

of capture [2]. Most widely used modalities in the medicinal field are X-ray, magnetic resonance imaging (MRI), ultrasonography (US), computed tomography (CT), and positron emission tomography (PET) [3–6]. This triggered and motivated the researchers to develop various robust encryption schemes [2, 3, 7–19].

Recent research uses efficient techniques such as watermarking [7–10], Arnold mapping [11, 12], chaotic mapping [13, 14], permutation [14, 15], compression [15], scrambling [16], genetic algorithm [2], swarm optimization [17], elliptic curve cryptography [18], and neural networks [19] individually or as a combination for hiding the medical information or authenticating the information. Researchers developed their encryption schemes and validated for different modalities of medical images to defy against attacks. The generic need of any cryptosystem is to keep high computation cost and time of breaking the cipher [7]. The entire above encryption schemes are good enough to maintain the generic requirements. Nevertheless, the scenario has changed, and one more requirement is also needed to be considered along with those two generic requirements.

Nowadays, computing systems are getting smarter and almost all the devices which use the Internet are embedded and low power systems in nature. This demanded the lightness of computation and added the lightweight computation of the scheme in the requirement list [20]. Based on the need of the hour, many lightweight encryption schemes have been proposed for general images [21–25] and medical images [4, 5, 17, 26]. Various exciting and unique techniques are introduced for encryption to make it light. The logic used in moves of the knight pawn in the ancient Indian game chess along with a genetic algorithm is used to encrypt the image [21]. This approach attains the secured encryption in four stages with the moderate performance of entropy and PSNR. Mondal et al. formulated a lightweight scheme; two pseudorandom numbers (PRNs) are used. One is to permute the plain image, and another is to encrypt it by deoxyribonucleic acid (DNA) computation. This method achieves better performance at the expense of computation complexity [22].

Javeed et al. presented a chaotic-based lightweight system which uses the chaotic oscillator for generating random numbers to scramble the pixels. Authors claimed that their system withstands for plaintext attacks and brute-force attacks. This system has the complexity of solving and makes use of second-order differential equations for generating the random numbers [23].

A hybrid method of encryption after compression is developed by Almalkawi et al. [24]. Here, 2D logistic chaotic and Henon chaotic maps are used for bit scrambling and pixel shuffling of the compressed image, respectively. This method needs more number of rounds for better encryption. A partial encryption method of encrypting the region of interest area is proposed by the Khashan and AlShaikh [25]. Authors used the edge detection principle to locate useful information in the region of interest, and information of edges are encrypted. This obviously reduces the overhead of computation, but it reveals other regions explicitly. Mortajez

et al. [26] developed a method of generating a key from the images and shuffling the pixel positions based the generated random numbers. The shuffled pixels are later encrypted by the sequence of logistic systems and XOR gates.

From the literature, it is evident that complex logics are needed to produce randomness. It also needs multiple rounds for achieving better encryption. In this work, we proposed a novel and unique scheme which avoids these overheads and attains good encryption.

1.1. The Significant Novelty and Contribution of Our Work in the Encryption and Decryption Process

- (1) A unique method of deriving random numbers for shuffling bit positions of the pixels is proposed.
- (2) Proposes a novel random error weighted (Hamming weight) shift approximate adder of less computational complexity for diffusing the bits in the pixel.
- (3) An effective key distribution strategy is proposed among the pixels within an image to ensure the randomness of encryption.
- (4) The two ways application of encryption process inherits itself the pixel confusion property and avoids the need for separate logic and computation complexity for the same.
- (5) Uses very simple arithmetic and logical operations, which makes the encryption and decryption process a lightweight (based on computational complexity). Cryptanalysis of the implemented scheme proves that the proposed scheme defy various attacks.

1.2. Organization. The rest of the manuscript is presented as follows: related works are discussed in Section 2. Section 3 describes the proposed encryption and decryption schemes. Implementation of the proposed scheme and its outputs are presented in Section 4. Security analysis of the proposed schemes is detailed in Section 5. The suitability of the proposed scheme for medical application is justified by comparing PSNR with existing works in Section 6. Section 7 concludes the article with the consolidated research outcomes.

2. Related Works

The tremendous growth of the IoT triggered the researchers to focus on the concept of lightweight cryptography recently, even though it existed well before IoT. Most of the lightweight works are described for general images and not for medical images due to its close correlation of pixels. This motivates and challenges the researchers. The development of lightweight computation without accuracy loss has acquired major attention. In this section, we presented closely related latest works which involve less computation for the encryption process.

A lightweight encryption scheme for implantable medical devices is presented by Belkhouja et al. [27]. This encryption scheme targeted information transfer through

the wireless medium. Authors claimed that they had ensured the patient's information by encrypting the data using a lightweight chaotic system generated keys. Abd El-Latif et al. [28] developed the simple encryption protocol with the new logic called controlled-NOT image. They performed encryption based on the NOT image generated from the logistic map. In addition to that, the method of key matrix generation from the embedding process is suggested to improve the encryption process. The encrypted image can only be decrypted by having both the logistic map and the key matrix.

The idea of hardware-dependent [29, 30] encryption schemes has started to evolve recently for better compatibility and security. Ravichandran et al. [29] implemented a low power medical image encryption scheme in the field programmable gate array (FPGA) using the penta-layer approach. Encryption of the medical image is done in five different layers. Each layer uses a different scheme of shuffling and scrambling and attains a secured encrypted image. High security is attained by producing hardware-dependent encryption in the fourth layer. Janakiraman et al. [30] created the hardware-dependent lightweight steganography scheme by considering resource constraints in the hardware for embedded applications. The authors explored the device-dependent implementation of the cryptography schemes, which exhibits robustness against various attacks.

Encryption methods for the secure transmission of DICOM images are presented [26, 31]. A chaotic secure encryption scheme with dynamic secret keys is developed by Mortajez et al. [26]. The system uses a periodic confusion strategy to encrypt the DICOM images. The scheme first extracts the key from the medical image, and pixels position has been permuted using the periodic confusion strategy. Then, pixels are encrypted based on random sequences of the logistic map and XOR operator. Based on the cryptanalysis, authors claimed that their scheme is able to withstand against statistical attacks. Manikandan and Amirtharajan [31] formulated the new way of scrambling and used a RC6 cipher encrypted approximate coefficients of the Harr wavelet transform for encrypting the medical images to withstand various attacks. With the sufficient key space, the algorithm effectively withstands the key hack.

Venkateswarlu [9] has implemented a fast medical image security algorithm for color medical images using both watermarking and encryption schemes for better security in each color channel. In this scheme, the patient's information, along with the smoothed key image, is embedded into the image color channels to produce a watermarked image. In the second stage, each watermarked color channel is separately encrypted to generate a final encrypted image. The reverse process is done on the decryption side. The author claims that it shows better resistance against key guessing attacks. The key which made the scheme robust is the actual set back of this method. Here, the image of size less or equal to the plain image needs to be used as a key, and it needs more bandwidth for transmission and needs to have more storage for the computation.

Madhusudan and Sakthivel [12] developed the image encryption approach that first represents image pixels in binary form. The binary values of the pixels have been shuffled by the two random numbers generated by the

Arnold map. The shuffled binaries are converted back to pixel integers and positions of the pixels are scrambled randomly with reference to the chaotic matrix to produce an encrypted image. The strength of the algorithm is justified by various security analyses. In this work pixels are not handled as an array; instead, it is converted into a sequence of numbers later diffusion and confusion are done. The shape conversion needs more logic to remember the indexes of each pixel when it is in the decryption stage, and it is an extra overhead.

Tamilarasi and Jawahar [17] designed a hybrid lightweight encryption algorithm with the swarm optimization technique (HLE-SO). HLE-SO combines the Paillier and KATAN methods to make it lightweight. The authors' utilized swarm optimization for managing key space which addresses the limitation of the key sizes of the lightweight encryption methods. They deployed the scheme for EEG medical data and simulated the algorithm using MATLAB. Authors mentioned that the KATAN algorithm used for lightweight features was subjected to 254 rounds.

Khashan and AlShaikh [25] presented a lightweight encryption scheme which encrypts a selected portion of the medical image. This scheme concentrates on the region of the interest portion of the medical image rather than useless black pixels around the actual medical image. To do that, this scheme first performs edge detection then encrypts the edges with the random keys generated by chaotic map using one-time pad algorithm. This method of selective portion encryption greatly reduces the computation overhead so that authors claim that it is a lightweight scheme. Authors claim their schemes robustness against various attacks. However, still, most of the portion is visible in the encrypted image that will lead to the guess of an image. Nevertheless, recovering entire information without the proper key is not possible.

Mubashar et al. [32] used a novel block chain based technique for the encryption of medical data along with the optimization algorithm to create a framework for the IoT based medical data archival system.

The existing works in the literature have its own merits and demerits. At the outset, frequent shuffling, more storage requirement, and a number of rotations are the major cons of the existing cryptosystems. Moreover, most of the works mentioned above are developed without considering hardware implementation complexities. Randomness and less complex computations are the desired pros of the cryptosystem. Our proposed architecture carefully handles the listed overheads and makes use of the novel weighted shift approximate adder to form a simple, lightweight, and secure encryption scheme.

3. The Proposed Scheme

This work is drafted to overcome the short comings of the existing lightweight cryptography schemes especially when they are implemented in portable devices. By carefully analyzing the needs, we proposed a scheme which has approximate adder as a core part and performs encryption and decryption process based on it. In this section, the research flow of the proposed scheme is explained from the design of

the novel proposed random error weighted shift approximate adder (WSAA) and its suitability for utilizing it in cryptography domain is analyzed. The encryption and decryption process are performed over various medical modality images using the MATLAB tool. Later, the strength of the proposed scheme is evaluated by applying various cryptanalysis methods and results of them are presented as proof of versatility of the proposed scheme with the less computation effort.

Our proposed random error approximate adder-based lightweight encryption and decryption scheme for securing remote medical data deploys the novel proposed random WSAA for bringing diffusion kind of property to the pixel values. In most of the cases, the diffused or permuted pixel and original pixel will have an equal Hamming weight. However, in our proposed scheme, weight after diffusion and before diffusion would not be equal. This gives additional security for the information in the diffusion phase.

3.1. Weighted Shift Approximate Adder. The proposed WSAA has two steps of computation. In the first step, an operand is circularly shifted toward right up to the number of positions equal to the Hamming weight of the given first operand. The weighted shift (WS) operand is added with the second operand and the shifted second operand in the second step. It can be observed that the Hamming weight of the sum is no way correlated to the source operand A or B. A sample calculation is presented in Figure 1.

The design expressions of the WSAA is given in the following equations (1)–(5):

$$\text{weighted shift}(A) = \text{circular shift}(A, \text{Hamming}(A)), \quad (1)$$

$$\text{Sum}_0 = \text{WS}_0(A) \oplus B_0 \oplus 0, \quad (2)$$

$$\text{Sum}_i = \text{WS}_i(A) \oplus B_i \oplus B_{i-1}, \quad (3)$$

$$\text{Carry}_{\text{in}} = \text{Carry}_0 = 0, \quad (4)$$

$$\text{Carry}_i = B_{i-1}. \quad (5)$$

In our propose, encryption scheme inputs for the WSAA is as follows:

$$A = \text{input pixel}, \quad (6)$$

$$B = \text{segmented key}, \quad (7)$$

$$\text{Carry}_i = \text{shifted segmented key}. \quad (8)$$

The segmented key is used as a second operand. The second operand is shifted toward left by one position. It is mentioned in equations (2), (3), and (7).

3.2. Error Characteristics of the Proposed WSAA. Error characteristics of the approximate adders describe the nature and accuracy of the results produced. In this work, the

proposed WSAA is aimed to produce more errors in random so that it could be used in encryption schemes. For the proposed 8-bit WSAA, the following error metrics [33, 34] are calculated:

$$\text{MED} = \frac{1}{2^{2n}} \sum_{i=1}^{2^{2n}} \text{ED}_i,$$

$$\text{NMED} = \frac{\text{MED}}{D}, \quad (9)$$

$$\text{ER} = \frac{\text{number of erroneous results}}{\text{total number of results}},$$

where ED_i is the error distance which equals to the absolute difference of i^{th} actual and i^{th} approximate result. MED is the mean error distance and NMED is the normalized mean error distance. D denotes the maximum result of 8-bit accurate addition, and ER indicates the error rate.

The calculated error parameters of the proposed 8-bit WSAA are presented in Table 1. From the characteristics, it is clear that the proposed adder can be able to produce 99.7% erroneous results and its MED is also high. Moreover, the produced results are also random in nature.

3.2.1. Statistical Characteristics of the WSAA. The probability of each value of output is $1/256$. Since 8-bit is only used to represent the result, the minimum value of the sum output is 0 and the maximum value of the output is 255. Each output value is produced by 256 unique combinations of inputs. This ensures the randomness of the produced output, and it minimizes the probability of guessing. Table 2 shows the few samples of combinations that produce 1 as the sum output.

In the conventional accurate adders, a specific number can be arrived as a sum result for a few related combinations and it is highly predictable. For instance, a resultant sum value 1 can be attained either by $0+1=1$ or by $1+0=1$. Similarly a bigger sum result may have “ n ” number of combinations to produce the sum as a result, yet it is easily predicted. However, in the proposed adder, the sum value is produced by various combinations of irrelevant numbers as shown in Table 2 and it is hard to predict the same.

Moreover, unlike the conventional and other approximate adders in the literature, the proposed WSAA has a peculiarity and it is noncommutative. Results of the same combination of inputs in a different order will give rise to different outputs, and the samples are given as proof in Table 3.

The proposed WSAA addition operation on a set S is said to be commutative if

$$\begin{aligned} \text{Input 1} + \text{Input 2} &= \text{Input 2} + \text{Input 1}, \\ &\text{for all Input 1} = \text{Input 2} \in S. \end{aligned} \quad (10)$$

WSAA addition operation on a set S is said to be noncommutative if

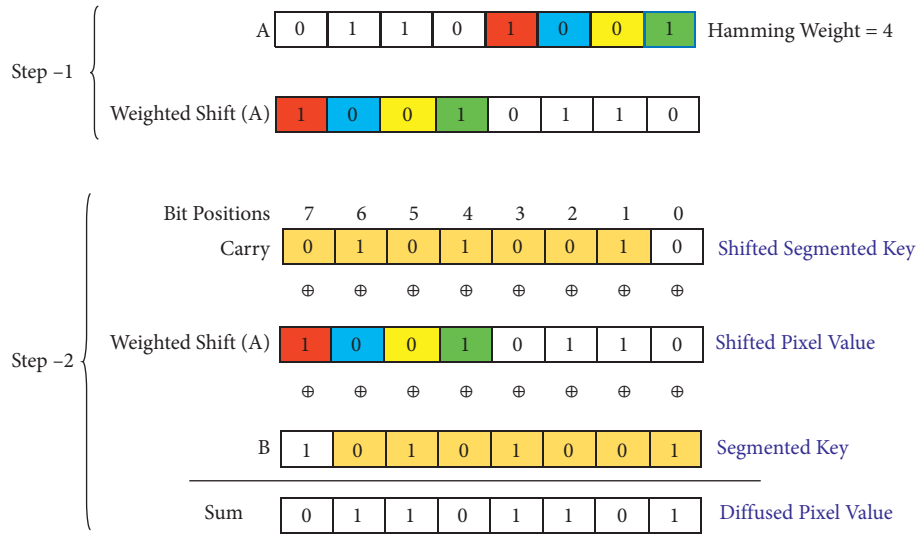


FIGURE 1: The proposed 8-bit weighted shift approximate adder (WSAA) steps with a sample calculation.

TABLE 1: Error characteristics of the proposed 8-bit WSAA.

| Adder | ER | MED | NMED |
|---------------------|-------|--------|-------|
| Proposed 8-bit WSAA | 0.997 | 149.25 | 0.293 |

TABLE 2: Samples of input combinations that produce the proposed WSAA sum output as 1.

| Input1 | Input2 | Actual sum | Approximate sum of WSAA |
|--------|--------|------------|-------------------------|
| 0 | 255 | 255 | 1 |
| 1 | 1 | 2 | 1 |
| 2 | 3 | 5 | 1 |
| 3 | 251 | 254 | 1 |
| 4 | 7 | 11 | 1 |
| 5 | 243 | 248 | 1 |
| 6 | 247 | 253 | 1 |
| 7 | 23 | 30 | 1 |
| 8 | 15 | 23 | 1 |
| 9 | 227 | 236 | 1 |
| 10 | 231 | 241 | 1 |

TABLE 3: Proof for the noncommutative property of the proposed WSAA.

| Input1 | Input2 | Actual sum | Approximate sum of WSAA |
|--------|--------|------------|-------------------------|
| 208 | 75 | 283 | 91 |
| 75 | 208 | 283 | 196 |
| 217 | 5 | 222 | 52 |
| 5 | 217 | 222 | 127 |
| 52 | 29 | 81 | 134 |
| 29 | 52 | 81 | 141 |
| 95 | 128 | 223 | 87 |
| 128 | 95 | 223 | 224 |
| 234 | 246 | 480 | 71 |
| 246 | 234 | 480 | 131 |

$$\text{Input 1} + \text{Input 2} \neq \text{Input 2} + \text{Input 1}, \quad (11)$$

for all Input 1 \neq Input 2 $\in S$.

