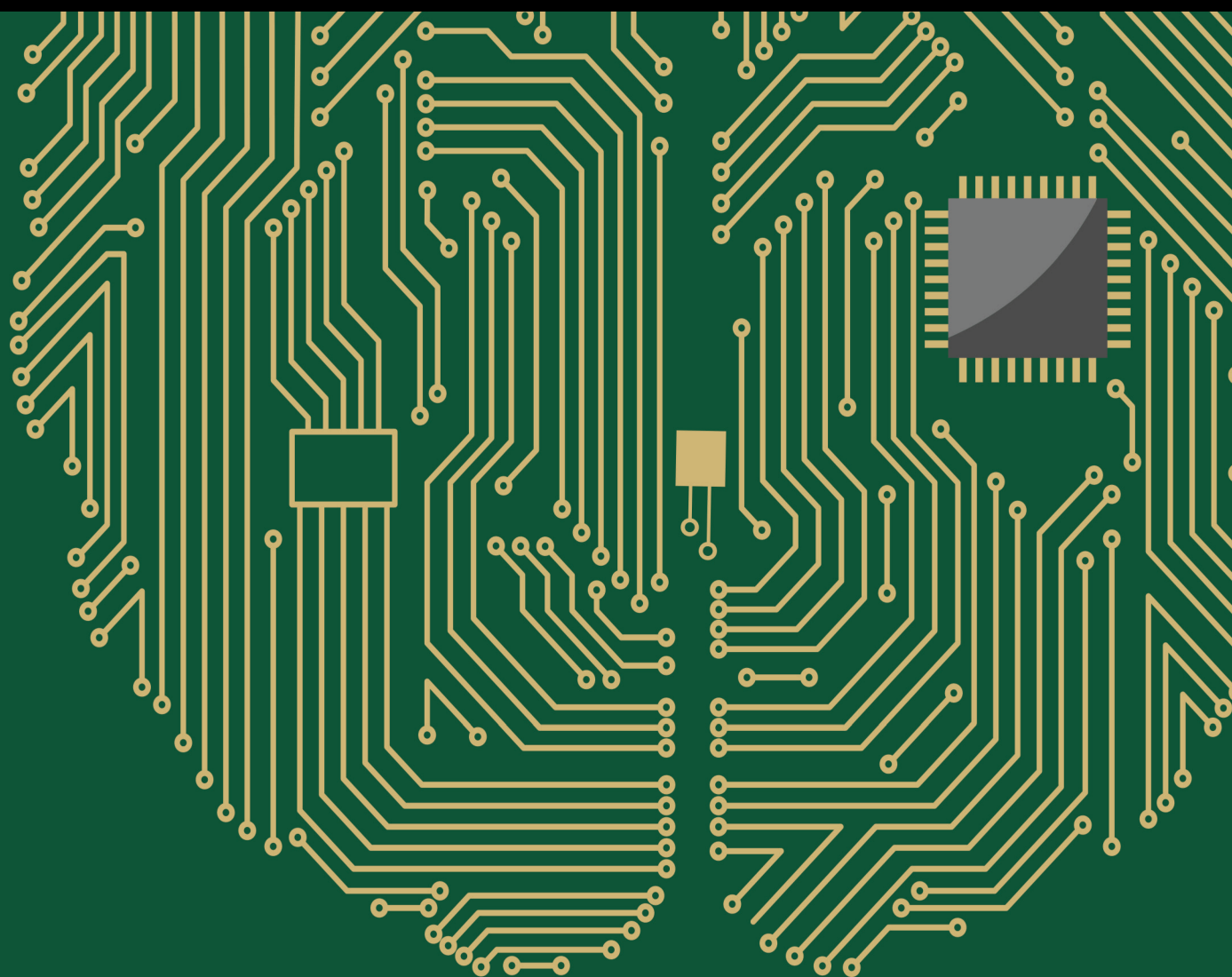


Computational Intelligence for Medical Internet of Things (MIoT) Applications

Lead Guest Editor: Ahmed A. Abd El-Latif

Guest Editors: Lo'ai A. Tawalbeh and Yassine Maleh





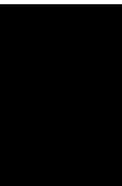
Computational Intelligence for Medical Internet of Things (MIoT) Applications

Computational Intelligence and Neuroscience

Computational Intelligence for Medical Internet of Things (MIoT) Applications

Lead Guest Editor: Ahmed A. Abd El-Latif

Guest Editors: Lo'ai A. Tawalbeh and Yassine Maleh



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in "Computational Intelligence and Neuroscience." 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

Andrzej Cichocki, Poland

Associate Editors

Arnaud Delorme, France
Cheng-Jian Lin , Taiwan
Saeid Sanei, United Kingdom

Academic Editors





Mohamed Abd Elaziz , Egypt
Tariq Ahanger , Saudi Arabia
Muhammad Ahmad, Pakistan
Ricardo Aler , Spain
Nouman Ali, Pakistan
Pietro Aricò , Italy
Lerina Aversano , Italy
Ümit Ağbulut , Turkey
Najib Ben Aoun , Saudi Arabia
Surbhi Bhatia , Saudi Arabia
Daniele Bibbo , Italy
Vince D. Calhoun , USA
Francesco Camastra, Italy
Zhicheng Cao, China
Hubert Cecotti , USA
Jyotir Moy Chatterjee , Nepal
Rupesh Chikara, USA
Marta Cimitile, Italy
Silvia Conforto , Italy
Paolo Crippa , Italy
Christian W. Dawson, United Kingdom
Carmen De Maio , Italy
Thomas DeMarse , USA
Maria Jose Del Jesus, Spain
Arnaud Delorme , France
Anastasios D. Doulamis, Greece
António Dourado , Portugal
Sheng Du , China
Said El Kafhali , Morocco
Mohammad Reza Feizi Derakhshi , Iran
Quanxi Feng, China
Zhong-kai Feng, China
Steven L. Fernandes, USA
Agostino Forestiero , Italy
Piotr Franaszczuk , USA
Thippa Reddy Gadekallu , India
Paolo Gastaldo , Italy
Samanwoy Ghosh-Dastidar, USA

Manuel Graña , Spain
Alberto Guillén , Spain
Gaurav Gupta, India
Rodolfo E. Haber , Spain
Usman Habib , Pakistan
Anandakumar Haldorai , India
José Alfredo Hernández-Pérez , Mexico
Luis Javier Herrera , Spain
Alexander Hošovský , Slovakia
Etienne Hugues, USA
Nadeem Iqbal , Pakistan
Sajad Jafari, Iran
Abdul Rehman Javed , Pakistan
Jing Jin , China
Li Jin, United Kingdom
Kanak Kalita, India
Ryotaro Kamimura , Japan
Pasi A. Karjalainen , Finland
Anitha Karthikeyan, Saint Vincent and the
Grenadines
Elpida Keravnou , Cyprus
Asif Irshad Khan , Saudi Arabia
Muhammad Adnan Khan , Republic of
Korea
Abbas Khosravi, Australia
Tai-hoon Kim, Republic of Korea
Li-Wei Ko , Taiwan
Raşit Köker , Turkey
Deepika Koundal , India
Sunil Kumar , India
Fabio La Foresta, Italy
Kuruva Lakshmana , India
Maciej Lawrynczuk , Poland
Jianli Liu , China
Giosuè Lo Bosco , Italy
Andrea Loddo , Italy
Kezhi Mao, Singapore
Paolo Massobrio , Italy
Gerard McKee, Nigeria
Mohit Mittal , France
Paulo Moura Oliveira , Portugal
Debajyoti Mukhopadhyay , India
Xin Ning , China
Nasimul Noman , Australia
Fivos Panetsos , Spain

Evgeniya Pankratova , Russia
Rocío Pérez de Prado , Spain
Francesco Pistolesi , Italy
Alessandro Sebastian Podda , Italy
David M Powers, Australia
Radu-Emil Precup, Romania
Lorenzo Putzu, Italy
S P Raja, India
Dr.Anand Singh Rajawat , India
Simone Ranaldi , Italy
Upaka Rathnayake, Sri Lanka
Navid Razmjooy, Iran
Carlo Ricciardi, Italy
Jatinderkumar R. Saini , India
Sandhya Samarasinghe , New Zealand
Friedhelm Schwenker, Germany
Mijanur Rahaman Seikh, India
Tapan Senapati , China
Mohammed Shuaib , Malaysia
Kamran Siddique , USA
Gaurav Singal, India
Akansha Singh , India
Chiranjibi Sitaula , Australia
Neelakandan Subramani, India
Le Sun, China
Rawia Tahrir , Iraq
Binhua Tang , China
Carlos M. Travieso-González , Spain
Vinh Truong Hoang , Vietnam
Fath U Min Ullah , Republic of Korea
Pablo Varona , Spain
Roberto A. Vazquez , Mexico
Mario Versaci, Italy
Gennaro Vessio , Italy
Ivan Volosyak , Germany
Leyi Wei , China
Jianghui Wen, China
Lingwei Xu , China
Cornelio Yáñez-Márquez, Mexico
Zaher Mundher Yaseen, Iraq
Yugen Yi , China
Qiangqiang Yuan , China
Miaolei Zhou , China
Michal Zochowski, USA
Rodolfo Zunino, Italy





Contents

ResNet-50 for 12-Lead Electrocardiogram Automated Diagnosis

Nizar Sakli, Haifa Ghabri, Ben Othman Soufiene , Faris. A. Almalki , Hedi Sakli , Obaid Ali , and Mustapha Najjari


Research Article (16 pages), Article ID 7617551, Volume 2022 (2022)

Internet of Medical Things (IoMT) and Reflective Belief Design-Based Big Data Analytics with Convolution Neural Network-Metaheuristic Optimization Procedure (CNN-MOP)

A. Sampathkumar , Miretab Tesfayohani , Shishir Kumar Shandilya , S. B. Goyal, Sajjad Shaukat Jamal, Piyush Kumar Shukla, Pradeep Bedi , and Meshal Albeedan



Research Article (14 pages), Article ID 2898061, Volume 2022 (2022)

Examining the Determinants of Patient Perception of Physician Review Helpfulness across Different Disease Severities: A Machine Learning Approach

Adnan Muhammad Shah , Wazir Muhammad , and KangYoon Lee 




Research Article (15 pages), Article ID 8623586, Volume 2022 (2022)

Ensemble of Deep Learning Based Clinical Decision Support System for Chronic Kidney Disease Diagnosis in Medical Internet of Things Environment

Suliman A. Alsuhibany, Sayed Abdel-Khalek, Ali Algarni, Aisha Fayomi, Deepak Gupta , Vinay Kumar, and Romany F. Mansour 

Research Article (13 pages), Article ID 4931450, Volume 2021 (2021)

CNN-Based Personal Identification System Using Resting State Electroencephalography

Yongdong Fan , Xiaoyu Shi , and Qiong Li 





Research Article (12 pages), Article ID 1160454, Volume 2021 (2021)

Secure Health Monitoring Communication Systems Based on IoT and Cloud Computing for Medical Emergency Applications

Ali I. Siam , Mohammed Amin Almaiah , Ali Al-Zahrani, Atef Abou Elazm, Ghada M. El Banby, Walid El-Shafai, Fathi E. Abd El-Samie, and Nirmeen A. El-Bahnasawy





Research Article (23 pages), Article ID 8016525, Volume 2021 (2021)

A Computer-Aided Diagnosis System Using Deep Learning for Multiclass Skin Lesion Classification

Mehak Arshad, Muhammad Attique Khan , Usman Tariq , Ammar Armghan, Fayadh Alenezi , Muhammad Younus Javed, Shabnam Mohamed Aslam , and Seifedine Kadry




Research Article (15 pages), Article ID 9619079, Volume 2021 (2021)

An IoMT-Enabled Smart Healthcare Model to Monitor Elderly People Using Machine Learning Technique

Muhammad Farrukh Khan, Taher M. Ghazal, Raed A. Said, Areej Fatima, Sagheer Abbas , M.A. Khan , Ghassan F. Issa, Munir Ahmad , and Muhammad Adnan Khan 


Research Article (10 pages), Article ID 2487759, Volume 2021 (2021)

A Blockchain-Based Federated Learning Method for Smart Healthcare

Yuxia Chang , Chen Fang , and Wenzhuo Sun 

Research Article (12 pages), Article ID 4376418, Volume 2021 (2021)

Research on PM2.5 Spatiotemporal Forecasting Model Based on LSTM Neural Network

Fang Zhao, Ziyi Liang, Qiyang Zhang, Dewen Seng , and Xiyuan Chen




Research Article (10 pages), Article ID 1616806, Volume 2021 (2021)



Medical Text Classification Using Hybrid Deep Learning Models with Multihead Attention

Sunil Kumar Prabhakar  and Dong-Ok Won 

Research Article (16 pages), Article ID 9425655, Volume 2021 (2021)






A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning

Baha Ihnaini , M. A. Khan , Tahir Abbas Khan, Sagheer Abbas , Mohammad Sh. Daoud, Munir

Ahmad , and Muhammad Adnan Khan 

Research Article (11 pages), Article ID 4243700, Volume 2021 (2021)

Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning

Rohit Bharti , Aditya Khamparia , Mohammad Shabaz , Gaurav Dhiman , Sagar Pande , and

Parneet Singh 

Research Article (11 pages), Article ID 8387680, Volume 2021 (2021)

Research Article

ResNet-50 for 12-Lead Electrocardiogram Automated Diagnosis

Nizar Sakli^{1,2}, **Haifa Ghabri**², **Ben Othman Soufiene**³, **Faris. A. Almalki**⁴,
Hedi Sakli^{1,2}, **Obaid Ali**⁵, and **Mustapha Najjari**⁶

¹EITA Consulting, 5 Rue du Chant des Oiseaux, Montesson 78360, France

²MACS Research Laboratory RL16ES22, National Engineering School of Gabes, Gabes University, Gabes 6029, Tunisia

³PRINCE Laboratory Research, ISITcom, Hammam Sousse, University of Sousse, Sousse 4023, Tunisia

⁴Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

⁵Ibb University, Department of Computer Science and Information Technology, Ibb, Yemen

⁶LR18ES34 PEESE, National Engineering School of Gabes, Gabes University, Gabes 6029, Tunisia

Correspondence should be addressed to Obaid Ali; obaid.ali2016@gmail.com

Received 2 November 2021; Revised 13 January 2022; Accepted 22 March 2022; Published 28 April 2022

Academic Editor: Yassine Maleh

Copyright © 2022 Nizar Sakli 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.

Nowadays, the implementation of Artificial Intelligence (AI) in medical diagnosis has attracted major attention within both the academic literature and industrial sector. AI would include deep learning (DL) models, where these models have been achieving a spectacular performance in healthcare applications. According to the World Health Organization (WHO), in 2020 there were around 25.6 million people who died from cardiovascular diseases (CVD). Thus, this paper aims to shed the light on cardiology since it is widely considered as one of the most important in medicine field. The paper develops an efficient DL model for automatic diagnosis of 12-lead electrocardiogram (ECG) signals with 27 classes, including 26 types of CVD and a normal sinus rhythm. The proposed model consists of Residual Neural Network (ResNet-50). An experimental work has been conducted using combined public databases from the USA, China, and Germany as a proof-of-concept. Simulation results of the proposed model have achieved an accuracy of 97.63% and a precision of 89.67%. The achieved results are validated against the actual values in the recent literature.

1. Introduction

Nowadays, the medical field requires new techniques and technologies in order to evaluate information objectively. According to data from the World Health Organization (WHO), cardiovascular diseases (CVD) represent the leading cause of death globally, where the CVDs account for more than 30% of global mortality each year, and it is estimated to reach around 130 million people by 2035 [1]. Therefore, researchers are developing new methods for preventing, detecting, and treatment of diseases related to the CVD. There are many types of cardiovascular abnormalities, while this study focuses on 26 anomalies, which will be cited later.

The electrocardiogram (ECG) is a recording of the electrical activity of the human heart, which is deemed as a

noninvasiveness and real-time exam. It is still one of the essential pillars of the diagnosis of cardiac problems. In recent years, the methods of analysing CVDs have been strengthened by the introduction of imaging procedures, especially the echocardiogram. However, this does not change the importance and usefulness of ECGs, and the parameters could be extracted from this signal. The number of leads on a typical ECG acquisition equipment divides it into 1-lead, 3-lead, 6-lead, and 12-lead ECG. The 12-lead ECG is the most often utilized kind in clinical practice due to its ability to concurrently capture the potential changes of 12 sets of electrode patches attached to the body in standardized places [2]. When comparing to other types of ECG acquisition equipment, 12-lead ECG provides more information on cardiac activity and is frequently utilized in hospital for diagnosis and treatment. In fact, many essential parameters

can be extracted from the ECG signal; for instance, the duration and patterns of the various waves, which are indicative of specific cardiac abnormalities.

Professional doctors frequently make ECG analysis and interpretation [3], which is heavily reliant on training, qualifications, experiences, and expertise; thus it is difficult to extract all information from ECG signals [4, 5]. In practice, manual detection of characteristic waves of the ECG signal and classification of heartbeats are difficult and tedious tasks, especially to analyse long-term recordings as Holter examination or ambulatory cases for continuous monitoring in intensive care and resuscitation wards.

With the progress of physical hardware technologies and algorithm, computer-assisted medical diagnoses (CAMD) have become vital in diagnosing CVDs. CAMD based on ECG signals can give professional suggestions or decide instantly by searching for characteristic patterns. It can help doctors make diagnoses and appears to be required due to the huge number of patients in critical care units where they need continuous monitoring. This is how CAMD looked to use the ECG signal to help in cardiac diagnosis. These systems should be easy to set up, upgradeable, accurate, durable, and dependable. The authors of [6] emphasised the importance of using optimization techniques to enhance efficiency for prediction in healthcare applications.

Over the past decades, many techniques for detecting CVDs have been proposed, where some of them are based on signal processing techniques and classification algorithms like support vector machines (SVMs). Deep neural network-based machine learning (ML) and convolutional neural networks (CNN) methods have lately emerged as efficient tools in large applications such as computer vision and natural language processing. Noticeably, coupling ML and DL with healthcare has brought up massive advantages and researchers are striving to find more innovative solutions.

This work aims to classify 27 classes, with ECG signals containing 26 types of CVDs and normal sinus rhythm. This classification where we used four databases contains 42511 ECG records to train, validate, and evaluate models such as CPSC 2018, CPSC 2018-Extra [7], PTB-XL [8], and Georgia [7]. The used dataset contains ECG 12-leads signals, which is a typical ECG set used in clinical cases and hospitals. It is trained with a model based on Residual Neural Networks-50 (ResNet-50) from CNN methods, which is known as one of the most efficient models in classification.

The rest of this paper is structured as follows. Section 2 presents an overview of related works in the literature; Section 3 represents background information on the interpretation of an ECG. Section 4 describes the proposed model and our simulation workflow. The proposed ECG classification model results are discussed in Section 5. Finally, Section 6 presents the conclusion and future works.

2. Related Work

DL is a subdivision of ML; ML is a subdivision of AI and AI is enabling the machine to act like a human. ML is a way for achieving AI using algorithms trained on data, while DL is inspired by the structure of the human brain or also known

as an artificial neural network. The features in ML are picked out with an expert in the domain, whereas in DL they are detected by the neural network without human intervention. That is why DL needs much higher volume of data to be trained to obtain best performance. AI has been shown in numerous experiments to be capable of automatically identifying anomalies registered by an ECG.

Generally, the databases used in papers about ECG diagnosis are public. The first one is from PhysioNet, Massachusetts Institute of Technology-Beth Israel Hospital (MITBIH) [9] which contained only 49 recordings with 30-minute length of each subject, including five classes, normal (*N*), ventricular ectopic (*V*), supraventricular ectopic (*S*), fusion (*F*), and unknown (*Q*). Enabio et al. [10] used MITBIH as a database for ECG classification [11–16]. The second database largely used is Physiological Signal Challenge 2018 (CPSC) [7] which is a public too. It comprises 687,712 lead ECG recordings including eight arrhythmias IAVB (1st degree AV block), AF (atrial fibrillation), LBBB (left bundle branch block), PAC (premature atrial contraction), RBBB (complete right bundle branch block), and SNR (sinus normal rhyme) [17–19]. The third one is Physikalisch Technische Bundesanstalt (PTB) [20] diagnostic database, which contains 54,912 lead ECG records from 290 individuals [21–23]. Selvalingam et al. [24] used private databases to predict ventricular arrhythmias with a DL model, CNN. In addition, Smith et al. [25] collected their data to interpret ECG arrhythmias.

Some studies instead have used more than one. For example, Li et al. [26] used five databases (FANTASIA, CEBSDB, NSRDB, STDB, and AFD). However, they do not combine the data to categorize ECG; instead, they test their model for each data set separately. Zhang et al. [27] used four databases, Acharya et al. [28] constructed 4 sets from a combination of three databases (MITBIH [9], FANTASIA [29], and BIDMC [30]). The study varied on using balanced and imbalanced ones. Wang et al. [31] used two databases (MIT-BIH [9], CPSC2018 [7]) to classify ECG with a recurrent neural network (RNN) model. Table 1 lists the different databases used in classifying the ECG signals.

In fact, in their workflows, ML methods consider four fundamental steps:

- (i) Signal preprocessing, which includes resampling, noise removal (e.g., band-pass filters), and signal normalization/standardization.
- (ii) Heartbeat segmentation, which entails detecting the R-peak (e.g., QRS complex) using algorithms like Pan and Tompkins algorithm [32], the open-source GQRS software supplied by the PhysioNet community.
- (iii) Feature extraction, which entails converting raw signals into features that are most suited to the job at hand (e.g., classification, prediction, and regression.).
- (iv) ECG signal analysis using traditional machine learning approaches such as multilayer perceptron (MLP) and decision trees.

TABLE 1: Overview of various databases using ECG classification.

Database	Subjects	Records	Duration	Frequency (Hz)	Leads	References
MITBIH [9]	47	48	30 min	360	2	[10–13]
CPSC 2018 [7]	6877	6877	6–60 sec	500	12	[14–16]
PTB [19]	290	549	Not specified	1000	12	[18–20]
Fantasia [29]	40	40	120 min	250	Not specified	[23–25]
BIDMC [30]	Not specified	53	8 min	125	2	[26]

Even though traditional ML algorithms with hand-crafted features have achieved good results for ECG analysis, deep neural network (DNN) methods with the power of automated features extraction and representation learning have demonstrated human-level performance in analysing biomedical signals [33].

DL approaches, on the other side, need a large quantity of data and many parameters to be learnt. Furthermore, most of the suggested methodologies and workflows for evaluating ECG signals are specific to the task, at hand, and cannot be applied to other biomedical topics. Various studies have classified ECG data using a DL approach. Ribeiro et al. [34] created an end-to-end DNN that is capable of identifying six ECG anomalies with a database of 2,322,513 ECG records. The detection accuracy ranges from 83.3% to 100%. This DL model achieves an overall accuracy of 97.57% for the prediction of CVDs. Ahsanuzzman et al. [35] investigated the classification and prediction of a single arrhythmia class, atrial fibrillation (AFib), using ECG signals. A hybrid long short-time memory (LSTM) and RNN was used for this task. Obeidat et al [36] classified six ECG beats classes using a hybrid DL model that combines CNN and LSTM. The hybrid model achieves accuracy and precision of 98.22% and 98.27%, respectively. Further, [37] stressed on utilizing an optimization method to improve efficiency in healthcare applications.

Adedinsewo et al. [38] constructed a CNN model for classifying arrhythmia type left ventricular systolic dysfunction (LVSD) where the attaining accuracy was 85.9%. Xiong et al. [39] decided to train 8528 ECG records from CPSC data, with ResNet-16 model achieving an accuracy of 82%. Zhang et al. [17] used CPSC2018 database, which contained 6877 ECG recordings to build a 34-layer ResNet 1D model in order to detect 9 distinct arrhythmias in 12-lead ECG signals. This model had a classification accuracy of 96.6% for ECG signals.

It can be said that the number of records used is a bit small to train a model of DL; however, as mentioned above, DL needs a much higher volume of data. In this study, we choose to combine four public databases to confirm the efficacy of the model proposed. In this paper, the proposed model has succeeded to diagnose the majority of 27 classes, including 26 CVDs and normal sinus rhythm, which will assist domain experts in identifying patient records, while other researches used ECG to classify just one or two anomalies [35, 38].

3. Background Knowledge

It is critical to comprehend electrical cardiac function, since the heart is a mechanical organ that ensures periodic contraction and relaxation. Cells grouped at the nodal level are

responsible for an electrical flow that spreads to nearby heart cells (myocardial). Following that, it recontacts to be able to expel blood from other organs.

3.1. ECG Principal. The ECG is a recording of the electrical activity of the heart, which is usually shown as a graph of voltage values vs. time. Electrodes are used to detect electrical changes caused by cardiac muscle cell depolarization and repolarization at a distance from the heart, through the skin. To note, an electrocardiograph is used in this examination. Figure 1 represents a simplified diagram of the conductive elements of the heart, which consists of conductive tissues which are the bundle of His, Bachmann’s bundle, the left and right bundle branches, the Purkinje fibres, and cardiac myocytes themselves. Contractile tissues are the atrial and ventricular wall myocytes. This figure is vital in showing the main components of the heart, so extracting data and signals can be done in more accurate way.

3.2. The Foundation of ECG Interpretation. ECG interpretation includes an assessment of the morphology (appearance) of the waves and intervals on the ECG curve. Therefore, ECG interpretation requires a structured assessment of the waves and intervals. Figure 2 shows a depolarization/repolarization phase of the heart that are represented electrocardiographically by various P waves, QRS, and T waves.

- (i) P wave: This is a result of atrial depolarization, which is initiated by the sinus node. Pacemaker cells at this node carry the signal to the right and left atria. The ECG demonstrates abnormal atrial repolarization.
- (ii) QRS complex: This is the average of the inner (endocardial) and outer (epicardial) cardiomyocyte depolarization waves. A typical QRS pattern is formed when endocardial cardiomyocytes depolarize somewhat earlier than the outer layers.
 - (a) The Q wave is the first negative deflection following the P wave. The Q is missing if the first deflection is not negative.
 - (b) The R wave is the positive deflection.
 - (c) The S wave is the negative deflection that occurs following the R wave.
- (iii) T wave: It indicates the ventricular repolarization. During the T wave, there is no action in the heart muscle.

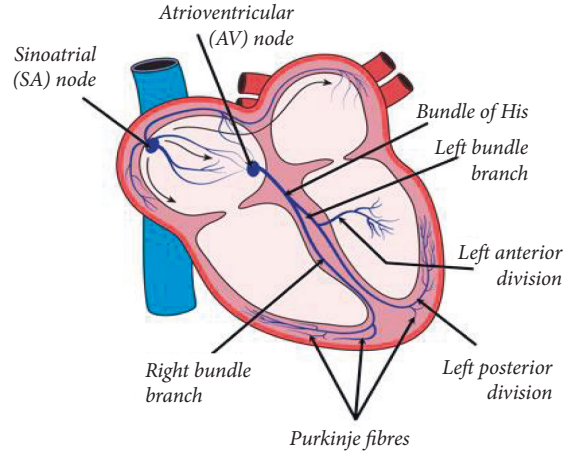


FIGURE 1: The conductive elements of the heart.

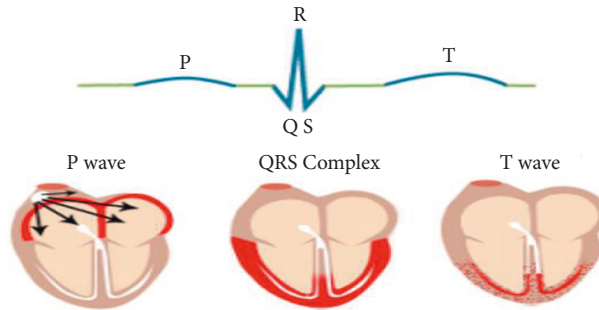


FIGURE 2: Depolarization/repolarization phases of the heart that are represented electrocardiographically by various P waves, QRS, and T waves.

Pathologies or abnormalities in ECG analysis are discovered and categorized based on their departure from normal cardiac rhythm. Normal sinus rhythm (NSR) refers to normal cardiac activity in which there is no deviation or change in the morphology of the ECG signal.

This paper focuses on classifying 27 classes of ECG signal; the classes are 1st Degree AV Block (IAVB), Low QRS Voltages (LQRSV), Right Axis Deviation (RAD), Atrial Fibrillation (AF), Nonspecific Intraventricular Conduction (NSIVCB), Atrial Flutter (AFL), Bradycardia (Brady), Complete Right Bundle Branch Block (CRBBB), Incomplete Right Bundle Branch Block (IRBBB), Left Anterior Fascicular Block (LANfb), Pacing Rhythm (PR), Right Bundle Branch Block (RBBB), Premature Atrial Contraction (PAC), Premature Ventricular Contractions (PVC), Sinus Arrhythmia (SA), Sinus Bradycardia (SB), Sinus Rhythm (SNR), Sinus Tachycardia (Stach), Supraventricular Premature Beats (SVPB), Left Axis Deviation (LAD), Prolonged Pr Interval (LPR), Prolonged Qt Interval (LQT), T Wave Abnormal (Tab), T Wave Inversion (Tinv), Left Bundle Branch Block (LBBB), Qwave Abnormal (Qab), and Ventricular Premature Beats (VPB). Figure 3 shows samples from each of the 27 ECG signal classes.

4. Proposed Model

This paper proposes a ResNet model with four databases to classify ECG signals. This section starts by presenting the architecture of model proposed and then highlighting our working method.

4.1. Proposed Model Architecture. In this paper, ResNet-50 is the proposed model for features extraction. In fact, it combines convolutional neural network for ECG diagnoses. Figure 4 illustrates an overview of the model architecture. Making the model training tractable has been assured by the residual blocks with shortcut connections. As input, the model takes an ECG signal $x \in \mathbb{R}^{n_{\text{samples}} \times 12}$. As outputs, the result of the multilabel classification is $\tilde{y} \in \mathbb{R}^{1 \times 27}$.

A 1D convolution layer (conv1D) was applied to these inputs, a batch normalization layer (BN), a rectified linear unit activation layer (ReLU), and a Max Pooling layer. Also, 16 residual blocks have been used to extract deep features. There are two types of residual blocks as follows:



(a)

FIGURE 3: Continued.

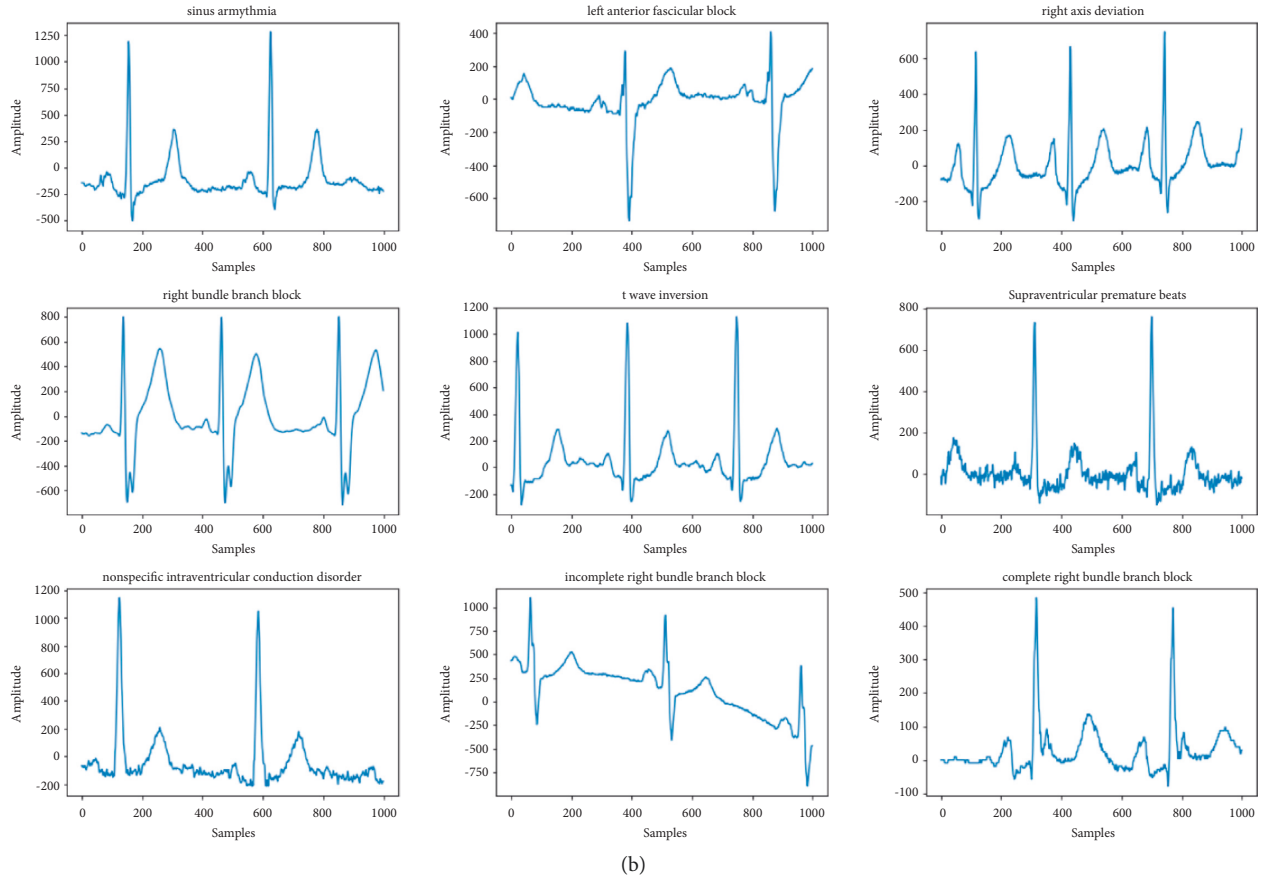


FIGURE 3: Samples of each class of ECG.

- (i) Res_Block_1 is composed of three Conv1d layers, three BatchNorm1d layers, and two ReLU activation layers. On the one hand, one Conv1d layer and one BatchNorm1d layer are used to match dimensions and skip connections on the other.
- (ii) Res_Block_2 is composed only of three Conv1d layers, three BatchNorm1d layers, and two ReLU activation layers.

The Conv1d layers are used for extracting features and the BatchNorm1d layers are used to make the model faster and stable. The ReLU layers are introduced to perform nonlinear activation. The features extracted by the residual blocks are pooled using Average Pooling, where the pooling results are collected and sent to the output layer, which uses the sigmoid activation function to produce predictions.

4.2. Dataset Characteristics. The used dataset in this work combines four public databases containing 42,511 recordings of 12-lead ECG. This type of ECG is the most used in clinical cases because of the large amount of information that it generates. These recordings are sampled at a frequency of 500 Hz. Table 2 describes the characteristics of each database.

The used dataset in this work contains 27 classes, where 26 classes are of CVDs and a class represents a normal heart state. Figure 5 shows the distribution of these classes on each

database. Figure 6 illustrates an overview of its distribution in the dataset where a problem of data imbalance and data insufficiency are noticed.

4.3. Simulation Workflow. Figure 7 illustrates the workflow of the proposed method that has been implemented in our study. Each step of this workflow will be explained in the following subsections.

4.3.1. Data Preprocessing. The length of the signals of the four databases varies from 6 seconds to 60 seconds. Therefore, it has been decided to uniform all the lengths n samples. Since the common length is 10 seconds, we set 5000 samples (10 s, 500 Hz as sampling rate). For ECGs recordings having a duration superior to 10 seconds, the first 10 s was kept. Otherwise, signals will be zero-padded until having 10 s as a duration. Figure 8 describes this preprocessing technique, where in this step, for the signal counting less than 5000 samples will be zero-padded to obtain 5000 samples. For signals containing more than 5000, samples above this value will be discarded.

Figure 9 demonstrates in more detail the technique of uniformly reducing the length of an ECG signal, in which we have a signal with a length of 7500 reduced to 5000 to train our model. Data preprocessing is explained as per Algorithm 1.