The above statements are justified with a few sample inputs, and it is listed in Table 3.

The proposed approximate adder’s error and statistical characteristics make this adder an excellent fit for the cryptography domain.

3.3. The Proposed Encryption Scheme. The proposed scheme uses a 256-bit key for the encryption process. This huge key is not directly used for encrypting a single pixel. Rather it is to be used as a key space to get the 8-bit segmented key for encrypting the 8-bit pixels of the medical images. Most of the modalities used in the medical imaging are in grayscale, so the size of the key for the encryption is chosen as 8-bit. For better encryption, pixels should undergo diffusion and confusion process. In our scheme, diffusion in the pixel bits is attained by means of the WSAA and confusion of the pixel position is done by proper distribution of the segmented keys to the pixel positions in a horizontal and vertical way. The distributed segmented keys (Sk) are used for encrypting the pixels in two rounds.

Steps involved in the proposed encryption scheme are given below.

Input: Plain Medical Image

Output: Encrypted (Ciphered) Medical Image

Step 1: A 256-bit secret private key (KEY) is taken for the encryption process, and it is circularly right-shifted to a certain number of bit positions equal to its Hamming weight.

Step 2: Key segmentation and distribution for the pixels in the medical image for the first round (horizontal) is done with a row number of pixels (R_n) using the following formula:

$$m = (R_n) \text{ modulus } 32, \quad (12)$$

$$Sk_{R_n}[7: 0] = KEY[(7 + m * 8): (m * 8)]. \quad (13)$$

Step 3: Each and every pixel in the specific row number will be passed along with its segmented key to the proposed WSAA and evaluated as per equations (1)–(8). This completes the diffusion process.

Step 4: Steps 1 to 3 are to be repeated till all the pixels in the medical image is covered. Once all the pixels are covered, then move to Step 5.

Step 5: Key segmentation and distribution for the row diffused pixels in the medical image for the second round (vertical) is done with a column number of pixels (Cn) using the following formula:

$$m = (Cn) \text{ modulus } 32, \quad (14)$$

$$\text{Sk}_{Cn}[7: 0] = \text{KEY}[(7 + m * 8): (m * 8)]. \quad (15)$$

Step 6: Each and every row diffused pixel in the specific column number will be passed along with its segmented key to the proposed WSAA and evaluated as per equations (1)–(8). This completes the confusion process. Repeat Steps 5 and 6 until all the pixels are processed.

The process from Step 1 to 6 outputs encrypted medical image with the best entropy and poor correlation between adjacent pixels.

3.4. The Proposed Decryption Scheme. The decryption process is the reverse process of the encryption process. Steps followed in the encryption scheme are carefully reprocessed with the valid 256-bit key to gain the original medical image back at the receiver end.

Input: Encrypted (CIPHERED) Medical Image

Output: Plain Medical Image

Step 1: A 256-bit secret private key (KEY) is taken for the encryption process, and it is circularly right-shifted to a certain number of bit positions equal to its Hamming weight.

Step 2: Key segmentation and distribution for the pixels in the medical image for the first round (vertical) are done with the column number of pixels (Cn) using equations (14) and (15).

Step 3: Each and every column encrypted pixel in the specific column number will be passed along with its segmented key to the proposed WSAA and evaluated as per equations (16) and (20). This solves the confusion process and restores the pixel to supply for Step 6.

$$\text{WS}_0(A) = \text{Sum}_0 \oplus B_0 \oplus 0, \quad (16)$$

$$\text{WS}_i(A) = \text{Sum}_i \oplus B_i \oplus B_{i-1}, \quad (17)$$

$$\text{Carry}_{in} = \text{Carry}_0 = 0, \quad (18)$$

$$\text{Carry}_i = B_{i-1}, \quad (19)$$

$$A = \text{circular shift}(\text{WS}(A), \text{Hamming}(\text{WS}(A))). \quad (20)$$

Step 4: Steps 1 to 3 are to be repeated until all the pixels in the medical image are covered. Once all the pixels are covered, move to Step 5.

Step 5: Key segmentation and distribution for the restored pixels of the medical image for the second round (horizontal) are done with the row number of pixels (Rn) using equations (12) and (13).

Step 6: Each and every restored pixel in the specific column number will be passed along with its segmented key to the proposed WSAA and evaluated as per equations (16)–(20). This resolves the diffusion process. Repeat Steps 5 and 6 until all the pixels are processed.

The process from Step 1 to 6 outputs decrypted medical image with the best PSNR value that matches with ideal values and makes our scheme best fit for practical applications.

4. Implementation of the Proposed Scheme

The proposed WSAA-based lightweight medical encryption and decryption scheme has been implemented and validated using MATLAB 2019b. The scheme is applied and validated for different modalities medical images such as X-ray, CT, MRI, ultrasound, and PET. One sample image for each modality is taken for the evaluation. Our proposed scheme works well irrespective of the sizes of the images. For the fair evaluation and testing, two different sizes of images (512×512) and (256×256) are taken for testing the developed scheme.

The results of the encryption and decryption process, along with its histogram, are presented in Figures 2 and 3, respectively. The histogram of the encrypted image is given to prove the ability of the proposed scheme in distributing the pixels over the wide range equally. Among the images presented in Figure 2, X-ray chest, CT chest, MRI brain, and ultrasonography fetal images are of size 512×512 , and the PET brain image is of size 256×256 .

5. Security Analysis

The proposed cryptosystem is validated for its basic functionality, and cryptanalysis is done against various attacks. Our proposed scheme withstands for all the types of statistical attacks. The performance against various attacks is described in this section.

5.1. Histogram Analysis. Histogram of an image indicates the frequency of pixels in the image. An attacker with the histogram of an image can be able to guess the nature and some information of the image. For our proposed scheme, the histogram analysis was made and presented in Figures 2 and 3. Figure 2 compares the histogram of the original medical image and the encrypted medical image. From the figure, it is proved that pixels have been distributed in the encrypted image uniformly to the wide range. This nature

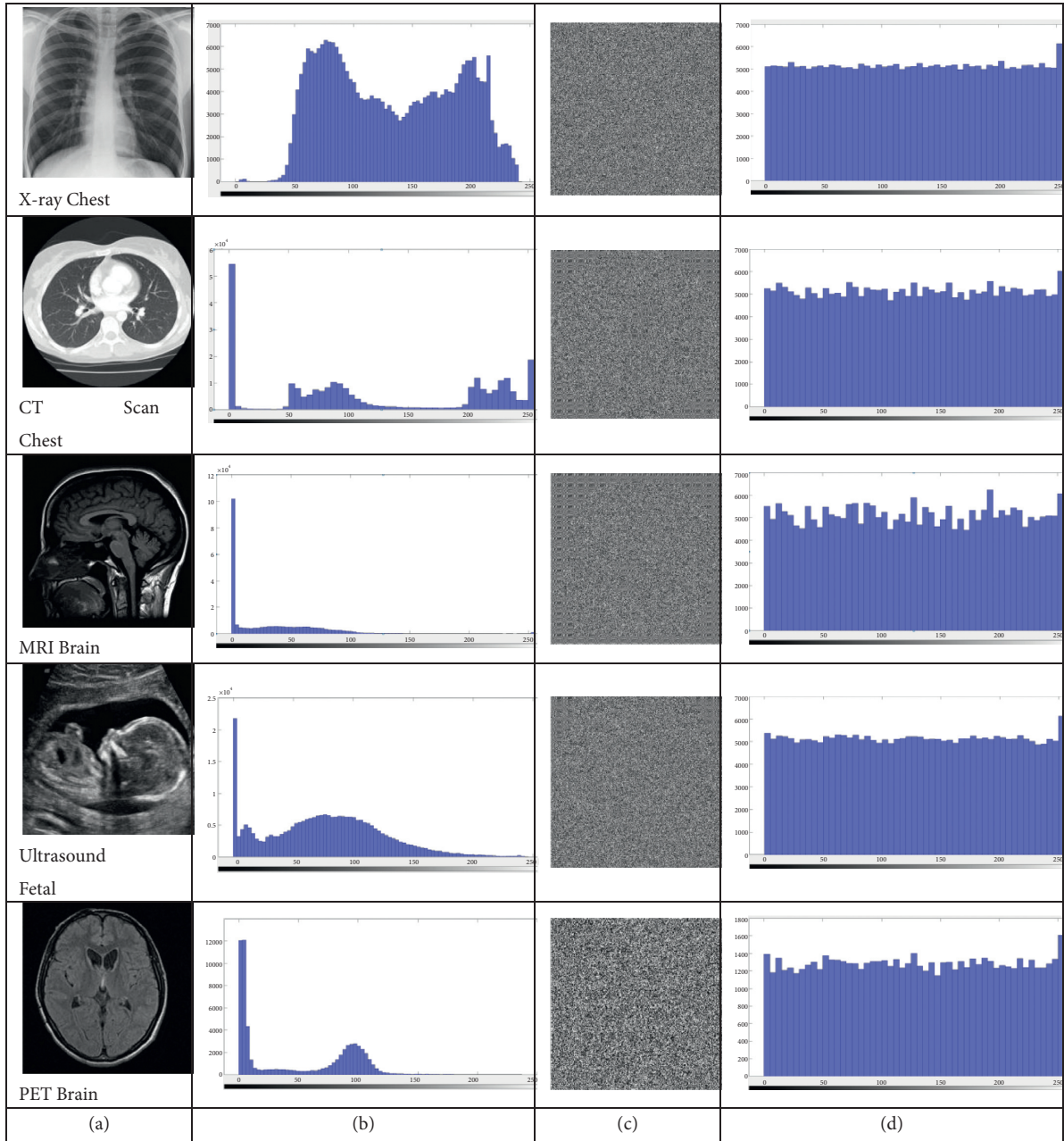


FIGURE 2: Implementation results of the proposed encryption scheme. From left to right column: (a) plain medical images subjected to encryption, (b) histogram of the plain medical images, (c) encrypted image of the plain images in the corresponding row, (d) histogram of the encrypted medical images.

creates a tough challenge for the attacker to gain any useful information from the encrypted image and makes statistical attack difficult.

5.2. Correlation Analysis. By nature, an adjacent pixel of any image has a close dependency with its neighbor pixels in all the direction. In fact, this property of an image only gives a smooth appearance to the image. This kind of adjacent pixel dependencies very high in the case of medical images having grayscale value, and it creates

really a great challenge for the cryptosystem to overcome this and create ciphered images. These dependencies aid the attackers to regain the original image with few iterations of guessing easily. Hence, the robustness of any cryptosystem could be analyzed easily with this correlation measurement.

$$\text{Cor}_{(a,b)} = \frac{\text{Cov}(a,b)}{\sqrt{\sigma(a)}\sqrt{\sigma(b)}} \quad (21)$$

where

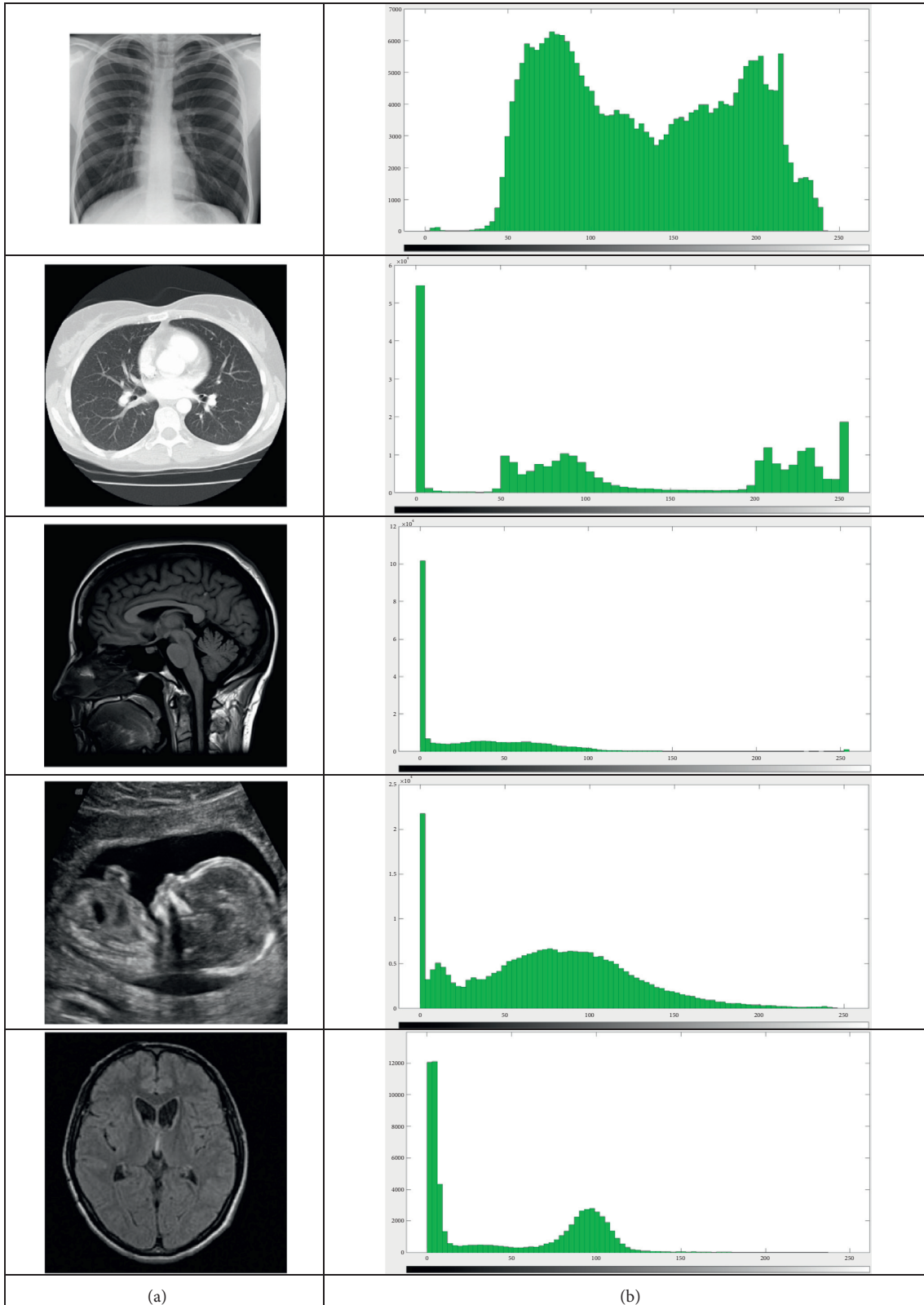


FIGURE 3: Implementation results of the proposed decryption scheme. (a) Decrypted medical images subjected to encryption, (b) histogram of the decrypted medical images.

$$\begin{aligned}\sigma(a) &= \frac{1}{N} \sum_{i=1}^N (a_i - E(a))^2, \\ \text{cov}(a, b) &= \frac{1}{N} \sum_{i=1}^N (a_i - E(a))(b_i - E(b)), \\ E(a) &= \frac{1}{N} \sum_{i=1}^N a_i.\end{aligned}\quad (22)$$

In the above equation, a and b are the adjacent pixels in the plain or ciphered image and N represents the number of pixels. Adjacent pixels are taken in a horizontal, vertical, and diagonal direction, and correlation has been calculated. The measured correlations at different adjacencies for the plain and encrypted medical image sample X-ray chest image are listed in Table 1. Pixel distribution of X-ray chest image is plotted before and after encryption, and it is given in Figure 4.

From the results shown in Figure 4 and Table 4, it is evident that the proposed encryption scheme minimizes the correlation among the adjacent pixels and distributes the pixels values of all the ranges evenly throughout the image. This makes statistical analysis tough to crack the image.

5.3. Key Space Analysis. The availability of high-end complex systems with graphical processing units (GPU) supports the brute-force attack. Hence, it is required to increase the key space more than 2^{128} to avoid the attack. However, lightweight cryptography limits the increase of the key length up to 256-bits. In our work, we carefully handled the key length and the number of bits used for the computation. The proposed scheme uses 256-bits as a secret key to confuse the attackers, and it uses 8-bit key internally. This increases the key space of the proposed scheme to 2^{256} for a single round. The proposed system uses two rounds. Thus, the total key space enlarged to 2^{512} , which is more than enough to resist the brute-force attack.

5.4. Key Sensitivity Analysis. The robust encryption scheme needs to be very sensitive to the changes in the key values. Any minimum variations in the key need to produce a different encrypted image in the encryption process, and the minimum variations should not decrypt any portion of the image information. Our proposed scheme has been evaluated for both the conditions with a single-bit variation in the key.

| | | | |
|--|---------|-----|----|
| Actual | key | K1 | is |
| 256'hF56C0062E818FFEA9F15E75DEF5EE81D- | | | |
| F656831C2C3E31B39C0C62BA5C51E | B4 | | |
| 1-bit | changed | key | K2 |
| 256'hF56C0062E818FFEA9F15E75DEF5EE81D- | | | is |
| F656831C2C3E31B39C0C62BA5C51EB5 | | | |

Encrypted image for different keys of 1-bit variation is shown in Figure 5. It is found that an encrypted image with key K2 is 98.458% different from the encrypted image with key K1. This proves that the proposed encryption scheme is

robust in producing randomness even for single-bit variations.

Results of sensitivity analysis of the proposed scheme at the decryption side with keys K1 and K2 are presented in Figure 6. Column (a) is the decrypted image with the actual key K1. Column (b) is the decrypted image with the fake key K2 of 1-bit change. The results indicate even a single-bit change in the key is not accepted by the proposed scheme, and it is highly sensitive to the original key.

5.5. Entropy Analysis. The unpredictability of the information related to images can be measured with entropy. Entropy analysis verifies the randomness and uncertainty of the encrypted image information. Entropy information of the images is calculated using Shannon's formula [35] given in equation (23), and it is represented in a number of bits.

$$H(p) = - \sum_{i=0}^{2^N-1} P(p_i) \log_2(P(p_i)). \quad (23)$$

In the above equation, $P(p_i)$ represents the probability of the pixel (p_i) and number of bits in this is represented by N . For a grayscale image of 256 levels, an ideal value of the uniformly distributed information is $N=8$ bits. Hence, it is always preferred to get the nearer entropy as that of the ideal case for the encrypted image. This ensures the uncertainty and randomness, and it is difficult to guess the encrypted pixel values. Deviation from this ideal value indicates the possibilities of disclosure of original information.

Measured entropy values of the plain images and encrypted sample images are listed in Table 5.

From Table 5, the entropy values of the ciphered images are very close to the ideal value 8. Since a fraction of information only available from the encrypted image, the proposed encryption scheme is secure against entropy attack.

5.6. Differential Attack Analysis. Differential attack is done to guess the key by analyzing the characteristics of the encrypted image, and this analysis corresponds to explore the strength of the encryption process against minute variations in the plain text or key [29]. Our proposed scheme is evaluated by the number of pixel change rate (NPCR) and unified average change in intensity (UACI). Analyzed values of various modalities of images are listed in Table 6.

5.7. Chosen Plain Text Attack Analysis. As the proposed encryption scheme uses xor logic gates in the adder circuits to encrypt the images, we attempted to do the chosen plain text attack to prove that our scheme is versatile and it is not revealing any information about the original image. The chosen plain text attack [29] is performed by following the equation

$$E_1(x, y) \oplus E_2(x, y) = I_1(x, y) \oplus I_2(x, y), \quad (24)$$

where I_1, I_2 are two plain images and E_1, E_2 are encrypted images of I_1, I_2 , respectively. In the above equation, if the

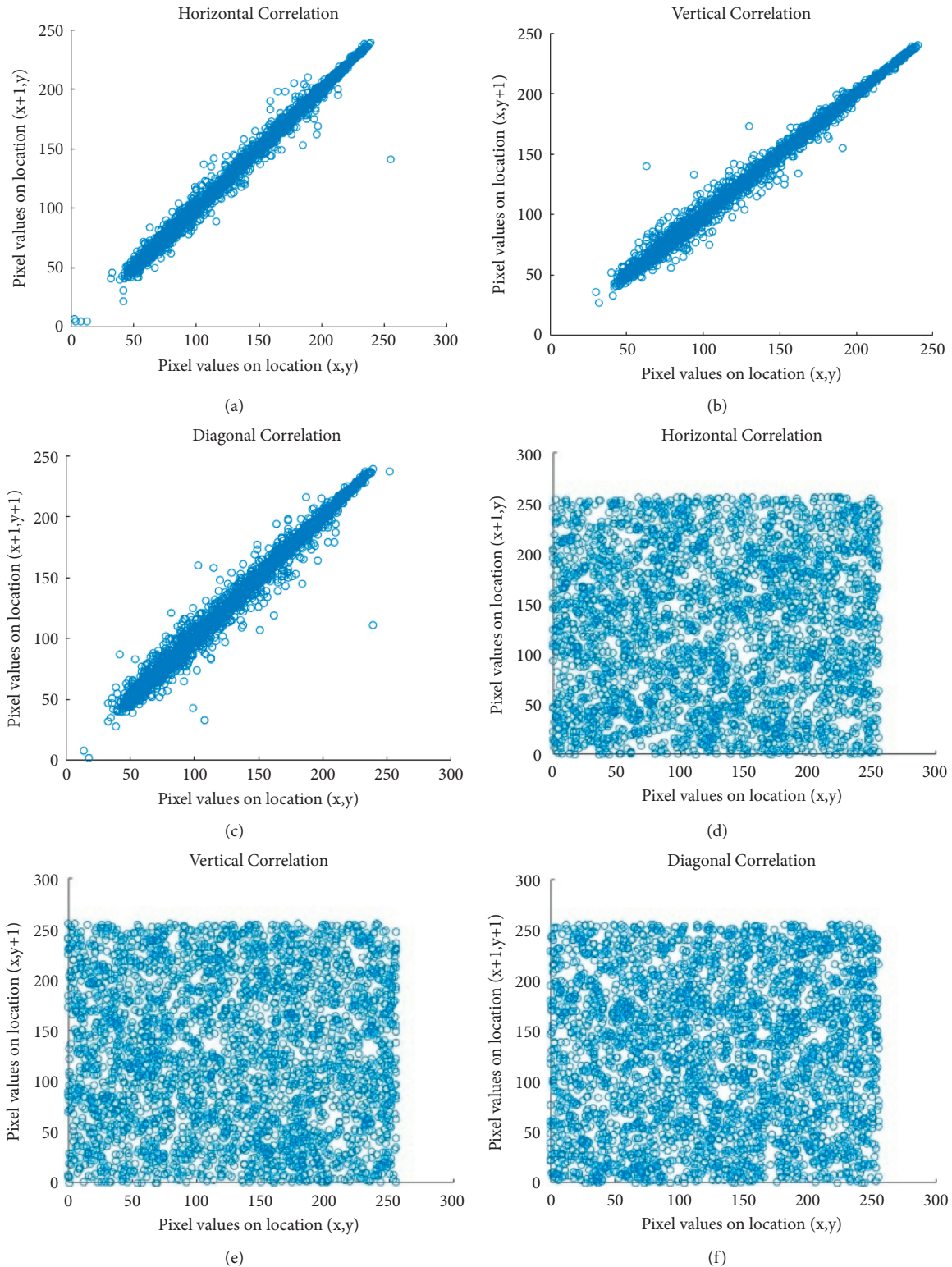


FIGURE 4: Correlation analysis results and plot of pixel distribution in the X-ray chest image sample. Pixel distributions plots are from left to right. For the plain image, (a) horizontal distribution, (b) vertical distribution, (c) diagonal distribution. Pixel distribution of encrypted image, (d) horizontal, (e) vertical, and (f) diagonal distributions.

statement is equal, then that encryption scheme is subjected to chosen plain text attack. Our encryption scheme successfully passed this analysis and produced unequal results.

5.8. PSNR Analysis and Comparison with Existing Works. This section presents the comparison of the peak signal-to-noise ratio of the decrypted image with the plain image.

TABLE 4: Correlation coefficients of different modality medical images.

| Sample medical image | Horizontal | | Vertical | | Diagonal | |
|----------------------|-------------|----------------|-------------|----------------|-------------|----------------|
| | Plain image | Ciphered image | Plain image | Ciphered image | Plain image | Ciphered image |
| X-ray chest | 0.9965 | -0.2898 | 0.9963 | -0.0205 | 0.9928 | -0.0109 |
| CT scan chest | 0.9943 | -0.0570 | 0.9895 | -0.0073 | 0.9873 | 0.0168 |
| MRI brain | 0.9777 | -0.1135 | 0.9833 | 0.0156 | 0.9581 | 0.0219 |
| Ultrasound fetal | 0.9938 | -0.0279 | 0.9863 | -0.0122 | 0.9809 | -0.0049 |
| PET brain | 0.9674 | -0.0766 | 0.9768 | 0.0316 | 0.9426 | 0.0194 |

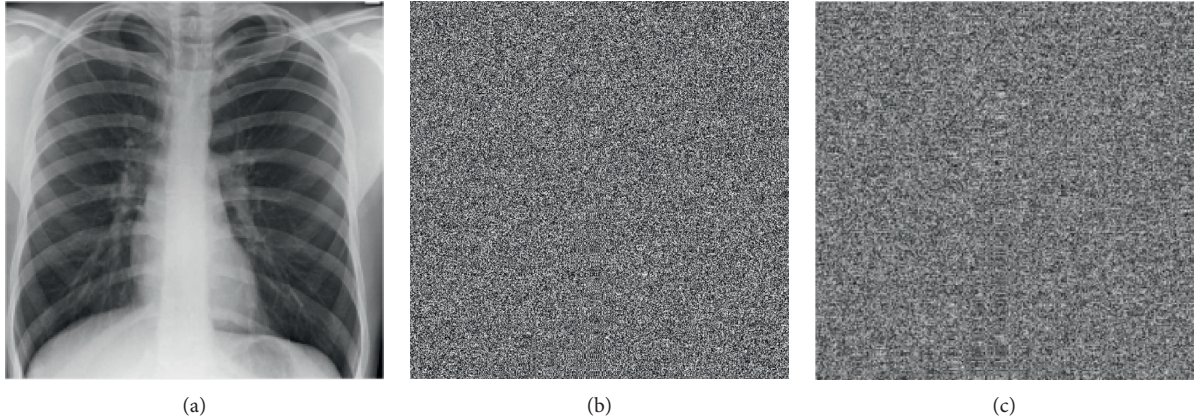


FIGURE 5: Key sensitivity test for single-bit variations in the key at encryption. (a) Plain X-ray chest image. (b) Encrypted image for actual key K1. (c) Encrypted image for 1-bit changed key K2.

Decrypted image and its histogram are shown in Figure 3. By comparing Figures 2 and 3, we can come to the conclusion that the decrypted image and its histogram exactly match the plain image and its histogram. This proves that no internal noises added during the encryption process. The PSNR value of the proposed scheme is calculated, and it is equal to the ideal value infinity. The PSNR value of the proposed work is compared with the average PSNR values of the existing works, and it is listed in Table 7.

6. Hardware Implementation

The proposed encryption scheme is implemented in a Field Programmable Gate Array (FPGA) to evaluate its lightweight and high-speed computation against various similar

works in the literature. The proposed encryption architecture is deployed in verilog HDL as an encoder of block size 256-bit. Here, single block 32-pixels are to be encrypted in parallel by scheduling the 8-bit key for each pixel from the 256-bit key. The proposed encryption architecture is simulated and synthesized in Xilinx ISE 14.7 for the XC5VLX330T-2 FPGA. The resource utilization and performance factors in comparison with existing lightweight encryption schemes are listed in Table 8.

From the above table, it is evident that the proposed WSAA encryption scheme has 112% higher throughput per area and occupies moderate area (LUT) compared to LEA-256 and utilizes 5.2% and 23.61% less area compared to LEA-192 and LEA-256, respectively. This proves that the proposed WSAA scheme is a lightweight and high-speed encryption scheme.

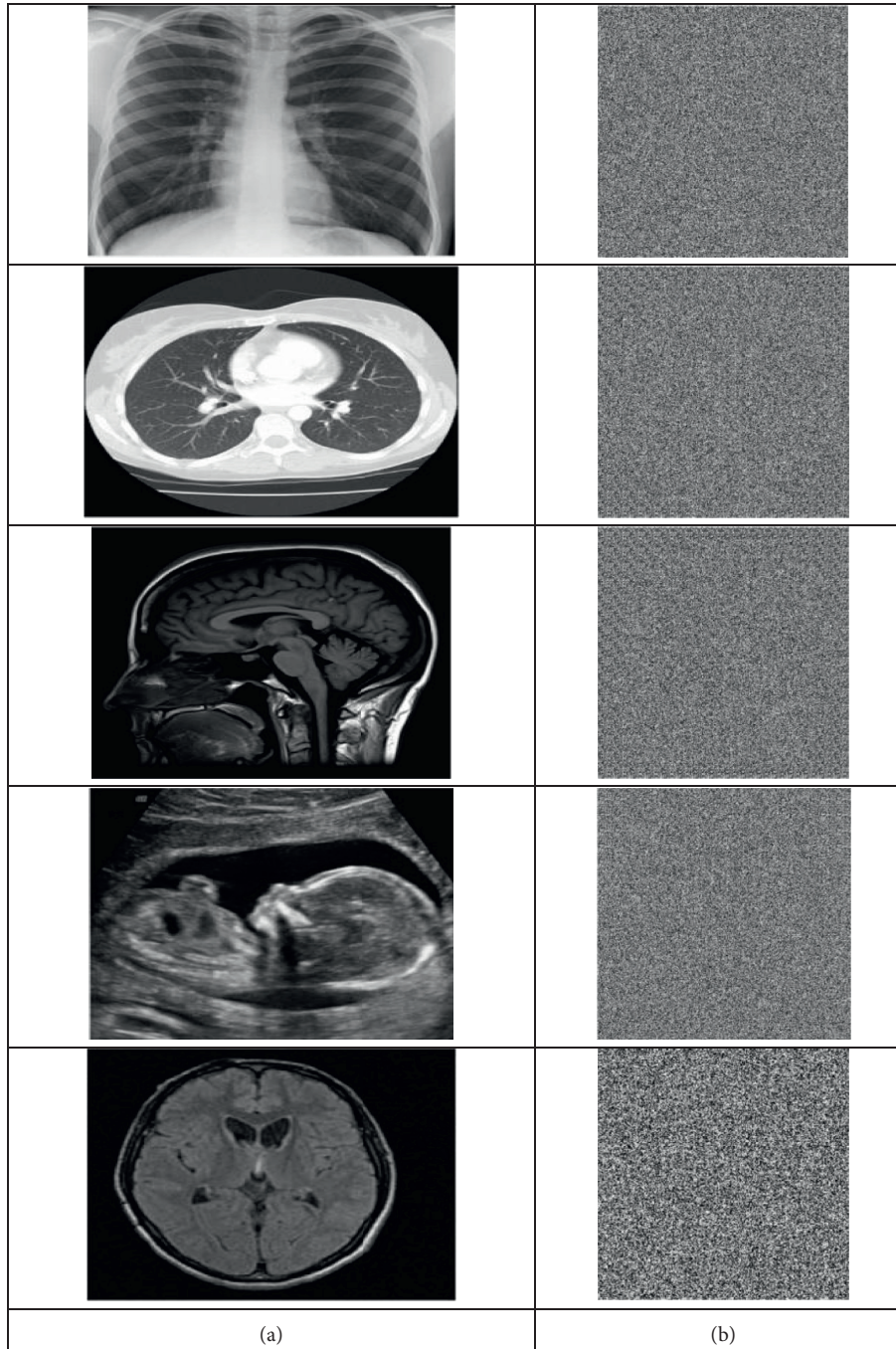


FIGURE 6: Key sensitivity test for single-bit variations in the key at decryption. (a) Decrypted image with actual key K1. (b) Decrypted image with 1-bit variation key K2.