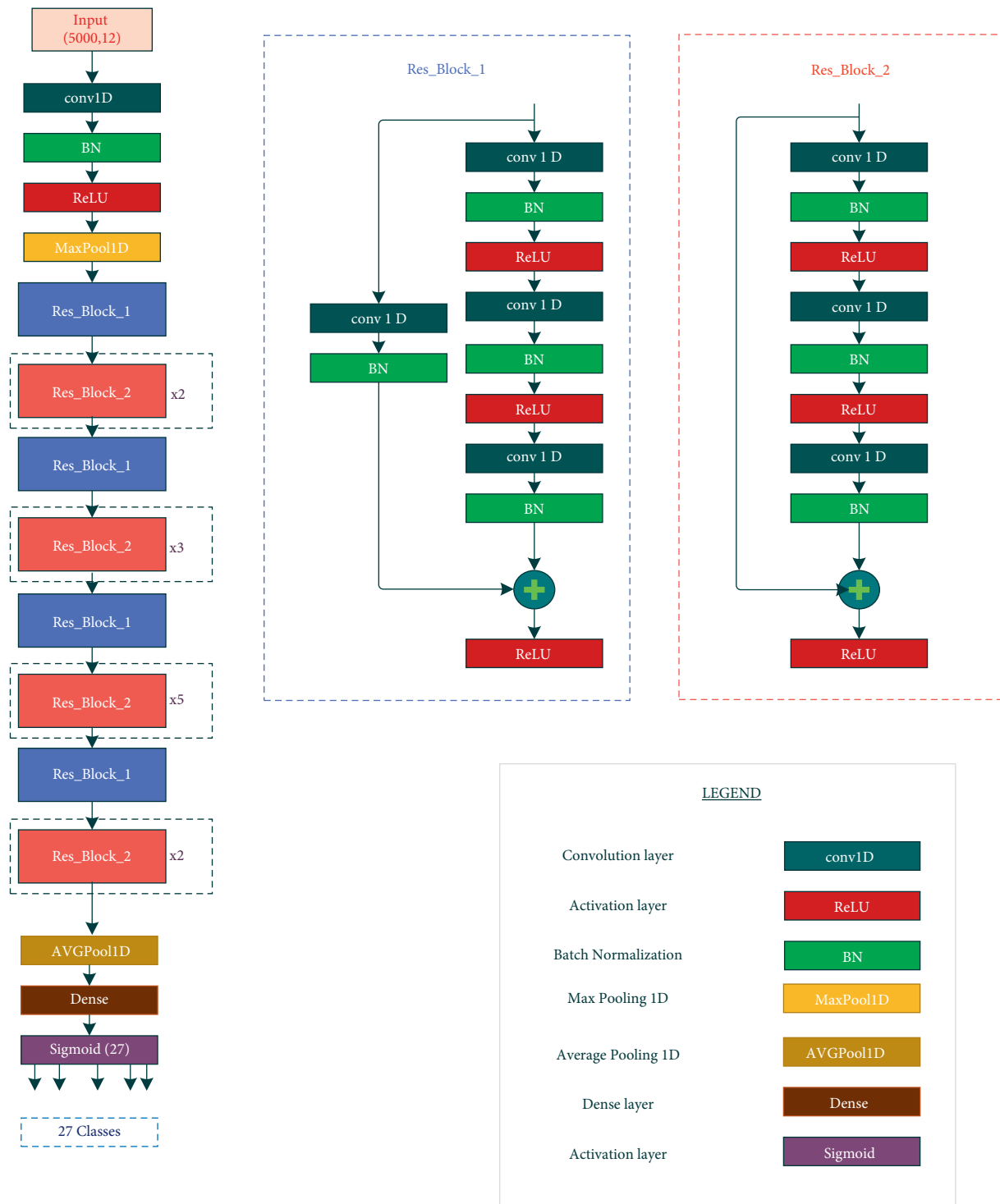


FIGURE 4: Presentation of the proposed model.

TABLE 2: Description of each database’s characteristics.

Database	Sources	Number of ECG recordings	Length of ECG recordings
CPSC 2018 [7]	China Physiological Signal Challenge in 2018	6877 (i) M: 3699 (ii) F: 3178	6 s–60 s
CPSC 2018 EXTRA [7]		3453 (i) M: 1843 (ii) F: 1610	6 s–60 s
PTB-XL [8]	Physikalisch Technische Bundesanstalt	21,837 (i) M: 11,379 (ii) F: 10,458	10 s
Georgia [7]	Georgia	10,344 (i) M: 5551 (ii) F: 4793	10 s

	- CPSC 2018	- CPSC 2018 - Extra	- PTB-XL	- Georgia
IABV (2394) -	722	106	797	769
AF (3458) -	1221	153	1514	570
AFL (313) -	0	54	73	186
Brady (277) -	0	271	0	6
CRBBB (683) -	0	113	542	28
IRBBB (1611) -	0	86	1118	407
LAnFB (1806) -	0	0	1626	180
LAD (6086) -	0	0	5146	940
LBBB (1041) -	236	38	536	231
LQRSV (556) -	0	0	182	374
NSIVCB (996) -	0	4	789	203
PR (299) -	0	3	296	0
PAC (1726) -	616	73	398	639
PVC (188) -	0	188	0	0
LPR (340) -	0	0	340	0
LQT (1513) -	0	4	118	1391
QAb (1013) -	0	1	548	464
RAD (427) -	0	1	343	83
RBBB (2400) -	1857	1	0	542
SA (1238) -	0	11	772	455
SB (2359) -	0	45	637	1677
SNR (20766) -	918	4	18092	1752
STach (2390) -	0	303	826	1261
SVPB (211) -	0	53	157	1
TAb (4673) -	0	22	2345	2306
TInv (1111) -	0	5	294	812
VPB (365) -	0	8	0	357

FIGURE 5: Pathologies distribution in each database.

4.3.2. *Data Augmentation.* As shown in Figures 4 and 5, the problem of data insufficiency and data imbalance is serious for CVDs. To deal with this issue, amplitude scaling was applied as a data augmentation technique. The creation of realistic data to prevent data scarcity is known as data

augmentation. Practically, it enhances the model robustness and lessens the fitting concerns against similar examples [14]. Amplitude scaling is the multiplication of ECG signals by a random factor α . This technique aims to compress or stretch the magnitude. The factor α is sampled from normal

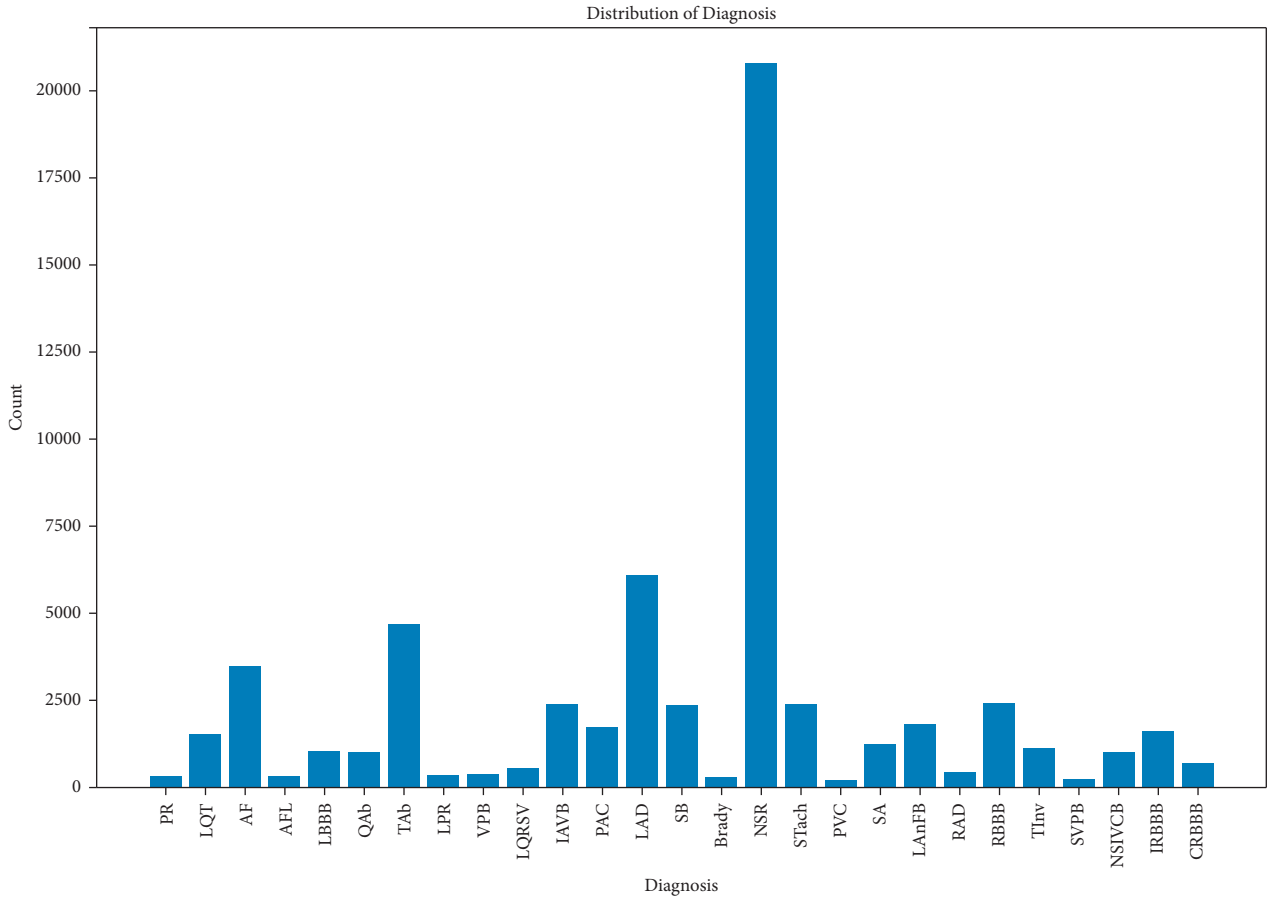


FIGURE 6: Histogram of pathology distribution in the dataset.

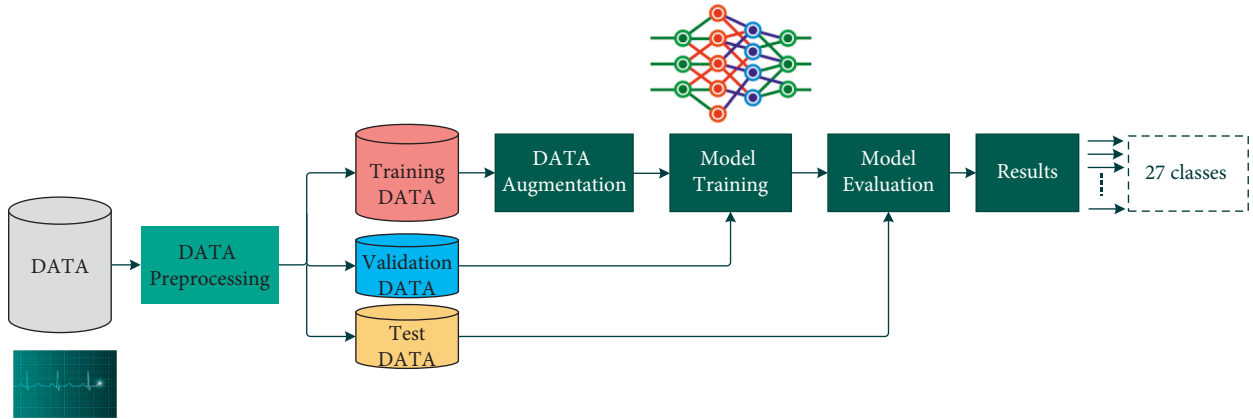


FIGURE 7: Work methodologies.

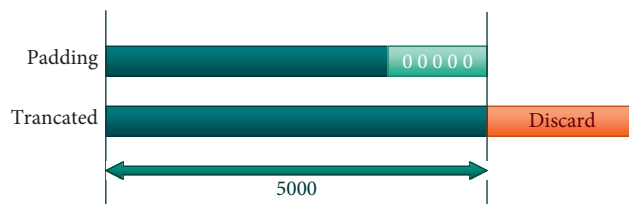


FIGURE 8: Preprocessing technique.

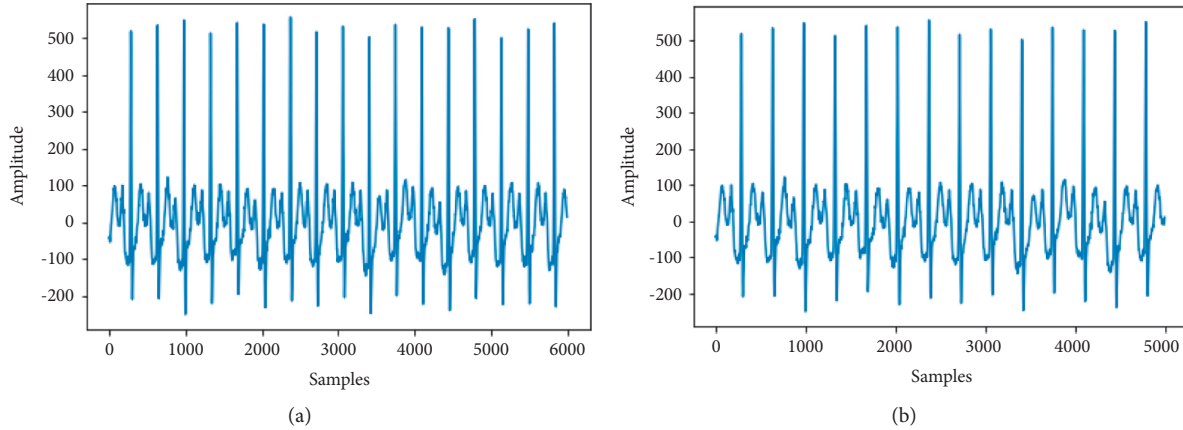


FIGURE 9: Preprocessing example.

```

Input: Training data
Output: Training data pre-processed
(1) X = ECG signal
(2) l = length (X)
(3) if l < 5000
(4) for I from 1 to 5000 do
(5) X [i] = 0
(6) end for
(7) else X = X [:5000]
(8) end if
(9) return X
(10) END

```

ALGORITHM 1: Data preprocessing.

```

Input: Training data pre-process
Output: Training data pre-process augmented
(1)  $\alpha = 0.1$ 
(2) X = ECG signal
(3) ScalingFactor = random.normal (loc = 1.0, scale =  $\alpha$ , size = (1, X.shape [1]))
(4) Noise = matmul (ones ((X.shape [0], 1)), scalingFactor)
(5) return X * Noise

```

ALGORITHM 2: Amplitude scaling.

distribution $N(1, 0.1)$. The algorithm of amplitude scaling algorithm is shown in Algorithm 2.

4.3.3. Data Split (Train, Validation, Test). As mentioned in Section 4.2 the dataset used comprises 42,511 ECG records. First, dataset has been split into two sets: test set and training and validation set in the ratio of 0.75 : 0.25. After this, 10-fold stratified cross-validation approach on the training and validation set was applied. This will return 10 stratified folds. These folds will be made by preserving the percentage of samples for each class. This forces the class distribution in each data split to match the distribution in the whole training dataset.

Generally, the training data is dedicated to train the model. The validation data is reserved for optimizing the model. Therefore, a search for the best parametrization without using the test data is done to measure the model performance and allow us to evaluate the model generalization ability. Finally, we obtain a test set and training and validation with 10,627 and 31,884 ECG records, respectively. In addition, the shapes of each training fold and validation fold are 25,507 and 6377 ECG records, respectively. Figure 10 illustrates an overview of this proposed method.

4.3.4. Training and Evaluation. The trial-and-error approach is used to determine the hyperparameters. In essence,

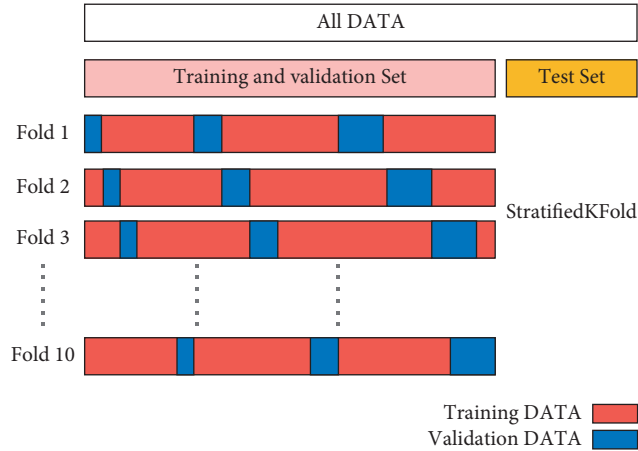


FIGURE 10: Split data method.

Adam with a learning rate of 10^{-3} is employed as the optimizer. The binary cross-entropy loss function was used. The optimal values of the hyperparameters of the deep neural network are as follows: the length of the 12-lead ECG input is set to 5000, the batch size is 32, and the number of epochs is equal to 100.

To reduce the learning rate, we used the learning rate scheduler with the following schedule:

$$\mathbf{lr} = \begin{cases} \text{lr}, & \text{epoch} < 10, \\ \text{lr} \cdot e^{-0.1}, & \text{epoch} \geq 10. \end{cases} \quad (1)$$

4.4. Evaluation Metric. In multiclassification problems, precision and accuracy are commonly used to assess the model's performance. The performance of an algorithm is often measured in terms of four variables for each record. These two performance indicators (accuracy and precision) can be calculated in equations (2) and (3)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (2)$$

$$\text{Precision} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (3)$$

where TP denotes True Positive, FP denotes False Positive, TN denotes True Negative, and FN denotes False Negative.

5. Results and Discussion

This section presents visual and descriptive discussion based on the proposed model. Additionally, a comparative table has been introduced to compare the proposed work against other studies cited in related works as per Table 3. To note, OVH Cloud has been used with the following characteristic, to train the proposed model.

- (i) Memory: 45 Go
- (ii) vCore: 8
- (iii) GPU: NVIDIA Tesla V100 16 Go

Precision and accuracy are generally used as two performance indicators to evaluate model performance in multiclassification model. In our situation, precision represents the probability that the model makes the correct prediction, while accuracy is defined as the ratio between the proportion of correct predictions made by the model and the number of total predictions.

In the training and validation phase, the obtained accuracy is 97.63% and 97.58%, respectively. In terms of precision, we obtained 89.67% and 88.85%, respectively. The loss value indicates how well or poorly the proposed model performs after each iteration. For the loss, 3.10^{-3} and $1.27 \cdot 10^{-2}$ for each phase were reached as can be seen in Table 4.

Because of using stratified 10 folds in the data-splitting step, in the transition from fold to another, the model undergoes a disorder until the stabilization in the last fold. We can observe that, after the 60th iteration, the model progressively converges to reach a stable accuracy, precision, and loss at the 100th iteration. Figures 11–13 demonstrate the evolution of these performance metrics.

It is important for disease diagnosis to improve performance metrics for the correct classification of cardiovascular diseases. ResNet-50 shows better classification performance in comparison to the other studies cited in related works as can be seen in the comparative Table 3.

In the evaluation of the proposed model performance, a normalized confusion matrix was created as can be seen in Figure 14, where each row refers to an actual class, while each column represents a predicted class. The proposed model performs well for NSR, RBBB, STach, TInv, AF, IRBBB, and LBbB classes. In effect, their percentage of correct predictions is higher than 80%. It performs moderately for CRBB, Brady, SA, PAC, PVC, and SB classes. Next comes NSIVCB, IAVB, LanFB, AFL, and RAD where the percentage of correct predictions is higher than 60%. For the rest of the classes, like QAb, LAD, and LPR, PR, the model performs badly. This problem of lower predictions is due to the data imbalance even though an amplitude scaling was applied.

TABLE 3: Results obtained by different research in relation to the proposed work.

Author	Year	Number of records	Model	Preprocessing	Number of classes	Accuracy (%)	Precision (%)
Antonio et al. [34]	2020	2,322,513	DNN	No	6		92.36
Ahsanuzzman et al. [35]	2020	48	LSTM and RNN	Yes	1	97.57	
Obeidat et al [36]	2021	2000	CNN and LSTM	Yes	6	98.22	98.26
Adedinsewo et al. [38]	2020	6613	CNN	No	1	85.9	74
Xiong et al. [39]	2020	8528	ResNet-16	Yes	4	82	
Dongdong et al. [19]	2021	6877	ResNet-34	Yes	9	96.6	
Proposed work	2021	42,511	ResNet-50	Yes	27	97.63	89.67

TABLE 4: Results of the proposed method.

Performance	Results	
	Training phase	Validation phase
Accuracy	97.63%	97.58%
Precision	89.67%	88.85%
Loss	3.10^{-3}	$1.27.10^{-2}$

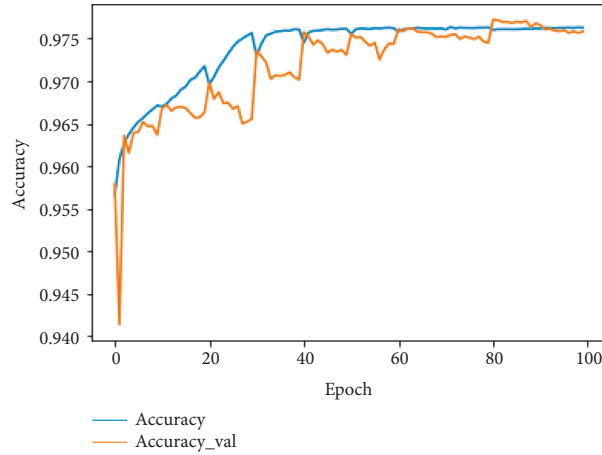


FIGURE 11: Evolution of training and validation accuracy.

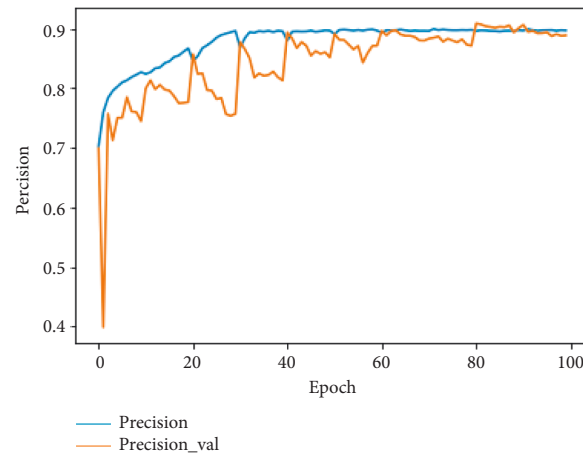


FIGURE 12: Evolution of training and validation precision.

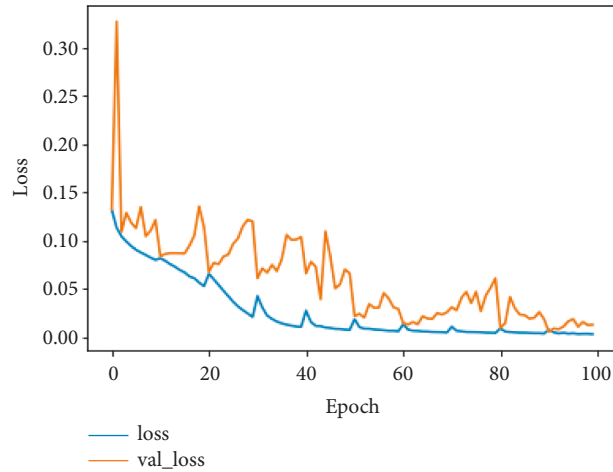


FIGURE 13: Evolution of the loss in the training and validation.

TABLE 5: Test results by the model proposed.

	Test 1	Test 2
Samples		
Incorrect	<div style="background-color: black; color: white; padding: 5px;"> Predicted: Sinus Rhythm Actual: sinus bradycardia </div>	<div style="background-color: black; color: white; padding: 5px;"> Predicted: 1st Degree AV Block Actual: left bundle branch block </div>
	Test 3	Test 4
Samples		
Correct	<div style="background-color: black; color: white; padding: 5px;"> Predicted: Sinus Rhythm Actual: sinus rhythm </div>	<div style="background-color: black; color: white; padding: 5px;"> Predicted: Atrial Fibrillation Actual: atrial fibrillation </div>

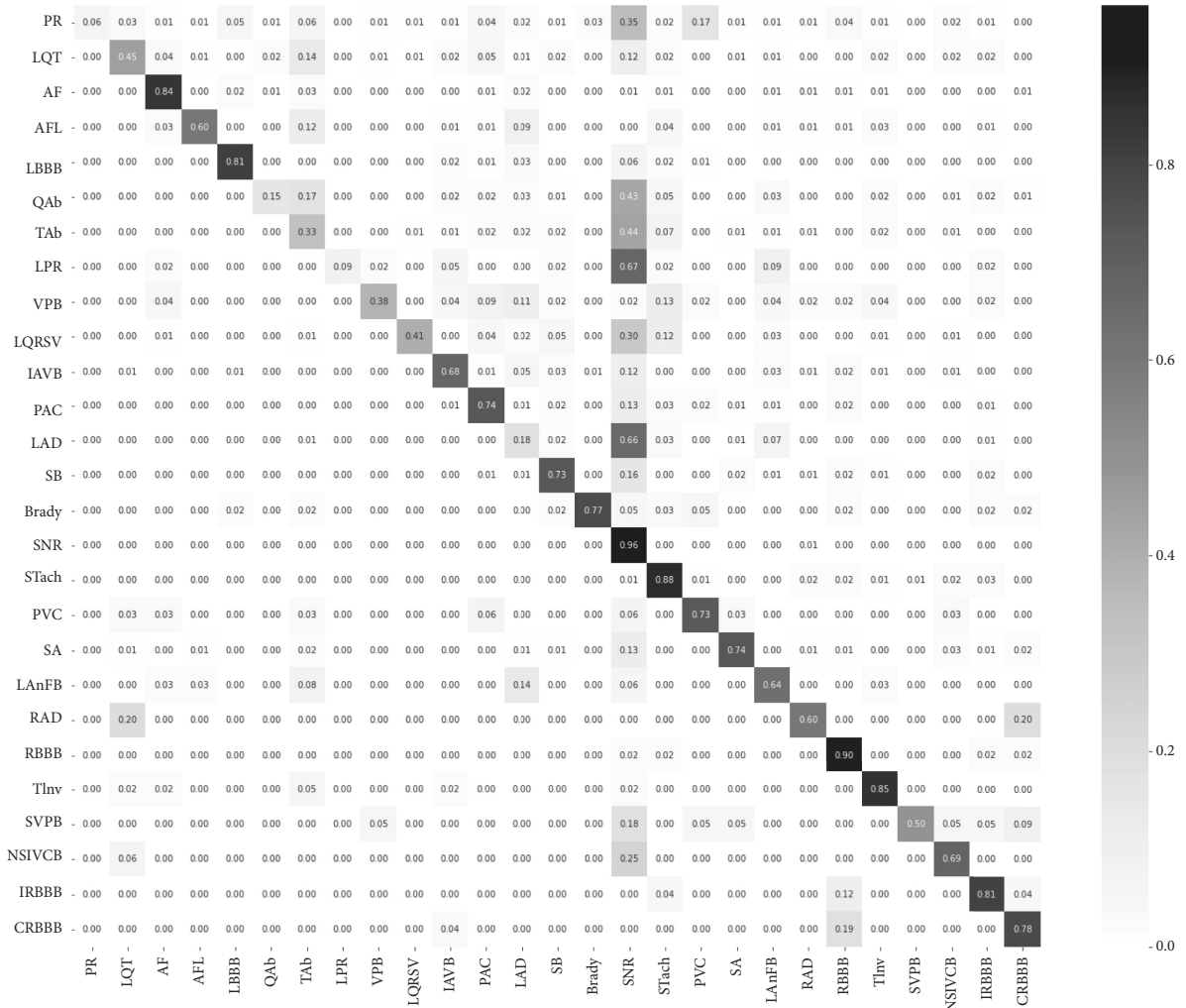


FIGURE 14: Confusion matrix.

Table 5 shows the test results of the proposed model including the incorrect samples (Tests 1 and 2) and correct (Tests 3 and 4) samples that were detected by our model, as well as its prediction and the current state of the ECG.

6. Conclusion and Future Work

An effective DL approach based on ResNet-50 has been presented in this paper to classify CVDs. The number of classes that have been considered were 27, where 26 belong to heart anomalies and 1 belongs to normal state. The dataset used in this study combines four datasets collected from three different countries. The achieved results proved the feasibility and the efficiency of the proposed model. The results, also, have been compared and validated against values in the recent published literature. However, the proposed model suffers from high computational complexity and low range of interpretability. Thus, as future research, the proposed approach will be improved to be ideally adapted for wider range of different healthcare applications.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the Deanship of Scientific Research at Taif University, Kingdom of Saudi Arabia, Taif, University Researchers Supporting Project no. TURSP-2020/265.

References

- [1] J. Emelia, S. S. V. Benjamin, W. Clifton et al., "Heart disease and stroke statistics—2018 update: a report from the American heart association," *Circulation*, vol. 137, pp. e67–e492, 2018.
- [2] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez et al., "An artificial intelligence-enabled ECG algorithm for the

- identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.
- [3] R. C. Schlant, R. J. Adolph, J. P. DiMarco et al., "Guidelines for electrocardiography: a report of the American college of Cardiology/American heart association task force on assessment of diagnostic and therapeutic cardiovascular procedures (committee on electrocardiography)," *Journal of the American College of Cardiology*, vol. 19, pp. 473–481, 1992.
 - [4] R. B. Schnabel, X. Yin, P. Gona et al., "50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study," *The Lancet*, vol. 386, no. 9989, pp. 154–162, 2015.
 - [5] W. G. Morrison and I. J. Swann, "Electrocardiograph interpretation by junior doctors," *Emergency Medicine Journal*, vol. 7, no. 2, pp. 108–110, 1990.
 - [6] F. A. Almalki, S. Ben Othman, F. A. Almalki, and H. Sakli, "EERP-DPM: energy efficient routing protocol using dual prediction model for healthcare using IoT," *Journal of Healthcare Engineering*, vol. 2021, Article ID 9988038, 15 pages, 2021.
 - [7] A. L. Goldberger, L. A. N. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet: PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, 2000.
 - [8] P. Wagner, N. Strodthoff, and R. D. Bousseljot, *PTB-XL, a Large Publicly Available Electrocardiography dataset*, *Scientific Data*, vol. 7, no. 154, 2020.
 - [9] Y. Wu, H. Guo, C. Chakraborty, M. Khosravi, S. Berretti, and S. Wan, "Edge computing driven low-light image dynamic enhancement for object detection," *IEEE Transactions on Network Science and Engineering*, p. 1, 2022.
 - [10] E. Jing, H. Zhang, Z. G. Li, Y. Liu, Z. Ji, and I. Ganchev, "ECG heartbeat classification based on an improved ResNet18 model," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6649970, 13 pages, 2021.
 - [11] M. Gowri Shankar and C. Ganesh Babu, "An exploration of ECG signal feature selection and classification using machine learning techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, 2020.
 - [12] A. Rajkumar, M. Ganesan, and R. Lavanya, "Arrhythmia Classification on ECG Using Deep Learning," in *Proceedings of the 2019 5th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 365–369, IEEE, Coimbatore, India, March 2019.
 - [13] V. Acharya, V. Ravi, T. D. Pham, and C. Chakraborty, "Peripheral blood smear analysis using automated computer-aided diagnosis system to identify acute myeloid leukemia," *IEEE Transactions on Engineering Management*, pp. 1–14, 2021.
 - [14] H. Alquran, A. M. Alqudah, I. Abu-Qasmieh, A. Al-Badarneh, and S. Almashaqbeh, "ECG classification using higher order spectral estimation and deep learning techniques," *Neural Network World*, vol. 29, no. 4, pp. 207–219, 2019.
 - [15] A. M. Alqudah, S. Qazan, L. Al-Ebbini, H. Alquran, and I. Abu Qasmieh, "ECG heartbeat arrhythmias classification: a comparison study between different types of spectrum representation and convolutional neural networks architectures," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–31, 2021.
 - [16] A. M. Alqudah, A. Albadarneh, I. Abu-Qasmieh, and H. Alquran, "Developing of robust and high accurate ECG beat classification by combining Gaussian mixtures and wavelets features," *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, no. 1, pp. 149–157, 2019.
 - [17] D. Zhang, X. Yuan, P. Zhang, and S. Yang, "Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram," *iScience*, vol. 24, 2021.
 - [18] D. Jia, W. Zhao, J. Hu, H. Wang, C. Yan, and Z. Li, "Detection of first-degree atrioventricular block on variable-length electrocardiogram via a multimodal deep learning method," in *Proceedings of the 2019 Computing in Cardiology (CinC)*, September 2019.
 - [19] R. He, Y. Liu, K. Wang et al., "Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM," *IEEE Access*, vol. 7, 2019.
 - [20] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung "der EKG-Signaldatenbank CARDIODAT der PTB über das Internet," *Biomedizinische Technik*, p. 1, 1995.
 - [21] W. Liu, M. Zhang, Y. Zhang et al., "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, 2017.
 - [22] H. Bhuyan, D. C. Chakraborty, S. Pani, and V. Ravi, "Feature and subfeature selection for classification using correlation coefficient and fuzzy model," *IEEE Transactions on Engineering Management*, pp. 1–15, 2021.
 - [23] J. Zhang, F. Lin, P. Xiong et al., "Automated detection and localization of myocardial infarction with stacked sparse autoencoder and TreeBagger," *IEEE Access*, vol. 7, 2019.
 - [24] A. Selvalingam, M. Alhuseini, A. J. Rogers et al., "Developing convolutional neural networks for deep learning of ventricular action potentials to predict risk for ventricular arrhythmias," *Circulation*, vol. 140, 2019.
 - [25] S. W. Smith, B. Walsh, K. Grauer et al., "A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation," *Journal of Electrocardiology*, vol. 52, pp. 88–95, 2019.
 - [26] Y. Li, Y. Pang, K. Wang, and X. Li, "Toward improving ECG biometric identification using cascaded convolutional neural networks," *Neurocomputing*, vol. 391, pp. 83–95, 2020.
 - [27] Y. Zhang, Z. Xiao, Z. Guo, and Z. Wang, "ECG-based personal recognition using a convolutional neural network," *Pattern Recognition Letters*, vol. 125, pp. 668–676, 2019.
 - [28] U. R. Acharya, H. Fujita, S. L. Oh et al., "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Applied Intelligence*, vol. 49, no. 1, pp. 16–27, 2019.
 - [29] N. Iyengar, C.-K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, pp. 1078–1084, 1996.
 - [30] D. S. Baim, W. S. Colucci, E. S. Monrad et al., "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, 1986.
 - [31] P. Wang, B. Hou, S. Shao, and R. Yan, "ECG arrhythmias detection using auxiliary classifier generative adversarial network and residual network," *IEEE Access*, vol. 7, 2019.
 - [32] S. Hadiyoso, K. Usman, and A. Rizal, "Arrhythmia detection based on ECG signal using android mobile for athlete and patient," in *Proceedings of the 2015 Third International Conference on Information and Communication Technology (ICoICT)*, pp. 166–171, Nusa Dua, Bali, Indonesia, May 2015.
 - [33] F. Murat, O. Yildirim, M. Talo, U. B. Baloglu, Y. Demir, and U. R. Acharya, "Application of deep learning techniques for

- heartbeats detection using ECG signals-analysis and review,” *Computers in Biology and Medicine*, vol. 120, 2020.
- [34] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixoa et al., “Automatic diagnosis of the 12-lead ECG using a deep neural network,” *Nature*, vol. 11, 2020.
- [35] S. M. Ahsanuzzaman, T. Ahmed, and M. Atiqur Rahman, “Low Cost, Portable ECG Monitoring and Alarming System Based on Deep Learning,” in *Proceedings of the 2020 IEEE Region 10 Symposium (TENSymp)*, IEEE, Dhaka, Bangladesh, June 2020.
- [36] Y. Obeidat and A. M. Alqudah, “A hybrid lightweight 1D CNN-LSTM architecture for automated ECG beat-wise classification,” *Traitement du Signal*, vol. 38, no. 5, pp. 1281–1291, 2021.
- [37] 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,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5594159, 18 pages, 2021.
- [38] D. Adedinsewo, R. E. Carter, Z. Attia et al., “An artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea,” *Nature Medicine*, vol. 9, pp. 707–715, 2020.
- [39] Z. Xiong, M. K. Stiles, and J. Zhao, “Robust ECG signal classification for detection of atrial fibrillation using a novel neural network,” in *Proceedings of the 2017 Computing in Cardiology (CinC)*, Rennes, France, September 2017.