TABLE 5: Measured entropy values of sample images before and after encryption.

| Modality | Sample image | Size | Entropy | |
|------------|--------------|------------------|-------------|-----------------|
| | | | Plain image | Encrypted image |
| X-ray | Chest | 512×512 | 7.5509 | 7.9913 |
| CT scan | Chest | 512×512 | 6.4985 | 7.9838 |
| MRI | Brain | 512×512 | 5.4162 | 7.9687 |
| Ultrasound | Fetal | 512×512 | 7.2543 | 7.9899 |
| PET | Brain | 256×256 | 5.9094 | 7.9850 |

TABLE 6: Measured entropy values of sample images before and after encryption.

| Modality | Sample image | Size | NPCR (%) | UACI (%) |
|------------|--------------|------------------|----------|----------|
| X-ray | Chest | 512×512 | 99.615 | 33.653 |
| CT scan | Chest | 512×512 | 99.258 | 33.374 |
| MRI | Brain | 512×512 | 98.995 | 33.391 |
| Ultrasound | Fetal | 512×512 | 99.579 | 33.521 |
| PET | Brain | 256×256 | 99.571 | 33.518 |

TABLE 7: Comparison of average PSNR values of the decrypted image.

| Scheme | Proposed | [9] | [15] | — |
|--------------|----------|-------|-------|---|
| PSNR (in dB) | ∞ | 78.35 | 36.78 | |

TABLE 8: Comparison of FPGA implementation results of the proposed WSAA with other lightweight schemes.

| Algorithm | Block size | Device name | LUT's | FFs | Slices | Cycle | Max. Frequency (MHz) | Throughput (Mbps) | Throughput per area |
|-------------------------|------------|-------------|-------|-----|--------|-------|----------------------|-------------------|---------------------|
| LEA-128 [36] | 128 | XC5VLX330T | 713 | 386 | — | 24 | 217.806 | 1161.6 | 0.198 |
| LEA-192 [36] | 128 | XC5VLX330T | 911 | 508 | — | 28 | 218.250 | 996.57 | 0.153 |
| LEA-256 [36] | 128 | XC5VLX330T | 1131 | 645 | — | 32 | 126.23 | 505 | 0.071 |
| Unified (sel-2'b0) [37] | 128 | XC5VLX330T | 735 | 832 | 273 | 33 | 292 | 1152 | 0.186 |
| LEA-256 [37] | 256 | XC5VLX330T | 440 | 812 | 250 | 33 | 340 | 1367 | 1.092 |
| Proposed WSAA | 256 | XC5VLX330T | 864 | 864 | 437 | 32 | 500.25 | 4002 | 2.316 |

7. Conclusion

The technological advancement and the present scenario mandate the remote diagnosis and secure transmission of medical information in the form of an image over the Internet from the data centre to another or from patient to doctors. The disturbances and attacks on the medical images are non-tolerable. This work proposed a unique approximate adder-based lightweight encryption and decryption scheme which tolerates to various adversarial attacks. The lightness of the scheme is featured by the proposed novel random error weighted shift approximate adder. The inheritance of the key distribution and rounding scheme within the image pixels contribute to the robustness of the proposed scheme against various attacks. The proposed scheme is implemented and tested for different modalities of medical images such as X-ray, CT, MRI, ultrasound, and PET. The implemented system is analyzed for the histogram, correlation, entropy, and key sensitivity. The developed scheme defies the statistical attacks and able to produce infinite PSNR at the decryption stage. It satisfies the ideal condition for any medical transmission system. Thus, our proposed lightweight approximate adder-based encryption and decryption scheme is the best fit for real-time secure remote monitoring of medical data.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] W. Cao, Y. Zhou, C. L. P. Chen, and L. Xia, "Medical image encryption using edge maps," *Signal Processing*, vol. 132, pp. 96–109, 2017.
- [2] H. Nematzadeh, R. Enayatifar, H. Motameni, F. G. Guimarães, and V. N. Coelho, "Medical image encryption using a hybrid model of modified genetic algorithm and coupled map lattices," *Optics and Lasers in Engineering*, vol. 110, no. October 2017, pp. 24–32, 2018.
- [3] R. Thanki and S. Borra, *Medical Imaging and its Security in Telemedicine Applications*, Springer International Publishing, Berlin/Heidelberg, Germany, 2019.
- [4] R. Tadeusiewicz and M. R. Ogiela, *Medical Image Understanding Technology*, pp. 470–479, Springer, Berlin/Heidelberg, Germany, 2004.
- [5] A. B. Wolbarst and W. R. Hendee, "Evolving and experimental technologies in medical imaging," *Radiology*, vol. 238, no. 1, pp. 16–39, 2006.
- [6] S. R. Cherry, "Multimodality imaging: beyond PET/CT and SPECT/CT," *Seminars in Nuclear Medicine*, vol. 39, no. 5, pp. 348–353, 2009.
- [7] A. Mahmood, T. Hamed, C. Obimbo, and R. Dony, "Improving the security of the medical images," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 9, pp. 137–146, 2013.
- [8] I. Jasmine Selvakumari Jeya and J. Suganthi, "RONI based secured and authenticated indexing of lung CT images," *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1–9, 2015.
- [9] I. B. Venkateswarlu, "Fast medical image security using color channel encryption," *Brazilian Archives of Biology and Technology*, vol. 63, pp. 1–8, 2020.
- [10] A. Sivaprakash, S. N. E. Rajan, and S. Selvaperumal, "Privacy protection of patient medical images using digital watermarking technique for E-healthcare system," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 15, no. 8, pp. 802–809, 2019.
- [11] A. Umamageswari and G. R. Suresh, "Security in medical image communication with arnold's cat map method and reversible watermarking," in *Proceedings of the IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT*, pp. 1116–1121, Nagercoil, India, March 2013.
- [12] K. N. Madhusudhan and P. Sakthivel, "A secure medical image transmission algorithm based on binary bits and Arnold map," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 5413–5420, 2020.
- [13] J. Zhang, D. Hou, and H. Ren, "Image encryption algorithm based on dynamic DNA coding and chen's hyperchaotic system," *Mathematical Problems in Engineering*, vol. 2016, pp. 1–11, 2016.
- [14] X. Chai, J. Zhang, Z. Gan, and Y. Zhang, "Medical image encryption algorithm based on Latin square and memristive chaotic system," *Multimedia Tools and Applications*, vol. 78, no. 24, pp. 35419–35453, 2019.

- [15] L. B. Zhang, Z. L. Zhu, B. Q. Yang, W. Y. Liu, H. F. Zhu, and M. Y. Zou, "Medical image encryption and compression scheme using compressive sensing and pixel swapping based permutation approach," *Mathematical Problems in Engineering*, vol. 2015, pp. 1–9, 2015.
- [16] Z. Hua, S. Yi, and Y. Zhou, "Medical image encryption using high-speed scrambling and pixel adaptive diffusion," *Signal Processing*, vol. 144, pp. 134–144, 2018.
- [17] K. Tamilarasi and A. Jawahar, "Medical data security for healthcare applications using hybrid lightweight encryption and swarm optimization algorithm," *Wireless Personal Communications*, vol. 114, no. 3, pp. 1865–1886, 2020.
- [18] M. Benssalah, Y. Rhaskali, and K. Drouiche, "An efficient image encryption scheme for TMIS based on elliptic curve integrated encryption and linear cryptography," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2081–2107, 2020.
- [19] A. Vizitiu, C. I. Niã, A. Puiu, S. Constantin, and L. M. Itu, "Applying deep neural networks over homomorphic encrypted medical data," *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–26, 2020.
- [20] Z. Chen, "A lightweight encryption algorithm for images," *Advances in Intelligent and Soft Computing*, vol. 137 AISC, pp. 235–241, 2012.
- [21] J. Kumar and S. Nirmala, "A new light weight encryption approach to secure the contents of image," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI*, pp. 1309–1315, Delhi, India, September 2014.
- [22] B. Mondal and T. Mandal, "A light weight secure image encryption scheme based on chaos & DNA computing," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 4, pp. 499–504, 2017.
- [23] A. Javeed, T. Shah, and A. Attaullah, "Lightweight secure image encryption scheme based on chaotic differential equation," *Chinese Journal of Physics*, vol. 66, no. July 2019, pp. 645–659, 2020.
- [24] I. T. Almalkawi, R. Halloush, A. Alsarhan, A. Al-Dubai, and J. N. Al-karaki, "A lightweight and efficient digital image encryption using hybrid chaotic systems for wireless network applications," *Journal of Information Security and Applications*, vol. 49, p. 102384, 2019.
- [25] O. A. Khashan and M. AlShaikh, "Edge-based lightweight selective encryption scheme for digital medical images," *Multimedia Tools and Applications*, vol. 79, no. 35-36, pp. 26369–26388, 2020.
- [26] S. Mortajez, M. Tahmasbi, J. Zarei, and A. Jamshidnezhad, "A novel chaotic encryption scheme based on efficient secret keys and confusion technique for confidential of DICOM images," *Informatics in Medicine Unlocked*, vol. 20, no. May, Article ID 100396, 2020.
- [27] T. Belkhouja, A. Mohamed, K. Abdulla, A. Ali, X. Du, and M. Guizani, "Lightweight encryption of wireless communication for implantable medical devices using Henon chaotic system," in *Proceedings of the 2017 International Conference on Wireless Networks and Mobile Communications, WINCOM*, Rabat, Morocco, November 2017.
- [28] A. A. Abd El-Latif, B. Abd-El-Atty, M. S. Hossain et al., "Efficient quantum information hiding for remote medical image sharing," *IEEE Access*, vol. 6, pp. 21075–21083, 2018.
- [29] D. Ravichandran, S. Rajagopalan, H. N. Upadhyay et al., "Encrypted biography of biomedical image - a pentalayer cryptosystem on FPGA," *Journal of Signal Processing Systems*, vol. 91, no. 5, pp. 475–501, 2019.
- [30] S. Janakiraman, K. Thenmozhi, J. B. B. Rayappan, and R. Amirtharajan, "Indicator-based lightweight steganography on 32-bit RISC architectures for IoT security," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31485–31513, 2019.
- [31] V. Manikandan and R. Amirtharajan, "On dual encryption with RC6 and combined logistic tent map for grayscale and DICOM," *Multimedia Tools and Applications*, vol. 80, pp. 1–30, 2021.
- [32] A. Mubashar, K. Asghar, A. R. Javed et al., "Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm," *Journal of Circuits, Systems, and Computers*, p. 2250010, 2021.
- [33] M. Nagarajan, A. Sasikumar, D. Muralidharan, and M. Rajappa, "Fixed point multi-bit approximate adder based convolutional neural network accelerator for digit classification inference," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 6, pp. 8521–8528, 2020.
- [34] R. Jothin, M. P. Mohamed, and C. Vasanthanayaki, "High performance compact energy efficient error tolerant adders and multipliers for 16-bit image processing applications," *Microprocessors and Microsystems*, vol. 78, Article ID 103237, 2020.
- [35] X. Hu, C. Jin, M. Alazab et al., "On the design of blockchain-based ECDSA with fault-tolerant batch verification protocol for blockchain-enabled IoMT," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [36] D. Lee, D.-C. Kim, D. Kwon, and H. Kim, "Efficient hardware implementation of the lightweight block encryption algorithm LEA," *Sensors*, vol. 14, no. 1, pp. 975–994, 2014.
- [37] Z. Mishra, P. K. Nath, and B. Acharya, "High throughput unified architecture of LEA algorithm for image encryption," *Microprocessors and Microsystems*, vol. 78, Article ID 103214, 2020.

Research Article

Blockchain and Business Process Management in Health Care, Especially for COVID-19 Cases

Ibrahim Abunadi ¹ and R. Lakshmana Kumar ²

¹Department of Information Systems, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia

²Centre of Excellence for Artificial Intelligence and Machine Learning, Hindusthan College of Engineering and Technology, Coimbatore, India

Correspondence should be addressed to Ibrahim Abunadi; iabunadi@psu.edu.sa

Received 8 September 2021; Revised 25 September 2021; Accepted 10 October 2021; Published 2 November 2021

Academic Editor: Thippa Reddy G

Copyright © 2021 Ibrahim Abunadi and R. Lakshmana Kumar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

From this global health disaster, the health profession is searching for a new technology to monitor and mitigate COVID-19 infections. Accurate and reliable data are needed to surveil and prevent the diffusion of the coronavirus. However, there is a lack of consistent data on existing technologies. Various entities—for example, medical labs and public hospitals—could provide data on patients with coronavirus infection. Yet, this information might not be precise since it is not supervised and recorded correctly. This paper proposes a Blockchain and Business Process Management (BBPM) system in healthcare to solve these problems. This system could be crucial in tracing coronavirus diffusion and discovering more dangerous patients and is extremely proficient at controlling the disease. BBPM can be utilized like a digital database that includes information concurrently utilized and distributed inside a wide, decentralized, and openly accessible network. Usage of blockchains in the medical sector is anticipated to revolutionize the industry in many regions, primarily because centralization in current health technologies inhibits information sharing and causes a lack of confidentiality. Furthermore, BBPM could help supervise the diffusion of coronavirus infection worldwide by utilizing blockchain networks on the mobile devices of individuals. Protecting patient information is one of the critical strengths of BBPM. Participating BBPM nodes can be governments, hospitals, testing labs, or patients. In addition, the digital ledger has a few documents, including patient reports, consequences, the condition of treatment, and a summary of discharge. The BBPM system is categorized into four processes. In the first process, the patient is analyzed and diagnosed by a testing lab to detect any early signs of COVID-19 infection. A sample of the patients was taken. If it is positive, the second process begins. In the second process, the patient is isolated and begins treatment for a minimum of 14 days. If the patient's health condition is improving, the third process begins. In the third process, the patients were retested for COVID-19. If the patient sample is negative, the patient is discharged, and an outline is created. During discharge, the patient pays the hospital for treatment. However, if the patient sample is positive, the isolation period is continued for another 14 days. In the last process, the details of total COVID-19 are confirmed, recovered, and death cases are conveyed to the government by the testing labs and hospitals. Patient records are stored for upcoming usage; their confidentiality is maintained when distributed on a larger scale. BBPM can guarantee the security and accuracy of patients' recorded data.

1. Introduction

The whole world is battling a new disease called COVID-19 and its mutations. It was initially discovered in Wuhan, Hubei Province, China [1]. An unprecedented spread of this virus has created many challenges that make the roots of human civilization tremble [2]. Some of the most apparent challenges are described below.

(i) Social exclusion: social exclusion is a method employed to slow the disease spread and “flatten the curve” of new cases with no licensed drugs or vaccines for COVID-19 treatment and prevention. However, most everyday activities, such as medical treatment, transportation, education, banking, and shopping, require physical contact. Apart from this, controlling physical contact can lead to social

isolation and unfavourable psychological consequences.

- (ii) False infodemic: the massive flow of fake data does not adhere to government policies, such as harmful self-medication or prophylactic treatments, panic behaviours, depressive disorders and social dissociation, motion restrictions, and restrictions on work and shopping hours. Furthermore, predictive models and evaluations of future claims based on such false data would be meaningless. Unfortunately, current sites and essential technologies cannot cope with this issue, becoming increasingly difficult to resist.
- (iii) Continuation of necessary government services: necessary government services, for example, public utilities (sanitation, electricity, water, and so on), salaries and pensions, tax gathering, marriage, birth and death registration, elections, and visa provisions are always anticipated to be obtainable at any time. Their continued distribution and management are more challenging, as citizens and government employees are under lockdown restrictions.
- (iv) Real-time data distribution: global data synchronization is a critical factor in combating the COVID-19 epidemic. Distributing necessary data, for instance, the number of infected patients, acute cases, recovered patients, deaths, and so on, should happen in real-time to raise public awareness, support quick action, and forecast future methods. However, technical challenges against the management of COVID-19 include misuse of data ownership, a lack of ways to verify data damage, the use of a single point-of-view, centralized data stores, and insufficient transparency in data transfer, as digital information is subject to security attacks.
- (v) Finance and charitable distribution: some banking institutions, such as the International Monetary Fund (IMF) and the World Bank, provide loans and grants to many countries to deal with the COVID-19 economic crisis. Such financial assistance should be shared transparently with those in need. However, due to corruption and a lack of correct automation systems, multiple countries have failed to receive such assistance [3]. In addition, citizens can be encouraged to donate if they can see the final use of the money donated.

In this worldwide health disaster, the health profession seeks novel technologies for monitoring and controlling the COVID-19 (coronavirus) epidemic. Therefore, accurate and reliable data are needed to supervise and prevent the spread of COVID-19. However, in current circumstances, there is not enough reliable data with the present technology, which may provide additional accurate data regarding the outbreak of COVID-19. Certain sources, such as medical labs and public hospitals, could present data on patients with coronavirus infection [4]. However, the information will not be trusted since it was not supervised or recorded and probably

not aggregated [5]. This paper proposes a blockchain and business process management system (BBPM) to track the spread of coronavirus and solve these issues quickly. The system identifies more at-risk patients and is extremely efficient in controlling the epidemic. It is described in this paper as the digital database. It includes data that can be utilized and distributed simultaneously on a comprehensive, decentralized, and openly accessible network. The use of telemedicine for people with diabetes in combating the COVID-19 epidemic has already been proven [6, 7]. Different techniques especially deep learning [8] can be used effectively to identify affected patients. In the healthcare field, deep learning [8] has been executed in numerous applications, for example, diabetic retinopathy detection [9] and lung nodule classification [10]. Many sources of medical images (for example, MRI, CT, and X-ray) create deep learning, a better technique to discover affected patients.

Blockchain is considered a digital ledger that distributes, decentralizes, and often stores public data [11]. Blockchain consists of three main parts: blocks, nodes, and miners. Chains consist of several blocks, each of which carries data, hash, and nonce information. Also, miners can construct a new chain block utilizing a procedure known as mining. Nodes are electronic devices that maintain a copy of the blockchain and keeps its network functioning.

Any user has a personal right to access a blockchain network for transfer transactions through the so-called consensus protocol. Blockchain uses a SHA256 hash to insert transactions. The NSA creates a SHA256 hash, and it is 64 characters substantial [11]. Although all transactions are recorded on a blockchain network, the public ledger does not change and is not manipulated [12]. Both transmissions are disseminated to different consumers throughout the network to transmit and modernize information [13]. A blockchain network can be copied to an individual location; for instance, in a similar capacity, networks of healthcare sharing or fractions of a global or regional information transfer scheme [14]. The blockchain data structure is a set of hierarchical blocks, illustrated in Figure 1. It shows an example of a blockchain containing n blocks. Each successive block includes the previous block's hash, timestamp, transaction data, the nonce number for the mining procedure, and other details required for the protocol to operate [15].

The blocks are attached in tuple format as the present block record values—for example, the preceding block hash, timestamp, Merkle root hash (in a blockchain network, the hash of the hashes of the entire transaction is called the Merkle root hash) and nonce number (nonce stands for “number only used once”; it is a random number utilized to secure private communications by avoiding replay attack) in its header [16]. Each block has two entities: a header and a body. The header includes the number of a block, the hash value of a previous block to protect the dependability of the chain, the hash of the present block body to preserve transaction information, honesty, nonce, timestamp, and the address of the block creator. All block bodies have numerous transactions [17]. Furthermore, distribution, durability, transparency, and audit capability are vital features of

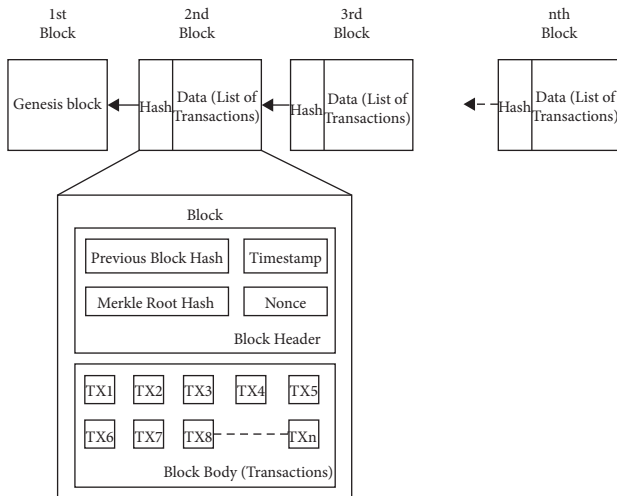


FIGURE 1: Structure of the blockchain.

blockchain [18, 19]. The types of blockchain are public, private, and consortium. Each archive is obtainable for the public in the public blockchain; thus, anybody can be involved in a consensus process [20]. Only a set of nodes are selected before the consensus method of a mutual network is required. In a private blockchain, only nodes from a solitary component are allowed to connect through the consensus process.

Figure 2 shows the participating blockchain nodes and documents of a distributed ledger in the proposed BBPM system. The participating BBPM nodes are testing laboratories, patients, government sites, and hospitals. In addition, the digital ledger has documents, including patient reports, consequences, the condition of treatment, and a summary of discharge.

The blockchain procedures involve the following steps: (a) collection of required data from participating blockchain nodes and (b) construction of source data converted to large amounts of information. Blockchain ensures the safety of the gathered information and assists in maintaining its confidentiality. Blockchain-protected information is examined by utilizing different solutions based on artificial intelligence. BBPM presents potential solutions to the COVID-19 epidemic, i.e., outbreak monitoring and medical supply chain management. It is utilized to set up quick, secure, and dependable data transfer with partners. Around the world, health centres and people have encountered a deficiency of medical equipment to battle the COVID-19 epidemic.

It is essential to build a BBPM to monitor COVID-19 transmission, as numerous recently implemented systems are vulnerable to hacking and cybercriminals. Table 1 illustrates the advantages of developing a BBPM-based solution instead of a conventional centralized solution in various areas, including fault tolerance, quality guarantee, data handling, and so on.

BPM is the discipline of enhancing a business process, modelling how it will perform in various situations, implementing enhancements, monitoring the advanced process, and continuously improving. A business procedure

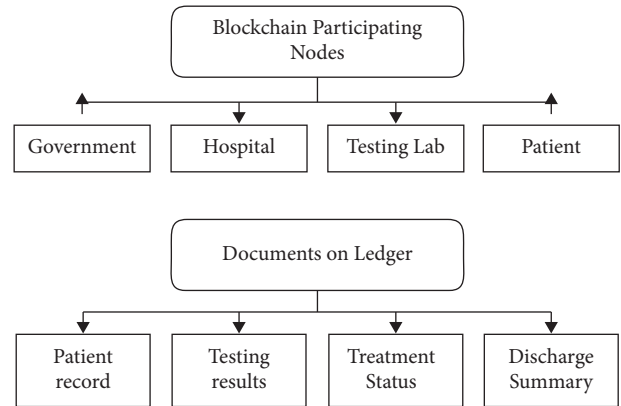


FIGURE 2: Blockchain participating nodes and documents on a ledger in a BBPM system.

is a set of behaviours that achieve particular organizational goals. BPM is not a one-time job but a continuous function involving continuous process reengineering. BPM frequently engages spontaneous jobs in any business process, and process enhancements could occur outside of automation. A well-implemented BPM could decrease waste, decrease errors, decrease time consumption, and create superior products and services. Thus, this paper combines blockchain technology with BPM to effectively combat the COVID-19 epidemic.

The significant findings of this study can be summarized as follows. First, the patients were tested by a testing laboratory according to the early signs of COVID-19 infection. A sample of the patient is then taken, and if it is positive, the patient is isolated for a minimum of fourteen days. The BBPM is utilized for treating and monitoring the patient during the isolation period. After that, the recovery stage begins, and the patient is retested. However, if the sample of the patient is negative, the patient is released, and a summary of discharge is developed.

The rest of the paper is arranged as follows. Section 2 reviews the associated work. Section 3 describes the system with a data model. The methodology of BBPM is explained in Section 4. The results and discussion are examined in Section 5; finally, the paper’s conclusion is presented in Section 6.

2. Related Work

2.1. *Blockchain in Healthcare.* The traditional healthcare system encounters a few issues, which include the following:

Interoperability: this is a way to exchange data between various data systems. Data should be exchanged and used for further applications. Healthcare systems’ key feature is their health information exchange (HIE) or common data distribution feature. With several EHR systems being deployed in various hospitals, they have a varying level of terminology, techniques, and functional capabilities, which means there is no universally defined standard. Moreover, medical records need to be swapped at a technical level so that more of the data can be used.

TABLE 1: Comparison of using a conventional centralized system and A BBPM system.

| Features | Conventional centralized system | BBPM system |
|-------------------|---|---|
| Fault tolerance | Huge hazard of a single point of breakdown | A distributed ledger is highly fault-tolerant |
| Quality guarantee | Administrators are required to authenticate data (data provenance does not apply) | Data could be tracked from its origin using cryptography technology |
| Transparency | Databases are not transparent | Data are stored on a distributed network |
| Data privacy | High chances of malicious cyber attacks | Data are stored using cryptography technology |
| Data integrity | The data can be changed | Data are unchanging and compatible |
| Control | Controlled by administrator (centralized) | Even decentralized in private blockchains |
| Data handling | Supports four primary functions: creating, reading, updating, and deleting | There are only read and write options |

Information asymmetry: the biggest issue that critics notice in the healthcare sector is data asymmetry, which means that one contributor has better access to data than another. Health systems and the public health sector suffer from this issue because physicians can access patient records, centralizing them. If patients need to access their medical records, they have to undergo a long and arduous procedure. The data are federalized to a solitary health sector, while their regulation is delegated merely to hospitals.

Medical data breaches: an information breach of health data includes the personal health data of any individual's EHR or medical billing data from their health insurance.

Numerous health care systems are not set up to meet patients' needs or address problems associated with incompetence and poor adaptation of these systems. Some literature further suggests that the usage of health applications has had negative effects on data processing. These issues justify finding a platform that could help turn the healthcare sector into a patient-centric one—for example, blockchain, which provides a safe, obvious platform and data honesty for patients' medical records.

Blockchain technology is increasingly strengthening healthcare and operational protocols and creating the perfect foundation for a competent, proof-based decision-making procedure. Table 2 shows the SWOT analysis for model adoption based on blockchain in the medical industry. This SWOT study underscores the key advantages and disadvantages of the usual method, while the chances and risks of adoption were identified.

Disintermediation, aimed at the absence of a central authority to collect, process, and verify information or construct and distribute samples, enables us to minimize time, faults, and prices due to the effectiveness of procedures to build and update a forecast model that supports supported risk management and clinical practice. Blockchain is a combined scheme; the procedures mentioned in it are automated and consistent.

The blockchain verifies the reliability of the dealings, and the information included therein cannot be altered while enhancing the security in which the operations occur. Furthermore, with the cryptography scheme, the consistency of the information disseminated over the entire network and the lack of centralized power create more confidence in the

scheme—for example, the requirement to maintain it between the contributors engaged in this procedure vanishes. Finally, the assurance between the contributors to the chain to cooperate in the execution and modernize incomplete methods is increasingly necessary through the general attention of the contributors in getting a precise, operational, and efficient forecast method.

A blockchain-based health-information-sharing network was proposed in [21]. The authors used two liberally linked blockchains to handle different health data types. They combined the storage of off-chain and authorization of on-chain to establish security and reliability criteria.

In [22], a radical user-centred healthcare information transfer technique recommended using channel creation by a decentralized and authorized blockchain to protect the privacy and improve personal security through a relationship program based on blockchain. Proof of legality with authenticity is retrieved for unspecified periods from a cloud database and embedded in the blockchain network to secure the privacy of EHR in all documents.

The safe and confidential Protected Health Information networking project based on blockchain discussed in [23] aimed to improve analysis in the e-Health program. A private consortium blockchain was formed by creating its data formation with consensus methods. The private ledger handles PHI, while the ledger society maintains a solid index of Protected Health Information.

In [24], smart contracts using blockchain were proposed to allow safe health sensor analysis and organization. They created a network using the Ethereum protocol and a private blockchain where the sensor connects to a computer that refers to the intelligent contract and records of each activity in the blockchain.

In [25], a system based on blockchain was established for secure, operational, and competent access for patients and physicians from third parties' clinical data while maintaining patient information confidentiality. Ethereum-based blockchain uses smart contracts to increase access control and encryption clarity, using modern cryptographic techniques for advanced security.

In [26], a new framework for storing blockchain-based clinical data was introduced. However, users need to keep valuable data permanently so that when there is an interruption, the originality of the data can be verified. Therefore, the authors used intelligent information administration methods and various cryptography techniques to secure consumer privacy.

TABLE 2: SWOT analysis for blockchain-based model adoption in healthcare.

| | Positive | Negative |
|----------|---|-----------------|
| Internal | <i>Strengths</i> | <i>Weakness</i> |
| | (i) Automation and disintermediation | |
| | (ii) Immutability | |
| | (iii) Hope | |
| | (iv) Transparency | |
| External | (v) Confidentiality | <i>Threats</i> |
| | <i>Opportunities</i> | |
| | (i) Greater cooperation among health system operators | |
| | (ii) Development of technical awareness and new expertise | |
| | (i) Resistance to change | |
| | (ii) Lack of expertise | |
| | (iii) Lack of confidence in the use of new technology by health workers | |

MedBlock, an information administration system based on blockchain, as discussed in [27], aims to manage patients' data. The centralized MedBlock database in this scheme enables the safe entrance and reporting of medical data. In addition, an advanced consensus procedure constructs a consensus on health reports with no substantial power expenditure or obstruction of the network.

In the mobile cloud, Nguyen et al. [28] recommended a novel EHR distributor framework that integrates a decentralized interplanetary file system (IPFS) with a blockchain. Notably, they developed a reliable system for controlling access using an intelligent agreement securing health records delivery between patients and health care providers. Thus, they provided an efficient solution for reliable communication in the mobile cloud while securing the necessary medical data against possible risks.

In [29], a review of the difficulties and feasibility of blockchain in healthcare applications was presented. First, they introduced the issues related to personal healthcare dealings with an organization, which will face challenges through its unique characteristics as a blockchain. They then give extended consideration to blockchain tools in the health sector and review previous work. At last, they explain the benefits and possible research chances for blockchain-associated technology to be utilized in the medical industry.

In [30], a complete outline of blockchain technology was presented. It presented a summary of blockchain structure and the advantages of blockchain, including protection [31]. In addition, they reviewed the application of blockchains in the medical industry. With the Health Information System, the authors explained how health records could be secured utilizing blockchain.

In [32], a shared program like blockchain in healthcare is described. It presented a secure data access system using blockchain to patients and doctors at a specific hospital. The security learning of their project shows that it maintains companies' honesty and resists famous attacks [33]. Subsequently, the execution consequences exemplify the feasibility of the proposed program.

MeDShare, proposed by Xia et al. [34], is a blockchain-based source of data auditing management for health data distributed in cloud storage among large data organizations. Transfers of data transmitted from one company to another are stored on MeDShare without damage. In addition, the program uses access control algorithms to effectively

supervise data behaviours caused by companies when discovering data breaches [35].

Wang et al. [36] recommended a specific EHR program using blockchain technology and attribute-based cryptography. The authors used ID-based and attribute-based encryption (IBE and ABE) [37], while [38] also used ID-based signature (IBS) to generate a digital signature for encrypted health data. It helps set up the project effectively and does not require various cryptography programs for various safety needs.

In [39], a frivolous blockchain structure for Health Data Management (HDM) was recommended to slow down calculations and interaction overhead compared to the network of Bitcoin by separating network providers within clusters and keeping a ledger copy per cluster. Their framework initiates the necessity for a canal that allows secure and confidential dealings inside a collection of network providers. Furthermore, they recommended a resolution to avoid fraud in the Bitcoin network. They demonstrate the effectiveness of the authors' recommended structure in providing security with secrecy to the Bitcoin network by analyzing different attacks. They further discussed how the authors recommended structural deals for discovering attacks.

In [40], an analysis of various solutions using blockchain was provided. First, the authors explored the recent sophisticated solutions that make smart devices compatible with blockchain in various industry 4.0 devices. After that, the benefits and drawbacks of traditional security resolutions are discussed in terms of their countermeasures.

Bach et al. [41] presented comparative research into the consensus mechanisms of blockchains. In particular, Ethereum currently uses the consensus protocol known as the Proof of Work (PoW). It is a system that allows a decentralized Ethereum network to agree to deal with balances of accounts. Thus, it allows consumers to avoid "double spending" their currency and ensures that the Ethereum chain is harder to attack or overwrite.

Ying et al. [42] used attribute power to deliver keys to information consumers in the Ciphertext Policy Attribute-Based Encryption (CP-ABE) model to attain precise access control for distributing health records in the cloud. Using test surroundings in an Ubuntu Linux desktop, they assessed the proposed program with a numerical simulation, and the decentralized usage capability was avoided. In the context of

blockchains, different studies have explored the potential of blockchains to sustain e-health information distribution.

Numerous research projects are currently being conducted on the utilization of standard blockchains in health care. The most current is MedRec [43]. The MedRec method uses the Ethereum platform to set up a decentralized health proof distribution scheme for smart contracts. It distributes medical reports among various health partners and patients with any other contributor that executes health or medical reports. For example, healthcare providers could add patient records at any time, but it is up to the patient to determine what information could be accessed by other providers. MedRec recommends two mining methods. The first uses ether as an effective method. In contrast, the second recommends utilizing compiled and anonymous information like a gift to motivate investigators. The primary node going to a block is allowed access to the desired information.