Research Article

Internet of Medical Things (IoMT) and Reflective Belief Design-Based Big Data Analytics with Convolution Neural Network-Metaheuristic Optimization Procedure (CNN-MOP)

A. Sampathkumar ¹, **Miretab Tesfayohani** ², **Shishir Kumar Shandilya** ³, **S. B. Goyal**,⁴
Sajjad Shaukat Jamal,⁵ **Piyush Kumar Shukla**,⁶ **Pradeep Bedi** ⁷, and **Meshal Albeedan**⁸

¹Department of Applied Cybernetics, Faculty of Science, University of Hradec Kralove, Hradec Kralove, Czech Republic

²Department of Information Technology, Dambi Dollo University, Dambi Dollo, Ethiopia

³School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India

⁴City University, Petaling Jaya 46100, Malaysia

⁵Department of Mathematics, College of Science, King Khalid University, Abha, Saudi Arabia

⁶Computer Science and Engineering Department, University Institute of Technology, Rajiv Gandhi Proudhyogiki Vishwavidyalaya, (Technological University of Madhya Pradesh), Bhopal-462033, Madhya Pradesh, India

⁷Galgotias University, Greater Noida, Uttar Pradesh, India

⁸Department of Computer Science, Faculty of Engineering and Technology, Liverpool John Moores University (LJMU), Liverpool L3 3AF, UK

Correspondence should be addressed to Miretab Tesfayohani; miremsc2011@gmail.com

Received 10 November 2021; Accepted 29 January 2022; Published 18 March 2022

Academic Editor: Anastasios D. Doulamis

Copyright © 2022 A. Sampathkumar 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 times, the Internet of Medical Things (IoMT) is a new loomed technology, which has been deliberated as a promising technology designed for various and broadly connected networks. In an intelligent healthcare system, the framework of IoMT observes the health circumstances of the patients dynamically and responds to backings their needs, which helps detect the symptoms of critical rare body conditions based on the data collected. Metaheuristic algorithms have proven effective, robust, and efficient in deciphering real-world optimization, clustering, forecasting, classification, and other engineering problems. The emergence of extraordinary, very large-scale data being generated from various sources such as the web, sensors, and social media has led the world to the era of big data. Big data poses a new contest to metaheuristic algorithms. So, this research work presents the metaheuristic optimization algorithm for big data analysis in the IoMT using gravitational search optimization algorithm (GSOA) and reflective belief network with convolutional neural networks (DBN-CNNs). Here the data optimization has been carried out using GSOA for the collected input data. The input data were collected for the diabetes prediction with cardiac risk prediction based on the damage in blood vessels and cardiac nerves. Collected data have been classified to predict abnormal and normal diabetes range, and based on this range, the risk for a cardiac attack has been predicted using SVM. The performance analysis is made to reveal that GSOA-DBN_CNN performs well in predicting diseases. The simulation results illustrate that the GSOA-DBN_CNN model used for prediction improves accuracy, precision, recall, F1-score, and PSNR.

1. Introduction

In the modern world, a better healthcare system is the main challenge for the growing population of the world. The vision of the Internet of Medical Things (IoMT) is to provide a better and more pervasive health monitoring system. The IoMT is the integration of medical devices through Wi-Fi and permits device-to-device (D2D) communication. In recent days, the most challenging issue is the time needed for web services. Three-dimensional (3D) video can be downloaded at sporadic intervals by keeping in mind the latest technological trends. The collected voluminous data with less delay are obtained for accurate data measurement. It will increase the device resource allocation ability and offers quicker speed for the heterogeneous networks. The IoMT comprises various heterogeneous networks, for instance, Wi-Fi, Bluetooth, ZigBee, and other cellular platforms. The D2D communication is the central part of the IoMT platform with high efficiency and reliability. The main traits of an intelligent healthcare system are to offer less delay and high throughput and reliability, which are very important for an effective and accurate diagnosis and consultation. The critical time analysis is the key parameter to be considered for emergency healthcare applications. The highly reliable and delay-tolerant communication and transmission of data was achieved through IoT-driven wearable devices [1].

Metaheuristic optimization approaches are employed in the partitioning clustering methods with which a dataset is partitioned as groups according to the specific measure considered as a fitness function [2]. This function has a greater impact on the nature of forming these groups. When an appropriate fitness function is selected, the process of partitioning is converted as an optimization problem. Here, partitioning is performed by minimizing the distance or maximizing the similarity between the patterns, or the frequency is optimized in N-dimensional space. These techniques are generally employed in various research works as they are able to clustering large-scale datasets, like signal/image processing for segmenting images, analyzing homogeneous users to classify groups, generating precise hidden equalizers, organizing humans effectively in robotics based on their actions, matching aftershocks in seismology from background conditions, obtaining high-dimensional data report, mining web text and recognizing image patterns in computer science domain, managing portfolio in control studies, classifying diseases in medical anthropology using the medical records of patients and investigations, distributing sensor nodes in a wireless sensor network for improving lifetime, and grouping publications in library based on the contents [3].

Nowadays, big data is extensively used for analyzing the huge data for the predictors, business people, and researchers to estimate the predictions with much accuracy than the traditional analysis. Big data is structured with the five dimension maps such as velocity, value, volume, veracity, and variety. Now the researchers are working to handle scalability and the high dimension of the databases with high processing needs. Volume defines the size of the data, and velocity defines the arrival of data streams

continuously from which most valuable information is gathered [4].

Moreover, in big data, the throughput, connectivity, and speed of computing have been enhanced for digital devices that improve the retrieval, progression, and fabrication of the data. Authenticity controls the information standard from different places. Variety designates the communicative path of information between different places, for example, the primary data incorporating the conventional structure. The example of data source consists of both structured traditional and relational data, along with that it also comprises of semi-structured, quasi-structured, and unstructured data such as sensor data, audio, video, text, and graph. In healthcare, big data helps in predicting epidemics, curing diseases, improving the survival rate, and avoiding unnecessary deaths. As the population in the world is increasing and the life span of individuals is long, there occur rapid changes in providing treatment, and many decisions are made based on these changes. With big data, patients take the right decision at the right time. With the data report of the patient, "proactive care" required for individuals is identified or the change required for them is analyzed so that degradation in their health can be avoided [5].

However, machine learning, as well as statistical approaches, undergoes a few necessary changes to maintain and follow some specific constraints while estimating high-dimensional data; moreover, existing issues are solved when input variables are reduced before data mining approaches are applied. Thus, there are two ways techniques are employed for dimensionality reduction in machine learning. While exploring the redundant input data or in feature selection, the essential variables are taken from the primary dataset. In the dimensionality reduction process, data are eliminated forming a new dataset with few input variables, where each column holds the combined input variables providing the same information as that of the input variables. In statistical modeling, this process is termed exploration [6]. The existing works are related to the data optimization with machine learning techniques, which does not improve the predictive accuracy with their optimization level.

The contribution of the work is as follows:

- (i) To present the metaheuristic optimization algorithm for big data analysis in the IoMT using gravitational search optimization algorithm (GSOA) and reflective belief network with convolutional neural networks (DBN-CNNs).
- (ii) To perform data optimization using GSOA for the collected input data. The input data were collected for the diabetes prediction with cardiac risk prediction based on the damage in blood vessels and cardiac nerves.

Big data in healthcare is being used to predict epidemics, cure disease, improve quality of life, and avoid preventable deaths. With the world's population increasing and everyone living longer, the models of treatment delivery are rapidly changing, and data are driving many of the decisions behind those changes. Big data can help patients make the right

decision in a timely manner. From patient data, analytics can be applied to identify individuals who need “proactive care” or need a change in their lifestyle to avoid health condition degradation. For example, patients in early stages of some diseases (e.g., heart failure often caused by some risk factors such as hypertension or diabetes) should be able to benefit from preventive care, thanks to big data.

In Section 2, the related research works are presented. Section 3 shows the proposed model for data optimization and diabetes data classification. Evaluation criteria are discussed in Section 4. The conclusion is finally presented in Section 5.

2. Related Works

The healthcare system involves machine learning (ML) approaches in the fields such as diagnosing, predicting, and surveillance. It is believed by the health monitoring agents that by using ML techniques, life can be saved [7]. Using SVM, diabetic nephropathy type 2 patients can be detected [8]. Moreover, earlier detection of abnormalities can be achieved using a decision tree-based approach when integrated with genotype and clinical data of several type 2 diabetic patients. Data can be classified gender-wise using support vector machine (SVM), Naïve Bayes (NB), decision tree (DT), and random forest (RF) approaches to produce satisfactory results. In [9], ML approaches are embedded with data mining profiling for gaining knowledge from the larger collection of information. Classification approaches like SVM, NB, RF, and logistic regression (LR) are used. In [10], for cardiovascular and cerebrovascular events, an automated prediction model was developed using heart rate variability (HRV) analysis. This was suitable for patients with 55 years and more who have higher risks. 10-fold cross-validation based on HRV features with data mining approaches such as NB, DT, RF, SVM, AdaboostM1 (AB), and multilayer perceptron neural network (MLPNN) were used for prediction. Reference [11] focused on antidiabetic drug failure and developed a prediction model to maintain an exponential increase in diabetic type 2 patients. SVM was used for training large-scale medical datasets. In [12], the risk of diabetic neuropathy was examined. When diabetes has been spread in the entire body, the nervous system is affected leading to cardiac arrest. Heart rate variability (HRV) was estimated using a multiscale Allan vector and the features of ECG helped in automatic detection using ML techniques. A graph-based machine learning system (GBMLS) was introduced to diagnose diabetic neuropathy effectively. In [13], a healthy and asthma patient was comparatively examined with alternative devices to obtain the feature vector of asthma patient with the guidelines of GINA. RF, AB-RF, and MLPNN were used to develop an ML prediction model. Specificity, sensitivity, and accuracy were the parameters considered. Leave-one-out (LOO) validation methods were used to train and test the dataset to eradicate the overfitting problem. In [14], a surveillance system was introduced to effectively monitor the impacts of dengue hemorrhagic fever (DHF) and the rate of *Aedes aegypti* mosquito infection using SVM, which depends on the

climate conditions and geographical area. In [15], a prediction system was designed for detecting influenza-like illness (ILI) at the earlier stages. Using NB and SVM classifiers not only produced better results but logistics regressions (LR) and sequential minimal optimization (SMO) also are suitable. In [16], an automatic coronary artery disease (CAD) diagnosis was introduced where this disease caused cardiac arrests. The diagnosis was made using tunable Q-wavelet transform (TQWT), and the heart rate signals were observed and monitored using raw ECG (electrocardiogram). With threefold cross-validation, least square support vector machine (LS-SVM) approach was used for classification. In [17], heart disease was predicted and analyzed using BagMOOV ensemble model with a multi-purpose-weighted voting approach. Reference [18] studied the mood disorder which is a psychological behavior of humans and investigated a psychiatry solution using ML techniques. Three ML approaches, namely, SVM, least average shrinkage and selection operator (LASSO), and relevance vector machine (RVM) were developed using MATLAB to predict the possibilities of a suicide attempt. In [19], risks, diagnosis, and prediction of breast cancer were examined with four ML approaches, namely, SVM, DT, NB, and k-nearest neighbor (KNN). In [20], the classification of UCI’s disease datasets was studied in which SVM integrated with endocrine-based particle swarm optimization (EPSO) and artificial bee colony (ABC) was employed. In [21], it was revealed that SVM with fruit fly optimization algorithm (FOA) was used in several medical datasets, namely, Wisconsin Parkinson’s dataset, breast cancer dataset, thyroid disease diagnosis, and Pima Indian diabetes dataset. In [22], a model for diabetes prediction utilizing machine learning approaches was predicted. The supervised machine learning algorithms used for the prediction model are decision tree, Naïve Bayes, artificial neural network, and logistic regression. Performance evaluation is performed in this method by employing parameters such as accuracy, recall, precision, and F1-score. In [23], the data-driven SmartWork system’s AI component was described, comprising the predictive models which are personalized and decision support tools. In these subsystems, long-term predictive models and data mining techniques are implemented for providing probabilistic prediction of particular risk indicators motivated towards decision-making and T2DM intervention, among other chronic conditions.

Based on the comparison discussed, it is not enough for both optimizations to improving the accuracy. So, this research aims to propose metaheuristic algorithm for data optimization of diabetic data and the predictive analysis for cardiac attack risk prediction. Here, the IoMT module is used for data collection and uses a gravitational algorithm for data optimization, and then the data classification is done using DBN-based CNN. Finally, the predictive analysis will be done using SVM on the basis of image predictive analysis.

3. System Model

This section discusses the proposed metaheuristic algorithms based on data optimization for diabetes data, which

leads to the predictive analysis of cardiac attack risk based on blood vessels and cardiac nerves and cardiac nerve damages. The dataset, obtained from public healthcare, contains more than 100,000 records comprising 55 attributes. Few among them are age, gender, race, number of procedures, number of medications, number of diagnoses, and readmissions.

The data have been initially collected using the IoMT module, and these data have been clustered for improving the optimization of data to be processed. Here we use metaheuristic algorithms for data optimization in which gravitational search optimization algorithm has been used. Then using these optimized data, the diabetic data have been classified for identifying the abnormal range. Here we establish a deep belief network (DBN), where the classification is carried out using CNN. By this classification, the normal range and abnormal range of diabetes have been classified. The normal diabetes range has been updated to the hospital database, and for the abnormal diabetes range, the cardiac nerve and blood vessel damages have been analyzed using SVM-based image predictive analysis. The proposed diabetic data analysis has been given in Figure 1. In the IOMT module, a dataset for diabetes is collected and the data from the dataset are clustered in the data clustering phase, and the optimization of data is done by metaheuristic data optimization by using GSOA. The optimized data are classified by the classification algorithm DBN-based CNN. From the classification results, the normal range of diabetes is sent to the hospital database, whereas an abnormal range of diabetes is sent to the analysis of predicting cardiac attack risk, which will be done by various parameters.

3.1. Metaheuristic-Based Gravitational Search Optimization Algorithm (GSOA). Gravitational search optimization algorithm (GSOA) is a stochastic population-based metaheuristic approach that was developed based on Newton's laws of gravity and motion [24]. Originally, the basic GSOA model was developed to find a solution for continuous optimization problems. A set of agents/objects were introduced in the search space within dimension to determine an optimal solution where the principle of Newton's laws was followed. Here, the position of every agent describes a candidate solution X_i , which is a vector in the search space. An agent whose performance is higher obtains more gravitational mass as heavier objects gain more attraction radius. In the life span of GSOA, X_i is adjusted successively by an agent with the positions of best agents in KGSOA adapting Newton's laws. To explain in detail, a system with s agents is assumed, where the position of the agent is given by

$$X_i = (x_i^1, \dots, x_i^d, \dots, x_i^n); \quad i = 1, 2, \dots, s, \quad (1)$$

where x_i^d represents the position of agent i in dimension d with search space dimension n . For every agent, the gravitational mass after estimating the current data fitness is computed as given in the following equations:

$$q_i(t) = \frac{\text{fit}_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)}, \quad (2)$$

$$M_i(t) = \frac{q_i(t)}{\sum_{j=1}^s a_j(t)}, \quad (3)$$

where $M_i(t)$ and $\text{fit}_i(t)$ are the gravitational mass and fitness value of i^{th} agent, respectively, at time t . The $\text{best}(t)$ and $\text{worst}(t)$ are given in the following equations:

$$\text{best}(t) = \min_{j \in (1, \dots, s)} \text{fit}_j(t), \quad (4)$$

$$\text{worst}(t) = \max_{j \in (1, \dots, s)} \text{fit}_j(t), \quad (5)$$

Agent acceleration is estimated by adding the forces of every agent present in the set of KGSA based on the gravitational law, which is determined by using equation (6) and agent acceleration which is estimated by the motional law is expressed in equation (7):

$$F_i^d(t) = \sum_{j \in K_{\text{best}}, j \neq i} \text{rand}_j G(t) \frac{M(t)M_i(t)}{R_{ij}(t) + z} \quad (6)$$

$$\cdot (x_j^d(t) - x_i^d(t)),$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)}$$

$$= \sum_{j \in K_{\text{best}}, j \neq i} \text{rand}_j G(t) \frac{M(t)M_i(t)}{R_{ij}(t) + z} (x_j^d(t) - x_i^d(t)), \quad (7)$$

where r and j are random numbers distributed evenly ranging between $[0, 1]$,

- (i) ε , a small value, helps to get rid of division by zero error when $R_{ij}(t)$ is zero.
- (ii) $R_{ij}(t)$ represents the Euclidean distance of agents i and j , defined as $\|X_i(t) - X_j(t)\|_2$.
- (iii) K_{best} indicates the set of first agents in KGSOA having best fitness value and higher gravitational mass, where KGSOA is the time function which is initially assigned a K_{initial} value and gets decreased with time.
- (iv) $G(t)$ is the gravitational constant which initially takes G_{initial} value and decreases with time till G_{end} is reached as given in the following equation:

$$G(t) = G(G_{\text{initial}}, G_{\text{end}}, t). \quad (8)$$

Then, agent velocity and next position are estimated using the following equations:

$$v_i^d(t+1) = \text{rand}_i * v_i^d(t) + a_i^d(t), \quad (9)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1). \quad (10)$$

3.2. Establishing DBN Architecture. Generally, when compared with a general neural network, deep neural network (DNN) is superior. In DNN, the input layer is separated

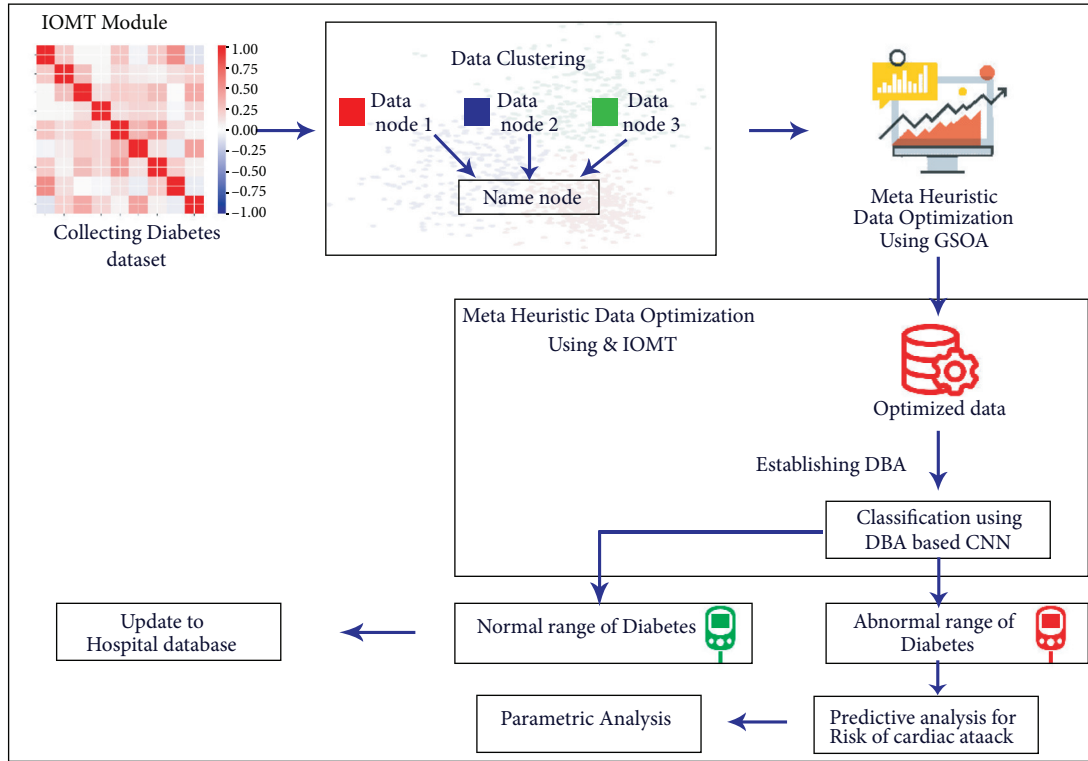


FIGURE 1: Proposed design for metaheuristic algorithms for data optimization with the IoMT.

from the output layer by several hidden layers. The method of training the networks differs. Particularly, in DNN, the unsupervised learning method is used for adjusting the weights in hidden layers, and the network is capable of identifying the optimal features from the input provided. Hence, DNN is flexible and enables high-order modeling of the nonlinear complex relationship between input provided to the network and output produced. The advantage of using DNN for learning the features and classifying data is proved with various pattern recognition applications like vision, speech, and natural language processing [24–26]. The outcomes observed have motivated the researchers to develop automated pattern recognition systems using DL methods. In spite of some challenges, developing a suitable training model for DNN is still a challenging one.

With the presence of several hidden layers in DNNs and numerous parameters, training has to be done with utmost care. In the process of training, the feature detection layers are randomly initialized. In DNN, a series of generative models, namely, a single visible input and a hidden layer are considered for initializing weights and are trained without taking discriminative information into account. At last, the standard back propagation approach is used to train DNN discriminatively. From the investigation, it is observed that the standard gradient-based random initialization approach used for initializing weights of the network produces very little performance in DNN containing more than two layers. As DNNs with several parameters and numerous hidden layers increase the computational complexity, training is much slower and even gets stuck with local minima providing unexpected results. Here, the parameters are initialized during

unsupervised pretraining, so that the process of optimization ends with local minima of the cost function. The architecture of DBN with RBM is given Figure 2 [24].

In RBM, the energy function E taking parameters v and h representing a pair of visible and hidden vectors, respectively, has the general form with weight matrix W as in

$$E(v, h) = -a^T v - b^T h - v^T W h, \quad (11)$$

where a and b indicate the bias weights for visible and hidden units, respectively. With v and h in terms of E , the probability distribution P is given by

$$P(v, h) = \frac{1}{Z} e^{-E(v, h)}. \quad (12)$$

Here the normalizing constant Z is given by

$$Z = \sum_{v, h} e^{-E(v, h)}. \quad (13)$$

Moreover, the probability of v over hidden units is the sum of above given equations and is given by

$$P(v) = \frac{1}{Z} \sum_h e^{-E(v, h)}. \quad (14)$$

Log-likelihood difference of training data in terms of W is estimated as

$$\begin{aligned} \sum_{n=1}^{n=N} &= \frac{\partial \log P(v^n)}{\partial w_{ij}} \\ &= v_i h_{j\text{data}} - v_i h_{j\text{model}}. \end{aligned} \quad (15)$$

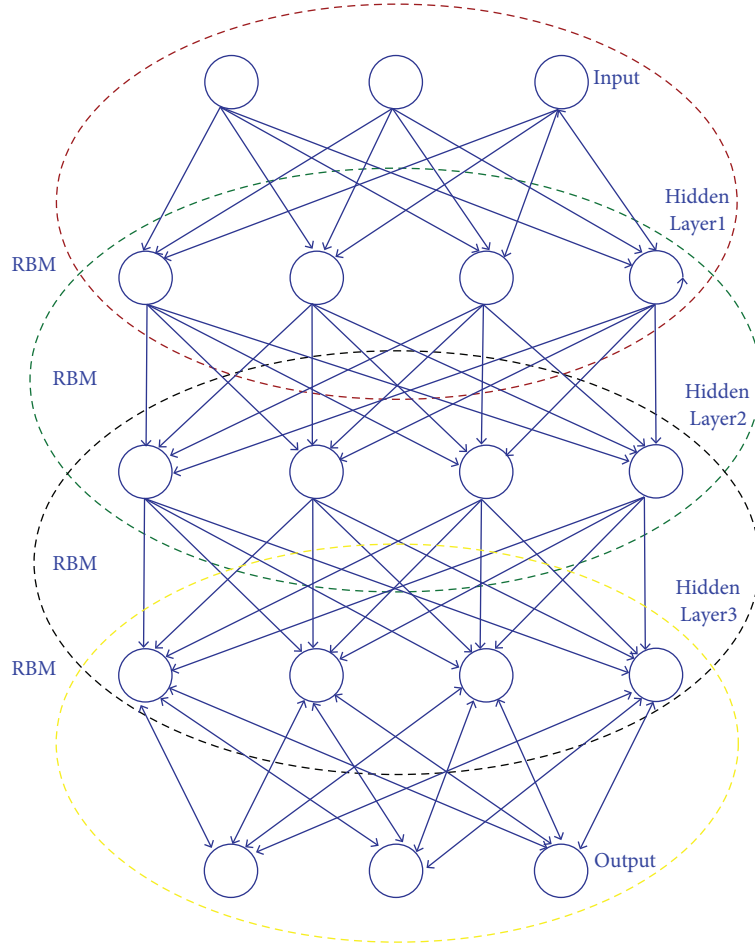


FIGURE 2: Architecture of the DBN with RBM.

where $v_i h_{j\text{data}}$ and $v_i h_{j\text{model}}$ represent the values expected for data and distribution model, respectively. For log-likelihood-based training data, network weights are computed using the learning rate ε as in

$$W_{ij} = \varepsilon (\langle v_i h_{j\text{data}} \rangle - \langle v_i h_{j\text{model}} \rangle) (v_i h_{j\text{data}}). \quad (16)$$

As neurons are not connected at either the hidden or the visible layers, it is possible to obtain unbiased samples from $v_i h_{j\text{data}}$. Further, the activation of hidden or visible units is conditionally independent for given h and c , respectively. For given v , conditional property is described in:

$$P(h|v) = \prod_j P(h_j|v), \quad (17)$$

where $h_j \in \{0, 1\}$ and the probability of $h_j = 1$ is given in

$$P(h_j = 1|v) = \sigma \left(b_j + \sum_i v_i W_{ij} \right). \quad (18)$$

Here the logistic function σ is specified as in

$$\sigma(x) = (1 + e^{-x})^{-1}. \quad (19)$$

Likewise, when $v_i = 1$, the conditional property is estimated by

$$P(v_i = 1|v) = \sigma \left(a_i + \sum_j W_{ij} h_j \right). \quad (20)$$

Generally, with $\langle v_i, h_j \rangle$, the unbiased sampling is not straightforward; however, it is applicable for reconstructing the first sampling of v from h and then Gibbs sampling is used for multiple iterations. With this Gibbs sampling, every unit of the hidden and visible layers is updated in parallel. At last, with $\langle v_i, h_j \rangle$, the proper sampling is computed by multiplying the expected and updated values of h and v . RBM weights can be used for initializing feedforward neural networks. Among several supervised and unsupervised approaches, this work focuses on DBN as it is commonly used for classifying diabetic data.

Convolutional neural network (CNN), which is deep learning knowledge-based neural network model, is a well-described and widely used technique for classifying images. It incorporates linear convolutional layer (conv), a fully connected layer (FC), this model consists of nonlinear function with an activation function above the linear function. This nonlinear function impinges on every component of input and pooling layer and minimizes the size of the final outcomes. Multiple perceptrons are used to analyze the image inputs, and it is trained with learnable weights and

bias value, which is used in several parts of images to segregate pixel values. One main advantage of using CNN is it uses a local spatial domain for the input images; in addition, it also shares a few sharable parameters and a lesser number of weights. This technique is predominately more efficient than other models due to less computation complexity and less usage of memory. The CNN architecture is shown in Figure 3 [25].

Convolutional layer: basically, the image taken as input is resized to $3 \times 224 \times 224$, which is the standard size for the CNN model. The resized image undergoes a stack of multiple layers as convolutional layers of various receptive fields. In the convolution layer, the basic operation is a convolution where several sequential mathematical operations are performed on convoluting sliding kernel matrix over the input matrix by which feature data are extracted and then those data are mapped to the successive layers. At each co-ordinate, an element-wise matrix multiplication carries out, and the outcomes are combined to obtain a feature matrix. Convolution recurrent is a distinguishable kind of linear model that assists in various platforms such as image processing, statistics, and physics. Convolution is estimated by more than one axis. Where, two-dimensional I is the input image and K kernel filter, respectively, then the convoluted image is calculated by

$$S(i, j) = \sum_m \sum_n I(m, n)k(i - m, j - n). \quad (21)$$

Pooling layer: the pooling layer, termed also as pool, is the successive layer of the convolutional layer whose main function is to minimize the spatial domain representation thereby reducing the network computations performed. Generally, in CNN, the pooling kernel is 2×2 in size with a stride of 2.

Fully connected layer: this FC layer is replicated in CNN due to convolution. $n1 \times n2$ is its usual size, where $n1$ and $n2$ are the size of the input tensor ($7 \times 7 \times 512$) and output tensor, respectively, which is normally an integer.

Dropout: this layer is abbreviated as ‘‘Drop.’’ Usually, it is employed for eliminating the input overfits; the basic role is to improve the conjecture of deep learning methods. Normally, it assigns the weights to the connected network nodes.

Softmax: normally represented as ‘‘ σ ,’’ is a deep learning model followed by many layers and the convolutional layer is followed by a ReLU layer which determines the nonlinearity of CNN and improves it.

The convolution layers present within both the pooling layers possess the equal channel number, kernel size, and stride. Actually, collecting two 3×3 convolution layers and three 3×3 convolution kernels is equal to a single 5×5 and 7×7 convolution layers, respectively. Stacking 2 or 3 small convolution kernels works much quicker than a single huge convolution kernel. Moreover, parameter numbers have

been minimized. ReLU layers, which are inserted among undersized convolution layers, are really useful.

The input data and their corresponding map images are $S = (S^{(1)} \dots S^{(N)})$ and $M = (M^{(1)} \dots M^{(N)})$, respectively. The major objective is to design a model which maps S to M with the help of some training data. This is modeled as a probabilistic approach by learning the model of distribution over labels which is represented by

$$P(n(M, i, w_m) | n(S, i, w_s)), \quad (22)$$

where $n(I, i, w)$ is a patch with $w * w$ size for image I , focused on pixel i . Here, $W_S W_S$ is preferred to be higher so that more contextual information can be extracted. Its functional form f is given by

$$\begin{aligned} f_t(s) &= \sigma(a_1(s)) \\ &= P\left(m_1 = \frac{1}{s_1}\right), \end{aligned} \quad (23)$$

where a_i and f_i represent the sum of input for the i^{th} output and significance of i^{th} output component, respectively. (x) , a logistic utility, is expressed by

$$(x) = \frac{1}{1 + \exp(-x)}. \quad (24)$$

CNN alongside softmax output unit is used for multi-class marking. The softmax output is a vector of size L , which demonstrates the conveyance more than potential marks for pixel i . Along these lines for multiclass marking, if the path from pixel i to output unit l is thought of, the recomposed condition is given in

$$\begin{aligned} f_{il}(s) &= \frac{\exp(a_{il}(S))}{z} \\ &= P(m_i = l | s), \end{aligned} \quad (25)$$

where $f_{il}(s)$ is the prediction probability, where pixel I is mapped to label j . The advantages of the proposed method are summarized as follows:

- (i) First, CNN possibly handles a huge amount of labelled data from various domains.
- (ii) Second, it is faster when paralleled with graphics processing unit (GPU). Hence this is also extended for a greater number of pixels. The kernel size has been minimized to simulate the training data, and the kernel size minimization is performed by the computational learning process of the proposed method.
- (iii) Every patch in training data has been given by initiative sigma. Due to a large number of training patches, optimization becomes complicated. This can be done using a binary classifier that uses minimum patches. Few of the hyper parameters have been altered to some extent. The analysis over sensitivity has been defined by hyper parameters for them to be tuned with higher accuracy.

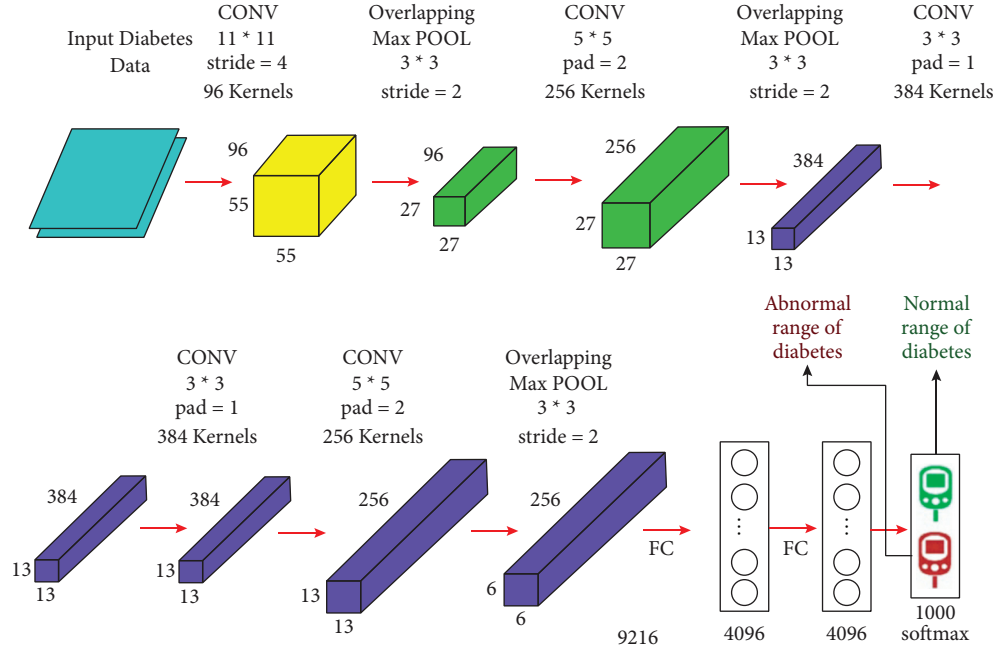


FIGURE 3: Architecture of the convolutional neural network.

The output from the softmax layer will be classified to be normal and abnormal ranges of diabetes, and the abnormal range of diabetes will be carried out for the predictive analysis of cardiac attack based on the cardiac nerve and blood vessel damage. This has been analyzed by using SVM-based image predictive analysis.

3.3. SVM-Based Image Predictive Analysis. SVM method is employed in classification and regression. With methods based on SVM, data points with identical properties are grouped together depicting the space. With linear SVM, for the data provided, p-dimensional vector is contemplated and divided by utmost p-1 planes called hyperplanes, which are employed to divide the space for classification and regression. The mathematical form of SVM is given. The line equation is outlined by

$$a_1 = a_2x + b. \quad (26)$$

In (26), “x” stands for the line’s slope and “b” stands for intersect as given in

$$a_1 - a_2x + b = 0. \quad (27)$$

Let $a' = (a_1, a_2)^T$ and $z' = (x, -1)$. Thus,

$$z' \cdot a' = 0. \quad (28)$$

The above equation obtained from two-dimensional vectors is known as the hyper lane equation. $a' = (a_1, a_2)^T$ is mentioned as z' as given in

$$z' = \frac{a_1}{\|a\|} + \frac{a_2}{\|a\|}. \quad (29)$$

Here,

$$\|a\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \dots + a_n^2}. \quad (30)$$

It is known that

$$\cos(g_1) = \frac{a_1}{\|a\|}, \quad (31)$$

$$\cos(g_2) = \frac{a_2}{\|a\|}.$$

Moreover, this can be expressed as equations (32)–(35):

$$z' = (\cos \cos(\|_1), \cos \cos(\|_2)), \quad (32)$$

$$z' \cdot a = \|z\| \|a\| \cos(\theta), \quad (33)$$

$$\theta = \theta_1 - \theta_2,$$

$$\begin{aligned} \cos(\theta) &= \cos(\theta_1 - \theta_2) \\ &= \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2), \end{aligned} \quad (34)$$

$$= \frac{z'_1 a_1}{z' \|a\|} + \frac{z'_2 a_2}{z' \|a\|} \frac{z'_1 a_1 + z'_2 a_2}{z'},$$

$$z' \cdot a' = \sum_{i=1}^n z'_i a_i. \quad (35)$$

For n-dimensional vectors, dot product is computed as follows: consider $f = y(z \cdot a + b)$ when $\text{sign}(f) > 0$, then it is a correct categorization, and for $\text{sign}(f) < 0$, it is incorrect. If dataset D is provided, then on a training dataset, f is calculated by

$$f_i = y_i(z' \cdot a + b). \quad (36)$$

Functional margin (F) on the dataset is estimated as given in

$$F = \min_{i=1\dots m} f_i. \quad (37)$$

Through the contrast between the hyperplanes, hyperplane having greatest F is chosen. For selecting optimal hyperplane, optimal z and b values have to be found. Lagrangian function L is given by the following equations:

$$L(z', b, \alpha) = \frac{1}{2} z' \cdot z' - \sum_{i=1}^m \alpha_i [y_i : (z' \cdot a_i + b) - 1], \quad (38)$$

$${}_b L(z', b, \alpha) = - \sum_{i=1}^m \alpha_i y_i = 0. \quad (39)$$

Thus z' is calculated as follows:

$$z' = \sum_{i=1}^m \alpha_i y_i a_i, \quad (40)$$

$$\sum_{i=1}^m \alpha_i y_i = 0.$$

In the following equation, replacing L ,

$$z'(\alpha, b) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j a_i a_j. \quad (41)$$

Thus, the following equation is obtained:

$$\max_{\alpha} = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j a_i a_j. \quad (42)$$

When a point lies over the hyperplane, it is categorized as +1 class describing those cardiac risks are identified while under the hyperplane, it is -1 class describing that no cardiac risks are identified.

4. Performance Analysis

For the performance evaluation of the proposed method, MATLAB software is used for implementation. The efficiency of the proposed method is evaluated with a few parametric measures like accuracy, precision, recall, and F1-score. By using the U-matrix error rate of topographical error, then quantization error has been estimated. The dataset chosen for estimation is the diabetic dataset. To test the performance of the model, data are randomly selected from the dataset.