2.2. Blockchain and Business Processes. We did not initially discover the application possibilities of blockchain for business processes. Many blockchains are now widely accepted in different domains to assist in the function of novel business processes. For example, a caterpillar proposed by López-Pintado et al. [44] is an open-source business process management system (BPMS) that runs on the summit of the Ethereum blockchain. Like any BPMS, Caterpillar assists in constructing procedure method events and lets consumers supervise the status of procedure events and perform their jobs. The uniqueness of Caterpillar is that the status of each processing event is kept in the Ethereum blockchain. In addition, the flow of work algorithm is executed by smart contracts created by the Business Process Model and Notation BPMN-to-Solidity compiler. The compiler assists in a broad range of BPMN configurations, containing consumer and service functions parallel to exclusive gateways, subprocesses, multiple event functions, and event handlers.

In [45], an automated BPM framework explored composing services in free commerce surroundings, and blockchain was investigated and presented to change and check trade trust. The BPM solution illustrates how blockchain technology could provide instant, dependable, and cost-efficient service and deliver quality services in workflow composition.

2.3. Use of Blockchain in Epidemic Situations. The possible use of a blockchain system for regulating and alleviating the COVID-19 situation is explained and explored in detail.

2.3.1. Data Management of Clinical Trials. Clinical trials must maintain information according to rules, for example, reports being open to shareholders, the confidentiality of reports, security, and immutability [46]. Blockchain records and makes real-time data available to physicians. It enhances data accuracy, facilitates data transfer, guarantees compliance, and provides an audit path for superior confidentiality and data security [47, 48].

Conducting clinical trials to implement the COVID-19 vaccine is a complex, time-consuming, and expensive process. The vaccine's clinical trials require close coordination and cooperation between the companies engaged and are frequently situated in geographically distributed areas. Researchers, regulators, donors, and pharmaceutical companies are instances of companies that have been seriously engaged in clinical trials to implement and administer the vaccine for COVID-19 successfully.

Traditional centralized clinical trial data management systems encounter several challenges, including compliance with course enrollment, limited effectiveness and clinical testing necessities, data confidentiality guarantees, compliance with clinical trial rules for participants' healthcare, and reliability of clinical trial data. Furthermore, centralized clinical trial management systems could provide multiple versions of clinical trial data to construct organizations' data pits. Consequently, they could guide the duplication of clinical trial information frequently recorded and handled by numerous companies.

Therefore, copying clinical test data makes it hard to access, implement, and examine consequences. Furthermore, centralization creates clinical trial data susceptible to changes by external hackers. Blockchain technology could help research institutes, and pharmaceutical companies protect clinical trial data's honesty when developing a vaccine. It ensures that a single and synchronized view of clinical trial data is obtainable to all accredited companies. Therefore, problems, such as duplication and discrepancy of clinical trial data because of the breakdown of previous centralized clinical trial management systems, could be successfully addressed.

Smart contracts could check a company's access rights before allowing the utilization of clinical trial data to protect data confidentiality and safety. The authorized contributors can digitally check that the clinical trial necessities have signed the consent form. Thus, anonymous data gathering and verifiable approval management will allow contributors to distribute their case records to authorized companies without disclosing their identities.

To keep contributors in a clinical trial, pharmaceutical companies generally provide appreciation tokens to contributors in cash or on gift cards. Smart contracts will help the payment process quickly by presenting an automated, transparent, and accountable system for converting cryptocurrencies. Accountability and transparency features ensure that data can only be utilized for the intended purpose, thus increasing user confidence.

2.3.2. Vaccine and Necessary Drug Supply Chain. Blockchains could effectively handle the healthcare supply chain, especially in epidemiological circumstances that engage important transactions across the globe. There may be instability in its distribution until an approved version of the vaccine is available to sell and advertise. Dishonest performance, for example, incorrect vaccinations, high pricing, and stock accumulation may be possible. These problems can be effectively managed to utilize a blockchain-based medical supply chain [49].

Symptomatic patients can contact remote health professionals through information technology infrastructure to reduce the risk of transmission of infectious viruses using advanced remote health practices, such as telehealth and telemedicine services. In addition, remote detection and treatment of patients could notably reduce patient access and staff limitations, thereby effectively controlling the quick rise in global COVID-19 cases of employment in remote health services.

Because they are governed and handled by a centralized authority, remote health systems can be susceptible to the point-of-failure issue, which finally infects the honesty and reliability of health records. The intrinsic features of the revolutionary blockchain technology could introduce various advantages to the remote healthcare industry. Significant features include setting up a source of electronic health records, checking the legitimacy of users, requesting patient data, verifying patient anonymity, and automating micro-payments to use remote health services.

These significant features help to demonstrate self-examination clinical instruments for COVID-19 testing successfully. Following a test result, individuals whose test results are negative should generally adopt self-isolation policies to minimize the spread of the virus to the community. There is a need for safe tracking of medical items for self-isolated individuals; blockchain technology brings changes to explicitly record the timestamped location data of medical items in the ledger. Guaranteeing social exclusion and wearing masks when engaging in business behaviours can help avoid the diffusion of COVID-19.

The growing number of confirmed cases of COVID-19 worldwide, particularly in areas with the highest virus transmission rate to estimate the diffusion of COVID-19, requires the administration of drugs without contact with patients. Aerial vehicles may be utilized to transmit medicines and medical supplies to distant patients. Also, aerial vehicles could help in transporting medical supplies to hospitals located in remote locations. For example, China has used aerial vehicles to deliver medicine from one city to another during the COVID-19 epidemic.

Blockchain allows the location of aerial vehicles to be tracked, checking the level of service provided, and computing the reputation score of an aerial vehicle using its effectiveness in a reliable, responsible, and transparent manner. By executing access to control protocols and identity management, blockchain technology reduces the feasibility of attacks by enemy vehicles. Furthermore, it invariably stores orders provided by the control room on aerial vehicles (for audit purposes to check noncompliance with published orders) with measures to discover human movements and reactions by cleaning highly infected areas.

A swarm consists of several autonomous aerial vehicles that work together to attain a general aim. For example, a crowd of aerial vehicles could utilize blockchain technology to achieve the most dependable global results by trading safely in a blockchain. To take another example, a voting system based on blockchain makes it possible to discover densely populated areas where aerial vehicles can be sprayed with disinfectants.

2.3.3. Communication Tracking. Health facilities are active inpatient contact monitoring systems; however, records obtained may be misused. The utilization of blockchain will offer data stability and authenticity [50]. Networks of blockchains could monitor patient behaviour and present current updates to infected fields. In addition, records could be created for affected and potential victims through contacts. Communication tracking poses challenges to privacy, as information has to be collected, fitted, and distributed. Guaranteeing the identity protection of users with COVID-19 introduces other issues. Although participation in the exam can ensure some control, we have not yet observed how we can guarantee that merely appropriate information is distributed. Blockchain could play a neutral role in a disseminated way to separate authorized solvers from mitigating users/patients and the identity of user and place data. It could present a resolution to protect the confidentiality of technological plans before adhering to rules within the centralized scheme.

Moreover, the integration of blockchain with anonymization and encryption techniques could also defend the individuality of users. Blockchain is nonregional and provides an appropriate worldwide usage platform for detecting and controlling the COVID-19 epidemic. This explicit aspect may prevent the public from deliberately misinforming officials or other third parties.

Respecting the social distancing orders provided by the government could notably reduce the social interaction of humans to avert the spread of COVID-19. Social exclusion is activated by a public health activity called digital contract tracing, which can break a person's chain of spreading the virus. Digital communication tracking constantly monitors the affected population to quickly and efficiently discover all social communications during the incubation period of infected COVID-19 patients. It primarily uses Bluetooth or GPS to use nearby data to discover social communications with a virus-infected person.

After coming in close contact with a confirmed COVID-19 case, exposed persons should be tested, supervised, and self-isolated. The clarity and consistency of the data ensure that users' health data, for example, COVID-19 test results, cannot be changed or removed by attackers or health workers. Furthermore, the General Data Protection Regulation (GDPR) protects the confidentiality of users' information through the confidentiality rules outlined in the Privacy Act. The positioning technique parameter refers to technologies that could be utilized to discover a user's location.

The coverage area parameter explains the geographical regions in which a COVID-19 patient could detect social interactions with another person. The heavyweight designs of the Contact Tracing application are intended to utilize computer resources aggressively when identifying and verifying social interactions between people. Conversely, lightweight application design enhances computer resources by leading users and providing essential features. Digital contact tracing users' necessity promises an extended battery life of devices and greater confidentiality, safety, and transparency of data related to COVID-19.

Preferably, the digital contact tracing solution could present greater confidentiality of data, an expanded coverage region, a lightweight application plan, better data protection and clarity, and battery-friendly functionality. However, ensuring the confidentiality of an individual's health data and, at least, COVID-19 false-positive events is a significant challenge for digital communication tracking solutions. Data confidentiality is protected by encrypting a person's location and contact history and averting the dissemination of personal health information to the public.

On the occasion of close contact with a COVID-19-affected patient, users can be informed about the latest social interaction without revealing the evidence of the affected person (e.g., their name). Digital contact tracking by smartphone applications such as Google-Apple's Contact Tracing and Singapore's Trace Together uses Bluetooth to discover the close physical contacts of a person affected by the virus. However, regarding the battery barriers of smartphones, Trace Together is not user-friendly. Furthermore, Google-Apple Contact Tracing does not reveal the location and identity of users to protect data confidentiality.

Blockchain-based solutions are less reliable because they are subject to data fraud through the application administrator, given the high confidentiality and sensitivity of users' information. Unchanging and decentralized blockchain technology is a feasible alternative to digital communication tracking. It protects the confidentiality of the user's information by allowing pseudo-anonymity. Digital contact tracking with a custom exposure matching mechanism can use the blockchain site to store social contact data and enable authorized users to access the information to protect its confidentiality. An external trusted network of servers is utilized in the provided system to create anonymous addresses for users to protect data confidentiality. The organization has executed some smart agreements, for example, registration of companies, COVID-19 testing, geodata processing, query processing and approval management, automation of services, and assurance that the evidence of individuals affected by COVID-19 has not been revealed to others.

As a private blockchain-based system, all entities are registered before making a transaction on the blockchain. Geodata processor contracts assure us that the duplicate data (e.g., the location data of a user with limited mobility) are not forwarded to the contact solver to speed up the contact tracing process. COVID-19 testing contracts assisted in recording COVID-19 test results on the blockchain for each employee. Subsequently, the consent management contract seeks to legalize the location data usage of the employees of an organization. The contact solver component of the contact tracing system leverages AI-based techniques for identifying social interactions among individuals. It informs users about possible risk levels based on many factors, such as distance, mobility, and total time spent during social interaction with a COVID-19 infected person.

In a blockchain-based system, all companies are registered before creating a transaction on the blockchain. The Geodata Processor Agreement guarantees that no duplicate information (e.g., the location information of a user with

limited mobility) is sent to the contact solution to expedite the contact tracking procedure. The COVID-19 test contract allows each worker to store the COVID-19 test outcomes in the blockchain. The consent management agreement aims to legitimize the use of location information by employees of a company. The Contact Solver element of the Contact Tracing System uses AI-based methods to identify social communications between individuals. COVID-19 informs users about potential risk stages using factors such as distance, movement, and total time spent during social communication with the victim.

2.3.4. Data Collection. Monitoring outbreaks by deploying, collecting, and retrieving data that respond effectively to the epidemic, understanding trends, and managing tests are vital resources. Blockchain's ability to confirm and store enduring real-time data confirms information reliability [51]. The utilization of the blockchain network presents the surveillance and communication infrastructure to assist in capturing, recording, and examining virus spread and control data.

2.3.5. Consumer Information Confidentiality. Those responsible for making the policy and health practitioners should obtain patient information through patient monitoring and other enhanced decision-making and discuss patient privacy problems. In these troubled days, a balance law must be enacted between registry management and user confidentiality management to enhance hope in the scheme. Blockchain is a viable resolution for maintaining and displaying patient data, monitoring patient processes, and setting up degrees of social isolation while protecting confidentiality.

2.3.6. Early Discovery of Susceptible People. Different triage systems based on AI can reduce patient concerns. The online bot would assist in comprehending the origin symptoms of early discovery and then guide them through preventive measurements, for example, social exclusion and hand sanitation. It would warn users about medical facilities if symptoms intensify [52]. The secrecy of patient data is vital to the protection of their social and personal values. Blockchain-based architecture could efficiently manage these safety and confidentiality problems.

2.4. Blockchain Cases of COVID-19 Epidemic. A few of the blockchain cases utilized to fight the COVID-19 epidemic are explained below.

Hyperchain is a platform based on donations created to support hospitals and governments donating to affected patients in China [53]. To solve the lack of facilities during this epidemic, numerous users could join the millions of nodes of hyperchain that could receive donated items and necessary medical equipment from factories.

PHBC: this platform based on blockchain is utilized for continuous and unknown checking of society and working places open to COVID-19 and dangerous

viruses [54]. A vital aspect of this platform based on blockchain is detecting the movement of noninfected individuals and controlling these individuals' return if they go to the affected regions.

VeChain: this is a platform using a blockchain created to supervise vaccine manufacture [55]. All behaviours associated with vaccine manufacture, from substance to codes to packages, are stored and recorded in a distributed ledger. Thus, it presents an efficient way to decrease the risk of possible changes to vaccine data.

Hashlog: this project was developed by Acoer, a Georgia-based health technology startup [56]. The Hashlog blockchain solution could be implemented through distributed blockchain ledger technology, ensuring logging and data visualization of coronavirus outbreaks from US Centers' public Data for Disease Control (CDC) and WHO.

2.5. Blockchain Security. The blockchain platform should guarantee the fundamental features of security: confidentiality, integrity, and availability that advantage the healthcare industry.

Confidentiality could be attained by ensuring that the application is on a personal blockchain and that users have restricted access. It will reflect the certification needed in the healthcare field, such as becoming a physician, with the proper qualifications needed. Likewise, in the business network, the accounts of the physicians should be constructed by the medical entity. In addition, blockchain network data should be allowed to protect confidentiality. In addition, contributors will have various roles and privileges. In addition, encryption should ensure that the data between the blockchain and the user is safe. Confidentiality is further mandatory in this business network; however, it directly battles data breaches and phishing attacks [44].

Integrity: integrity is about ensuring that information is reliable and accurate. Blockchain attains this in two different methods: (1) hashing and (2) shared distributed ledger. Strong collision-resistant and safe hashing algorithms should be utilized to ensure integrity. Likewise, privacy and access control ensure that the data are reliable by controlling the number of individuals who could damage the data.

Availability: significantly, there is dependable and simple access to data on the blockchain. Ensuring that the network of blockchains is fault-tolerant decreases the number of failed links to information in the blockchain. In addition, the data in a blockchain are a shared ledger; thus, there is a variety of copy information that ensures that the data do not disappear. The network of blockchains should run on the newest version of HyperLedger to ensure that errors do not impact the system's availability.

Berdik et al. [45] presented an extensive review of the use of blockchain as a service for applications within today's data systems. This review provides the reader with an in-depth

foresight on how blockchains can help to protect and handle today's data systems. The review includes detailed reports on the various examples of blockchain studies and applications presented through the investigation group and their implications for blockchain and other applications or their use in scenarios. A few of the very significant discoveries this review highlights are the framework of blockchain and the latest cloud and edge computing examples that are important in allowing the extensive adaptation and implementation of blockchain technologies for novel players in today's unparalleled vibrant global market.

3. System Model

This section discusses the BBPM system model and the notations utilized in this model. Here, Figure 3 illustrates the system model. It has four entities.

Participating nodes in BBPM are testing laboratories, patients, government sites, and hospitals. In addition, the digital ledger has documents, including patient reports, consequences, treatment conditions, and a summary of discharge. Figure 4 demonstrates the necessary steps utilized in BBPM to trace and store the data's active COVID-19 patients. (1) Patient visits a testing lab. (2) First, the patients were analyzed by a testing lab according to the early signs of COVID-19. The testing lab is an important node in the network of blockchains. It acts as a miner. (3) The patient sample is taken, and if it is negative, the patient may be discharged; a summary of discharge is also constructed. (4) However, if the result of the patient sample is positive, the patient is isolated for a minimum of 14 days. (5) During the time of isolation, BBPM is used to treat and monitor the patient. (6) After that, the recovery stage begins, and the patient is retested for COVID-19 after 14 days. (7) If the patient dies during treatment, the body of the deceased will be disposed. (8) After recovery, the patient pays the hospital for treatment. (9) Details of total COVID-19 confirmed cases are informed to the government through the testing lab. (10) Details of recovered and death cases are informed to the government through the hospital. The government records patient data for upcoming usage; its confidentiality is maintained and provided when required to be demonstrated on a large scale. Figure 4 demonstrates the workflow of the BBPM scheme.

BBPM provides a guarantee of the accuracy of the patient's stored information. Table 3 shows the notations used in the proposed system model.

4. Proposed Methodology

This section discusses the Blockchain and Business Process Management (BBPM) system in health care methodology. One of the major issues has been the requirement for current information on the epidemic and the spread of COVID-19. BBPM helps solve this problem more efficiently. One significant benefit of this system is that it provides provable and safe information utilizing its peer-to-peer networking features and distributed ledger technology. This technology plays a vital role in recording patient data on COVID-19

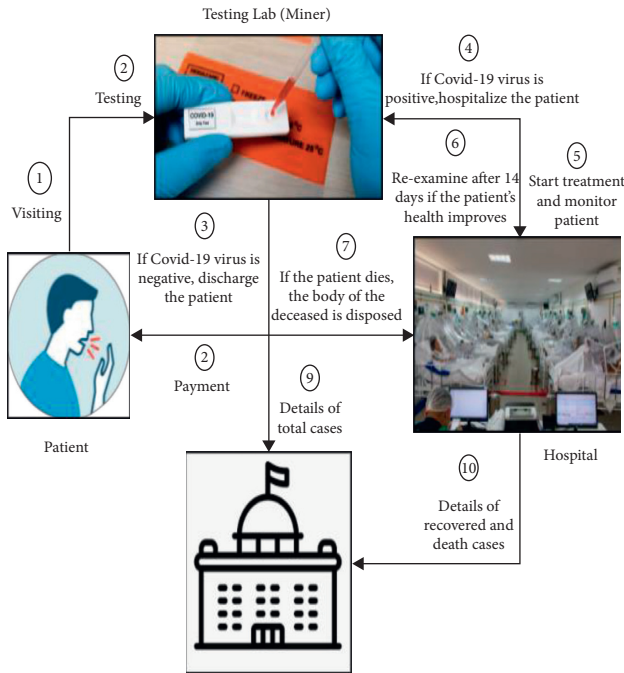


FIGURE 3: Proposed system model.

signs, locations, and records of health situations during the infection. BBPM helps distribute, encrypt, and securely record digital transactions. It is anticipated to revolutionize calculations in numerous regions, mostly where centralization is unnatural; moreover, privacy is necessary. By establishing a network of blockchains on citizens’ mobile devices, it could be improved internationally to monitor the spread of COVID-19.

In general, there are three kinds of nodes: miner nodes, full nodes, and light nodes. A miner node could suggest blocks and contain a whole blockchain history. Full nodes contain the entire blockchain history, although without presenting novel blocks. Meanwhile, light nodes depend on the full node blockchain history. In the BBPM scheme, the miner nodes are the testing lab, and the hospital is a full node; the patient and government also play the role of light nodes. In the BBPM system, many patients, testing labs, and hospitals are obtainable. Thus, control of access is essential. Their BBPM scheme presents access control. The collection, use or disclosure of personal health data without the consent of individuals is generally called as unauthorized access or “snooping.” Unauthorized access involves viewing personal health data in electronic data systems and can be triggered by a number of factors, including individuals’ conflicts, interests, personal gain, or concerns about their health and well-being. As a health data protector, we must take reasonable steps to guarantee that personal health data are protected against theft, loss and unauthorized access and disclosure, and that records containing data are protected against unauthorized copying, alteration or removal. We must take reasonable steps to guarantee that personal health data are not collected without authority and that records of personal health data are retained, altered, and disposed of in a secure manner. Protecting privacy should be integrated

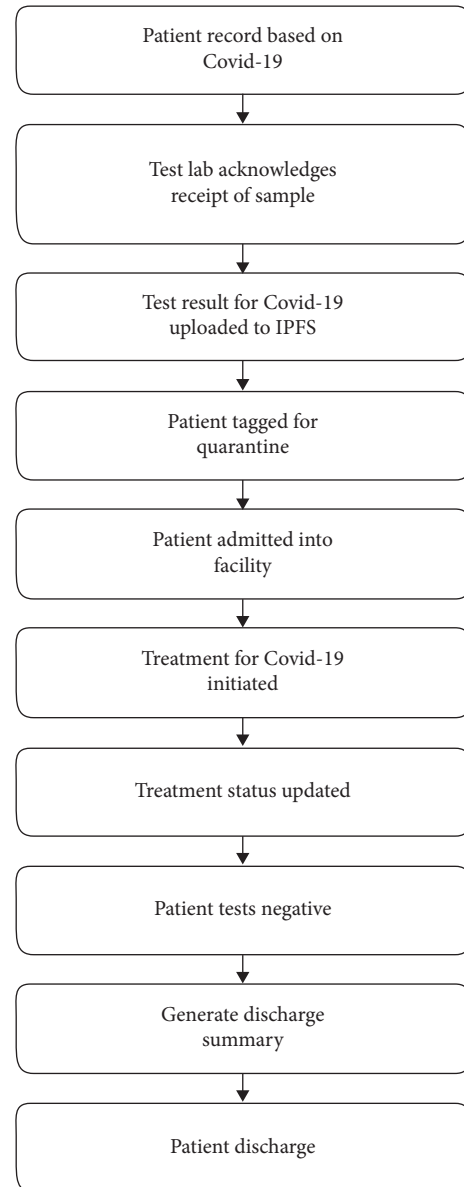


FIGURE 4: Workflow of the BBPM system.

TABLE 3: Notations.

| Notation | Description |
|----------|----------------------|
| BH | Block header |
| BB | Block body |
| Bid | Block ID |
| EB | Encrypted block body |
| DS | Discharge summary |
| DD | Discharge date |
| PD | Patient details |
| TD | Treatment details |

into the provision of health care and should be embedded in the culture of each health care system. To overcome this problem, access control is proposed in this BBPM.

At BBPM, the patient could merely see his or her digital ledger; moreover, nobody else could see his or her ledger.

Furthermore, a testing lab can only see the ledger of the patients it has tested. Like the testing lab, the hospital can only see the ledger of the patients it has treated. Finally, the government cannot look at any patient's ledger. They can only view statistical information, such as total, recovered, and death cases.

4.1. Algorithm Design. Algorithm 1 explains the proposed Blockchain for Business Process Management (BBPM) in healthcare.

4.2. Benefits of BBPM System in the COVID-19 Pandemic. This section discusses the benefits of the proposed BBPM system in the COVID-19 pandemic, as shown in Figure 5.

4.2.1. Enhancing Transparency When Treating Affected Patients. Transparency is one of the most significant aspects of BBPM. It is essential to protect personal data and information regarding patients undergoing treatment. The spread of false information on social websites creates fear against untested data. The ability to verify BBPM data and modernize current information could present a viable way to guarantee the analysis of data accuracy. Thus, it could aid the change from an organization powered by interoperability to patient-centred interoperability.

4.2.2. Traceability. Diagnosis refers to the monitoring of affected patients. Controlling the spread of the coronavirus is essential. With BBPM, one could trace the activities of affected patients; present modern information regarding total confirmed, recovered, and death cases; and report direct combating efforts. This tracking could be completed through the transaction of the blockchain network's storing and tracking capabilities.

4.2.3. Enhanced Healthcare Protection. BBPM operates on platforms that deal with reasonable and excellent healthcare security. It could present a viable solution for monitoring coronavirus outbreaks to defend numerous patients from this infection. It monitors affected patients by regular testing to ensure timely and appropriate treatment. If hospitals have a safe and dependable health record database, it will reduce the risk of misdiagnosis.

4.2.4. Record and Exchange Treatment-Associated Data. Keeping a record of the gathered information and transmission of treatment-associated information are some of the very serious and difficult jobs during the COVID-19 epidemic. From July 2020 to June 2021, an average of 3,343,448 health records was breached each month. The BBPM system could maintain an incorruptible, decentralized, and obvious record of patient information. BBPM lets healthcare providers, doctors, and patients distribute similar data rapidly and securely.

4.2.5. Enhancing the Recovery of Affected Patients. Timely treatment can improve the recovery rate of affected patients. In addition, BBPM helps to monitor isolated cases in hospitals effectively. In a COVID-19 infection, the blockchain patient's symptoms, location, and historical health status can be recorded with high privacy. The data block spreads over distributed networks of end-users, governments, and health professionals.

4.2.6. Disease Control. To control and prevent the spread of infection, effective and accurate disease monitoring is essential. BBPM could be utilized worldwide to monitor the spread of COVID-19 infection in humans. In the COVID-19 pandemic, BBPM must sustain the victims of the virus by immutably storing patient disease signs.

5. Experimental Results

The effectiveness of the proposed work is examined in this section. The intensity of COVID-19 was high that the World Health Organization (WHO) had to announce COVID-19 as an epidemic within a week of its complete growth. The greatest difficulty many governments face is a shortage of accurate methods for diagnosing recently affected cases and predicting the danger of the coronavirus pandemic. Therefore, this paper proposes the Blockchain and Business Process Management system to resolve this COVID-19 disaster. This experiment uses a blockchain constructed utilizing POJO in Java. Blockchain makes dealing easier among unreliable groups. The blockchain is a collection of blocks, including many transactions. Each block is hashed, a hash is added, the hash is reconnected, and the hash is reshaped until there is a hash and Merkle root. Each block stores the hash of an ex-block by linking the blocks. Thus, it ensures that a block will not alter without altering the adjacent blocks. However, the experiment grasps a string of data that contains anything you can envision, including smart contracts based on the style of Ethereum. The experiment driving this BBPM system also calculates the performance of the BBPM system with a primary evaluation metric, namely, execution time.

5.1. Execution Time. Execution time (ET) is explained as the period (in a sec) between the transaction confirmation (TC) and its execution (TE) in the blockchain network shown in the following equation:

$$ET = TE - TC. \quad (1)$$

The time of execution increases as the number of transactions is raised. These transactions perform a variety of operations contained in the smart contract algorithm as explained in Algorithm 1. For example, when merely one consumer is utilizing the system, operations at a time, including testing records, hospital allocation records, treatment and monitoring records, re-examination records, corpse disposal records, discharge summary records, payment records, and statistics records, it would take 1 min 20 seconds, 30 seconds, 50 seconds, 1 min 10 seconds,

```

Testing Lab (Miner)
Res = Take a sample of the patient and examine and diagnose COVID-19 in the testing lab
if (Res == positive) then
  Patient registration
  The private key and public key generation for a patient using RSA algorithm
  Select hospital for patient isolation and treatment
  BB = Generate block body using patient details with hospital name, testing lab name, and date of testing
  EB = Encrypt BB based on the public key of the hospital
  Hash = Generate hash based on HmacSHA1 algorithm
  Timestamp = Get current date and time
  Nonce = Generate random numbers
  BH = Generate block header using previous block hash, timestamp, and nonce
  Block = Add BH with EB
  Upload Block to Blockchain
Else
  DS = Generate Discharge Summary of the patient
  DD = Get current date and time//Discharge Date
  BB = Generate block body using patient details with hospital name, testing lab name, date of testing, DS, and DD
  EB = Encrypt BB based on the public key of patient
  Hash = Generate hash based on HmacSHA1 algorithm
  Timestamp = Get current date and time
  Nonce = Generate random numbers
  BH = Generate block header using previous block hash, timestamp, and nonce
  Block = Add BH with EB
  Upload Block to Blockchain
End
Hospital
The hospital can only access the ledger of the patients who should be treated
EB = Extract EB from a block of the blockchain
BB = Decrypt EB based on the private key of the hospital
PD = View details of patient, testing lab name, and date of testing from BB
TD = Get details of treatment using patient monitoring
if (the patient health improves) then
  //Take a sample of the patient and re-examine the request to the testing lab
  BB = Generate block body using PD and TD
  EB = Encrypt BB based on the public key of testing lab
  Hash = Generate hash based on HmacSHA1 algorithm
  Timestamp = Get current date and time
  Nonce = Generate random numbers
  BH = Generate block header using previous block hash, timestamp, and nonce
  Block = Add BH with EB
  Upload Block to Blockchain
else//if the patient is death
  Dispose of the body of the patient
End
Patient
The patient can only see his or her digital ledger
EB = Extract EB from a block of the blockchain
BB = Decrypt EB based on the private key of patient
PD = View details of patient, testing lab name, date of testing, hospital name, treatment details, and discharge summary from BB
Government
Government can only see the details of statistics to ensure the patient's privacy

```

ALGORITHM 1: Blockchain and Business Process Management (BBPM) algorithm.

40 seconds, 30 seconds, 25 seconds, and 15 seconds, respectively, for these operations to be performed. This time would increase when 100 consumers are utilizing the system concurrently. The experiment assessed the effectiveness of the BBPM system using a comparison between the average size of EHRs or blocks and the average execution time for

accessing EHR from centralized storage or accessing blocks from the blockchain using existing BSF-EHR [57] and the proposed BBPM algorithms, as shown in Figure 6.

Figure 6 compares centralized storage and the existing BSF-EHR algorithm, and the proposed BBPM algorithm works quickly. Furthermore, Figure 7 shows the

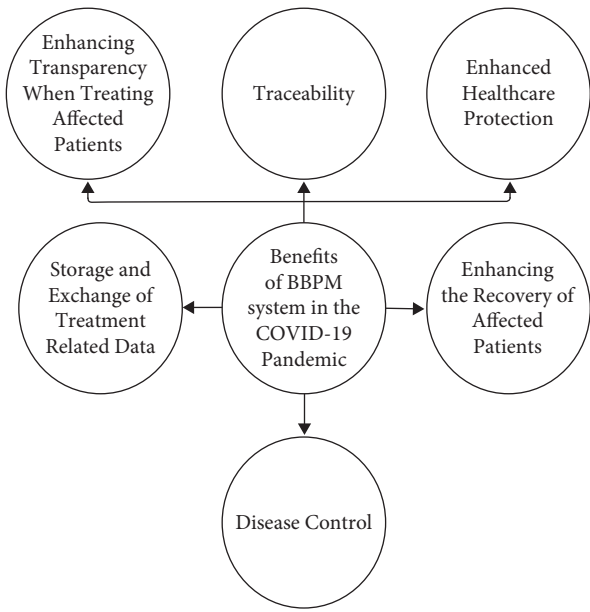


FIGURE 5: Benefits of BBPM system in the COVID-19 pandemic.

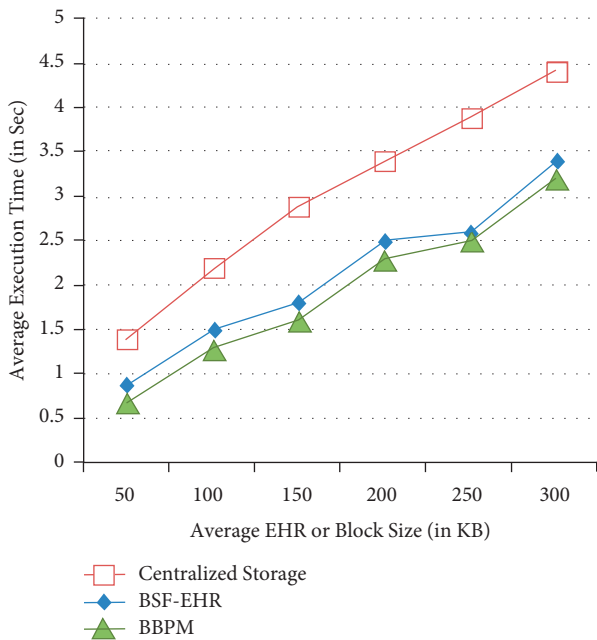


FIGURE 6: Average EHR or block size vs. average execution time.

effectiveness of the BBPM scheme using a comparison between the number of user requests and execution time for accessing EHR from centralized storage [28] or accessing blocks from the blockchain using the existing BSF-EHR and the proposed BBPM algorithms.

Figure 7 concludes that the proposed BBPM algorithm quickly responds to any user request compared with centralized storage and the existing BSF-EHR algorithm. Figure 8 compares the blockchain hash generation time of different algorithms using blockchain in healthcare, specifically Shynu et al. [58], the BSF-EHR algorithm of Abunadi et al. [57], and the proposed BBPM.

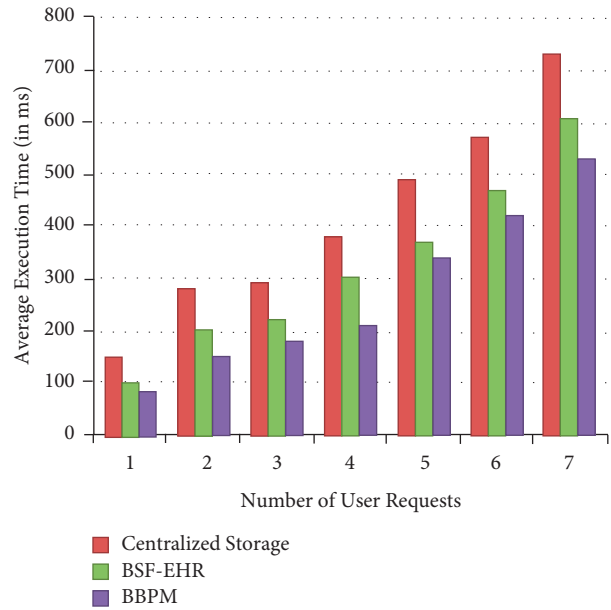


FIGURE 7: A comparison between the number of user requests and average execution time.