4.1. Quantization Error. During training, the required amount of computation level is reduced. Moreover, it estimates the performance level by applying stochastic quantization evaluating the value of the gradients. It regulates the standard distance between every input vector of the node and its winner.

$$Q_e = \frac{1}{N} \sum_{j=1}^N \|x_j - w_j\|. \quad (43)$$

In the above equation, w_j indicates the weight vector of winner for input x_j . Thus, the expected Q_e has to be small.

4.2. Neuron Utilization N_U . This specifies the ratio of the winner neurons of a single input or more inputs in vector map.

$$N_U = \frac{1}{nm} \sum_{i=1}^{mm} u_i. \quad (44)$$

Here by (45), if neuron i is the winner, then $u_i = 1$ or $u_i = 0$. Therefore, N_U closer to 1.0 is more expected. The error calculation is shown in Figures 4(a) and 4(b)

4.3. U-Matrix. The U-matrix represents the cluster structure of the map illustrating the distances of the adjacent neurons. Output classes of the winning neurons of the U-matrix and winning classes with class names are shown in Figures 5(a) and 5(b)

4.4. Accuracy. This gives the ratio of the instances classified correctly which is estimated by

$$\text{accuracy rate} = \frac{\text{true positive} + \text{true negative}}{\text{total instances}} * 100. \quad (45)$$

4.5. Precision. This is the measure which reveals the ratio of data transmitted in the network with intrusion. This parameter estimates the correctness and quality of classification which is determined by

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}. \quad (46)$$

4.6. Recall. This metric presents the proportion of real positives that are correctly predicted positives given by

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}. \quad (47)$$

4.7. F1-Score. This is generally the mean value of precision and recall. Moreover, statistical measure is used in F1-score to calculate the performance rate of individual classifier of FN and FP as given in (49). Precision judges the accuracy, while recall detects sample instances with respect to faulty or nonfaulty attributes.

$$F1 \text{ score} = \frac{2X \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (48)$$

Table 1 shows some of the observations from the instances of diabetic datasets, which are then classified. The performance measures of various techniques of KNN, NN, and ANN are considered for estimating the efficiency of the proposed GSOA-DBN_CNN technique. Table 1 presents the comparative analysis of accuracy, precision, recall, and F1-score, which are represented in percentage.

Figure 6 graphically represents the comparison for various methods in terms of accuracy. It is observed that the proposed GSOA-DBN_CNN technique produced more

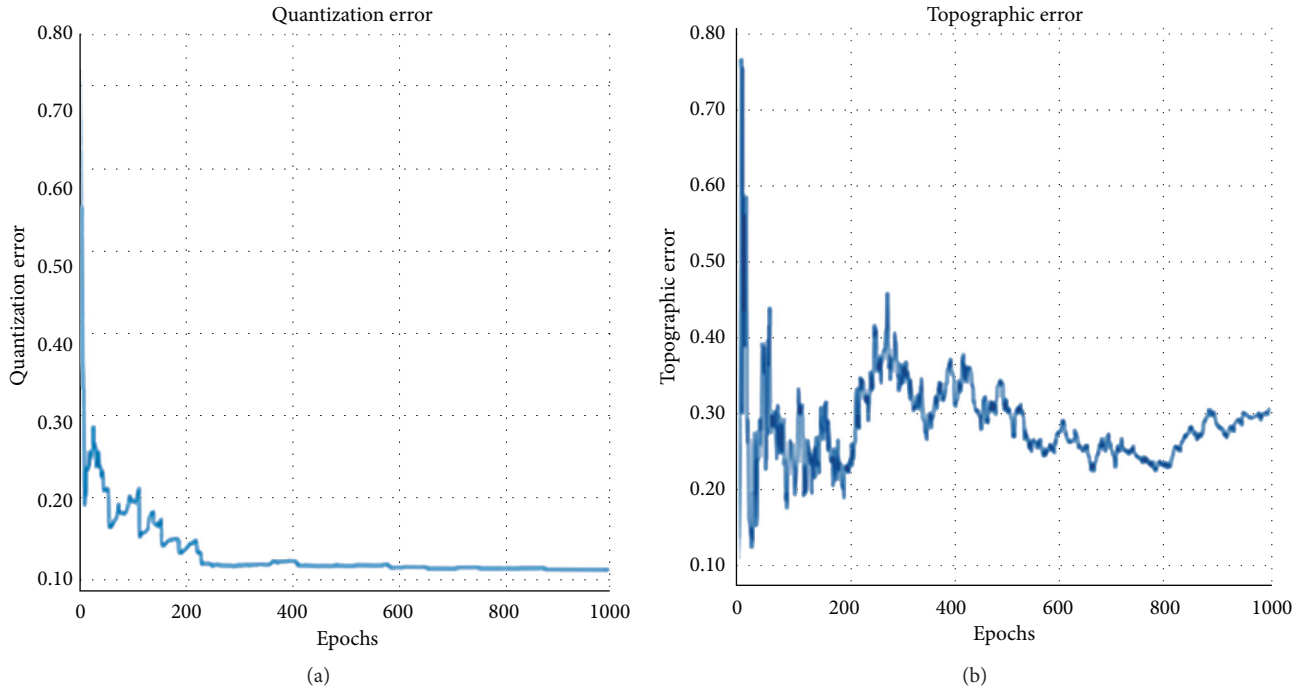


FIGURE 4: (a) Quantization error calculation and (b) topographic error calculation.

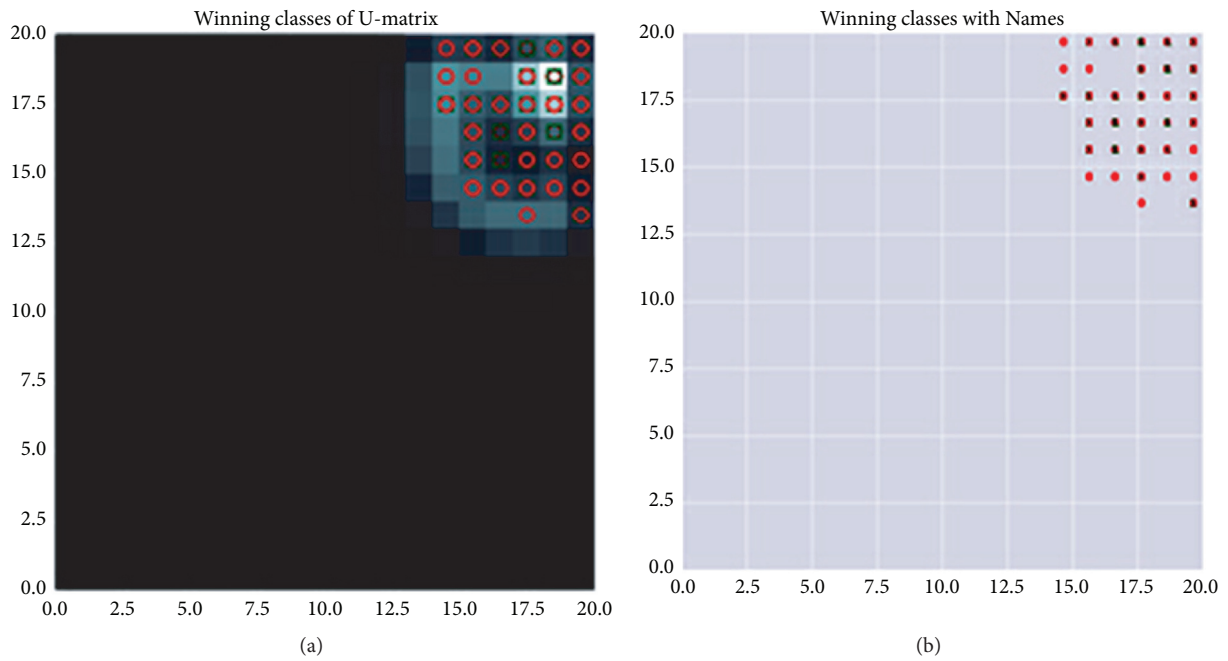


FIGURE 5: (a) Output classes of the winning neurons of the U-matrix and (b) winning neurons with class names.

TABLE 1: Comparison of performance.

Parameters	KNN	NN	ANN	GSOA-DBN_CNN
Accuracy	94	96	97	98
Precision	89	92	94	96
Recall	87	88	91	94
F1-score	82	86	88	92

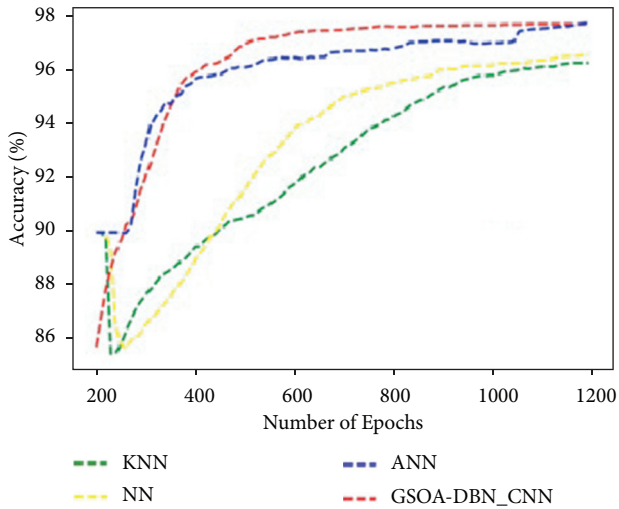


FIGURE 6: Comparison of accuracy.

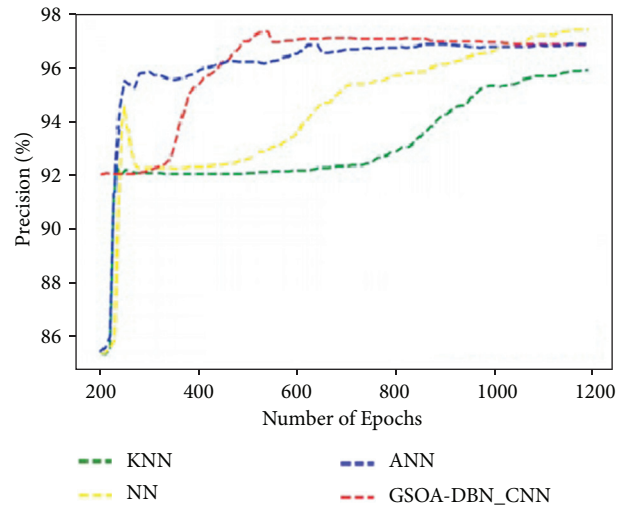


FIGURE 7: Comparison of precision.

accuracy than the existing techniques. Accuracy has been improved for proposed technique as 98%.

Figure 7 graphically represents the comparison for various methods in terms of precision. It is proved that the proposed GSOA-DBN_CNN technique produced higher precision than the existing techniques. By comparing precision analysis with existing technique, the proposed technique has been improved as 96%.

Figure 8 graphically represents the comparison for various methods in terms of recall. It is noticed that the proposed GSOA-DBN_CNN technique produced maximum recall than the existing techniques. Recall has been improved as 94% for proposed technique when compared with existing technique.

Figure 9 graphically represents the comparison for various methods in terms of F1-score. It is discovered that the proposed GSOA-DBN_CNN technique achieves improved F1-score as 92% than the existing techniques.

PSNR graph comparison for proposed GSOA and existing ACO, WOA, GA, and HA is shown in Figure 10.

Figure 10 graphically represents the comparison for various methods in terms of PSNR comparison for data optimization. It is inferred that the proposed GSOA-DBN_CNN technique achieves improved 0.2 PSNR than the existing techniques.

The comparative results are described in the Table 2, and the overall analyses show that the proposed GSOA-DBN_CNN achieved accuracy higher than 2% of ANN, 3% of NN, and 4% of KNN. The precision achieved is more than 2% of ANN, 3% of NN, and 6% of KNN. Recall obtained is higher than 2% of ANN, 3% of NN, and 4% of KNN, similarly the F1-score achieved is 2% of ANN, 3% of NN, and 3% of KNN. The proposed GSOA-DBN_CNN technique achieves improved 0.2 PSNR than the existing AWO, WOA, and GA techniques which is shown in Figure 11.

From the experimental results, it is observed that the proposed GSOA-DBN_CNN technique achieves accuracy of 98%, 96% of precision, recall of 94%, and PSNR of 0.2 higher than the existing ACO, WOA, and GA techniques. Although

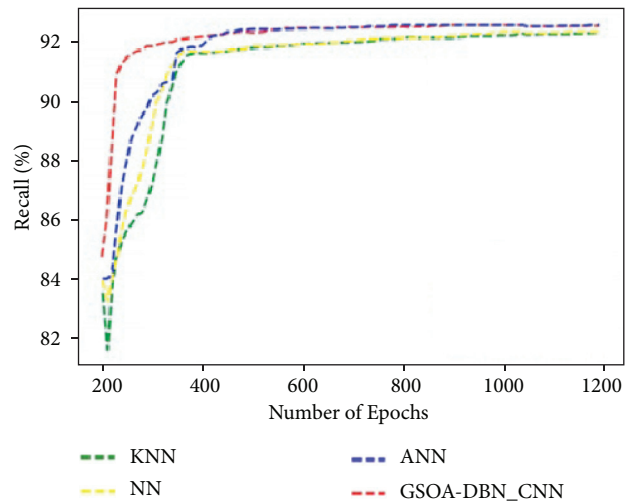


FIGURE 8: Comparison of recall.

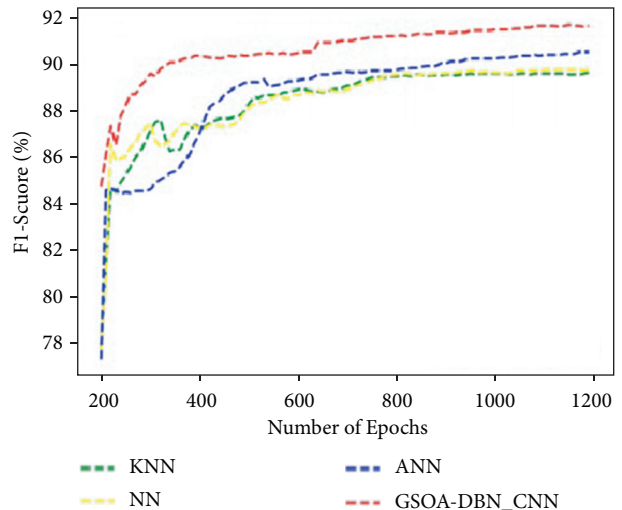


FIGURE 9: Comparison of F1-score.

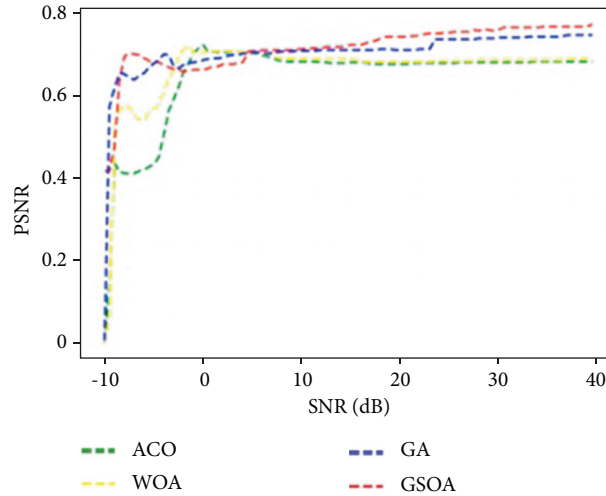


FIGURE 10: PSNR comparison for data optimization.

TABLE 2: Overall result comparison.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN	94	92	90	88
NN	95	95	91	88
ANN	96	96	92	90
GSOA-DBN_CNN	98	98	94	92

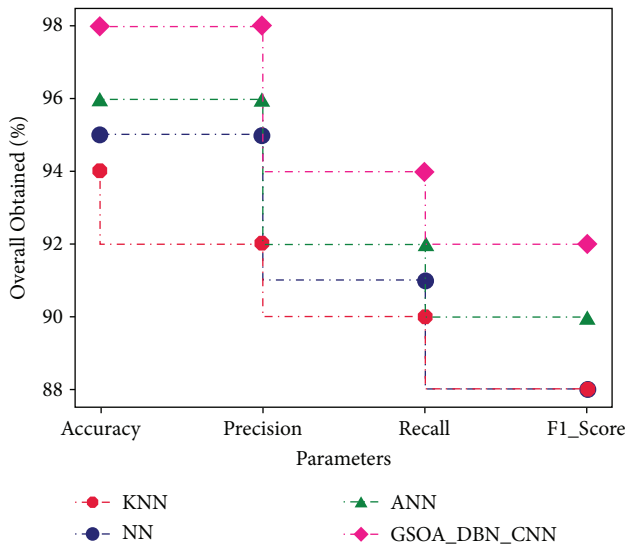


FIGURE 11: Overall comparison.

the proposed GSOA-DBN_CNN model achieves better results, it lacks efficiency in some areas, which is to be improved.

5. Conclusion

The efficiency of the metaheuristic optimization algorithms is proved by solving several issues related to text clustering. But, the trapping of local optima is possible as the focus is on global search rather than on local search, that is exploration instead of exploitation. In this study, a metaheuristic algorithm based on data optimization is proposed for diabetes

data, which leads to the predictive analysis of cardiac attack risk based on blood vessels and cardiac nerves and cardiac nerve damages. For data optimization, GSOA is used since those metaheuristic optimization algorithms are robustly feasible. Then, a deep belief architecture (DBA) is established, where the classification is carried out using CNN. By this classification, the normal range and abnormal range of diabetes have been classified. The normal diabetes range has been updated to the hospital database, and for the abnormal diabetes range, the cardiac nerve and blood vessel damages have been analyzed using SVM-based image predictive analysis. Further, the method has also been compared with ACO, WOA, GA, and HA. The results reveal that the proposed GSOA-DBN_CNN technique is better in terms of every parameter considered for comparison against the existing techniques. The accuracy achieved by the proposed technique is 98%, precision attained is 96%, the recall has been improved up to 94%, and F1-score obtained by 92% when compared with existing KNN, ANN, and NN. The comparative results are described as follows: the proposed GSOA-DBN_CNN achieved accuracy higher than 2% of ANN, 3% of NN, and 4% of KNN. The precision achieved is more than 2% of ANN, 3% of NN, and 6% of KNN. Recall obtained is higher than 2% of ANN, 3% of NN, and 4% of KNN, similarly, the F1-score achieved is 2% of ANN, 3% of NN, and 3% of KNN. The proposed GSOA-DBN_CNN technique achieves an improved 0.2 PSNR than the existing AWO, WOA, and GA techniques. The data optimization has been improved when compared with existing optimization techniques in terms of PSNR. In the future, hybrid deep learning methods can be utilized to further improve the efficiency of the model.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

This study was performed as a part of the employment of the authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors extend their gratitude to the deanship of scientific research at King Khalid University for funding this work through a research group program under Grant no. R.G.P. 1/399/42.

References

- [1] G. U. Maheswari, R. Sujatha, V. Mareeswari, and E. P. Ephzibah, "The role of metaheuristic algorithms in healthcare," in *Machine Learning for Healthcare*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2020.
- [2] H. Firdaus, S. I. Hassan, and H. Kaur, "A comparative survey of machine learning and meta-heuristic optimization algorithms for sustainable and smart healthcare," *African Journal of Comput. ICT Ref. Format*, vol. 11, no. 4, pp. 1–17, 2018.
- [3] S. Murugan, S. Jeyalakshmi, B. Mahalakshmi, G. Suseendran, T. N. Jabeen, and R. Manikandan, "Comparison of ACO and PSO algorithm using energy consumption and load balancing in emerging MANET and VANET infrastructure," *Journal of Critical Reviews*, vol. 7, no. 9, 2020.
- [4] A. Abugabah, A. A. AlZubi, F. Al-Obeidat, A. Alarifi, and A. Alwadain, "Data mining techniques for analyzing healthcare conditions of urban space-person lung using metaheuristic optimized neural networks," *Cluster Computing*, vol. 23, no. 3, pp. 1781–1794, 2020.
- [5] A. Saha, C. Chowdhury, M. Jana, and S. Biswas, "IoT sensor data analysis and fusion applying machine learning and metaheuristic approaches," *Enabling AI Applications in Data Science*, pp. 441–469, 2021.
- [6] P. Suganya and C. P. Sumathi, "A novel metaheuristic data mining algorithm for the detection and classification of Parkinson disease," *Indian Journal of Science and Technology*, vol. 8, no. 14, p. 1, 2015.
- [7] A. Doulamis, N. Doulamis, A. Angeli et al., "A non-invasive photonics-based device for monitoring of diabetic foot ulcers: architectural/sensorial components & technical specifications," *Inventions*, vol. 6, no. 2, p. 27, 2021.
- [8] J. Li, L. S. Liu, S. Fong et al., "Adaptive swarm balancing algorithms for rare-event prediction in imbalanced healthcare data," *PloS one*, vol. 12, no. 7, Article ID e0180830, 2017.
- [9] S. Chatterjee, J. Byun, K. Dutta, R. U. Pedersen, A. Pottathil, and H. Xie, "Designing an Internet-of-Things (IoT) and sensor-based in-home monitoring system for assisting diabetes patients: iterative learning from two case studies," *European Journal of Information Systems*, vol. 27, no. 6, pp. 670–685, 2018.
- [10] S. Fong, S. Deb, X.-S. Yang, and J. Li, "Feature selection in life science classification: metaheuristic swarm search," *IT Professional*, vol. 16, no. 4, pp. 24–29, 2014.
- [11] N. D. Hoang, "Image processing-based pitting corrosion detection using metaheuristic optimized multilevel image thresholding and machine-learning approaches," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6765274, 19 pages, 2020.
- [12] I. M. El-Hasnony, S. I. Barakat, and R. R. Mostafa, "Optimized ANFIS model using hybrid metaheuristic algorithms for Parkinson's disease prediction in IoT e," *IEEE Access*, vol. 8, pp. 119252–119270, 2020.
- [13] R. Mukherjee and U. M. Diwekar, "Multi-objective optimization of the TEG dehydration process for BTEX emission mitigation using machine-learning and metaheuristic algorithms," *ACS Sustainable Chemistry & Engineering*, vol. 9, no. 3, pp. 1213–1228, 2021.
- [14] R. Singh, "Nature inspired based meta-heuristic techniques for global applications," *The International Journal of Computer Applications & Information Technology*, vol. 12, no. 1, pp. 303–309, 2020.
- [15] A. Tayal, A. Solanki, and S. P. Singh, "Integrated frame work for identifying sustainable manufacturing layouts based on big data, machine learning, meta-heuristic and data envelopment analysis," *Sustainable Cities and Society*, vol. 62, Article ID 102383, 2020.
- [16] J. Mallick, S. Alqadhi, S. Talukdar et al., "Risk assessment of resources exposed to rainfall induced landslide with the development of GIS and RS based ensemble metaheuristic machine learning algorithms," *Sustainability*, vol. 13, no. 2, p. 457, 2021.
- [17] S. Kaul and Y. Kumar, "Nature-Inspired metaheuristic algorithms for constraint handling: challenges, issues, and research perspective," in *Constraint Handling in Metaheuristics and Applications*, Springer, New York, NY USA, 2021.
- [18] T. M. Le, T. M. Vo, T. N. Pham, and S. V. T. Dao, "A novel wrapper-based feature selection for early diabetes prediction enhanced with a metaheuristic," *IEEE Access*, vol. 9, pp. 7869–7884, 2020.
- [19] V. Mazaheri and H. Khodadadi, "Heart arrhythmia diagnosis based on the combination of morphological, frequency and nonlinear features of ECG signals and metaheuristic feature selection algorithm," *Expert Systems with Applications*, vol. 161, Article ID 113697, 2020.
- [20] A. Sampathkumar, S. Murugan, M. Sivaram, V. Sharma, K. Venkatachalam, and M. Kalimuthu, "Advanced energy management system for smart city application using the IoT," *Internet of Things in Smart Technologies for Sustainable Urban Development*, pp. 185–194, 2020.
- [21] V. Manikandan, K. Gowsic, T. Prince, R. Umamaheswari, B. F. Ibrahim, and A. Sampathkumar, "DRCNN-IDS approach for intelligent intrusion detection system," in *Proceedings of the 2020 International Conference on Computing and Information Technology (ICCIT-1441)*, pp. 1–4, IEEE, Tabuk City, Saudi Arabia, September 2020.
- [22] A. Sharma, K. Guleria, and N. Goyal, "Prediction of diabetes disease using machine learning model," in *Proceedings of the International Conference on Communication, Computing and Electronics Systems*, pp. 683–692, Springer, Coimbatore, India, October 2021.
- [23] N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis, and K. Moustakas, "Machine learning tools for long-term type 2 diabetes risk prediction," *IEEE Access*, vol. 9, pp. 103737–103757, 2021.

- [24] L. Abualigah, A. H. Gandomi, M. A. Elaziz et al., “Advances in meta-heuristic optimization algorithms in big data text clustering,” *Electronics*, vol. 10, no. 2, p. 101, 2021.
- [25] A. A. Sayed, M. M. Abdallah, A. M. Zaki, and A. A. Ahmed, “Big data analysis using a metaheuristic algorithm: twitter as case study,” in *Proceedings of the 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, pp. 20–26, IEEE, Aswan, Egypt, February 2020.
- [26] Y. Zhang and F. Liu, “An improved deep belief network prediction model based on knowledge transfer,” *Future Internet*, vol. 12, no. 11, p. 188, 2020.

Research Article

Examining the Determinants of Patient Perception of Physician Review Helpfulness across Different Disease Severities: A Machine Learning Approach

Adnan Muhammad Shah ^{1,2}, Wazir Muhammad ², and KangYoon Lee ¹

¹Department of Computing Engineering, Gachon University, Seoul 13120, Republic of Korea

²Department of Physics, Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL 33431-0991, USA

Correspondence should be addressed to KangYoon Lee; keylee@gachon.ac.kr

Received 17 September 2021; Revised 3 January 2022; Accepted 13 January 2022; Published 26 February 2022

Academic Editor: Lo'ai Tawalbeh

Copyright © 2022 Adnan Muhammad 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.

(1) *Background.* Patients are increasingly using physician online reviews (PORs) to learn about the quality of care. Patients benefit from the use of PORs and physicians need to be aware of how this evaluation affects their treatment decisions. The current work aims to investigate the influence of critical quantitative and qualitative factors on physician review helpfulness (RH). (2) *Methods.* The data including 45,300 PORs across multiple disease types were scraped from *Healthgrades.com*. Grounded on the signaling theory, machine learning-based mixed methods approaches (i.e., text mining and econometric analyses) were performed to test study hypotheses and address the research questions. Machine learning algorithms were used to classify the data set with review- and service-related features through a confusion matrix. (3) *Results.* Regarding review-related signals, RH is primarily influenced by review readability, wordiness, and specific emotions (positive and negative). With regard to service-related signals, the results imply that service quality and popularity are critical to RH. Moreover, review wordiness, service quality, and popularity are better predictors for perceived RH for serious diseases than they are for mild diseases. (4) *Conclusions.* The findings of the empirical investigation suggest that platform designers should design a recommendation system that reduces search time and cognitive processing costs in order to assist patients in making their treatment decisions. This study also discloses the point that reviews and service-related signals influence physician RH. Using the machine learning-based sentic computing framework, the findings advance our understanding of the important role of discrete emotions in determining perceived RH. Moreover, the research also contributes by comparing the effects of different signals on perceived RH across different disease types.

1. Introduction

Understanding patient preferences of service quality is vital for the healthcare industry and healthcare providers to develop optimal strategies to improve patients' quality of care [1]. With the growing popularity of physicians rating websites (PRWs), better information can be obtained regarding factors influencing patients' choices of selecting the right doctor [2]. Unlike traditional surveys used to collect information on patients' preferences and treatment experiences, physician online reviews (PORs) offer a rich source of knowledge without the interventions by researchers or healthcare organizations [3]. Recent studies have shown that

PORs are specific type of word of mouth that plays a crucial part in the patients' decision-making [4]. PORs are a significant source of knowledge for many patients who are looking for a good doctor [5]. They see these PRWs as an important source for finding the best doctor [6, 7]. These PORs offer authentic information for patients' wellbeing but likewise contribute to an evolving relationship between doctors and their patients [8, 9].

Using feature engineering to identify helpful reviews, users could be able to reduce the search cost. Although PORs alleviate the overall choice burden on users, they also trigger many issues, such as presenting misguided or inappropriate information [10]. Hence, it is critical to explore review

helpfulness (RH) by identifying the characteristics of highly helpful reviews. In the previous research, the review assessment was mainly focused on quantitative characteristics of a review (e.g., rating, valence, and sentiment polarity) [11, 12]. Shah et al. [13] indicated that quality of service in terms of dispersion of online reviews significantly influences RH. Researchers also indicated that reviews for popular services get more helpful votes [10, 14]. Moving away from quantitative measures, more recent research focused on qualitative measures (e.g., readability, word count, and emotions) to evaluate RH [11, 12, 15, 16]. In considering the multiple review types and related key issues, however, helpfulness is quite a nuanced term, as quantitative measures of reviews are equally useful, whereas others might consider qualitative characteristics as more helpful [15]. Fang et al. [17] indicated that text readability significantly influences perceived RH. Mauro et al. [18] revealed that review wordiness is a meaningful predictor of RH. The study of Malik and Hussain [19] indicated that discrete emotions are the most dominant emotions with greater influence on perceived RH. Ren and Hong [20] found a significant relationship between negative emotions and RH. This reveals that various aspects of information and behaviors are helpful in the desirable purchase decision-making process [21–23].

For quite a while, the research topic of RH has drawn academic interest in the search and experience goods context [19, 20, 24]. The idea of RH has also gained researchers' attention in the healthcare domain, which is categorized as credence goods. In comparison to other goods, credence goods are distinct since the quality of credence goods cannot be determined even after the utility has been consumed [25]. Evaluating the effectiveness of PORs in comparison to product and service reviews is challenging, as the provider defines the utility impact of the goods, creating an asymmetric information state. The findings from previous studies revealed that, for patients, the helpfulness of PORs plays a pivotal part in their treatment decision-making process and for a physician to improve the quality of care [13, 22, 23, 26]. However, RH was not thoroughly investigated as a function of quantitative and qualitative measures concurrently in credence goods context (healthcare). This research was therefore conceived to expand earlier studies on the RH by exploring not only the quantitative factors (valence and volume), but qualitative characteristics of reviews as well such as readability features, wordiness features, and discrete emotions. Hence, this study classifies different review attributes (concept-level prospective) and service characteristics and implicitly assesses patients' distinct emotions by employing sentic computing model proposed by Cambria et al. [27] to compute the physician RH.

In addition, patients' perception of physician service quality varies across two main disease types: serious diseases (high disease severity) and mild diseases (low disease severity). Disease severity determines how severe the effects of a type of illness are [28–30]. Because the disease severity is normally associated with unspecified factors such as unexpected mortality, increased treatment costs, and prolonged hospital stays, patients with more severe disease tend to be more anxious about the quality of treatment they

receive from their physicians than those with less severe disease [8]. Our approach shows that the review- and service-related signals are highly associated with the helpfulness count of PORs that affect decision-making of patients who suffer from different disease types (high-illness severity vs. low-illness severity). Using the plethora of information in terms of PORs provided by *Healthgrades.com* and users of the online healthcare services they choose for their cure (PORs), we applied a text mining and econometric approach to determine the signaling mechanisms that affect patients' treatment choice of different physicians. Additionally, secondary data analysis and machine learning classification approaches were used to construct more accurate models for predicting physician RH. As a result of these considerations, this study will attempt to answer the subsequent research questions (RQs):

RQ1. What effect do the various review- and service-related signals have on the physician RH?

RQ2. How do specific emotions (joy, sadness, surprise, trust, anger, anticipation, disgust, and fear) associated with review-related signals influence physician RH implicitly?

RQ3. What role does disease type have in the relationships between review- and service-related signals and physician RH?

This work contributes to the digital health literature by differentiating between the two kinds of signals (review- and service-related signals) and exploring their effects on online physician RH. Drawing on signaling theory, the research contributes to the conceptualization and interpretation of the RH features from both quantitative and qualitative perspectives. The dataset, including 45,300 PORs from *Healthgrades*, examined ten hypotheses. The proposed model was successfully validated, and critical components that would make an opinion relevant to readers were discovered. The study findings have added to related literature by offering more comprehension of the structural characteristics (quantitative and qualitative) of reviews and their effect on RH [18, 19, 31]. The findings indicate that both the review- and service-related signals significantly and positively influence perceived RH. Second, following an examination of the differential impacts on physician RH, we examine the effect of distinct emotions (sentiments) on RH. Despite the fact that extant research has been conducted on the role of emotions expressed in PORs [22, 32, 33], the research on the domain knowledge-based specialized discrete emotion analysis of online textual reviews has been neglected. This study examines the impact of two-sided discrete emotions embedded in PORs on physician RH. The results show that two-sided reviews (positive and negative discrete emotions) significantly influence the perceived RH. Third, this study utilized the idea of environmental uncertainty, which has been referred to as disease type in the virtual healthcare market setting in prior strategic studies [22, 26, 28, 33, 34]. This study extends the scope of prior research and advances signaling theory by examining how the effects of review- and service-related signals on

physician's RH differ according to the type of disease being treated. The findings revealed a significant positive moderating effect of disease type in the relationships between review wordiness, service quality, service popularity, and perceived RH. We focus on multimethod analysis, including implicit and domain knowledge-based specialized sentic computing emotion analysis and econometric approach to predict physician RH. The proposed multimethod model shows an excellent performance with a classification accuracy of 91.12%.

2. Theoretical Background

RH refers to how many times the review has been voted as useful by other reviewers in order to guide purchase decisions [35]. The confounding variation in reviews posted for one commodity makes it challenging for customers to evaluate helpful entities. The sheer number of competitive goods and overwhelmed data make it difficult for consumers' online decision-making. PRWs like RateMDs and Healthgrades were the pioneer platforms of helpful voting to mitigate the problem.

Particularly, this feature explicitly leverages crowdsourcing to determine the helpfulness of reviews. A question follows each review, "*Was this review helpful?*" Consumers who have read reviews might vote by clicking the option: Yes or No. Reviews receiving optimistic or crucial helpfulness votes are followed by notes, for instance, "*This review was useful to 6 out of the 9 people who read it.*" The most helpful satisfactory and negative reviews for a single commodity finally top the list of reviews on rating sites. Online rating sites rank customer feedback based on helpfulness score, which minimizes the customers' time to find valuable information [36]. Helpful voting, together with customer feedback, provides a broad range of information in which researchers can look at factors that could influence consumers' purchasing decisions in an e-shopping setting [37]. Consequently, improving the RH is positively linked to product sales, mainly if they are favorable [14]. The helpfulness voting feature has been of great scientific interest over time. For example, the main components of RH such as review valence [12, 38], volume [14], depth of a review [15, 16, 39], linguistic features [23, 40], readability [11], and emotions [41, 42] have been used to predict RH.

Many PRWs have set up peer-review systems that let people make healthcare decisions based on whether they found a review helpful [13]. For example, *Healthgrades.com* offers a service that presents the top two most-rated reviews submitted by online health consumers to assist other customers in evaluating the quality of physician care. These helpful votes are used as a proxy for review diagnosticity, allowing for the separation of helpful and unhelpful reviews [38]. To put it another way, the helpful information contained in a review may help the health consumers to evaluate the attributes of the physician service quality. This means that Internet information sources with more useful reviews can help patients feel more confident about their consultation intentions [43]. Consistent with this perspective in healthcare, the patients' behavior and interests also shift

across different disease type. For example, a patient with a high-risk disease receives different levels of care than a patient with low-risk disease. Serious disease patients could be more vulnerable to the healthcare quality than patients suffering from common diseases. Previous research has shown that individual health condition has a major impact on their decision to visit a healthcare professional [8, 28, 44].

The literature on RH focuses on the economics of knowledge and how it changes the purchaser buying decision-making process in order to lessen purchase uncertainty associated with the product [38]. The claim made in prior investigations is complemented by signaling theory, which provides a theoretical framework for explaining the differential influence of signals in PORs. In this investigation, we used signaling theory to describe the relationship between signals enclosed in PORs and RH. According to signaling theory, signals contribute to diminishing the information asymmetry between two transaction contributors. Spence [45] demonstrated that information asymmetry exists between various exchange groups when information is exchanged. Signals are important in an online environment because they help minimize the information gap as spatial and temporal gaps make information asymmetry worse between distinct partners [46]. Signal receivers are critical components of the signaling cycle. As a result, the *sender* communicates with the *receiver* via information (*signals*), which the receiver perceives as useful information [47]. The substantial impact of Internet information on stakeholders' decision-making demonstrates that the more the knowledge an individual possesses, the more improved the decision s/he will make [48].