Figure 8 concludes that the BBPM algorithm takes less time for hash generation than other algorithms as BBPM used lightweight hash generation algorithm, namely, HmacSHA1. This lightweight hash generation algorithm takes less time for hash generation than others. The experiment calculates the execution time for accessing health records, from demanding information to getting information. In centralized storage, health records are recorded on a centralized server. If the hospital desires to access a patient’s health record, it creates a health record demand. Now, we note the present time (PT1), and they send the health record demand to a centralized server. After receiving the health record demand from the hospital, the centralized server search also obtains the patient’s health record and transmits it to the requested hospital. We then note the present time again (PT2). Thus, the execution time for using health records = (PT2 – PT1) secs.

Moreover, the execution time is frankly relative to the size of the health record. If the size of the health record is very large, the time taken to access the health record is important. On the other hand, if the size of the health record is small, the time taken for accessing the health record is little. Therefore, the execution time varies depending on the size of the health record. At BBPM, each hospital and testing lab maintains the blockchain. This blockchain contains the EHR of any patient in an encrypted format. After decryption, the testing lab or hospital can access the EHR of the patients it has tested or treated. Compared with centralized storage and the existing BSF-EHR, the proposed BBPM takes the smallest amount of time. The experiment’s consequences regarding various sizes of health records show that BBPM is better than centralized storage based on execution time. This outcome also demonstrates the efficiency of the BBPM system. Table 4 shows an evaluation of BBPM by comparison with some associated works [57].

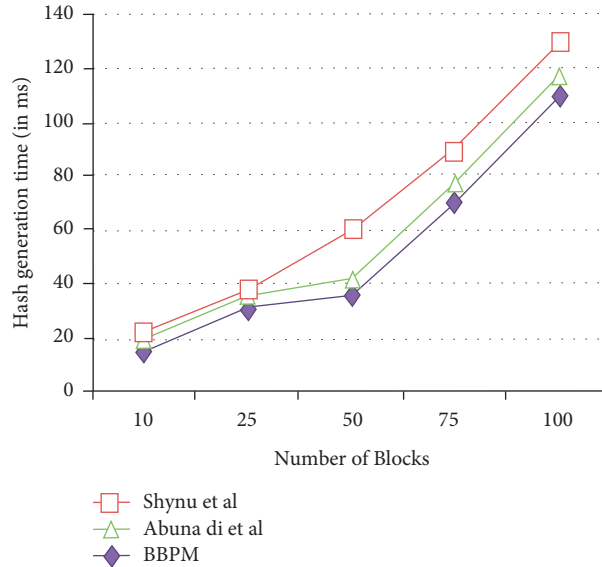


FIGURE 8: Comparison between the number of blocks and hash generation.

TABLE 4: Comparison of BBPM with some associated works.

| Associated works | Decentralized access | User authentication | Identity management | Data privacy | Flexibility | Availability | Integrity |
|--------------------|----------------------|---------------------|---------------------|--------------|-------------|--------------|-----------|
| Ying et al. [42] | No | Yes | Yes | Yes | No | No | Yes |
| Ramani et al. [32] | Yes | Yes | No | Yes | No | No | Yes |
| Xia et al. [34] | Yes | Yes | Yes | Yes | No | Yes | Yes |
| Liang et al. [22] | Yes | No | No | Yes | Yes | Yes | Yes |
| BBPM | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

6. Conclusion

This paper presented a Blockchain and Business Process Management System (BBPM) for combating the COVID-19 epidemic. The existing blockchain or business process management system is used separately to mitigate the COVID-19 pandemic. But, this paper integrates the blockchain and business process management system for COVID-19 epidemic mitigation. In this way, the benefits of both technologies can be obtained simultaneously. The main role of the BBPM system is to assist in managing the diffusion of this epidemic. The system could assist us during this epidemic disaster by presenting the advanced resolution, explosion monitoring, user privacy protection, donation monitoring, and secure daily operations. The BBPM system should minimize network delays by offering a safe environment for recording and transmitting sensitive data. A lightweight blockchain plan is essential in the medical industry to improve information confirmation and transactional communication. Creating modified ledgers that could be located on neighbouring servers in the blast region increases blockchain performance. BBPM system consumes a lot of energy because every transaction needs robust hardware resources. Scalability is the main limitation of this BBPM system. Another drawback of this BBPM system is the complexity of blockchain and the need for a comprehensive network of users. In the future, to deal with the above problems, a novel, energy-efficient and scalable BBPM

system is needed. Furthermore, the final mixture of the BBPM system with other growing techniques, such as big data and artificial intelligence, will efficiently manage deadly epidemics similar to the coronavirus.

Data Availability

The data that support the findings of this study are unavailable in any public repositories.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the Prince Sultan University for its support, which facilitated the publication of this paper.

References

- [1] S. S. Vedaei, A. Fotovvat, M. R. Mohebbian et al., "COVID-SAFE: an IoT-based system for automated health monitoring and surveillance in post-pandemic life," *IEEE Access*, vol. 8, pp. 188538–188551, 2020.
- [2] F. Hu, J. Liu, L. Li, M. Huang, and C. Yang, "IoT-based epidemic monitoring via improved gated recurrent unit model," *IEEE Sensors Journal*, 2021.

- [3] H. Wang, J. Tan, and X. Li, "Global NO₂ dynamics during the COVID-19 pandemic: a comparison between two waves of the coronavirus," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4310–4320, 2021.
- [4] S. Nisar, M. A. Zuhaib, A. Ulyasar, and M. Tariq, "A privacy-preserved and cost-efficient control scheme for coronavirus outbreak using call data record and contact tracing," *IEEE Consumer Electronics Magazine*, vol. 10, no. 2, pp. 104–110, 2020.
- [5] X. Chen, S. Jiang, Z. Li, and B. Lo, "A pervasive respiratory monitoring sensor for COVID-19 pandemic," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 2, pp. 11–16, 2020.
- [6] X. Ding, D. Clifton, N. Ji et al., "Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic," *IEEE reviews in biomedical engineering*, vol. 14, pp. 48–70, 2020.
- [7] A. Romanovs, E. Sultanovs, E. Buss, Y. Merkurjev, and G. Majore, "Challenges and solutions for resilient telemedicine services," in *Proceedings of the 2020 IEEE 8th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, pp. 1–7, Vilnius, Lithuania, April 2021.
- [8] S. Bhattacharya, P. K. Reddy Maddikunta, Q.-V. Pham et al., "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey," *Sustainable cities and society*, vol. 65, Article ID 102589, 2021.
- [9] T. R. Gadekallu, N. Khare, S. Bhattacharya et al., "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, 2020.
- [10] S. H. Ebenuwa, M. S. Sharif, M. Alazab, and A. Al-Nemrat, "Variance ranking attributes selection techniques for binary classification problem in imbalance data," *IEEE Access*, vol. 7, pp. 24649–24666, 2019.
- [11] I. Ezzine and L. Benhlima, "Technology against COVID-19 A blockchain-based framework for data quality," in *Proceedings of the 2020 6th IEEE Congress on Information Science and Technology (CiSt)*, pp. 84–89, Agadir-Essaouira, Morocco, June 2021.
- [12] S. Singh, A. S. Hosen, and B. Yoon, "Blockchain security attacks, challenges, and solutions for the future distributed iot network," *IEEE Access*, vol. 9, pp. 13938–13959, 2021.
- [13] G. Karame and S. Capkun, "Blockchain security and privacy," *IEEE Security & Privacy*, vol. 16, no. 4, pp. 11–12, 2018.
- [14] A. S. Musleh, G. Yao, and S. M. Muyeen, "Blockchain applications in smart grid-review and frameworks," *IEEE Access*, vol. 7, pp. 86746–86757, 2019.
- [15] H. Halpin and M. Piekarska, "Introduction to security and privacy on the blockchain," in *Proceedings of the 2017 IEEE European Symposium on Security and Privacy Workshops (EuroSec&PW)*, pp. 1–3, Paris, France, April 2017.
- [16] Z. Liehuang, G. Feng, S. Meng et al., "Survey on privacy-preserving techniques for blockchain technology," *Journal of Computer Research and Development*, vol. 54, no. 10, Article ID 2170, 2017.
- [17] Y. Lu, "Blockchain: a survey on functions, applications and open issues," *Journal of Industrial Integration and Management*, vol. 3, no. 4, Article ID 1850015, 2018.
- [18] A. Lamba, S. Singh, S. Balvinder, N. Dutta, and S. Rela, "Mitigating IoT security and privacy challenges using distributed ledger-based blockchain (DL-BC) technology," *International Journal for Technological Research in Engineering*, vol. 4, no. 8, 2017.
- [19] M. Attaran, "Blockchain technology in healthcare: challenges and opportunities," *International Journal of Healthcare Management*, pp. 1–14, 2020.
- [20] J. H. Lee, "BIDaaS: blockchain-based ID as a service," *IEEE Access*, vol. 6, pp. 2274–2278, 2017.
- [21] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma, and J. He, "BlocHIE: A BLOCKchain-based platform for healthcare information exchange," in *Proceedings of the 2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 49–56, Taormina, Italy, June 2018.
- [22] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–5, Montreal, Canada, October 2017.
- [23] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-Health systems via consortium blockchain," *Journal of Medical Systems*, vol. 42, no. 8, 2018.
- [24] K. N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson, and T. Hayajneh, "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring," *Journal of Medical Systems*, vol. 42, no. 7, p. 130, 2018.
- [25] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustainable Cities and Society*, vol. 39, pp. 283–297, 2018.
- [26] H. Li, L. Zhu, M. Shen, F. Gao, X. Tao, and S. Liu, "Blockchain-based data preservation system for medical data," *Journal of Medical Systems*, vol. 42, no. 8, p. 141, 2018.
- [27] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "Medblock: efficient and secure medical data sharing via blockchain," *Journal of Medical Systems*, vol. 42, no. 8, 2018.
- [28] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for secure EHRs sharing of mobile cloud based E-health systems," *IEEE Access*, vol. 7, pp. 66792–66806, 2019.
- [29] T. McGhin, K.-K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: research challenges and opportunities," *Journal of Network and Computer Applications*, vol. 135, pp. 62–75, 2019.
- [30] S. B. Wagh and J. K. Murthy, "Securing health care data for medical research using blockchain technology," *Journal of Advancement in Electronics Design*, vol. 1, no. 3, pp. 17–23, 2018.
- [31] I. Abu-Elezz, A. Hassan, A. Nazeemudeen, M. Househ, and A. Abd-Alrazaq, "The benefits and threats of blockchain technology in healthcare: a scoping review," *International Journal of Medical Informatics*, vol. 142, Article ID 104246, 2020.
- [32] V. Ramani, T. Kumar, A. Bracken, M. Liyanage, and M. Ylianttila, "Secure and efficient data accessibility in blockchain-based healthcare systems," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 206–212, Abu Dhabi, UAE, 2018 Dec 9.
- [33] K. M. Khan, J. Arshad, and M. M. Khan, "Secure digital voting system based on blockchain technology," *International Journal of Electronic Government Research*, vol. 14, no. 1, pp. 53–62, 2018.
- [34] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "MeDShare: trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14757–14767, 2017.

- [35] H. Es-Samaali, A. Outchakoucht, and J. P. Leroy, "Blockchain-based access control for big data," *International Journal of Computer Networks and Communications Security*, vol. 5, no. 7, p. 137, 2017.
- [36] H. Wang and Y. Song, "Secure cloud-based EHR system using attribute-based cryptosystem and blockchain," *Journal of Medical Systems*, vol. 42, no. 8, p. 152, 2018.
- [37] K. Balasubramanian and M. Rajakani, "Implementation of algorithms for identity based encryption and decryption," *International Journal of Chemical Reactor Engineering*, vol. 1, no. 1, pp. 52–62, 2019.
- [38] J. Sun, L. Ren, S. Wang, and X. Yao, "Multi-keyword searchable and data verifiable attribute-based encryption scheme for cloud storage," *IEEE Access*, vol. 7, pp. 66655–66667, 2019.
- [39] L. Ismail, H. Materwala, and S. Zeadally, "Lightweight blockchain for healthcare," *IEEE Access*, vol. 7, pp. 149935–149951, 2019.
- [40] U. Bodkhe, S. Tanwar, K. Parekh et al., "Blockchain for industry 4.0: a comprehensive review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020.
- [41] L. M. Bach, B. Mihaljevic, and M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1545–1550, IEEE, Opatija, Croatia, May 2018.
- [42] Z. Ying, L. Wei, Q. Li, X. Liu, and J. Cui, "A lightweight policy preserving EHR sharing scheme in the cloud," *IEEE Access*, vol. 6, pp. 53698–53708, 2018.
- [43] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: using blockchain for medical data access and permission management," in *Proceedings of the 2016 2nd International Conference on Open and Big Data (OBD)*, pp. 25–30, Vienna, Austria, August 2016.
- [44] A. Le Bris and W. El Asri, *State of Cybersecurity & Cyber Threats in Healthcare Organizations*, ESSEC Business School, Cergy, France, 2016.
- [45] D. Berdik, S. Otoum, N. Schmidt, D. Porter, and Y. Jararweh, "A survey on blockchain for information systems management and security," *Information Processing & Management*, vol. 58, no. 1, Article ID 102397, 2021.
- [46] D. R. Wong, S. Bhattacharya, and A. J. Butte, "Prototype of running clinical trials in an untrustworthy environment using blockchain," *Nature Communications*, vol. 10, no. 1, pp. 917–918, 2019.
- [47] D. G. Glover and J. Hermans, "Improving the traceability of the clinical trial supply chain," *Applied Clinical Trials*, vol. 26, no. 11/12, pp. 36–38, 2017.
- [48] I. A. Omar, R. Jayaraman, K. Salah, M. C. E. Simsekler, I. Yaqoob, and S. Ellahham, "Ensuring protocol compliance and data transparency in clinical trials using Blockchain smart contracts," *BMC Medical Research Methodology*, vol. 20, no. 1, pp. 1–17, 2020.
- [49] A. Chawla and S. Ro, "Coronavirus (COVID-19)—is blockchain a true saviour in this pandemic crisis," 2020, <https://thelivinglib.org/coronavirus-covid-19-is-blockchain-a-true-savior-in-this-pandemic-crisis/>.
- [50] M. M. Arifeen, A. Al Mamun, and M. Shamim Kaiser, "Blockchain-enabled contact tracing for preserving user privacy during COVID-19 outbreak," pp. 1–11, 2020, <http://www.preprints.org>.
- [51] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, "Digital technology and COVID-19," *Nature Medicine*, vol. 26, no. 4, pp. 459–461, 2020.
- [52] T. P. Mashamba-Thompson and E. D. Crayton, "Blockchain and artificial intelligence technology for novel coronavirus disease-19 self-testing," *Diagnostics*, vol. 10, no. 4, p. 198, 2020.
- [53] M. Shuaib, S. Alam, M. S. Nasir, and M. S. Alam, "Immunity credentials using self-sovereign identity for combating COVID-19 pandemic," *Materials Today: Proceedings*, 2021.
- [54] I. Zhou, I. Makhdoom, M. Abolhasan, J. Lipman, and N. Shariati, "A blockchain-based file-sharing system for academic paper review," in *Proceedings of the 2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pp. 1–10, IEEE, Gold Coast, Australia, December 2019.
- [55] Nasdaq, *VeChain Announces Blockchain Vaccine Tracing Solution for China*, <https://www.nasdaq.com/articles/vechain-announces-blockchain-vaccine-tracing-solution-china-2018-08-16>, 2018.
- [56] H. Hedera, "Acoer coronavirus tracker, powered by hedera hashgraph, now freely available to general public with added clinical trial data," *Hashgraph Hedera*, vol. 11, no. 2, pp. 1–6, 2020.
- [57] I. Abunadi and R. L. Kumar, "BSF-EHR: blockchain security framework for electronic health records of patients," *Sensors*, vol. 21, no. 8, Article ID 2865, 2021.
- [58] P. G. Shynu, V. G. Menon, R. L. Kumar, S. Kadry, and Y. Nam, "Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing," *IEEE Access*, vol. 9, pp. 45706–45720, 2021.