Signaling theory contributes to the reduction of information asymmetry between physicians and patients. People with less knowledge about the credibility of a healthcare provider tend to find information from people who know a lot about it. Thus, peer perspectives can help evaluate information quality and minimize information asymmetry [28]. Previous researchers have used the signaling theory to explore numerous signals in healthcare [25, 28]. While earlier research has largely focused on the sender's insight while neglecting the receiver's opinion, the bond between various signal-related elements could substantially impact the receiver's experience as RH. Signaling theory [28] states that the receiver (*patient*) requires supplementary information (*signals*) regarding the quality of healthcare to reduce information asymmetry prior to contacting their provider.

Signaling theory posits that uncertainty in the environment could have a big impact on how people process signals. According to signaling theory, the impact of various signal transmissions on a patient's choice differs between various settings [25]. The signaling environment is crucial in determining which signal to employ, and the strength of the signal is controlled by the signaling ecosystem in which it operates [47]. The intended recipients of the signal are users who are interested in learning more about healthcare services like RH. Keeping in view the user-generated content, the availability of PORs may have an impact on the examination of which elements influence RH under varied illness circumstances (disease severity).

Although the main objective of the above discussion was to improve the framework of the online review to promote more helpfulness votes, little has been done to explore how the interplay of review- and service-related factors is related to POR's helpfulness. No consistent conclusions have been drawn yet regarding the important factors influencing POR's helpfulness. In addition, the role of the disease type in the RH is yet to be widely explored. Our work fills these knowledge gaps.

3. Hypotheses Development

This work proposes research hypotheses to examine the influence of various review- and service-related signals on physician RH. The former considers three features: readability, wordiness, and discrete emotions, while the latter takes into account the service quality and popularity. Lastly, this work examines the extent to which patients' assessments of a physician RH vary among distinct disease types. Figure 1 sets out the research model.

The readability of a review implies how easily a reader may comprehend a piece of writing. Online reviews must be comprehensible when used as an input variable in order to make buying decisions [19]. Extant research indicated that the level of readability is how well an individual follows the product information [11]. Readability has been identified as a significant component in customers' perception of online information on virtual networks. A review that is sufficiently readable is deemed more beneficial to consumers than one that is excessively lengthy and contains several typographical errors, making it difficult to read [12, 49]. Following readable reviews could help patients save a search and cognitive costs by finding the right information easier [23]. Hence, we hypothesize that the more understandable the text is on the health rating platform, the more useful the review is.

Hypothesis 1 (H1). The higher readability of PORs is positively related to higher RH votes.

The idea of a review wordiness is usually thought of as the amount of information in a review that is detailed or long [23, 50]. It has been demonstrated by researchers that decision-makers capacity to comprehend information is hampered when the amount of data is either excessively high or excessively low. Insufficient information has a negative impact on buyer decisions [39]. Previous studies have also revealed that extreme or knowledge overload can have an adverse effect on the RH in some people [15]. The length of the review is regarded to be an important predictor of RH [16, 38]. According to previous research, the wordiness of a review is considered helpful and acts as directly proportional to the amount of knowledge generated by a review. However, in the event of excessive repetition of the concept, misconceptions, and needless details, wordiness may lead to a poor assessment of helpfulness [15]. Hence, following longer reviews when it provides extensive information may diminish patients' search costs due to enhanced information diagnosticity [26]. Therefore, we have the following hypothesis.

Hypothesis 2 (H2). Review wordiness is positively related to higher RH votes.

Emotions stated in PORs are relevant because they influence patients' clinical decisions [22, 28]. Emotions have been described as an appraisal of a shift in the feelings of a person [51]. Reviews that include both positive and negative sentiments about the appraisal of products or services are the best possible sources of information [51–53]. Researchers have claimed mixed findings of the emotions embedded in a review in positive-negative continuum with some researchers finding positive emotions are more useful [19, 22, 42], while other groups of scholars indicated negative emotions as more diagnostic and helpful [20, 54]. A review is considered as more significant because it gives clear signals about whether or not a service should be taken into account. Consequently, the specific emotion signals (i.e., joy, sadness, surprise, trust, anger, anticipation, disgust, and fear) have an effect on how health consumers perceive the RH [22]. Hence, we hypothesize the following.

Hypothesis 3 (H3). Discrete emotions embedded in PORs are positively related to higher RH votes.

Using service characteristics, capabilities, and features, consumers analyze and form opinions regarding the real level of service they receive [55]. Service quality indicates how users think about the dominance or weakness of the services they use [10]. Patients' perceptions of service quality are shaped by information regarding the quality of the service they get from their peers [56]. For example, patients prefer to consult a physician with a higher star rating for the quality of care [8]. Service rating has been shown to have a positive correlation with RH [37, 41]. Keeping in view the scope of study, physicians' ratings show whether or not people have an excellent or negative perception of their doctors [33]. People are more likely to be drawn to high-quality services, and they are also more inclined to give positive feedback about their experiences [57]. Hence, we hypothesize the following.

Hypothesis 4 (H4). Physicians with exceptional service quality are positively related to higher RH votes.

The popularity of service can be judged by the number of individuals who have talked about it and/or expressed interest in purchasing it [14]. In online health rating platforms, patients considered the amount of PORs to reflect the market or the reputation of a service based on how many people used it [25]. Patients' perceptions of the level of popularity on PRWs may support them in evaluating the quality of their treatment and predicting service delivery [56]. Moreover, a high volume of PORs enhances the likelihood of obtaining correct information that can assist patients in assessing the quality of healthcare services [58]. Researchers discussed that the more the information a patient has, the more likely s/he will make a better decision [59]. A well-known service entices additional users to read and vote on customer reviews. This means that patients may be more confident in assessing the quality and outcome of treatment if many individuals have already reviewed it. Hence, we hypothesize the following.

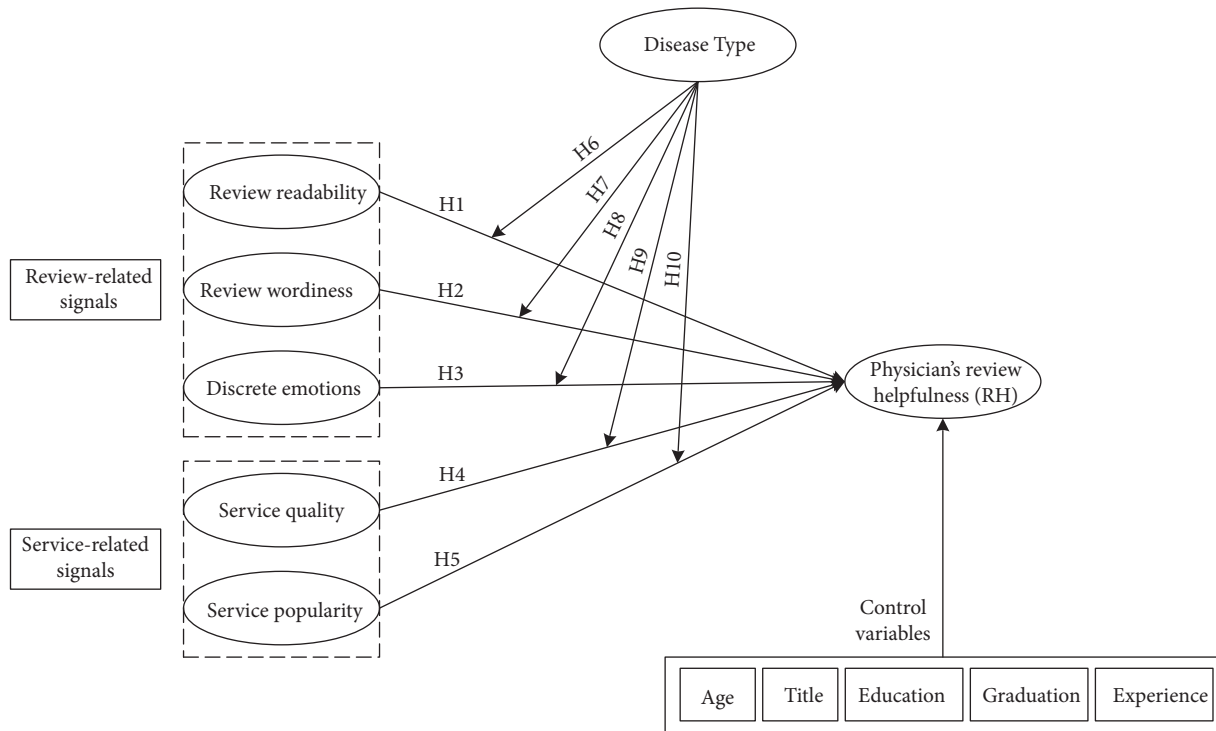


FIGURE 1: Research model.

Hypothesis 5 (H5). Service popularity is positively related to higher RH votes.

Information asymmetry theory states that the efficiency of signals transmission is dependent on the degree of environmental uncertainty. According to the findings of Siering et al. [47], signals have minimal effect on RH when the information environment uncertainty is low. On the other hand, relationships thrive in high uncertainty about information. Given that transmitting signals might help minimize uncertainty, it is evident that the usefulness of the signal is proportional to the degree of uncertainty.

It has also been assumed that the disease type as a feature moderates the association between readability and RH. In the services sector, a more readable review can be assessed more simply than a review with spelling mistakes and ambiguous words [17, 50]. Similarly, patients who suffer from serious diseases expect that information embedded in a review should be more readable than those who suffer from mild diseases. As follows, we posit that readable information about healthcare quality provided in PORs is more helpful for health consumers when they assess severe illnesses.

By adding confidence in the consumers' decision-making, longer reviews may be perceived as more helpful in the purchase process [38]. The detailed information provides further explanation about service and the context where the service was used. Wordiness has varying effects on the RH across different environments [10]. The added content in PORs is more likely to deliver crucial evidence about how the service is consumed and how it relates to alternatives [15]. Therefore, we assumed that patients need more detailed information about the quality of service for serious diseases than those who suffer from moderate illnesses.

This study also takes into account the disease type as a moderator of the relationship between individual emotions and RH. Patients with diverse illness conditions may require varying degrees of healthcare quality [8]. Those with a severe disease (high disease severity) may require a higher standard of care than patients with mild disease (low disease severity) [30]. As a result, we presume that specific emotions involved in PORs will be considered to provide comprehensive information, including service details for serious diseases compared to mild diseases [13, 23]. In this vein, a review that includes both good and negative emotions is likely to be more beneficial for serious illness, as these PORs are less distressing to readers who do not agree with the stated opinion. Keeping in view the earlier discussion, we hypothesize the following.

Hypothesis 6 H6. The moderating role of disease type in the relationship between the readability of a review and RH is stronger for serious diseases than it is for mild diseases.

Hypothesis 7 H7. The moderating role of disease type in the relationship between depth of a review and RH is stronger for serious diseases than it is for mild diseases.

Hypothesis 8 H8. The moderating role of disease type in the relationship between sentiment strength of a review and RH is stronger for serious diseases than it is for mild diseases.

Moving further, we hypothesize that if the disease severity is low, the influence of popularity and quality signals on physician RH will be minimal, owing to the fact that they do not properly minimize uncertainty. In comparison, physician RH will be strongly affected by both these signals if

the disease severity is high since it greatly decreases the related uncertainty. Furthermore, patients with serious diseases are more likely to seek medical help from a doctor who offers high-quality and well-known services. Patients who have had a positive experience with a popular health service that is high-quality are more inclined to recommend the service to others and write reviews about it.

As a result of enhanced service awareness among health consumers, the likelihood of receiving high-quality service and helpful reviews increases because high-risk diseases necessitate a greater degree of service than low-risk diseases; high-risk diseases are connected with popular and high-quality services. Consequently, evaluating the characteristics of popular and high-quality services requires more effort than evaluating the traits of less popular and low-quality services. [60]. Hence, we hypothesize the following.

Hypothesis 9 (H9). The moderating role of disease type in the relationship between service quality and RH is stronger for serious diseases than it is for mild diseases.

Hypothesis 10 (H10). The moderating role of disease type in the relationship between service popularity and RH is stronger for serious diseases than it is for mild diseases.

4. Research Methodology

4.1. Research Context and Data Collection. The data was collected from an online health rating platform (*Healthgrades*) from March 15–21, 2019. Data preprocessing was performed in the form of filtering physician description, review posting date, online reviews, quantitative ratings, and helpfulness count. Online reviews are given further consideration to find out the readability score (six readability tests) and review wordiness using number of concepts in each review in the dataset from multiword expressions, such as “hospital corridor,” “operation theatre equipment,” or “physician appointment,” matched from SenticNet3 [27]. A hybrid sentic computing framework based on the text mining methodology was used to analyze the number of concepts from SenticNet3 linked to each specific emotion. In order to assess the overall performance of the proposed model in predicting RH, a number of regression analyses and text classifications were performed on data that had been filtered and cleaned up. The proposed methodology is shown in Figure 2.

A Web crawler was developed and programmed in Python 3.6 to retrieve the physician web pages shown as search results for each provider. The current study chose 10 different types of online reviews on the basis of disease mortality rate taken from the U.S. health static book 2017 [61] and 4 metropolitan states (California, New York, Texas, and Florida). According to data from the State Medical Boards, these states constitute the largest number of physicians with active board licenses. After omitting 236 reviews because of no helpful votes, 45,300 reviews were used for further analysis in total. The following information was gathered and included in the analysis, such as doctor

specialty, title, education, experience, graduation year, review date, overall rating, number of ratings, patients’ comments, and user responses (helpfulness or usefulness votes).

4.2. Variables Measurements and Statistical Modeling

4.2.1. Review Helpfulness. A cumulative helpfulness vote is calculated on PRWs, and it is derived from the votes of other reviewers who rate the helpfulness level assigned to each individual review. When a POR gets more *helpfulness* votes, the review’s *helpfulness* value rises as a result. *Review helpfulness* variable is assumed to be continuous and assessed as the ratio of helpful/useful votes to total votes. Nonvoted reviews were culled from our database in order to diminish the noise.

4.2.2. Review Readability. Researchers revealed that RH could be influenced by the readability of online reviews [19, 49]. Ghose and Ipeiritos [49] revealed that the degree to which reviews contain subjectivity, knowledge, readability, and linguistic accuracy has an effect on their perceived usefulness. Six types of *readability* methods were explored for each review in order to assess its readability (refer to Table 1).

4.2.3. Review Wordiness. The amount of concepts in a review is used to determine review wordiness [15]. Earlier research has established a substantial correlation between the review depth and RH [38, 62]. *Wordiness* is calculated using the sentic computing framework and SenticNet3 to measure the number of concepts in a review from multiword expressions. SenticNet 3 has previously been used to determine review wordiness as multiword expressions that make online content viral [15, 28].

4.2.4. Service Quality and Service Popularity. Based on previous research [40], a collection of service-associated attributes is included, for instance, (1) *service quality* (i.e., service review valence) [63] and (2) *service popularity* (i.e., review volume), [14].

4.2.5. Disease Type. Following [28], in disease type as a dummy variable, serious diseases as high-risk diseases are labeled as 0, while mild diseases as low-risk diseases are termed 1.

4.2.6. Control Variables. We incorporate control variables to adjust for review- and physician-specific effects. As a control variable, the *Review Age* is provided to represent the distinctive qualities of a review [18]. The age of a review is how long it has been since it was written on online rating platform [39, 64]. Physician *title*, *education*, *graduation year*, and *experience* are the attributes displayed at the physician level. The *title* dummy variable measures the physician’s professional title in the healthcare facility where s/he works

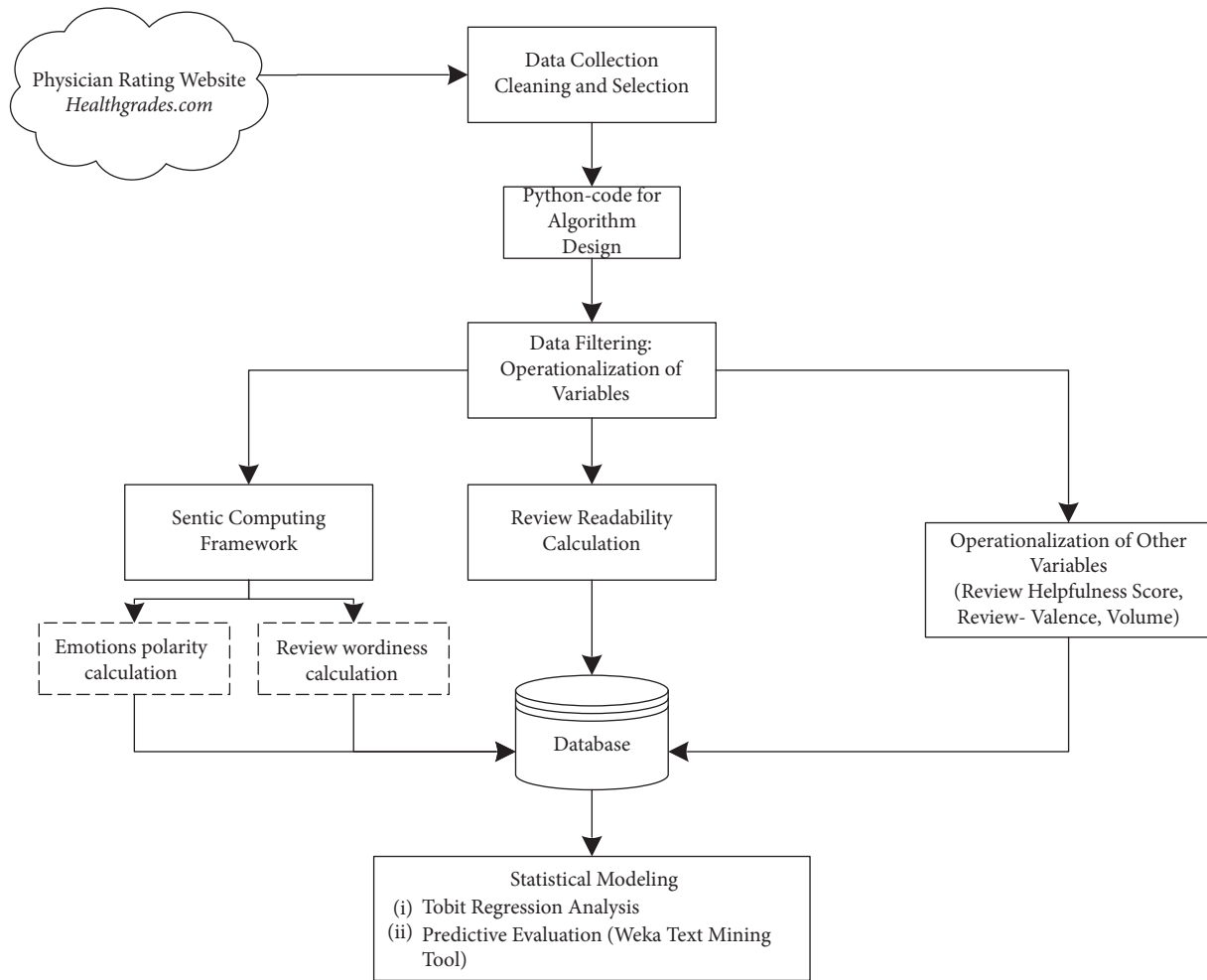


FIGURE 2: Research methodology followed.

[8]. We measured the *education* as a dummy variable using the medical school rank from where the physician graduated and *graduation year* as a categorical variable reflecting the number of years since s/he graduated [28]. Previous research has shown that physicians who graduated from prestigious medical schools are more likely to be rated highly. In comparison, the rating probability is lower for those younger physicians who graduated recently [28]. Finally, *experience* is defined as the period of time (in years) a physician has been in practice. In the past, more experienced physicians have been seen to receive higher ratings [8]. The description of variables and their measurements are listed in Table 1.

4.3. Sentic Computing Framework for Review Wordiness and Specific Emotions. Sentiment mining is a complex process that requires a thorough understanding of the goals of the study. Sentic computing has been applied to a variety of cognitively motivated tasks, including the classification of certain emotions (positive or negative) in natural language text [65]. Integrating knowledge-based methodologies and statistical methods, the hybrid approach to sentic computing and sentiment analysis is able to recognize emotions and

calculate sentiments from the text [28]. The concept mining and emotion classification procedures used in this investigation are depicted in Figure 1.

We undertake preliminary textual data preprocessing, which includes the following: (i) Stop words have been removed (i.e., the, an, and, a, an, etc.). (ii) Use the WordNetLemmatizer function and the WordNet Python natural language toolkit to convert a word to its base structure. (iii) Remove any unnecessary letters (i.e., thanks aloooot). (iv) Question words should be filtered (i.e., which, whose, where, etc.) and any unique characters are excluded (&, #, \$, etc.). (v) Finally, the entire text document is transformed to lower case.

We fragmented the review text into clauses first. Each verb and its corresponding noun phrase are deemed to extract one or more concepts. To make sentences more organized, the input text is chunked using Stanford Chunker [66]. Next, a semantic parser first breaks sentences into clauses and then employs a tree structure to divide clauses into noun and verb chunks [67]. Moreover, a two steps' procedure is followed for clause normalization: First, the Stanford lemmatization algorithm combined with WordNetLemmatizer function from WordNet NLTK is employed to normalize the *verb* chunks and to identify multiword

TABLE 1: Variable description.

Variable	Variable name	Description	Measurement
Dependent variable	Review helpfulness	Review helpfulness refers to ratio of the number of helpful votes a review received to the total votes evaluating helpfulness of that review	Helpfulness
Independent variables	Review readability	Readability is the amount of efforts and educational level required to understand an online review, which is measured by the (1) automated readability index (ARI), (2) Coleman-Liau index (CLI), (3) Flesch-Kincaid grade level (FKGL), (4) Flesch-Kincaid reading ease (FKRE), (5) Gunning's Fog index (GFI), and (6) simple measure of gobbledygook (SMOG) readability index of review text	Readability
	Review wordiness	Review wordiness is the total number of concepts in a review, measured through concept extraction process of the sentic computing framework	Wordiness
	Review emotions	Review emotion is the average polarity of a review, which is measured by the average percentage of the number of positive and negative discrete emotions embedded in a review	Emotions
Moderating variable	Service quality	Service quality reflects the tone or preference of users expressed in positive, negative, or neutral opinion for the service, which is measured by the review valence as average number of rating-stars a service receives	Quality
	Service popularity	Service popularity reflects the number of users discussing the service, which is measured by the review volume as number of reviews received by a service	Popularity
Control variables	Disease type	The disease mortality rate in which a patient suffered from	Disease type
Control variables	Review age	For how long a review has been written on a PRW	Age
	Physician title	A professional title of a physician practicing in healthcare facility	Title
	Physician education	Rank of a medical school the physician graduated from	Education
	Physician graduation	Number of years since a physician graduated	Graduation
	Physician experience	Number of years since a physician is in practice	Experience

expressions. Once the noun phrases have been converted into bigrams, they are processed using part-of-speech (POS) (<https://nlp.stanford.edu/software/tagger.html>) patterns to extract concepts as previously performed by Cambria and Hussain [68]. In addition, the event concept is captured by a parse graph; matches between the object concept and normalized verb chunks are explored in SenticNet3 [69]. Concepts are converted into a vector space model (VSM), where each concept is characterized by a point in a one-dimensional vector space corresponding to a vocabulary phrase.

The affective knowledge is represented using a multi-dimensional VSM. Concept Net and WordNetAffect have been used to create an Affective Space, a multidimensional vector space that illustrates lexical representations of affective knowledge. Affective space portrays the semantic and affective connections that exist between two concepts and allows for quick and effective analogical reasoning between them [68]. Equation (1) characterizes each document d , C_i denotes a concept in d , and f_i denotes the concept frequency in d .

$$d = \{(C_1, f_1), (C_2, f_2), \dots, (C_i, f_i)\}. \quad (1)$$

The cosine similarity between a concept's (C_i) vector space representation (\vec{f}_i) and the vector space

representation of the positive (\vec{f}^+) and negative context terms (\vec{f}^-) was obtained earlier, whereas, n denotes the total number of concepts in a document.

$$\begin{aligned} \text{Cosine_sim}(C_{xy_i^+}) &= \frac{\vec{f}_i \cdot \vec{f}_i^+}{|\vec{f}_i| \cdot |\vec{f}_i^+|} = \frac{\sum_{i=1}^n f_{xi} \times f_{yi}}{\sqrt{\sum_{i=1}^n f_{xi}^2} \sqrt{\sum_{i=1}^n f_{yi}^2}}; \\ \text{Cosine_sim}(C_{xy_i^-}) &= \frac{\vec{f}_i \cdot \vec{f}_i^-}{|\vec{f}_i| \cdot |\vec{f}_i^-|} = \frac{\sum_{i=1}^n f_{xi} \times f_{yi}}{\sqrt{\sum_{i=1}^n f_{xi}^2} \sqrt{\sum_{i=1}^n f_{yi}^2}}. \end{aligned} \quad (2)$$

Following the computation of concept similarity, a series of candidate concepts C^+ having a minimal cosine similarity $\{\text{Cosine_sim}(C_{xy_i^+}), \text{Cosine_sim}(C_{xy_i^-})\}$ is obtained. Using machine learning and the Hourglass of Emotions model developed by Plutchik in his research on human emotions, the framework categorizes emotions into different categories [42, 51] as sentic labels are constructed to denote each concept in VSM, and sentic API is used to forecast the comparable sentic levels for the eight emotion dimensions (positive and negative), suggested by the Cambria et al. [70]. If a match occurs, then the value of the particular emotional dimension is incremented. This procedure is repeated for all terms retrieved from the phrases in the review text to compute the emotions score using the following equations:

Positive emotions were measured as :

$$\begin{aligned} \text{Joy} &= \left(\frac{\text{Number of words associated with joy}}{\text{Number of concepts embedded in a review}} \right) \times 100, \\ \text{Surprise} &= \left(\frac{\text{Number of words associated with surprise}}{\text{Number of concepts embedded in a review}} \right) \times 100, \\ \text{Anticipation} &= \left(\frac{\text{Number of words associated with anticipation}}{\text{Number of concepts embedded in a review}} \right) \times 100, \\ \text{Trust} &= \left(\frac{\text{Number of words associated with trust}}{\text{Number of concepts embedded in a review}} \right) \times 100, \end{aligned} \quad (3)$$

Negative emotions were measured as :

$$\begin{aligned} \text{Angry} &= \left(\frac{\text{Number of words associated with angry}}{\text{Number of concepts embedded in a review}} \right) \times 100, \\ \text{Anxiety} &= \left(\frac{\text{Number of words associated with anxiety}}{\text{Number of concepts embedded in a review}} \right) \times 100, \\ \text{Sadness} &= \left(\frac{\text{Number of words associated with sadness}}{\text{Number of concepts embedded in a review}} \right) \times 100, \\ \text{Disgust} &= \left(\frac{\text{Number of words associated with disgust}}{\text{Number of concepts embedded in a review}} \right) \times 100. \end{aligned}$$

4.4. Empirical Analyses. According to the descriptive statistics listed in Table 2, the average quality rating is 4.59, reflecting the maximum users who expressed positive sentiments about the service quality of healthcare provider. The average readability score of reviews is 9.74 based on 69.80 average length for these reviews. In addition, each doctor has an average of 308.11 reviews. Also, an average opinion score of a review is 0.79, an average RH score is 0.84, and an average review life is found to be 1682 days. Our data set contains 89 percent of medical doctors. A significant number of doctors are recent graduates of the top 100 medical schools in the U.S.

An important decision at this point was the regression model to use, given the dependent variable's limited low and high extremes. In line with previous research [38, 47], we used the TOBIT regression model because the sample and dependent variable were both censored (*Helpfulness*) [71, 72], based on the ratio of helpful votes to the overall vote

count (ranges from 0 to 1) [38]. As a result, the dependent variable RH meets minimum dependent variables (censored data) requirements. This means the dependent variable does better than the censored value, which means the standard model of linear regression can be used [71].

The data were analyzed with STATA software, and the likelihood ratio and Efron's pseudo-R-square value were used to determine the goodness of fit [73]. Furthermore, the empirical analyses must be adjusted using the log-transformation in order to improve the fit of the variables in the empirical model and to adjust for overdispersion. We applied logarithmic transformation [74] to variables such as *helpfulness*, *wordiness*, *quality*, *popularity*, and *age*. To avoid having logarithms of zeros, the value of 1 is added to these variables [75]. All variables utilized to predict the RH are listed in equation (4). The description and measurement of variables are provided in Sections 4.2 and 4.3 and Table 1.

$$\begin{aligned} \ln(\text{Helpfulness}_i) &= \beta_0 + \beta_1 \text{Readability}_i + \beta_2 \ln \text{Wordiness}_i + \beta_3 \text{Emotions}_i + \beta_4 \ln \text{Quality}_i + \beta_5 \ln \text{Popularity}_i + \\ &\beta_6 (\text{Readability}_i \times \text{DiseaseType}_i) + \beta_7 (\text{Wordiness}_i \times \text{DiseaseType}_i) + \beta_8 (\text{Emotions}_i \times \text{DiseaseType}_i) + \\ &\beta_9 (\ln \text{Quality}_i \times \text{DiseaseType}_i) + \beta_{10} (\ln \text{Popularity}_i \times \text{DiseaseType}_i) + \beta_{11} \text{Controls}_i + \mu_i. \end{aligned} \quad (4)$$

TABLE 2: Variables descriptive statistics.

Variables	Variable name	Min	Max	Mean	Std. Dev
Dependent variable	Review helpfulness	0.5	1.0	0.84	0.22
Independent variables	Review readability	1	12	9.74	1.02
	Review wordiness	8	94	69.80	2.11
	Review emotions	0	1.0	0.79	0.24
	Service quality	1	5	4.59	1.26
	Service popularity	3	412	308.11	3.41
Moderating variable	Disease type	0	1	0.462	0.19
Control variables	Review age	0	1826	1682	91.23
	Physician title	0	1	0.89	0.23
	Physician education	0	1	0.86	0.29
	Physician graduation	0	26	4.70	0.62
	Physician experience	0	25	4.10	0.54

4.5. Classification Techniques and Evaluation Metrics. The data mining software Weka 3.8.5 was employed in this study, and the classification model was constructed using a support vector machine (SVM), linear regression (LR), random forest (RF), and gradient boost decision tree (GBDT). We chose these models because previous studies used these models successfully and achieved excellent classification results [18, 31].

SVM is based on statistical learning theory and is now one of the most successful approaches for analyzing high-dimensional datasets and is extensively used to perform classification tasks [76]. The fundamental notion of SVM is the application of structural risk minimization, which reduces boundary error through induction while minimizing overall risk. Once the data are transferred to a higher-dimensional space, they are separated by a hyperplane. A hyperplane-projected subspace can map a new instance, which can then be allocated to the majority class in that subspace.

Regression analysis refers to the statistical technique used to analyze data. Its objective is to ascertain the degree of correlation between two or more variables and build a mathematical model for forecasting the outcome. LR is a nonlinear regression model that attempts to predict how likely an event will happen by fitting data to a logistic function. This allows inputs with any value to be converted and confined to a value between 0 and 1.

A RF is a technique for ensemble learning that was established through the construction of numerous DTs [77]. Training an RF involves bagging bootstrap cases and then selecting a random subset of features. Following that, a set of DTs is generated using each bootstrap instance set containing a subset of features. After the set of trees has been built, the majority class of individual trees can be used to make a prediction about samples that have yet to be seen.

Gradient boost decision tree algorithm systematically adds weak learners in such a way that each new learner matches the preceding step’s residuals, hence improving the model [78]. The final model combines the outcomes of each phase to produce a strong learner. Gradient boosted decision trees technique makes use of decision trees as weak learners to achieve better results. The residuals are detected using a loss function. It is worth mentioning that when a new tree is

added to the model, the current trees remain unchanged. The residuals from the existing model are well-fit by the decision tree that has been introduced. The effectiveness of applied learning classifiers is assessed using two assessment measures (f-measure and accuracy). These metrics are mathematically defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F1 &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \end{aligned} \quad (5)$$

5. Results

Prior to conducting a tobit analysis, this research conducted several diagnostic tests to determine the model’s heteroscedasticity and multicollinearity. The maximum variance inflation factors (VIF) range between 1.53 and 5.51, much below the cut-off value of 10, indicating that multicollinearity is not an issue at the moment [79]. Furthermore, the correlation values between variables suggest that our dataset is free of multicollinearity (0.90 and higher) [50]. In addition, we calculated standard errors for our models that were consistent with heteroscedasticity (see Table 3). With a relatively substantial likelihood ratio, our model attained goodness of fit ($p \leq 0.001$) and McKelvey and Zavoina [73] Efron’s Pseudo R^2 value of 0.083.

In particular, when looking at the major effects for review-related signals in Table 3, the findings of the regression analysis suggest that the *readability* coefficient is positively significant ($\beta = 0.174$, $p < 0.05$). As a result, *H1* is supported, which is in line with earlier research findings [19]. Following that, the results indicate a substantial positive coefficient for *wordiness* ($\beta = 0.320$, $p < 0.05$), which supports *H2*, consistent with the findings of Mudambi and Schuff [38], but negates the results of Qazi et al. [15]. Furthermore, the

TABLE 3: Heteroscedasticity compatible results of hypotheses testing.

Constructs	B	Std. error	<i>p</i> value	<i>t</i> value
(Constant)	1.993	0.071	0.006**	30.401
Age	0.010	0.015	0.016*	0.041
Title	0.036	0.025	0.002**	0.087
Education	0.465	0.970	0.000***	1.432
Graduation	0.150	0.533	0.312	1.196
Experience	0.028	0.078	0.007**	0.125
Readability	0.174	0.006	0.043*	3.423
Wordiness	0.320	0.041	0.030*	2.732
Emotions	0.013	0.001	0.000***	1.230
Quality	0.125	0.015	0.045*	1.014
Popularity	0.232	0.033	0.004**	2.006
Disease type	0.024	0.041	0.012**	0.013
Readability × disease type	0.028	0.010	0.366	0.750
Wordiness × disease type	0.018	0.009	0.038*	0.900
Emotion × disease type	0.040	0.004	0.256	0.430
Quality × disease type	0.080	0.019	0.036*	0.548
Popularity × disease type	0.053	0.010	0.030*	0.430
Efron's R^2	0.084	Log-likelihood	-2745.618	
Likelihood ratio	429.631		$p \leq 0.001$, $df = 8$	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

significant positive coefficient for *emotions* ($\beta = 0.013$, $p < 0.001$) demonstrates that the association will be stronger when both positive (including joy, sadness, trust, except surprise) and negative emotions (including anger, disgust, fear, except anticipation) are included in a review, accepting *H3*. These findings are in line with earlier research [51]. We find support for *H4* when we looked at the service-related signals ($\beta = 0.125$, $p < 0.05$), emphasizing that when a POR focuses on the quality of service *quality*, its helpfulness value increases. These findings corroborate prior findings [14]. Moreover, we uncover evidence supporting *H5*, which hypothesizes a considerable association between service popularity and RH. In particular, the statistically significant positive coefficient of service *popularity* ($\beta = 0.232$, $p < 0.01$) demonstrates that there is a positive association between the number of reviews and helpful votes obtained by the doctor. These findings are consistent with those of Zhang and Lin [40].

When it comes to the moderating effect of illness severity, the results show that they strongly support the use of interaction terms (*Wordiness × Disease Type*), (*Quality × Disease Type*) and (*Popularity × Disease Type*) as *H7*, *H9*, and *H10*, indicating that having more words in a review, as well as its quality and popularity of a physician service embedded in the review, has a more significant positive impact on how people think about the RH for serious diseases than for mild diseases. However, our findings do not provide support for *H6* and *H8*. There is no evidence found to support the hypothesis that disease type moderates the influence of readability or emotion on the perceived RH. Readability and emotion both contribute equally to the perceived RH for different disease conditions. Furthermore, the results for the control variables are

consistent with earlier findings [28]. The coefficient (β) for age, title, education, and experience is significant for both disease conditions.

To maximize the practical value of our research, we used a text mining strategy to estimate the efficacy and performance of our suggested multimethod model by taking into account all of the signals from a review and service at the same time. For the purpose of testing different classification models, the data mining software Weka 3.8.5 was employed. Based on the number of helpful votes received by a review, we classify our reviews as “helpful” and “not helpful” groups and employ strategies to develop the classification model detailed in Table 4. In particular, a review is considered helpful if it receives at least one vote. The model estimation is performed using the training data and the model is validated using the holdout sample to see if the model is useful for physicians and to predict physician RH. This practice contributes to the avoidance of overfitting.

We classified PORs as helpful or not helpful based on review- and service-related attributes using well-known machine learning methods. The predictive models are built and 10-fold cross-validation is used to compare the accuracy of the predictive models across all experiments. The hybrid set of review- and service-related features (combination of readability, wordiness, emotions, quality, and popularity) is used to train four different learning algorithms. The results of experiments using a hybrid set of features are summarized in Table 4.

Using a hybrid set of features to predict physician RH, PORs dataset delivers 73.10% accuracy and 73.14% f-measure with SVM classifier. Next, using a hybrid set of review- and service-related features, the PORs dataset outputs 75.22% accuracy and 75.11% f-measure with the LR classifier, 81.13% accuracy and 81.15% f-measure with the RF, and 91.12% accuracy and 91.63% with GBDT classifier. The experimental results demonstrate that the overall performance of the model is rather promising, demonstrating the applicability of the suggested hybrid features in terms of accuracy and f-measure metrics for RH prediction. The proposed hybrid features model developed by this study clearly outperforms the other models, successfully classifying 91.12% of all cases correctly. A series of experiments are carried out with the machine learning classifier because it demonstrated the best performance compared to the models analyzed in the previous studies [16, 31, 40]. All of the evaluation parameters indicate that our suggested model performs well in terms of predicting physician RH.

6. Discussion

PRWs offer patients a place to talk about their healthcare experiences or write reviews online [80]. On social media networks, patients post thousands of reviews and remarks about their experiences, making it difficult to discern which reviews are useful and which are not. Therefore, it is vital to investigate characteristics and establish a study technique for implicitly classifying helpful reviews rapidly and reliably. For

TABLE 4: Measuring classification performance and comparing models.

Algorithms	Accuracy	Precision	Recall	<i>F</i> -measure	Previous models	Accuracy	Precision
Support vector machine	73.10	73.05	73.25	73.14	Lee et al. [31]	84.30%	—
Linear regression	75.22	74.12	76.19	75.11	Eslami et al. [16]	69.0%	—
Random forest	81.13	80.35	82.03	81.15	Zhang and Lin [40]	—	85.19%
Gradient boost decision tree	91.12	91.07	92.18	91.63	Proposed model	91.12%	—

this study, we collected a vast number of PORs from Healthgrades.com. A further investigation was carried out into the impact of three review- and two service-related signals on perceived RH. Both the review- and service-related signals on PRWs significantly influence perceived RH, in line with previous studies [10, 13, 23]. Inspired by earlier studies [25, 28], we examined the moderating influence of disease type in the connection between review- or service-related signals and RH. The impact of wordiness, quality, and popularity on the perception of RH is moderated by the disease type. Moreover, we used data mining approaches to construct multiple classification models for the assessment of RH. The results indicate that our proposed model performs exceptionally well at classifying and predicting physician RH.

The findings show that our proposed model for review- and service-related signals successfully influences perceived RH, which has various theoretical implications. From a theoretical point of view and to the best of our knowledge, this research is the first to explore the effects of review- and service-related signals on online physician RH.

By and large, these findings advance the state of the art in estimating helpfulness, which has historically concentrated on the local features of reviews [38, 47, 81]. The unique thing about our work is that we look at patient feedback in a wider user or service-related context, taking into account a broader set of determinants. Moreover, this study advances signaling theory by demonstrating that the signaling environment (disease type) has an effect on review- and service-related signals. The findings fill a gap in the existing knowledge by identifying the characteristics of the helpful POR across a wide range of disease conditions. The results revealed a significant moderating relationship between wordiness, quality, popularity, and perceived RH. The dichotomy between a review- or service-related signals and perceived RH is predicted to be interpreted by disease type. Our proposed model, which is predictive in nature, enables online healthcare providers to prioritize the most helpful reviews. Our research findings confirm the practical significance of the suggested approach by indicating the classification performance while predicting RH. Researchers also claim that this is the first attempt to use a multimethod approach including implicit and domain knowledge-based specialized emotion analysis methodologies to predict physician RH. This study demonstrates the critical relevance of RH as a future source of health informatics research.

The findings of this research have a number of implications in practice. For the designers of Internet health rating websites, our approach includes a recommendation mechanism that suggests useful reviews for patients on

PRW. It is possible to automatically identify the wants and requirements of patients while they explore the PRWs by searching for high to low levels of physician expertise, top-rated physicians, verified physicians, and selecting doctors based on their star ratings. We anticipate that this intelligent recommendation system will result in more helpful physician reviews based on the readability, depth, specific emotions, service quality, and popularity-related attributes of the service. Recommendation systems like this would save patients time and money when they use PRWs to find helpful reviews for providers they were about to visit. For healthcare providers, administrators and patients are able to view the useful assessments of their healthcare services on health rating websites and online forums. Individuals are eager to search useful/helpful reviews that might assist them in determining the worth of healthcare services or choosing the best physician they want to visit. Furthermore, when reading a vast collection of reviews, patients and their caregivers frequently experience substantial cognitive processing costs. The contribution of this study provides the opportunity for healthcare practitioners and administrators to reduce the cognitive processing costs associated with PORs in order to improve their organization.

This study has certain limitations. First, it is quite challenging to choose an acceptable dataset for analyzing the determinants affecting online physician RH, given only a small percentage of reviews receive a helpful vote from reviewers. Although it is well-known and ranks among the highest-trafficked platforms in the U.S., other platforms that allow users to post reviews should also be taken into account. Second, future research could incorporate reviewers' behavior and its effect on predicting physician RH. Finally, future studies into physician RH should use more advanced and efficient text mining algorithms, such as deep learning (Neural Networks).

Data Availability

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

Disclosure

This manuscript is the revised extended version of our manuscript "Exploring the Impact of Review and Service-related Signals on Online Physician Review Helpfulness: A Multi-Methods Approach" published in Pacific Asia Conference on Information Systems (PACIS) 2020, conference proceedings. However, more than 80% new material has been added to the previous conference version.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Dr. Adnan Muhammad Shah contributed to investigation, formal analysis, writing, reviewing, and editing the manuscript. Dr. Wazir Muhammad contributed to data curation, reviewing, and editing, and supervision. Dr. KangYoon Lee contributed to funding acquisition, methodology, and allocation of resources. All authors have read and approved the final version of the manuscript.

Acknowledgments

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2017-0-01630) supervised by the IITP (Institute for Information & Communications Technology Promotion) and was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1F1A1057663).

References

- [1] A. Shah, X. B. Yan, and S. A. Shah, "Tracking patients healthcare experiences during the covid-19 outbreak: topic modeling and sentiment analysis of doctor reviews," *Journal of Engineering Research*, vol. 9, no. 3A, 2021.
- [2] F. Rothenfluh and P. J. Schulz, "Physician rating websites: what aspects are important to identify a good doctor, and are patients capable of assessing them? A mixed-methods approach including physicians' and health care consumers' perspectives," *Journal of Medical Internet Research*, vol. 19, no. 5, p. e127, 2017.
- [3] A. M. Shah, X. Yan, S. A. A. Shah, and G. Mamirkulova, "Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 7, pp. 2925–2942, 2020.
- [4] Y. A. Hong, C. Liang, T. A. Radcliff, L. T. Wigfall, and R. L. Street, "What do patients say about doctors online? A systematic review of studies on patient online reviews," *Journal of Medical Internet Research*, vol. 21, no. 4, Article ID e12521, 2019.
- [5] A. M. Shah, X. Yan, S. J. Shah, and S. Khan, "Use of sentiment mining and online nmf for topic modeling through the analysis of patients online unstructured comments," in *Proceedings of the 6th International Conference for Smart Health (ICSH 2018)*, Wuhan, China, July, 2018.
- [6] P. J. Schulz and F. Rothenfluh, "Influence of health literacy on effects of patient rating websites: survey study using a hypothetical situation and fictitious doctors," *Journal of Medical Internet Research*, vol. 22, no. 4, Article ID e14134, 2020.
- [7] A. M. Shah, W. Muhammad, K. Lee, and R. A. Naqvi, "Examining different factors in web-based patients' decision-making process: systematic review on digital platforms for clinical decision support system," *International Journal of Environmental Research and Public Health*, vol. 18, no. 21, Article ID 11226, 2021.
- [8] N. Lu and H. Wu, "Exploring the impact of word-of-mouth about Physicians' service quality on patient choice based on online health communities," *BMC Medical Informatics and Decision Making*, vol. 16, no. 1, p. 151, 2016.
- [9] A. M. Shah, X. Yan, S. Khan, W. Khurram, and Q. R. Khan, "A multi-modal approach to predict the strength of doctor-patient relationships," *Multimedia Tools and Applications*, vol. 80, no. 15, Article ID 23207, 2021.
- [10] A. M. Shah, X. Yan, S. Khan, and S. J. Shah, "Exploring the impact of review and service-related signals on online physician review helpfulness: a multi-methods approach," in *Proceedings of the Twenty-Fourth Pacific Asia Conference on Information Systems*, Dubai, UAE, June, 2020.
- [11] J. P. Singh, S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. Kumar Roy, "Predicting the "helpfulness" of online consumer reviews," *Journal of Business Research*, vol. 70, pp. 346–355, 2017.
- [12] Y. H. Hu, K. Chen, and P. J. Lee, "The effect of user-controllable filters on the prediction of online hotel reviews," *Information & Management*, vol. 54, no. 6, pp. 728–744, 2017.
- [13] A. M. Shah, X. Y. Yan, S. J. Shah, and S. Khan, "Predicting the factors that influence helpfulness of online physician reviews," in *Proceedings of the 12th China Summer Workshop on Information Management*, Qingdao, Shandong China, June, 2018.
- [14] R. Filieri, F. McLeay, B. Tsui, and Z. Lin, "Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services," *Information & Management*, vol. 55, no. 8, pp. 956–970, 2018.
- [15] A. Qazi, K. B. Shah Syed, R. G. Raj, E. Cambria, M. Tahir, and D. Alghazzawi, "A concept-level approach to the analysis of online review helpfulness," *Computers in Human Behavior*, vol. 58, pp. 75–81, 2016.
- [16] S. P. Eslami, M. Ghasemaghaei, and K. Hassanein, "Which online reviews do consumers find most helpful? A multi-method investigation," *Decision Support Systems*, vol. 113, pp. 32–42, 2018.
- [17] B. Fang, Q. Ye, D. Kucukusta, and R. Law, "Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics," *Tourism Management*, vol. 52, pp. 498–506, 2016.
- [18] N. Mauro, L. Ardissono, and G. Petrone, "User and item-aware estimation of review helpfulness," *Information Processing & Management*, vol. 58, no. 1, Article ID 102434, 2021.
- [19] M. S. I. Malik and A. Hussain, "Helpfulness of product reviews as a function of discrete positive and negative emotions," *Computers in Human Behavior*, vol. 73, pp. 290–302, 2017.
- [20] G. Ren and T. Hong, "Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews," *Information Processing & Management*, vol. 56, no. 4, pp. 1425–1438, 2019.
- [21] H. Li, H. Liu, and Z. Zhang, "Online persuasion of review emotional intensity: a text mining analysis of restaurant reviews," *International Journal of Hospitality Management*, vol. 89, Article ID 102558, 2020.
- [22] A. Muhammad Shah and K. Lee, "The role of emotions intensity in helpfulness of online physician reviews," *Intelligent Automation & Soft Computing*, vol. 31, no. 3, pp. 1719–1735, 2022.
- [23] A. M. Shah, M. Ali, A. Qayyum et al., "Exploring the impact of linguistic signals transmission on patients' health consultation choice: web mining of online reviews," *International Journal of Environmental Research and Public Health*, vol. 18, no. 19, p. 9969, 2021.

- [24] A. M. Shah, X. Yan, S. Tariq, and M. Ali, "What patients like or dislike in physicians: analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach," *Information Processing & Management*, vol. 58, no. 3, Article ID 102516, 2021.
- [25] J. Li, J. Tang, D. C. Yen, and X. Liu, "Disease risk and its moderating effect on the E-consultation market offline and online signals," *Information Technology & People*, vol. 32, no. 4, pp. 1065–1084, 2019.
- [26] A. M. Shah, X. Yan, S. A. A. Shah, and R. Ullah, "Exploring important aspects of service quality while choosing a good doctor," *International Journal of Healthcare Information Systems and Informatics*, vol. 16, no. 4, pp. 1–23, 2021.
- [27] E. Cambria, D. Olsher, and D. Rajagopal, "Sentinet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Québec, Canada, July 2014.
- [28] A. M. Shah, X. Yan, S. A. A. Shah, S. J. Shah, and G. Mamirkulova, "Exploring the impact of online information signals in leveraging the economic returns of physicians," *Journal of Biomedical Informatics*, vol. 98, Article ID 103272, 2019.
- [29] X. Cao, Y. Liu, Z. Zhu, J. Hu, and X. Chen, "Online selection of a physician by patients: empirical study from elaboration likelihood perspective," *Computers in Human Behavior*, vol. 73, pp. 403–412, 2017.
- [30] S. Chen, X. Guo, T. Wu, and X. Ju, "Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis," *Information Processing & Management*, vol. 57, no. 5, Article ID 102253, 2020.
- [31] P. J. Lee, Y. H. Hu, and K. T. Lu, "Assessing the helpfulness of online hotel reviews: a classification-based approach," *Tele-matics and Informatics*, vol. 35, no. 2, pp. 436–445, 2018.
- [32] A. M. Shah, R. A. Naqvi, and O. R. Jeong, "Detecting topic and sentiment trends in physician rating websites: analysis of online reviews using 3-wave datasets," *International Journal of Environmental Research and Public Health*, vol. 18, no. 9, p. 4743, 2021.
- [33] Q. Chen, J. Jin, and X. Yan, "Understanding online review behaviors of patients in online health communities: an expectation-disconfirmation perspective," *Information Technology & People*, 2021.
- [34] A. M. Shah, R. A. Naqvi, and O. R. Jeong, "The impact of signals transmission on patients' choice through E-consultation websites: an econometric analysis of secondary datasets," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, p. 5192, 2021.
- [35] Y. Heng, Z. Gao, Y. Jiang, and X. Chen, "Exploring hidden factors behind online food shopping from amazon reviews: a topic mining approach," *Journal of Retailing and Consumer Services*, vol. 42, pp. 161–168, 2018.
- [36] A. Y. K. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," *Computers in Human Behavior*, vol. 54, pp. 547–554, 2016.
- [37] A. H. Huang, K. Chen, D. C. Yen, and T. P. Tran, "A study of factors that contribute to online review helpfulness," *Computers in Human Behavior*, vol. 48, pp. 17–27, 2015.
- [38] S. M. Mudambi and D. Schuff, "Research note: what makes a helpful online review? A study of customer reviews on Amazon.com," *MIS Quarterly*, vol. 34, no. 1, pp. 185–200, 2010.
- [39] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics," *Decision Support Systems*, vol. 81, pp. 30–40, 2016.
- [40] Y. Zhang and Z. Lin, "Predicting the helpfulness of online product reviews: a multilingual approach," *Electronic Commerce Research and Applications*, vol. 27, pp. 1–10, 2018.
- [41] M. Lee, M. Jeong, and J. Lee, "Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website," *International Journal of Contemporary Hospitality Management*, vol. 29, no. 2, pp. 762–783, 2017.
- [42] A. Felbermayr and A. Nanopoulos, "The role of emotions for the perceived usefulness in online customer reviews," *Journal of Interactive Marketing*, vol. 36, pp. 60–76, 2016.
- [43] N. Alodadi and L. Zhou, "Predicting the helpfulness of online physician reviews," in *Proceedings of the 2016 IEEE International Conference on Healthcare Informatics (ICHI)*, Chicago, IL, USA, October, 2016.
- [44] H. Yang, X. Guo, and T. Wu, "Exploring the influence of the online physician service delivery process on patient satisfaction," *Decision Support Systems*, vol. 78, pp. 113–121, 2015.
- [45] A. M. Spence, *Market signaling: informational transfer in hiring and related screening processes (harvard economic studies)*, Harvard University Press, Cambridge, MA, USA, 1974.
- [46] H. Li, Y. Fang, Y. Wang, K. H. Lim, and L. Liang, "Are all signals equal? Investigating the differential effects of online signals on the sales performance of E-marketplace sellers," *Information Technology & People*, vol. 28, no. 3, pp. 699–723, 2015.
- [47] M. Siering, J. Muntermann, and B. Rajagopalan, "Explaining and predicting online review helpfulness: the role of content and reviewer-related signals," *Decision Support Systems*, vol. 108, pp. 1–12, 2018.
- [48] B. L. Connelly, S. T. Certo, R. D. Ireland, and C. R. Reutzel, "Signaling theory: a review and assessment," *Journal of Management*, vol. 37, no. 1, pp. 39–67, 2011.
- [49] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498–1512, 2011.
- [50] Z. Liu and S. Park, "What makes a useful online review? Implication for travel product websites," *Tourism Management*, vol. 47, pp. 140–151, 2015.
- [51] X. Wang, L. Tang, and E. Kim, "More than words: do emotional content and linguistic style matching matter on restaurant review helpfulness?" *International Journal of Hospitality Management*, vol. 77, pp. 438–447, 2019.
- [52] S. N. Ahmad and M. Laroche, "How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis," *International Journal of Electronic Commerce*, vol. 20, no. 1, pp. 76–111, 2015.
- [53] A. M. Shah, X. Yan, A. Qayyum, R. A. Naqvi, and S. J. Shah, "Mining topic and sentiment dynamics in physician rating websites during the early wave of the covid-19 pandemic: machine learning approach," *International Journal of Medical Informatics*, vol. 149, Article ID 104434, 2021.
- [54] G. Craciun, W. Zhou, and Z. Shan, "Discrete emotions effects on electronic word-of-mouth helpfulness: the moderating role of reviewer gender and contextual emotional tone," *Decision Support Systems*, vol. 130, Article ID 113226, 2020.
- [55] Y. Liu and H.-f. Hu, "Online review helpfulness: the moderating effects of review comprehensiveness," *International Journal of Contemporary Hospitality Management*, vol. 33, no. 2, pp. 534–556, 2021.
- [56] H. Wu and N. Lu, "How your colleagues' reputation impact your patients' odds of posting experiences: evidence from an

- online health community,” *Electronic Commerce Research and Applications*, vol. 16, pp. 7–17, 2016.
- [57] G. Gao, B. N. Greenwood, B. N. Greenwood, R. Agarwal, and J. S. McCullough, “Vocal minority and silent majority: how do online ratings reflect population perceptions of quality?” *MIS Quarterly*, vol. 39, no. 3, pp. 565–589, 2015.
- [58] H. Yang, H. S. Du, and W. Shang, “Understanding the influence of professional status and service feedback on patients’ doctor choice in online healthcare markets,” *Internet Research*, vol. 31, no. 4, pp. 1236–1261, 2021.
- [59] L. Chen, A. Rai, and X. Guo, “Physicians’ online popularity and price premiums for online health consultations: a combined signaling theory and online feedback mechanisms explanation,” in *Proceedings of the 6th International Conference on Information Systems*, Fort Worth, TX, USA, December, 2015.
- [60] H. Wu and N. Lu, “Service provision, pricing, and patient satisfaction in online health communities,” *International Journal of Medical Informatics*, vol. 110, pp. 77–89, 2018.
- [61] Cdc, “Deaths and Mortality,” 2017, <https://www.cdc.gov/nchs/data/hus/2017/019.pdf>.
- [62] J. Wu, “Review popularity and review helpfulness: a model for user review effectiveness,” *Decision Support Systems*, vol. 97, pp. 92–103, 2017.
- [63] S. Lee, S. Lee, and H. Baek, “Does the dispersion of online review ratings affect review helpfulness?” *Computers in Human Behavior*, vol. 117, Article ID 106670, 2021.
- [64] B. Gao, N. Hu, and I. Bose, “Follow the herd or Be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews,” *Decision Support Systems*, vol. 95, pp. 1–11, 2017.
- [65] E. Cambria, “Affective computing and sentiment analysis,” *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.
- [66] L. A. Ramshaw and M. P. Marcus, “Text chunking using transformation-based learning,” in *Natural Language Processing Using Very Large Corpora*, S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann, and D. Yarowsky, Eds., Springer, Dordrecht, Netherlands, 1999.
- [67] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, “Sentic computing for patient centered applications,” in *Proceedings of the 10th International Conference on Signal Processing*, Beijing, China, October, 2010.
- [68] E. Cambria and A. Hussain, *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*, Springer, New York, NY, USA, 2015.
- [69] D. Rajagopal, E. Cambria, D. Olsher, and K. Kwok, “A graph-based approach to commonsense concept extraction and semantic similarity detection,” in *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, Brazil, May, 2013.
- [70] E. Cambria, P. Gastaldo, F. Bisio, and R. Zunino, “An elm-based model for affective analogical reasoning,” *Neurocomputing*, vol. 149, pp. 443–455, 2015.
- [71] C. Ekstrand and T. E. Carpenter, “Using a Tobit regression model to analyse risk factors for foot-pad dermatitis in commercially grown broilers,” *Preventive Veterinary Medicine*, vol. 37, no. 1, pp. 219–228, 1998.
- [72] D. N. Gujarati and D. C. Porter, *Basic Econometrics*, McGraw-Hill, New York, NY, USA, 2003.
- [73] R. D. McKelvey and W. Zavoina, “A statistical model for the analysis of ordinal level dependent variables,” *Journal of Mathematical Sociology*, vol. 4, no. 1, pp. 103–120, 1975.
- [74] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*, Nelson Education, Toronto, ON, Canada, 2016.
- [75] K. Hendricks and A. Sorensen, “Information and the skewness of music sales,” *Journal of Political Economy*, vol. 117, no. 2, pp. 324–369, 2009.
- [76] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, Berlin, Germany, 1999.
- [77] H. Tin Kam, “Random decision forests,” in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, Canada, August, 1995.
- [78] S. Calzavara, C. Lucchese, and G. Tolomei, “Adversarial training of gradient-boosted decision trees,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, November, 2019.
- [79] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis: A Global Perspective*, Pearson Prentice Hall, Upper Saddle River, NJ, USA, 2010.
- [80] A. M. Shah, X. Yan, S. Tariq, and S. Khan, “Listening to the patient voice: using a sentic computing model to evaluate physicians’ healthcare service quality for strategic planning in hospitals,” *Quality and Quantity*, vol. 55, no. 1, pp. 173–201, 2021.
- [81] H. Hong, D. Xu, G. A. Wang, and W. Fan, “Understanding the determinants of online review helpfulness: a meta-analytic investigation,” *Decision Support Systems*, vol. 102, pp. 1–11, 2017.

Research Article

Ensemble of Deep Learning Based Clinical Decision Support System for Chronic Kidney Disease Diagnosis in Medical Internet of Things Environment

Suliman A. Alsuhibany,¹ Sayed Abdel-Khalek,^{2,3} Ali Algarni,⁴ Aisha Fayomi,⁴ Deepak Gupta ,⁵ Vinay Kumar,⁶ and Romany F. Mansour ⁷

¹Department of Computer Science, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia

²Department of Mathematics, Faculty of Science, Taif University, Taif, Saudi Arabia

³Department of Mathematics, Faculty of Science, Sohag University, Sohag, Egypt

⁴Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

⁵Department of Computer Science & Engineering, Maharaja Agrasen Institute of Technology, Delhi, India

⁶Department of Computer Engineering and Application, GLA University, Mathura, Uttar Pradesh, India

⁷Department of Mathematics, Faculty of Science, New Valley University, El-Kharga 72511, Egypt

Correspondence should be addressed to Romany F. Mansour; romanyf@sci.nvu.edu.eg

Received 15 November 2021; Revised 9 December 2021; Accepted 16 December 2021; Published 27 December 2021

Academic Editor: Ahmed A. Abd El-Latif

Copyright © 2021 Suliman A. Alsuhibany 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.

Recently, Internet of Things (IoT) and cloud computing environments become commonly employed in several healthcare applications by the integration of monitoring things such as sensors and medical gadgets for observing remote patients. For availing of improved healthcare services, the huge count of data generated by IoT gadgets from the medicinal field can be investigated in the CC environment rather than relying on limited processing and storage resources. At the same time, earlier identification of chronic kidney disease (CKD) becomes essential to reduce the mortality rate significantly. This study develops an ensemble of deep learning based clinical decision support systems (EDL-CDSS) for CKD diagnosis in the IoT environment. The goal of the EDL-CDSS technique is to detect and classify different stages of CKD using the medical data collected by IoT devices and benchmark repositories. In addition, the EDL-CDSS technique involves the design of Adaptive Synthetic (ADASYN) technique for outlier detection process. Moreover, an ensemble of three models, namely, deep belief network (DBN), kernel extreme learning machine (KELM), and convolutional neural network with gated recurrent unit (CNN-GRU), are performed. Finally, quasi-oppositional butterfly optimization algorithm (QOBOA) is used for the hyperparameter tuning of the DBN and CNN-GRU models. A wide range of simulations was carried out and the outcomes are studied in terms of distinct measures. A brief outcomes analysis highlighted the supremacy of the EDL-CDSS technique on exiting approaches.

1. Introduction

Internet of Things (IoT) aim is to interconnect and develop the connected things by computer network. Instead of utilizing higher energy consumption gadgets like phones, tabs, and machines [1], currently, some objects like air conditioners and room fresh units are computed by microcontroller utilizing sensing devices and provide the experimental result almost embedded in regular devices. IoT

integrated with cloud computing (CC) method is highly beneficial while developing applications. A monitoring technique can be developed by incorporating cloud and IoT to monitor the infected people even if they are at distance that is highly employed by medical examiners [2]. Generally, IoT techniques keep on using cloud platforms for enhancing the efficacy in terms of data storage, programming abilities, computing, and utilization of resources efficiently. As well, cloud computing (CC) system attains superior experience

from IoT by extending range to manage real time and provide distinct services dynamically in nature. Integration of cloud platform and IoT serves higher efficiency when compared to other cloud based methods. Some regions where this incorporation is employed are armed forces, hospitality, banking sectors, and home appliances. Among different applications, healthcare and medicinal are some of the stimulating research which assists rapid growth in medicinal and sensing gadgets [3].

CC aimed at utilizing IoT devices to transmit patient data (unstructured, structured, and semistructured healthcare information) on cloud platforms for configuring big data of patients [4]. Consequently, CC technique in medical services assists stakeholders in managing patient medical records, retrieving big data of patients, application of disease predictions, telemedicine, etc. [5]. Recently, the entire world faces difficulties of public health problems of chronic diseases like CKD which subsequently increases healthcare costs [6]. Based on threats and increases in the costs of CKD treatment, particularly in developing nations, earlier CKD predictions become a major challenge for healthcare centres and physicians in this country. Currently, network technologies and biomedical sensors offer a wide-ranging development in the fields of IoT [7], as a scheme where the smart healthcare device with unique identifiers could be communicated and connected for receiving and sending essential healthcare multimedia data to earlier detection in severe conditions, for example, threatening chronic disease such as CKD [8].

For effective and accurate rereading of the patient's healthcare state, each essential healthcare parameter and data gathered by IoT sensor could be analyzed via ML methods in prediction model which is demonstrated as an efficient solution in earlier medical detection. Data mining approaches like classification approaches as effective tool is extensively employed in anomaly detection and disease prediction in extensive research as an efficient method [9]. Based on the developments of utilizing current IoT devices and biomedical sensors, several health tracking and smart medical care schemes were introduced [10]. Mostly, the present research aims at earlier detection of some chronic diseases including diabetes mellitus, heart disease, and CKD where several factors influencing chronic disease have been used. But, considering each essential characteristic needed for predicting diseases, performance of prediction method, and the execution time still remains a challenge [11].

This study presents an ensemble of deep learning based clinical decision support systems (EDL-CDSS) for CKD diagnosis in the IoT environment. The EDL-CDSS technique involves the design of Adaptive Synthetic (ADASYN) technique for outlier detection process. In addition, an ensemble of three models such as deep belief network (DBN), kernel extreme learning machine (KELM), and convolutional neural network with gated recurrent unit (CNN-GRU) takes place. Besides, quasi-oppositional butterfly optimization algorithm (QOBOA) is used for the hyperparameter tuning of the DBN and CNN-GRU models. To examine the improved CKD detection outcomes of the EDL-CDSS technique, an extensive experimental analysis is carried out.

2. Literature Review

This section offers a brief survey of recently developed CKD classification models. Arulanthu and Perumal [12] presented an optimum cloud and IoT based DSS for CKD detection. The presented technique utilizes SA based FS method with RMSProp Optimizer based LR method named SA-RMSPO-LR to categorize the presence of CKD from healthcare information. Noor et al. [13] introduced a primary healthcare scheme for the treatments of Chronic Renal Disease/CKD patient with the idea of IoT that consist of the nutrition of food mostly focusing on the levels of salt intake, patient log, activity level, water intake, monitoring the sleep pattern, etc. for providing necessary improvement required for the advancement of their healthcare status.

Arulanthu and Perumal [14] introduced an online medical decision support system (OMDSS) for predicting CKD. The proposed algorithm includes a group of phases, i.e., classification, data gathering, and preprocessing of healthcare information for CKD prediction. Additionally, the parameters of adaptive learning rate optimization, LR, and Adaptive Moment Estimation (Adam) methods were employed. Bhaskar and Manikandan [15] presented a novel sensing method for the automatic diagnosis of CKD. The salivary urea concentration is observed for detecting the diseases. A novel sensing method is presented for monitoring the urea stages in the saliva samples. Moreover, analyzing the raw signals attained from the sensor, they have executed a 1D-DL-CNN method that is integrated by an SVM classification. Abdelaziz et al. [16] designed a hybrid intelligent method for CKD prediction based cloud-IoT by utilizing 2 smart technologies that are NN and LR. NN is employed for predicting CKD. LR is utilized for determining crucial factors that influenced CKD.

Ma et al. [17] proposed a Heterogeneous Modified Artificial Neural Network (HMANN) for earlier segmentation, diagnosis, and detection of CKD failure on the IoMT environment. Moreover, the presented method is categorized as an SVM and MLP with a BP method. In Hosseinzadeh et al. [18], a diagnostic predictive method for CKD and its seriousness is presented which employs IoT multimedia dataset. Because the factors affecting CKD are very large and the amount of IoT multimedia information is generally large, choosing dissimilar features based on physician's medical experience and observation as well as earlier researches for CKD in various types of multimedia datasets is performed for assessing the performance measure of CKD predictions and its determination level through dissimilar classification methods.

3. The Proposed Model

The common structure of the presented method has been demonstrated in Figure 1. The major component in the presented method is benchmark CKD dataset, IoT devices embedded in the patients, patient healthcare records, cloud database server (CDS), security system, data collection module, disease prediction module, and CKD diagnosis. The wearable IoT healthcare devices are considered IoT devices that are

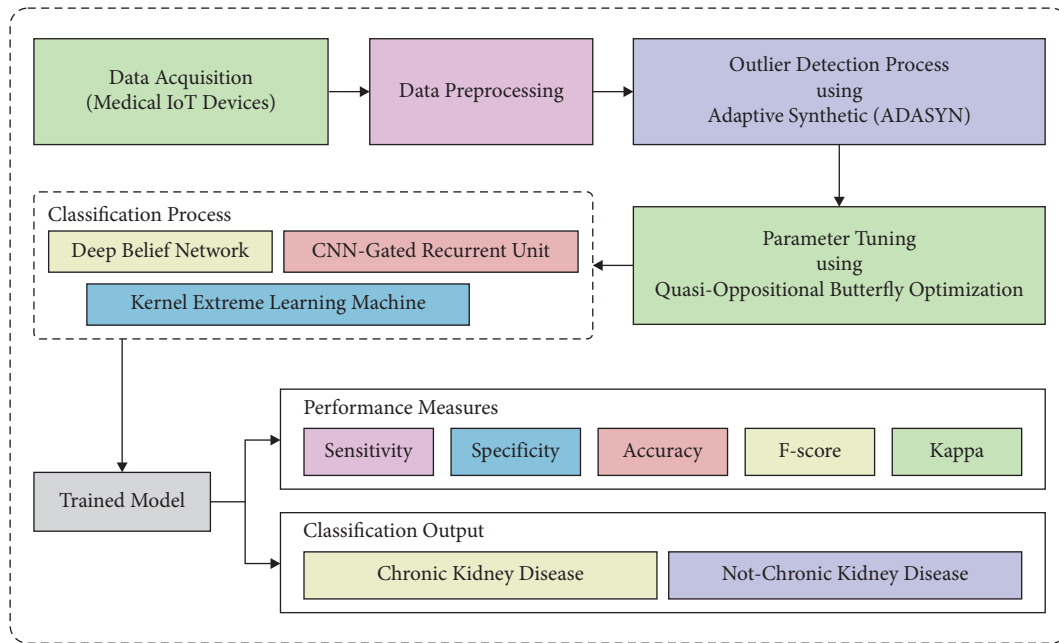


FIGURE 1: Overall process of EDL-CDSS technique.

located in the person's body. The standard CKD datasets from the UCI repository are employed. The healthcare datasets contain the previous information of the patient details that are collected from the clinics. This dataset is stored in the CDS. The security system recollects the information from the data collecting model. This information would be stored in a private way by using different stages of format conversion, information retrieval, and data integration. The secure information would be again stored in the CDS and it would be retrieved when it is needed. The disease prediction module and CKD diagnosis employ ensemble DL classifiers for the detection of CKD. In the training stage, the medical information from the CKD dataset and patient healthcare records are utilized for training the EDL-CDSS technique.

3.1. Data Acquisition. The presented architecture has 3 different types of information. In this phase, the medical information of the patients would be collected by using wearable IoT devices that are working by sensor nodes. The wearable devices are located on the patient's body to gather the certain patient's medical data frequently in a certain time period. Generally, the IoT devices in the human body check each sensed medical information either it is normal or not. The presented method uses 4G mobile network for transferring the sensed medical information to the CDS. Furthermore, the CKD dataset from UCI repository is applied for mapping the actual data that is created using the IoT gadgets. In addition, patients' healthcare details are applied for mapping the real information created using the individual patient information.

3.2. Data Preprocessing. For providing effectual efficiency with minimum cost to data mining procedures, the quality of data is optimum. The value missing from the database has

been filled under the entire CKD datasets. In few cases, if the continuous features exist, these techniques are synchronized for building discrete traits. It has any noisy and missing values from all samples. In order to improve the performance of medicinal information, a novel data was preprocessed.

3.3. ADASYN Based Outlier Detection. In this study, ADASYN is used to remove outliers, which is an extension of SMOTE. ADASYN has been found to be useful in medical imaging applications for detection of premature delivery, retinal health diagnosis, and diagnosis of focal liver lesions. Even though ADASYN is dependent on SMOTE, compared to borderline-SMOTE, ADASYN generates distinct synthetic instances for the minority class based on its distribution and for the borderline samples. Besides that, SMOTE offers equal opportunity for all the minority instances to get elected while in ADASYN the selection procedure is depending on the minority class distribution. The synthetic instance is generated according to the majority nearest neighbours through the KNN process. The method employs weighted dissemination for different minority class samples as per their complexity levels in training [19]. It generates additional synthetic instances for samples from minority class that is strenuous for training compared to those cases which is easy to train. The method initiates by evaluating the level of class imbalance. Then, it evaluates the amount of synthetic instances that need to be generated for the minority class by detecting KNN as per the Euclidean distance in n-dimension space. According to the ratio of density spread, the technique estimates the amount of synthetic data instances needed to be generated for the minority class.

3.4. *Ensemble Classification.* At this stage, the outliers removed data are fed into the ensemble classification model, which comprises three models, namely, KELM, DBN, and

CNN-GRU. These three feature vectors can be defined as follows:

$$\begin{aligned} f_{\text{KELM}_{1 \times n}} &= \{\text{KELM}_{1 \times 1}, \text{KELM}_{1 \times 2}, \text{KELM}_{1 \times 3}, \dots, \text{KELM}_{1 \times n}\}, \\ f_{\text{DBN}_{1 \times m}} &= \{\text{DBN}_{1 \times 1}, \text{DBN}_{1 \times 2}, \text{DBN}_{1 \times 3}, \dots, \text{DBN}_{1 \times m}\}, \\ f_{\text{CNN-GRU}_{1 \times l}} &= \{\text{CNN-GRU}_{1 \times 1}, \text{CNN-GRU}_{1 \times 2}, \text{CNN-GRU}_{1 \times 3}, \dots, \text{CNN-GRU}_{1 \times l}\}. \end{aligned} \quad (1)$$

In addition, the derived individual features are combined into a single vector, using the following equation:

$$\text{Fused (features vector)}_{1 \times q} = \sum_{i=1}^3 f_{\text{KELM}_{1 \times n}}, f_{\text{DBN}_{1 \times m}}, f_{\text{CNN-GRU}_{1 \times l}} \quad (2)$$

where f represents fused vectors (1×1186).

3.4.1. *KELM Model.* The output function of ELM in case of one output node is

$$f(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x) = \beta \cdot h(x), \quad (3)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ denotes the output weight vector. $G(a_i, b_i, x)$ indicates the output of i th hidden layer, as well as the node variable, is arbitrarily created. $h(x) = [G(a_1, b_1, x), \dots, G(a_L, b_L, x)]^T$ represent the output vector of hidden layer in relation to the input. Afterwards the kernel function can be determined by

$$\Omega_{\text{ELM}} = HH^T: \Omega_{\text{ELM}} = h(x_i) \cdot h(x_j) = K(x_i, x_j). \quad (4)$$

The output function of ELM classifiers is expressed by

$$\begin{aligned} f(x) &= h(x)H^T \left(\frac{I}{\lambda} + HH^T \right)^{-1} T \\ &= \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix} \left(\frac{I}{\lambda} + \Omega_{\text{ELM}} \right)^{-1} T, \end{aligned} \quad (5)$$

in which I represents the identity matrix, λ indicates the normalization coefficient, and T signifies the trained set label [20]. Afterwards using this model, we do not want to know the certain form of the feature map $h(x)$ but utilize the kernel function to resultant computation. Therefore, the arbitrarily generated bias and weights are evaded, and it is not necessary to set the amount of hidden layer L .

3.4.2. *DBN Model.* Since Hinton was developed in 2006, the DBN deployment is made up of stacked RBM. First, the network employs the Contrastive Divergence (CD)

method to unsupervised training of stacked RBM, later applying the BP method for fine-tuning the node parameter in the whole DBN network. Mainly, DBN training consists of fine-tuning and pretraining. The pretraining phase employs all the layers of RBM to implement unsupervised training on unlabeled sample data and simultaneously applies the CD method for tuning all the layers of RBM parameter. Afterwards, the pretraining DBN estimates the network error of all the layers by using the BP method and adjusts the parameters of all the layers by using BP method, for realizing the global fine-tuning of the node weight of the whole DBN network. The undirected graph method consists of hidden layer and visible layer. All the layers have multiple nodes. RBM is an energy based method, and energy function can be determined as hidden layer $h = (h_j)_m$ and visible layer $v = (v_i)_n$:

$$E_\theta(v, h) = - \sum_{i=0}^{n_v} a_i v_i - \sum_{j=0}^{n_h} b_j h_j - \sum_{i=0}^{n_v} \sum_{j=0}^{n_h} v_i w_{ij} h_j, \quad (6)$$

where a_i represents the bias of i th neurons from visible layer and b_j indicates the bias of j th neurons under the hidden layer; $\theta = [w = (w_{ij})_{n \times m}, a = (a_i)_n, b = (b_j)_m]$ signifies the parameter of RBM method; w_{ij} shows the connection weight among the hidden layer h_j and visible layer v_i ; nh denotes the amount of hidden layers and n_v is the amount of visible layers.

$$P_\theta(v, h) = \frac{1}{Z_\theta} \exp(-E_\theta(v, h)). \quad (7)$$

Among others, Z_θ represents the standardization factor as follows:

$$Z_\theta = \sum_v \sum_h \exp(-E_\theta(v, h)). \quad (8)$$

Once the v state of neurons over the visible layer is provided [21], the probabilities of j th neurons h_j from the hidden layer was initiated (by the probability 1) as

$$P_{\theta}(h_j = 1 | v) = \text{sigmoid} \left(b_j + \sum_{i=0}^{n_v} w_{ij} v_i \right). \quad (9)$$

Once the h state of neuron on hidden layer is provided, the probability that i th neuron v_i from the visible layer has initiated (by the probability 1) is

$$P_{\theta}(v_i = 1 | h) = \text{sigmoid} \left(a_i + \sum_{j=0}^{n_v} w_{ji} h_j \right), \quad (10)$$

in which, $\text{sigmoid}(x) = (1 + \exp(-x))^{-1}$ represent the activation function, whereas x is in the range of zero and one.

3.4.3. CNN-GRU Model. CNN is a multilayer neural network that is made up of multiple pooling, fully connected, and convolution layers, with the pooling and convolution layers. Furthermore, the FC layers are employed for classifying the CKD from the extracted feature map. All the layer inputs are interconnected to preceding layer output, and the outputs pass to following layer. The parameter of the networks is trained and shared by BP technique. The LSTM consists of update, input, and forget gates. In the meantime, the GRU is simple when compared to the LSTM because it has only 2 gates that are updated and reset gates. This gate is employed for determining either the data is beneficial or not. In GRU, W_z , W_r , and W imply the update gate, reset gate, and candidate data, correspondingly. Besides being fully connected with the neuron in the preceding hidden neuron at time t , all the neurons from GRU hidden layers are FC for each neuron under the present hidden layer at time $t - 1$. The l^{th} hidden layer output is calculated as follows:

$$h_t^l = (1 - z_t) * h_{t-1}^l + z_t * h_t^l, \quad (11)$$

where z_t represents an update gate, and h_t^l indicates the candidate memory data. Furthermore, z_t is provided by equation (12) which controls how many data of the preceding and present memory would be added/forgotten.

$$z_t = \text{sigmoid}(W_z \cdot [h_{t-1}^l, h_t^{l-1}]). \quad (12)$$

The candidate data value h_t^l is estimated as

$$h_t^l = \tanh(W \cdot [r_t * h_{t-1}^l, h_t^{l-1}]), \quad (13)$$

in which r_t determined by equation (15) is the GRU reset gate that effectively reset the data in the memory [22].

$$r_t = \text{sigmoid}(W_r \cdot [h_{t-1}^l, h_t^{l-1}]). \quad (14)$$

Sigmoid and tanh represent activating functions in the following:

$$\begin{aligned} \text{sigmoid}(x) &= \frac{1}{1 + e^{-x}}, \\ \text{tanh}(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}}. \end{aligned} \quad (15)$$

Compared to the CNN-FC layer, the estimated data by the GRU comprises historical state, wherein the neuron

value at time t can be described by the information in the preceding layer at time t , and it is defined by the information stored in the GRU cell at time t (equation (11)).

The GRU network replaces CNN-FC layer to convert the classification process as to sequential task, where the classification result of all the feature maps is included in the following feature map classification in the similar hidden layer for improving the CNN detection accuracy. The presented CNN-GRU framework has been demonstrated in Figure 2. In CNN-GRU architecture, the pooling, original CNN input, and convolution layers parameters with sizes are not changed to implement the features extraction. All the output feature maps of the l^{th} convolution layer are calculated by

$$O_{\text{conv}}^l = \text{sigmoid} \left(\sum_{i=1}^M \text{convn}(a_i^{l-1} * k_{ij}^l) + b_j^l \right), \quad (16)$$

in which a_i^{l-1} represent the i th feature maps in the overall feature maps M of the $l - 1^{\text{th}}$ preceding layer, and k_{ij}^l represents the l^{th} layer kernel. The values at the location of (m, n) in the convolutional process are determined by

$$\begin{aligned} \text{convn}(a_i^{l-1}, k_{ij}^l)[m][n] \\ = \sum_{w=0}^{k-1} \sum_{l=0}^{k-1} k_{ij}^l[w][l] * a_i^{l-1}[m+w][n+l], \end{aligned} \quad (17)$$

where $k = 5$ denotes the convolution kernel width. During pooling layer, all the output feature map values at the location of (m, n) are estimated by utilizing average pooling technique as follows, in which a_j indicates the j^{th} feature map of $l - 1$ convolution layer, and $K = 2$ denotes the pooling kernel width:

$$\text{aver_pool}_j^l[m][n] = \frac{1}{4} \sum_{l,w=0}^{K-1} a_j^{l-1}(m+l, n+w). \quad (18)$$

Next, each feature of all the feature maps in the last CNN pooling layers is interconnected to a respective GRU method. This implies that there are overall twelve GRU networks, and all the GRU network sets 3 layers, involving an output neuron of ten layers, an input neuron of twenty-five layers, and a hidden neuron of fifty layers. Lastly, the GRU output is activated using softmax function determined in equation (19) to categorize the CKD. In the testing stages, each GRU output has been chosen for finding the final detection outcome.

$$y_k = \frac{\exp(w_k h)}{\sum_k \exp(w_k \cdot h)}. \quad (19)$$

3.5. QOBOA Based Hyperparameter Optimization. To optimally tune the hyperparameter involved in the DBN and CNN-GRU models, the QOBOA is utilized and thereby boosts the overall CKD classification performance. BOA [23] is a population based nature simulated optimization method, depending on the food hunting system of

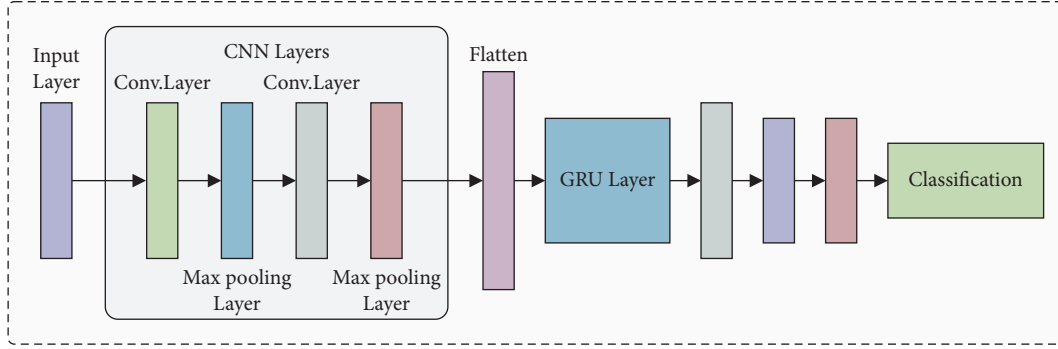


FIGURE 2: Framework of CNN-GRU model.

butterflies. Once butterflies move from one place to another, it releases a fragrance with intensity, i.e., transferred over the distance. The other butterflies could find this fragrance and be attracted to it according to the intensity level of fragrance. Once the butterfly senses the optimal butterfly fragrance it begins to move toward it. This procedure is called a local search. It produces fragrance with intensity once it begins to move. Another butterfly was attracted to it as per its level of fragrance. It can be determined by

$$f = c \times I^a, \quad (20)$$

where c represents a sensor modality, I indicates the fragrance level, and a denotes the degree of fragrance absorptions. The 2 main phases of the process are given in the following. Every butterfly releases fragrance once it starts moving and another butterfly is attracted to it as per its level of intensity of fragrance. This procedure is described in the following:

$$x_i(t+1) = x_i(t) + (\alpha^2 \times g^* - x_i(t)) \times f_i. \quad (21)$$

While $x_i(t)$ indicates a vector that denotes the butterfly (solution) at iteration t , g^* represents the total optimal solution, α shows an arbitrary value within $[0, 1]$, and f_i denotes a fragrance of j^{th} butterfly. Once the butterfly fails to find the fragrance of another butterfly, it could arbitrarily move in the searching space. The procedure is described in the following:

$$x_i(t+1) = x_i(t) + (\alpha^2 \times x_j(t) - x_k(t)) \times f_i. \quad (22)$$

In the equation, $x_j(t)$, $x_k(t)$ indicate 2 vectors that show two distinct butterflies in a similar population. A p switching probability was utilized in BOA for switching from general global search for intensive local search.

For improving the efficiency of the model based on optimal solution and convergence the concept of opposition based learning was presented in BOA [24]. A variables quasi-opposite value of a candidate solution is randomly considered among the mirror point of the parameter and the midpoint of the searching space. It can be expressed in the following equation:

$$\begin{aligned} x_{i,j}^q &= \text{rand}(a, b), \\ a &= \frac{x_{i,j}^{\min} + x_{i,j}^{\max}}{2}, \\ b &= x_{i,j}^{\min} + x_{i,j}^{\max} - x_{i,j}. \end{aligned} \quad (23)$$

Here, $x_{i,j}^q$ denotes the j^{th} dimension of j^{th} candidate solution; $x_{i,j}^{\min}$, $x_{i,j}^{\max}$ indicate the minimal and maximal values of $x_{i,j}$, and $x_{i,j}^q$ represents quasi-opposite value of $x_{i,j}$.

4. Results and Discussion

The EDL-CDSS technique is simulated using Python 3.6.5 tool and the results are examined using benchmark CKD dataset, which is publically available at https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease. The dataset includes 400 samples with 25 attributes. Among the available samples, 250 samples fall into CKD category and the residual of 150 samples come under Not-CKD category. The features involved in the CKD dataset are given in Figure 3.

Figure 4 shows the set of confusion matrices obtained by the EDL-CDSS technique on the test CKD data under five runs. The results show that the EDL-CDSS technique has classified the CKD and Not-CKD instances correctly. For instance, with run-1, the EDL-CDSS technique has identified 240 instances into CKD and 241 instances into Not-CKD. Meanwhile, with run-2, the EDL-CDSS manner has identified 244 instances into CKD and 240 instances into Not-CKD. Eventually, with run-3, the EDL-CDSS method has identified 244 instances into CKD and 238 instances into Not-CKD. Moreover, with run-4, the EDL-CDSS system has identified 242 instances into CKD and 241 instances into Not-CKD. Furthermore, with run-5, the EDL-CDSS manner has identified 240 instances into CKD and 243 instances into Not-CKD.

Table 1 offers the overall CKD classification performance of the EDL-CDSS technique under five runs. Figure 5 demonstrates the sens_y and spec_y analysis of the EDL-CDSS technique under five test runs. The figure reported that the EDL-CDSS technique has gained increased values of

S. No	Features	Datatype	Features Description	Unit of Measurement
1	Age	Numerical	Age	Years
2	bp	Numerical	Blood Pressure	mm/Hg
3	sg	Nominal	Specific Gravity	1.005, 1.010, 1.015, 1.020, 1.025
4	al	Nominal	Albumin	0, 1, 2, 3, 4, 5
5	su	Nominal	Sugar	0, 1, 2, 3, 4, 5
6	rbc	Nominal	Red Blood Cells	Normal, Abnormal
7	pc	Nominal	Pus Cell	Normal, Abnormal
8	pcc	Nominal	Pus Cell clumps	Present, Not_Present
9	ba	Nominal	Bacteria	Present, Not_Present
10	bgr	Numerical	Blood Glucose Random	mgs/dl
11	bu	Numerical	Blood Urea	mgs/dl
12	sc	Numerical	Serum Creatinine	mgs/dl
13	sod	Numerical	Sodium	mEq/L
14	pot	Numerical	Potassium	mEq/L
15	hemo	Numerical	Haemoglobin	gms
16	pcv	Numerical	Packed Cell Volume	0, 1, 2...
17	wbcc	Numerical	White Blood Cell Count	cells/cumm
18	rbcc	Numerical	Red Blood Cell Count	millions/cumm
19	htn	Nominal	Hypertension	Yes, No
20	dm	Nominal	Diabetes Mellitus	Yes, No
21	cad	Nominal	Coronary Artery Disease	Yes, No
22	appet	Nominal	Appetite	Good, Poor
23	pe	Nominal	Pedal Edema	Yes, No
24	ane	Nominal	Anemia	Yes, No
25	Class	Nominal	CKD, Not_CKD	CKD, Not_CKD

FIGURE 3: Attributes involved in CKD dataset.

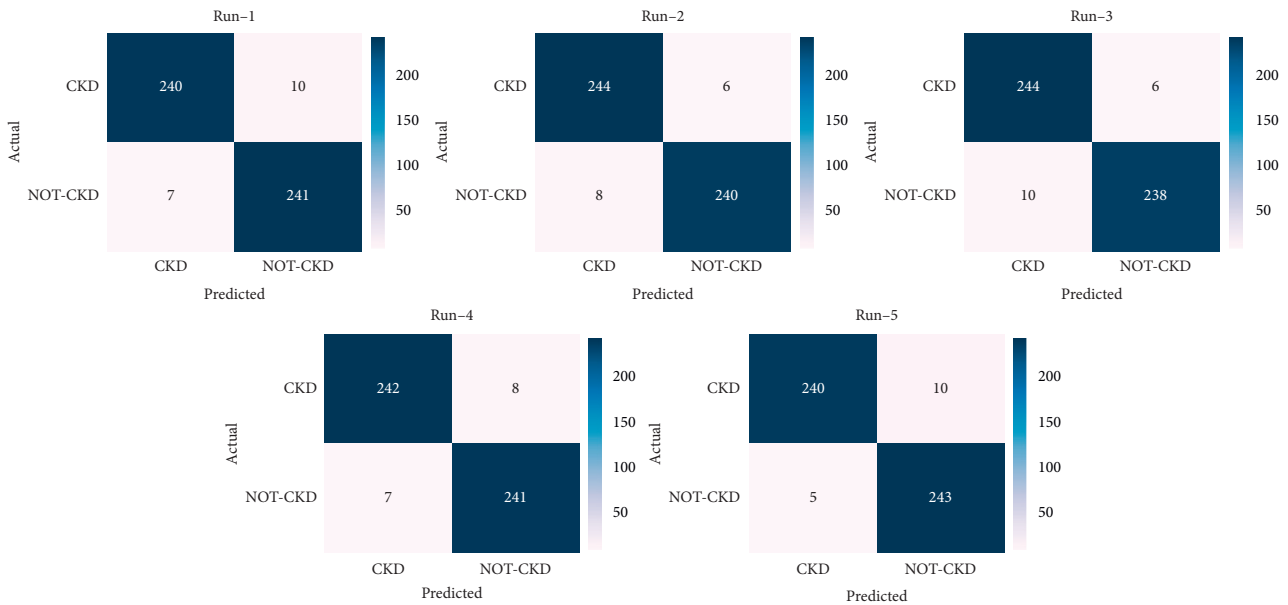


FIGURE 4: Confusion matrix analysis of EDL-CDSS technique with different runs.

TABLE 1: Result analysis of EDL-CDSS technique with five runs in terms of distinct measures.

No. of runs	Sensitivity	Specificity	Accuracy	F-score	Kappa
Run-1	0.9600	0.9718	0.9659	0.9658	0.9545
Run-2	0.9760	0.9677	0.9719	0.9721	0.9623
Run-3	0.9760	0.9597	0.9679	0.9683	0.9568
Run-4	0.9680	0.9718	0.9699	0.9699	0.9598
Run-5	0.9600	0.9798	0.9699	0.9697	0.9600
Average	0.9680	0.9702	0.9691	0.9692	0.9587

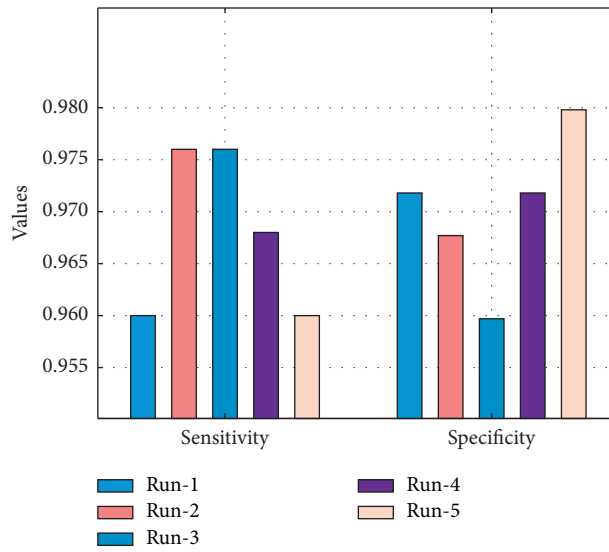


FIGURE 5: Sensitivity and specificity analysis of EDL-CDSS method with different runs.

$sens_y$ and $spec_y$. For instance, with run-1, the EDL-CDSS technique has attained $sens_y$ and $spec_y$ of 0.9600 and 0.9718, respectively. Besides, with run-3, the EDL-CDSS technique has resulted in $sens_y$ and $spec_y$ of 0.9760 and 0.9597, respectively. Lastly, with run-5, the EDL-CDSS technique has accomplished $sens_y$ and $spec_y$ of 0.9600 and 0.9797, respectively.

Figure 6 determines the F_{score} and kappa analysis of the EDL-CDSS system under five test runs. The figure stated that the EDL-CDSS approach has reached improved values of F_{score} and kappa. For instance, with run-1, the EDL-CDSS technique has gained F_{score} and kappa of 0.9658 and 0.9545 correspondingly. Also, with run-3, the EDL-CDSS system has resulted in F_{score} and kappa of 0.9683 and 0.9568 correspondingly. At last, with run-5, the EDL-CDSS methodology has accomplished F_{score} and kappa of 0.9697 and 0.9600, respectively.

Figure 7 inspects the $accu_y$ analysis of the EDL-CDSS technique on the test dataset. The results show that the EDL-CDSS technique has obtained improved CKD classification performance with the maximum $accu_y$ of 0.9659, 0.9719, 0.9679, 0.9699, and 0.9699 under five test runs, respectively.

The ROC analysis of the EDL-CDSS technique on test dataset is demonstrated in Figure 8. The figure obviously shows that the EDL-CDSS manner has the ability to accomplish increased classification performance with the maximal ROC of 99.7260.

Figure 9 depicts the accuracy analysis of the EDL-CDSS method on test dataset. The outcomes displayed that the EDL-CDSS algorithm has accomplished improved performance with increased training and validation accuracy. It can be noticed that the EDL-CDSS technique has reached increased validation accuracy over the training accuracy.

Figure 10 showcases the loss analysis of the EDL-CDSS approach on test dataset. The outcomes established that the EDL-CDSS system has resulted in a proficient outcome with the minimum training and validation loss. It can be stated that the EDL-CDSS system has offered lower validation loss over the training loss.

Finally, a detailed comparative results analysis of the EDL-CDSS manner with recent methods takes place in Table 2 [25]. A brief comparative $sens_y$ and $spec_y$ analysis of the EDL-CDSS approach with existing ones is provided in Figure 11. The figure reported that the DT model has obtained least performance with the $sens_y$ and $spec_y$ of 0.9060 and 0.8944. Along with that, the MLP system has reached slightly enhanced outcome with the $sens_y$ and $spec_y$ of 0.9251 and 0.9305. In line with that, the ACO, FNC, KELM, CNN-GRU, and D-ACO models have obtained moderately closer CKD classification performance with nearer values of $sens_y$ and $spec_y$. Though the DBN model has resulted in near optimal outcome with the $sens_y$ and $spec_y$ of 0.9618 and 0.9686, the proposed EDL-CDSS method has depicted the other approaches with the superior $sens_y$ and $spec_y$ of 0.9680 and 0.9702.

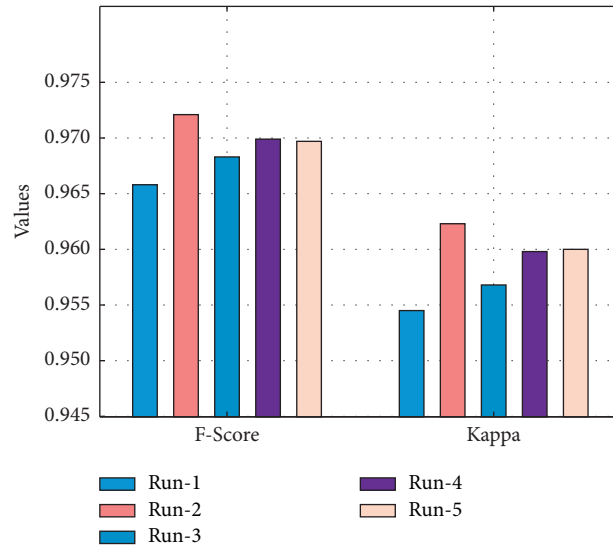


FIGURE 6: F-score and kappa analysis of EDL-CDSS method with different runs.

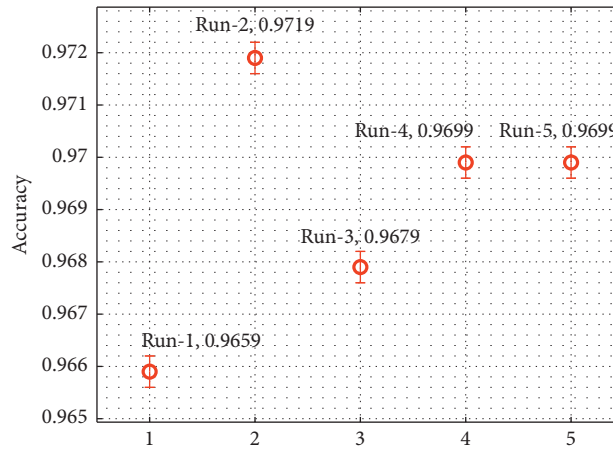


FIGURE 7: Accuracy analysis of EDL-CDSS method with different runs.

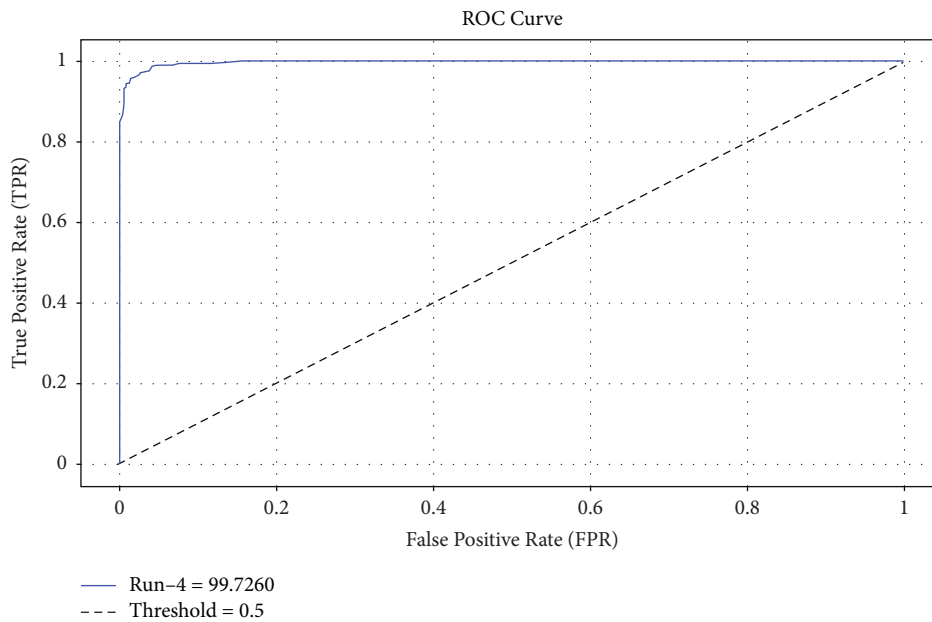


FIGURE 8: ROC analysis of EDL-CDSS method.

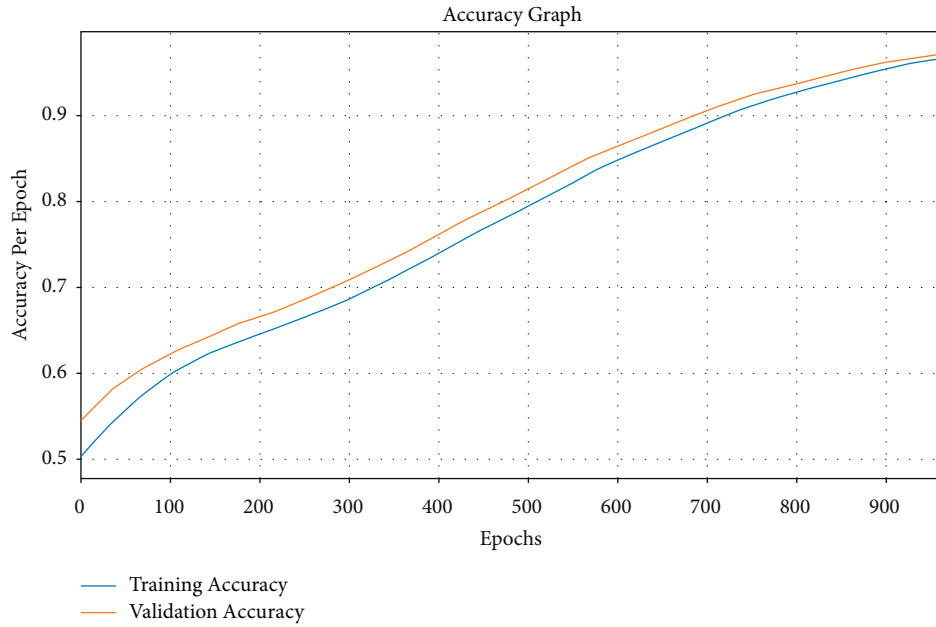


FIGURE 9: Accuracy graph analysis of EDL-CDSS method.

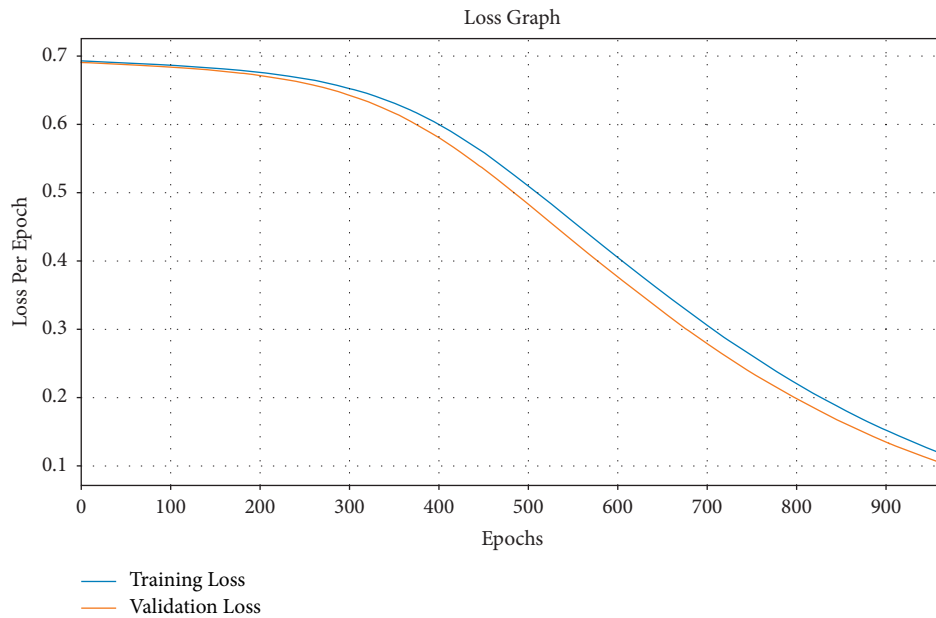


FIGURE 10: Loss graph analysis of EDL-CDSS method.

A detailed comparative acc_y and F_{score} analysis of the EDL-CDSS algorithm with existing ones is offered in Figure 12. The figure stated that the ACO approach has attained worse performance with the acc_y and F_{score} of 0.8778 and 0.9073. Besides, the DT technique has achieved somewhat increased results with the acc_y and F_{score} of 0.9026 and 0.9241. Likewise, the MLP, D-ACO, FNC, KELM, and CNN-GRU methodologies have reached moderately closer CKD classification performance with closer values of acc_y and F_{score} . But the DBN technique has

resulted in near optimum outcome with the acc_y and F_{score} of 0.9643 and 0.9651; the projected EDL-CDSS system has portrayed the other approaches with higher acc_y and F_{score} of 0.9691 and 0.9692.

After examining the above mentioned tables and figures, it can be ensured that the EDL-CDSS method has the capability of proficiently detecting the presence of CKD in the IoT environment. The enhanced performance of the proposed model is due to the inclusion of ensemble classification and QOBOA based hyperparameter tuning.

TABLE 2: Comparative analysis of various classifiers on CKD dataset.

Models	Performance measures			
	Sensitivity	Specificity	Accuracy	F-score
EDL-CDSS	0.9680	0.9702	0.9691	0.9692
DBN model	0.9618	0.9686	0.9643	0.9651
CNN-GRU model	0.9601	0.9653	0.9627	0.9659
KELM model	0.9599	0.9624	0.9611	0.9660
FNC model	0.9597	0.9612	0.9605	0.9665
D-ACO	0.9614	0.9354	0.9530	0.9620
MLP classifier	0.9251	0.9305	0.9265	0.9424
Decision tree	0.9060	0.8944	0.9026	0.9241
ACO	0.8910	0.8487	0.8778	0.9073

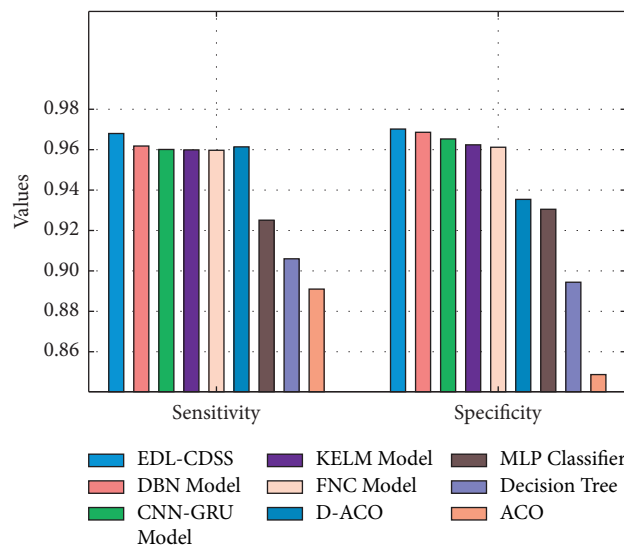


FIGURE 11: Comparative analysis of EDL-CDSS technique with existing methods.

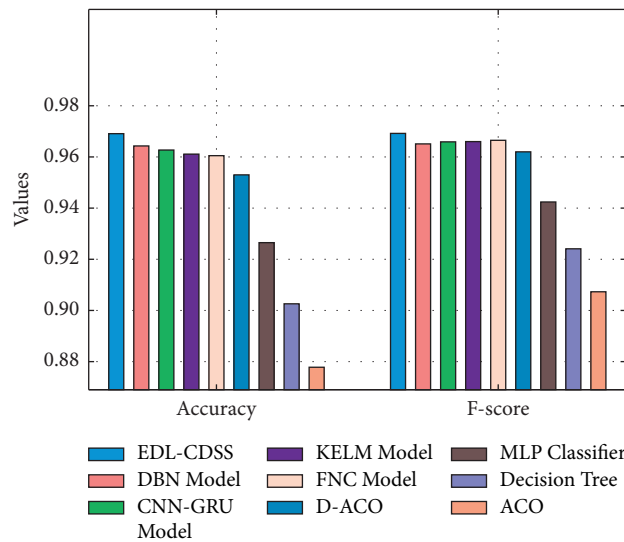


FIGURE 12: Accuracy and F-score analysis of EDL-CDSS technique with existing methods.

5. Conclusion

In this study, a novel EDL-CDSS approach was derived for CKD detection and classification in the IoT enabled cloud environment. The proposed EDL-CDSS technique encompasses distinct stages of operations, namely, data gathering preprocessing, ADASYN based outlier detection, ensemble classification, and QOBOA based hyperparameter tuning. For classification process, KELM, DBN, and CNN-GRU models are employed in which the hyperparameters of the DBN and CNN-GRU models are optimally tuned by the use of QOBOA. To examine the improved CKD detection outcomes of the EDL-CDSS approach, a huge range of simulations are implemented and the outcomes are studied with respect to various measures. A detailed comparative results analysis highlighted the supremacy of the EDL-CDSS manner on the existing approaches. Therefore, the proposed EDL-CDSS technique can be utilized as an effective tool for CKD diagnosis. In future, the proposed model can be applied to detect other diseases such as heart disease, diabetes, etc.

Data Availability

The dataset used in this paper is publicly available via the following link: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease.

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors.

Consent

Not applicable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The manuscript was written through contributions of all authors. All the authors have given approval to the final version of the manuscript.

Acknowledgments

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under Grant no. KEP-PhD-48-130-38. The authors therefore acknowledge with thanks DSR technical and financial support.

References

- [1] K. Velu, P. Arulanthu, and E. Perumal, *Energy reduction stratagem in smart homes using association rule mining*, in *Proceedings of the International Conference on Innovative*

Data Communication Technologies and Application, Springer, pp. 188–193, Cham, Switzerland, 2019.

- [2] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [3] J. R. Lambert, P. Arulanthu, and E. Perumal, "An effect of machine learning based classification algorithms on chronic kidney disease," *International Journal of Innovative Technology Exploring Engineering*, vol. 9, no. 3, pp. 2250–2256, 2020.
- [4] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, "An edge computing based smart healthcare framework for resource management," *Sensors*, vol. 18, no. 12, pp. 1–22, 2018.
- [5] R. Hamza, Zheng Yan, Khan Muhammad, Paolo Bellavista, and Faiza Titouna, "A privacy-preserving cryptosystem for IoT E-healthcare," *Information Science*, vol. 527, pp. 493–510, 2020.
- [6] P. Asghari, A. M. Rahmani, and H. Haj Seyyed Javadi, "A medical monitoring scheme and health-medical service composition model in cloud-based IoT platform," *Transactions Emerging Telecommunication Technologies*, vol. 30, no. 6, pp. 1–15, 2019.
- [7] P. Kaur, R. Kumar, and M. Kumar, "A healthcare monitoring system using random forest and internet of things (IoT)," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19905–19916, 2019.
- [8] A. Hussain, R. Wenbi, A. L. da Silva, M. Nadher, and M. Mudhish, "Health and emergency-care platform for the elderly and disabled people in the smart city," *Journal of Systems and Software*, vol. 110, pp. 253–263, 2015.
- [9] S. Tuli, N. Basumatary, S. S. Gill et al., "HealthFog: an ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments," *Future Generation Computer Systems*, vol. 104, pp. 187–200, 2020.
- [10] M. Hamim, S. Paul, S. I. Hoque, 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)*, January 2019.
- [11] A. Shrivastava, S. K. Sahu, and H. Hota, "Classification of chronic kidney disease with proposed union based feature selection technique," in *Proceedings of the 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, India, March, 2018.
- [12] P. Arulanthu and E. Perumal, *Intelligent Chronic Kidney Disease Diagnosis System Using Cloud Centric Optimal Feature Subset Selection with Novel Data Classification Model*, 2021.
- [13] A. Noor, A. Banerjee, M. F. Ahmad, and M. N. Uddin, "An IoT based mhealth platform for chronic kidney disease patients," in *Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1–6, IEEE, Dhaka, Bangladesh, May 2019.
- [14] P. Arulanthu and E. Perumal, "An intelligent IoT with cloud centric medical decision support system for chronic kidney disease prediction," *International Journal of Imaging Systems and Technology*, vol. 30, no. 3, pp. 815–827, 2020.
- [15] N. Bhaskar and S. Manikandan, "A deep-learning-based system for automated sensing of chronic kidney disease," *IEEE Sensors Letters*, vol. 3, no. 10, pp. 1–4, 2019.
- [16] A. Abdelaziz, A. S. Salama, A. M. Riad, and A. N. Mahmoud, "A machine learning model for predicting of chronic kidney disease based internet of things and cloud computing in smart

- cities,” in *Security in Smart Cities: Models, Applications, and Challenges*, pp. 93–114, Springer, Berlin, Germany, 2019.
- [17] F. Ma, T. Sun, L. Liu, and H. Jing, “Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network,” *Future Generation Computer Systems*, vol. 111, pp. 17–26, 2020.
- [18] M. Hosseinzadeh, J. Koochpayehzadeh, A. O. Bali et al., “A diagnostic prediction model for chronic kidney disease in internet of things platform,” *Multimedia Tools and Applications*, vol. 80, no. 11, pp. 16933–16950, 2021.
- [19] B. Abraham and M. S. Nair, “Computer-aided diagnosis of clinically significant prostate cancer from MRI images using sparse autoencoder and random forest classifier,” *Bio-cybernetics and Biomedical Engineering*, vol. 38, no. 3, pp. 733–744, 2018.
- [20] J. Li, C. Hai, Z. Feng, and G. Li, “A transformer fault diagnosis method based on parameters optimization of hybrid kernel extreme learning machine,” *IEEE Access*, vol. 9, pp. 126891–126902, 2021.
- [21] Z. Wang, Y. Zeng, Y. Liu, and D. Li, “Deep belief network integrating improved kernel-based extreme learning machine for network intrusion detection,” *IEEE Access*, vol. 9, pp. 16062–16091, 2021.
- [22] V. Nguyen, J. Cai, and J. Chu, “Hybrid CNN-GRU model for high efficient handwritten digit recognition,” in *Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition*, pp. 66–71, Beijing, China, August 2019.
- [23] S. Arora and S. Singh, “Butterfly optimization algorithm: a novel approach for global optimization,” *Soft Computing*, vol. 23, no. 3, pp. 715–734, 2019.
- [24] N. R. Maliseti and V. K. Pamula, “Performance of quasi oppositional butterfly optimization algorithm for cluster head selection in WSNs,” *Procedia Computer Science*, vol. 171, pp. 1953–1960, 2020.
- [25] M. Elhoseny, K. Shankar, and J. Uthayakumar, “Intelligent diagnostic prediction and classification system for chronic kidney disease,” *Nature Scientific Reports*, vol. 9, 2019.

Research Article

CNN-Based Personal Identification System Using Resting State Electroencephalography

Yongdong Fan , Xiaoyu Shi , and Qiong Li 

School of Cyberspace Science, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Qiong Li; qiongli@hit.edu.cn

Received 3 September 2021; Accepted 16 November 2021; Published 13 December 2021

Academic Editor: Yassine Maleh

Copyright © 2021 Yongdong Fan 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 a biometric characteristic, electroencephalography (EEG) signals have the advantages of being hard to steal and easy to detect liveness, which attract researchers to study EEG-based personal identification technique. Among different EEG protocols, resting state signals are the most practical option since it is more convenient to operate than the other protocols. In this paper, a personal identification system based on resting state EEG is proposed, in which data augmentation and convolutional neural network are combined. The cross-validation is performed on a public database of 109 subjects. The experimental results show that when only 14 EEG channels and 0.5 seconds data are employed, the average accuracy and average equal error rate of the system can reach 99.32% and 0.18%, respectively. Compared with some existing representative works, the proposed system has the advantages of short acquisition time, low computational complexity, and rapid deployment using market available low-cost EEG sensors, which further advances the implementation of practical EEG-based identification systems.

1. Introduction

As society enters digital era, identification has become vital in people's work and life. Traditional identification technologies, such as password and hardware token, may be forgotten, lost, or stolen, resulting in identity leakage or identification failure [1]. Such problems can be avoided by using biometric identification techniques, such as face, fingerprints, and gait, which have been widely studied [2–4]. However, face image is easy to be captured, fingerprint may attach to the surface of many objects unconsciously [5, 6], and gait can be recorded and analysed unknowingly, which may be exploited by malicious attacks. In addition, liveness detection is not easy to achieve for these biometrics. The brain signal represented by electroencephalography (EEG) based biometric technique may solve such problems and has become a prominent personal identification method [7].

EEG is a noninvasive imaging technique that records brain electrical signals generated by neurons. First, EEG acquisition requires special measuring devices and electrodes are placed on the surface of the scalp of subjects. When electrodes fail to touch the skin surface of the brain, the

quality of EEG signals degrades rapidly [8], which increases the challenge to steal EEG. Second, the brain is one of the best protected organs in the human body, so brain biometrics are not easy to be damaged by external factors. Third, if subjects died, their brain electrophysiological signals could not be generated anymore, so EEG is one of the main clinical indicators for detecting brain death [9]. Finally, brain biometrics can be elicited by numerous distinct brain systems, which makes it possible to change the stored EEG characteristics using a distinct form of brain activity and response [1]. In conclusion, EEG has the advantages of being difficult to steal and damage, easy to detect liveness, and replaceability. Meanwhile, low-cost EEG sensor systems provide an opportunity to implement practical EEG-based identification systems. Compared with medical-grade sensor systems (e.g., Neuroscan 64-channel system, as illustrated in Figure 1(a)), low-cost sensor systems (e.g., EMOTIV EPOC [10], as illustrated in Figure 1(b)) has a gap in the accuracy [11], but their smaller sizes make them more convenient to wear and more acceptable to users. Besides, low-cost sensor systems help to increase the size of subjects and thus avoid system performance failures caused by a lack of sample diversity [12].

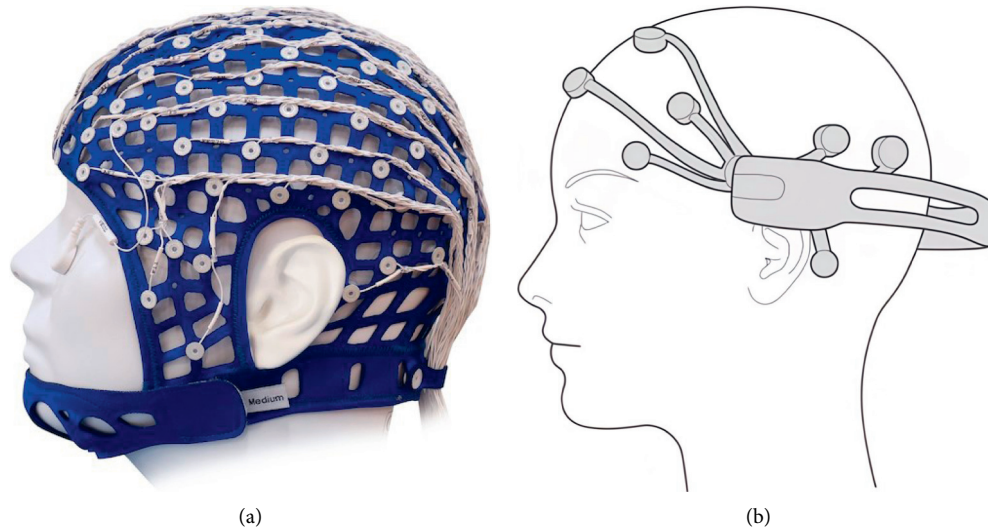


FIGURE 1: Typical medical-grade EEG sensor and low-cost EEG sensor. (a) Typical medical-grade sensor, Neuroscan 64-channel Quick cap and (b) typical low-cost sensor, EMOTIV EPOC.

The previously proposed identification schemes based on EEG can be categorized into three groups according to EEG protocols: resting states, cognitive tasks, and tasks with external stimuli [13]. In resting states [14], subjects are instructed to relax completely in a quiet environment, and EEG signals of eyes-open or eyes-closed are recorded. In cognitive tasks, such as motor imagery [15], mental workload [16, 17], and driving fatigue [18], subjects normally need to be trained and are required to complete specific tasks according to external cues while collecting their EEG signals [19]. In the tasks evoked by external stimulation, such as visual evoked potential [20] and acoustic stimuli [21], some additional devices are usually necessary to create and collect the appropriate stimulation. Compared with the other two tasks, resting states basically do not need training for subjects and are user-friendly, which has been favored by researchers.

In the present identification study of resting state EEG, most schemes are based on feature extraction. D. L. Rocca et al. [22] proposed a novel approach that focused on spectral coherence-based connectivity between different brain regions and used a Mahalanobis distance-based classifier to identify 10-second EEG signals in 2014. M. Frascini et al. [23] proposed a scheme that Phase Lag Index was used to calculate a weighted connectivity matrix; then, the nodal eigenvector centrality was computed, and finally genuine-impostor matching scores were computed to identify 12-second EEG signals in 2015. M. Garau et al. [24] proposed the fusion of the above two by feature-level and matching scores-level approaches in 2016, in which an equal error rate of 1.42% was achieved on 12-second eyes-open EEG signals. J.-H. Kang et al. [25] combined 10 single-channel features (seven spectral and three nonlinear) and 10 multichannel features by conducting network analysis into a set of EEG features, and finally a distance-based classifier for authentication was built in 2018. With the rise of deep learning, T. Schons et al. [26] applied CNN to learn the

features of resting state EEG in 2018, in which a sliding window of 12 seconds with a stride of 0.125 seconds was performed. The above scheme has two disadvantages in practical applications. (1) The acquisition time is mostly 10 or 12 seconds, which is too long in real-time identification [27]. (2) Recording 64-channel data relies on medical-grade sensors, which have difficulty in user acceptance and cost compared to low-cost sensors. To solve the above problems, Y. Sun et al. [27] proposed a system based on 1D-Convolutional Long Short-Term Memory Neural Network (1D-Convolutional LSTM) in 2019, which allowed only 16-channel EEG signals and 1-second acquisition time. However, this scheme also has two problems. (1) Introducing LSTM into the network will inevitably increase the computational complexity, which has disadvantages in training time and model loading time [27]. (2) Empirically selected channels cannot match market available low-cost sensors, which is not ideal to implement practical systems.

Recently, data augmentation is increasingly used with EEG, which promises to increase the accuracy and stability of EEG classification [28]. Data augmentation generates new samples by transforming existing samples, including noise addition, sampling, recombination of segmentation, Generative Adversarial Network, and so on. For resting state EEG, the commonly used data augmentation algorithm is sliding window, as depicted in Figure 2(a). Since there are no trigger signals in resting state EEG, the fixed window is normally applied to segment the data along the time boundary to generate training samples, as depicted in Figure 2(b). For example, fixed windows of 12 seconds and 1 second were, respectively, used in literature [23, 27]. The sliding window is a generalization form of the fixed window, which includes two attributes window length and stride, and is also suitable for sample segmentation of resting state EEG. For example, data augmentation based on a sliding window of 12 seconds with a stride of 0.125 seconds was implemented in literature [26]. Compared with the fixed window,

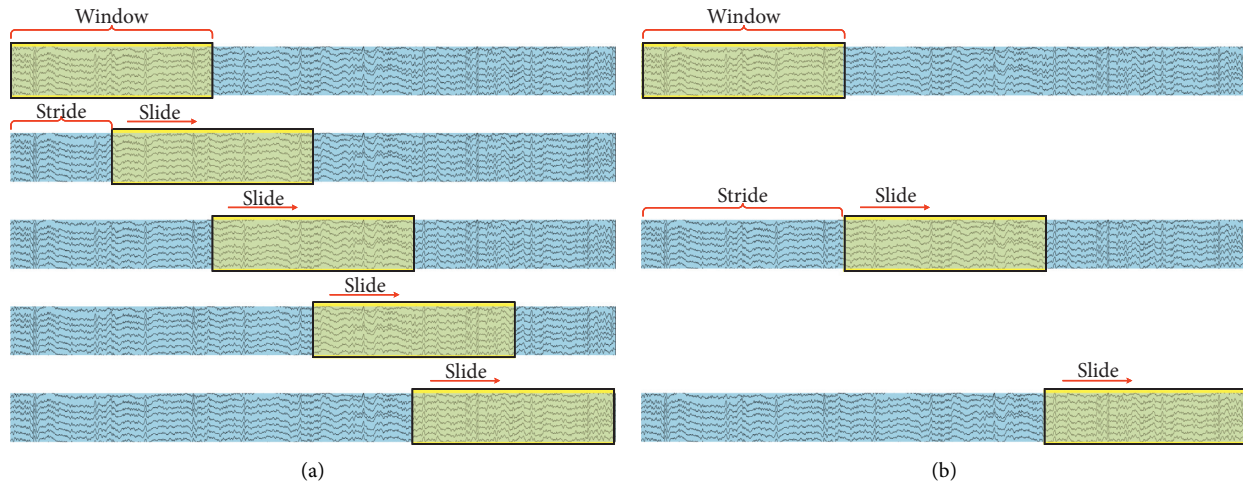


FIGURE 2: The sliding window and the fixed window. (a) The sliding window and (b) the fixed window.

the sliding window approach creates more samples. However, the present studies on personal identification have not reported the influence of sliding window on the performance.

For identification applications based on resting state EEG, a personal identification system using the CNN model (referred to below as ICAConvNet) is proposed, which applies a sliding window of 0.5 seconds for data augmentation and is validated on the PhysioNet dataset. Experimental results show that the sliding window is effective. When only 14 channels are used, the average Rank-1 accuracy is 99.32% and the average equal error rate can be as low as 0.18%, the performance of which is close to 64 channels. In summary, the proposed system has the advantages of short acquisition time, low computational complexity, and rapid deployment using market available low-cost sensors.

The rest of this paper is arranged as follows: the detailed research methods of the proposed system are introduced in Section 2, including dataset, preprocessing, data augmentation, network architecture, and experimental setup; the experimental results and discussion are given in Section 3; some conclusions are drawn in Section 4.

2. Methodology

2.1. Overview. Figure 3 shows the overview of the proposed identification system based on resting state EEG. First, in preparation stage, the resting state EEG data are preprocessed and augmented and then are divided into training and testing sets. Second, in enrollment stage, the training sets are trained by ICAConvNet, and the resting state EEG characteristics of all subjects are learned and stored in the system. Finally, in identification stage, test samples are identified by the trained network model in turn, and predicted identities are output. It is worth mentioning that it requires only 0.5-second resting state EEG to achieve rapid identification after enrollment stage.

2.2. Data Preprocessing. In order to preserve the original information and learn EEG features as much as possible, the filtering operations are not performed in data preprocessing.

EEG data are commonly a multichannel time series with several tens or even hundreds of sampling electrodes, which are a two-dimensional matrix data structure. Since the magnitude of EEG signals is usually small, in order to avoid gradient explosion and improve the convergence rate in deep learning, Z-score standardization is performed before neural network training, where the mean μ and standard deviation σ of all signals are calculated for each subject separately, and then a scaling is executed as indicated in the following equation:

$$\text{Output}_{i,j} = \frac{\text{Input}_{i,j} - \mu}{\sigma}, \quad (1)$$

where i , j , μ , and σ refer to the channel, the position in the time dimension, the mean, and standard deviation of all signals, respectively.

2.3. Data Augmentation Based on Sliding Windows. The segmented EEG samples for network input can be viewed as a two-dimensional matrix of Channels \times Points. In our work, a sliding window of 0.5 seconds with a stride of 0.25 seconds is adopted, and the data sampling rate is 160 Hz, so Channels and Points are set to 64 and 80, respectively. In the experimental section, the impact of sliding windows using different lengths and various strides on system performance will be discussed. Sample segmentation based on sliding window is provided in Algorithm 1.

2.4. Neural Network Architecture. Independent component analysis (ICA) [29, 30] is applicable to the problem of blind source separation and is widely used in the analysis of brain signals. Therefore, the observed EEG signals can be separated by ICA, and each separated signal may provide certain identity features. ICA algorithm is based on the following assumptions: the observed matrix X is linearly weighted by the independent

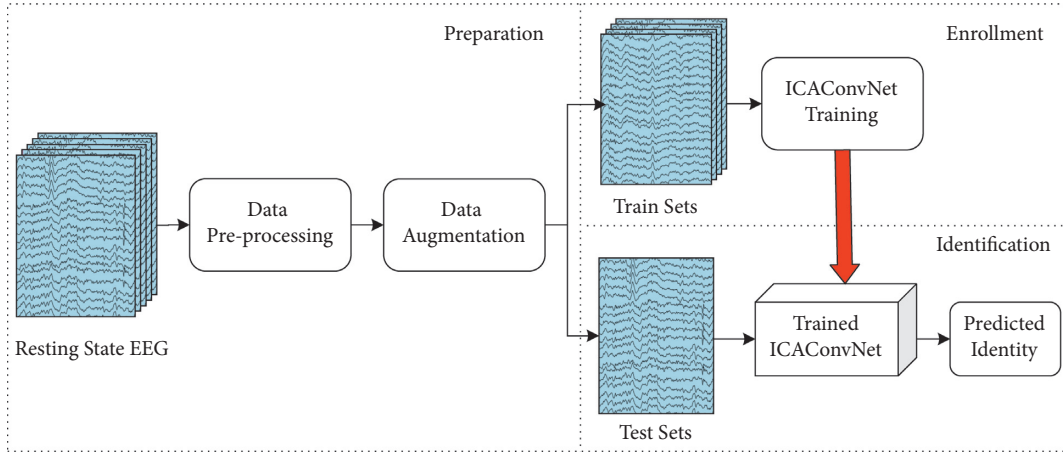


FIGURE 3: Overview of the proposed personal identification system.

```

Input: resting state EEG data  $D$  that are a two-dimensional matrix of channels  $\times$  points, window length  $L$ , stride  $S$ 
Output:  $Samples$ 
(1)  $Samples = \emptyset$ ;
(2)  $index = 0$ ;
(3) While  $index + L \leq len(D.Points)$  do
(4)  $sample_{start} = index$ ;
(5)  $sample_{end} = index + L$ ;
(6)  $Samples.append(D[:, sample_{start}: sample_{end}])$ ; //Create a sample
(7)  $index = index + S$ ;
(8) Return  $Samples$ ;

```

ALGORITHM 1: Sample segmentation based on sliding windows.

component matrix S and the mixed matrix A , as given in (2). The goal of ICA is to obtain a separation matrix W according to X so that the signal matrix Y obtained by W acting on X is the optimal approximation of the independent component matrix S , as expressed in the following equation:

$$X = AS, \quad (2)$$

$$\begin{aligned} Y &= WX = WAS, \\ A &= inverse(W). \end{aligned} \quad (3)$$

Corresponding to EEG, the matrices X , S , and A refer to the multichannel time series of subjects, the signal sources inside the brain, and the relationship matrix between the internal signal sources. Inspired by the above ICA algorithm, the collected multichannel EEG signals can also be separated to obtain the approximate original EEG signals of the internal signal sources, and then convolution is used to learn the biometric characteristics of the approximate internal signal sources for identity identification. The details of neural network architecture called ICACConvNet are as follows.

2.4.1. ICA Stage. A matrix with random initial weights is provided as the separation matrix W , which is multiplied by the neural network input (matrix X) to obtain the

approximate internal signal sources matrix Y , as shown in equation (3). Taking Y as the input of the subsequent convolution operations, the weights of W are iteratively optimized via backpropagation. The ICA stage may be regarded as a kind of implementation of ICA algorithm using neural network.

2.4.2. Convolution Stage. A typical multilayer convolution neural network, including multiple convolution layers and pooling layers, is expected to extract the biometric characteristics of the signal sources inside the brain and compress the parameter scale. The convergent parameters of convolutional kernels are obtained after several iterations of optimization.

2.4.3. Output Stage. The EEG biometric features learned during the convolution stage are combined and outputted. First, the features are flattened and then selected by multiple fully connected layers. Finally, the recognition results are outputted by Softmax function. In order to improve the generalization performance of the system, the dropout function is added between the fully connected layers.

The idea of combining ICA and neural network has been used in functional magnetic resonance imaging [31], but the ICA stage is generally relatively independent from neural

network, and the weight optimization of W has nothing to do with network. The proposed system integrates ICA into neural network architecture, and the weight optimization of W depends on the back propagation of network.

The network architecture is plotted in Figure 4 and implemented using PyTorch [32]. Hyperparameter tuning is performed on the eyes-open session of PhysioNet dataset by the grid search method, when a sliding window of 0.5 seconds with a stride of 0.25 seconds is adopted. After comprehensive consideration of accuracy, complexity, and training time, the main hyperparameters are finally selected as follows. In the ICA stage, the number of internal signal sources is 64. In the convolution stage, three 2-dimensional convolution layers with ELU activation and three max pool layers are adopted. In the output stage, two fully connected layers are established and the dropout rate is 0.5. The final output normalization function is Softmax. The loss function is cross entropy loss, and the optimizer is Adam with a learning rate of $3 * 10^{-3}$. The loss function of neural network is shown in equation (4), where N represents the number of samples, O represents the number of identity labels, y_{ij} is a function (if the true label of sample i is equal to j , take 1; otherwise, take 0), and p_{ij} represents the prediction probability that the label of sample i is j . All parameters of neural network are shown in Table 1, where P , C , and O represent the number of sample points, the number of sample channels, and the number of identity labels, respectively:

$$Loss = - \sum_i^N \sum_j^N y_{ij} \log(p_{ij}). \quad (4)$$

2.5. EEG Dataset and Channel Selection. The EEG signals used to verify the proposed system are obtained from a public database called PhysioNet EEG Motor Movement/Imagery Dataset [33–35]. The dataset is available free of charge and all data were collected using a 64-channel BCI2000 system with a sampling rate of 160 Hz. 109 subjects performed 14 different sessions, consisting of two resting baseline sessions and three groups of four motor or motor imagery tasks (T1–T4). Two resting state sessions are chosen, one for 1 minute with eyes-open (EO) and one for 1 minute with eyes-closed (EC).

In recent years, low-cost EEG sensors have made great progress [11], which further increases the possibility that EEG-based identification systems will be used in practical applications. In order to evaluate whether the proposed system works well on these commercial EEG sensors, a series of experiments are conducted using 14, 32, and 64 channels, respectively. The selected 14-channel and 32-channel are based on the EMOTIV EPOC X 14 Channel Mobile Brainwear® [10] and EMOTIV EPOC Flex EEG Brainwear® system [36]. Note that the 14 channels of EMOTIV EPOC X (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) are all contained in the original 64 channels of PhysioNet dataset. Because the 32 channels of EMOTIV EPOC Flex do not correspond precisely to the original 64

channels, four channels have been reselected (i.e., FT7, TP7, TP8, and FT8). The selected channels in experiments are highlighted in red in Figure 5.

2.6. Experimental Setup. To test whether the proposed system can meet the requirements of identification, two experiments were performed on PhysioNet datasets.

2.6.1. Data Augmentation Experiment. The first experiment validates the importance of data augmentation on the system performance. The segmented samples generated by sliding windows with different lengths and strides are trained and tested in turn. The data of the first experiment are all from the eyes-open session, in which the first 48s are divided as the training set and the last 12s as the testing set.

2.6.2. Channel Selection Experiment. The second experiment examines the effects of 14, 32, and 64 channels on the system performance using 5-fold cross-validation. In this experiment, eyes-open session, eyes-closed session, and the union of two resting state sessions are checked in turn.

It should be noted that in both experiments, a sliding window is executed after the training set and testing set are divided. After 1000 training epochs, the experimental results generally tend to be stable, so the termination condition is set as 2000 epochs. The batch size of the training sets is 64. The system performance is evaluated using Rank-1 accuracy, false rejection rate (FRR), false acceptance rate (FAR), and equal error rate (EER). Rank-1 accuracy is used to evaluate the performance in identification scenarios, which is the probability of correctly identifying a user's identity. FRR, FAR, and EER are used to evaluate the performance in authentication scenarios, where the system determines whether a user matches his or her claimed identity [27].

3. Results and Discussion

3.1. Experimental Results. In the data augmentation experiment, the performance of system using different sliding windows is shown in Table 2. A sliding ratio of 0.5 means that the overlap rate of window is 50%, and a sliding ratio of 1 means that a fixed window is used. When fixed windows were applied, the Rank-1 accuracy of 0.25 seconds, 0.5 seconds, 1 second, and 2 seconds was 99.39%, 98.89%, 94.80%, and 60.86%, respectively. When sliding windows with an overlap rate of 50% were used, the corresponding Rank-1 accuracy was 99.40%, 99.51%, 99.04%, and 92.74%, respectively, and the equal error rates were also improved to 0.15%, 0.06%, 0.19%, and 1.67%. The results show that sliding windows achieved a better performance than fixed windows. Interestingly, the performance of a sliding window of 1 second with a stride of 0.5 seconds was like that of a fixed window of 0.5 seconds, which may be because the number of samples eventually generated by these two segmentation schemes was almost the same. Test accuracy and training loss

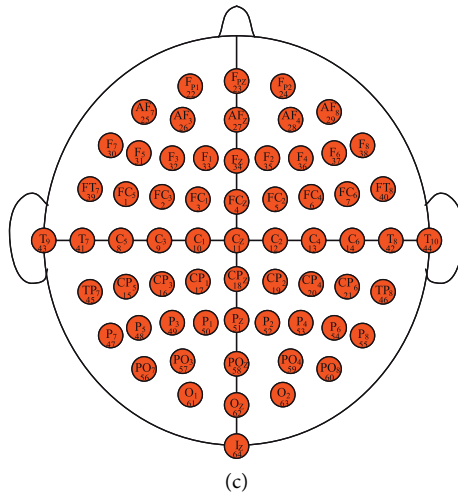
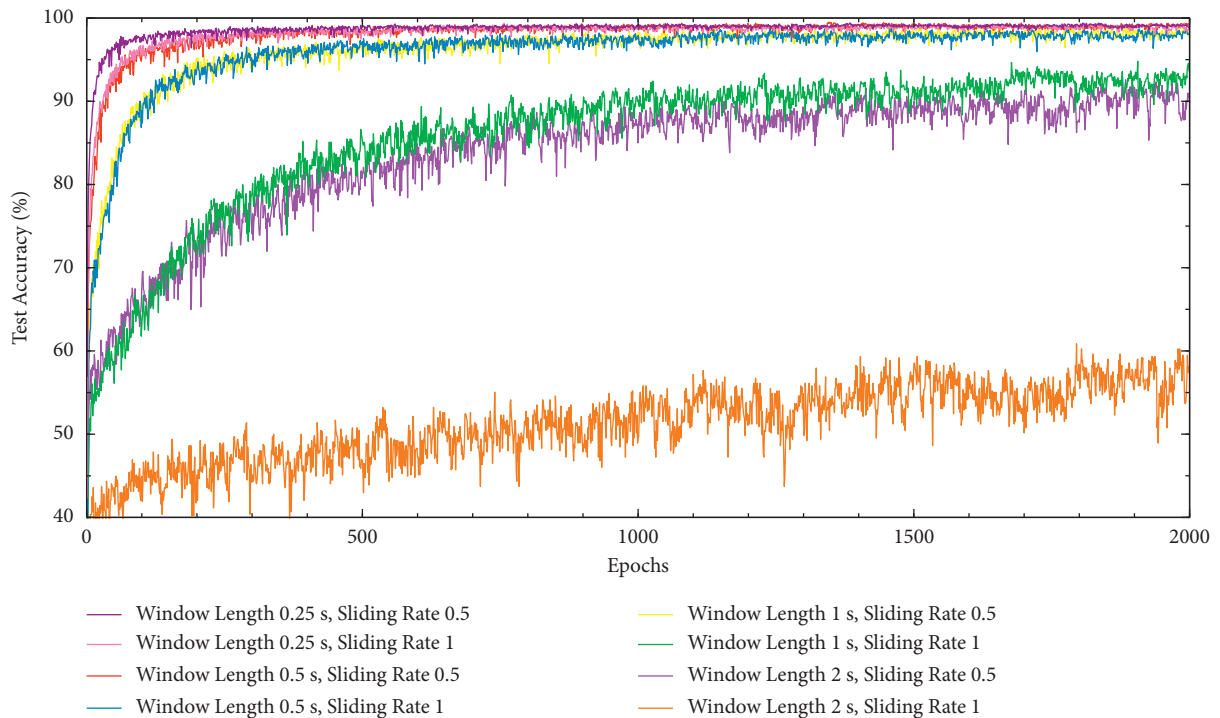


FIGURE 5: Electrode positions on scalp and their corresponding channels (red represents selected channels, and white represents unused channels). (a) 14 channels, (b) 32 channels, and (c) 64 channels.

TABLE 2: Comparison of the performance of the proposed personal identification systems with different sliding windows.

Window (s)	Sliding ratio	Scale of training set	Training time (min)	Rank-1 (%)	FRR (%)	FAR (%)	EER (%)
0.25	0.5	41747	560	99.40	0.15	0.15	0.15
	1	20928	309	99.39	0.17	0.17	0.17
0.5	0.5	20819	273	99.51	0.06	0.07	0.06
	1	10464	136	98.89	0.19	0.19	0.19
1	0.5	10355	147	99.04	0.20	0.18	0.19
	1	5232	74	94.80	1.38	1.38	1.38
2	0.5	5123	85	92.74	1.67	1.68	1.67
	1	2616	43	60.86	10.55	10.55	10.55

Bold values indicate the best performance.



(a)

FIGURE 6: Continued.

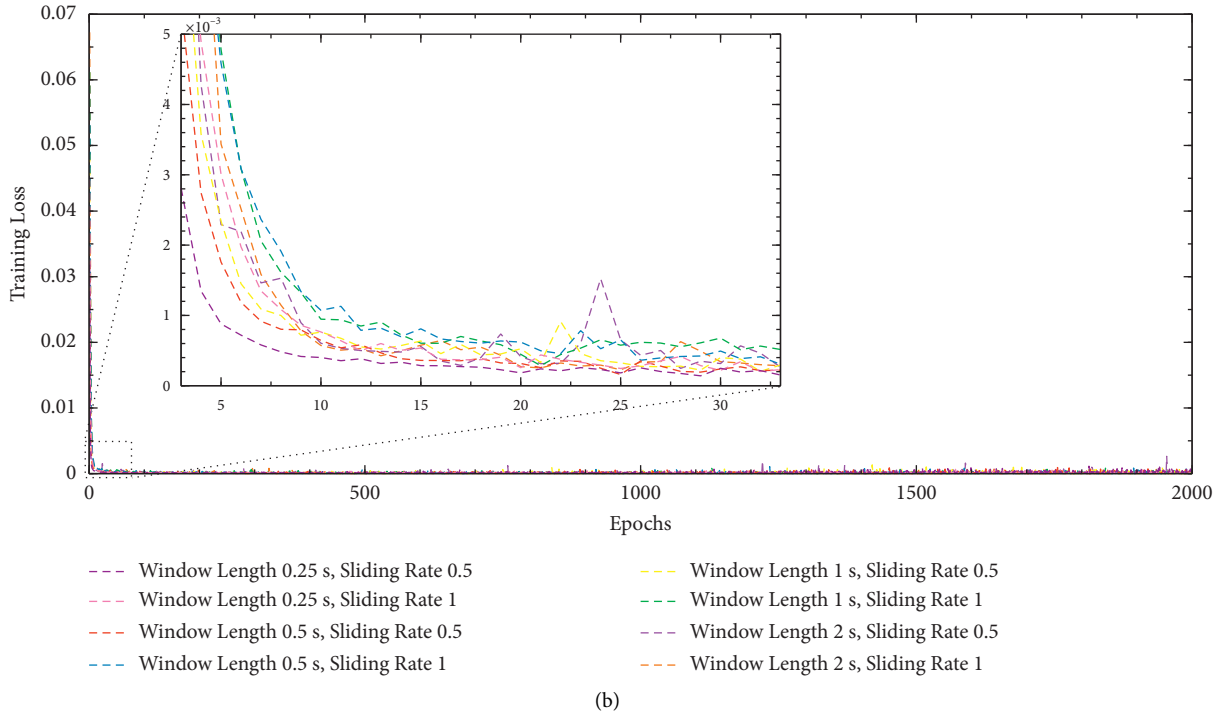


FIGURE 6: Test accuracy and training loss curves for different sliding windows. (a) Testing accuracy and (b) training loss.

curves of different windows are plotted in Figure 6. After 2000 rounds of training, the first three augmentation schemes were all in the first echelon and Rank-1 accuracy can reach over 99%, which indicated that the scheme using a window of 0.5 seconds with a stride of 0.25 seconds can achieve excellent performance with less training time. Figure 7 shows the detection error trade-off (DET) curves of the data augmentation experiment, that is, the relationship between FAR, FRR, and EER, were represented by threshold changes. The smaller EER meant a better performance in authentication scenarios. The scheme using a window of 0.5 seconds with a stride of 0.25 seconds also achieved the best EER of 0.06%. To some extent, for ICACONVNet-based identification system, more samples may bring certain performance gains. In addition, sliding windows with the same window length helped to reduce the acquisition time of the enrollment stage. In summary, a sliding window of 0.5 seconds with a stride of 0.25 seconds is suitable for the proposed system. If the window length is further reduced to 0.25 seconds, the performance is not improved, and the training time is significantly increased.

In the channel selection experiment, referring to comparison results of data augmentation experiment, a sliding window of 0.5 seconds with a stride of 0.25 seconds was used. The performance of 14, 32, and 64 EEG channels in different sessions is shown in Table 3. The selected 14 channels and 32 channels are based on market available EMOTIV EPOC X and EMOTIV EPOC Flex, as plotted in Figure 5. According to the experimental results, the performance of 64 channels was generally the best, followed by 32 channels and 14 channels slightly worse. However, although the number of channels was as low as half or even less than a quarter, the

performance of the system did not suffer a significant degradation. Therefore, the proposed system had an excellent identification performance even with very few EEG electrodes and can be used to build a practical identification system using low-cost EEG sensors. In addition, the average Rank-1 accuracy can achieve more than 99% in the cross-validation of different sessions, which indicated that the proposed system was effective and robust.

3.2. Comparison with Related Works. Our work was compared with the performance of other EEG-based identification systems using PhysioNet dataset, as shown in Table 4. It should be noted that the results selected were from the 14 channels of the union session in the channel selection experiment. The specific result of 5-fold cross verification was that Rank-1 accuracy was 98.152216%, 99.573588%, 99.599431%, 99.819098%, and 99.457294% and EER were 0.465176%, 0.155298%, 0.077230%, 0.064428%, and 0.144111%. The schemes proposed by M. Fraschini et al. [23], M. Garau et al. [24], and T. Schons et al. [26] require 64 EEG channels and 12-second signal segments, which may mean a long wait during the enrollment and identification stages for users. A similar situation exists in the approaches proposed by D. L. Rocca et al. [22] and J.-H. Kang et al. [25]. In the work of S. Yang et al. [37], the accuracy of 99% is achieved on T1-T4 tasks with 9-channel data; however, the window time is increased to 30 seconds. In the 16-channel system proposed by Y. Sun et al., only 1 second of EEG is needed to complete the work, but introducing LSTM into network architecture will inevitably increase the computational complexity, thus increasing the training time required for

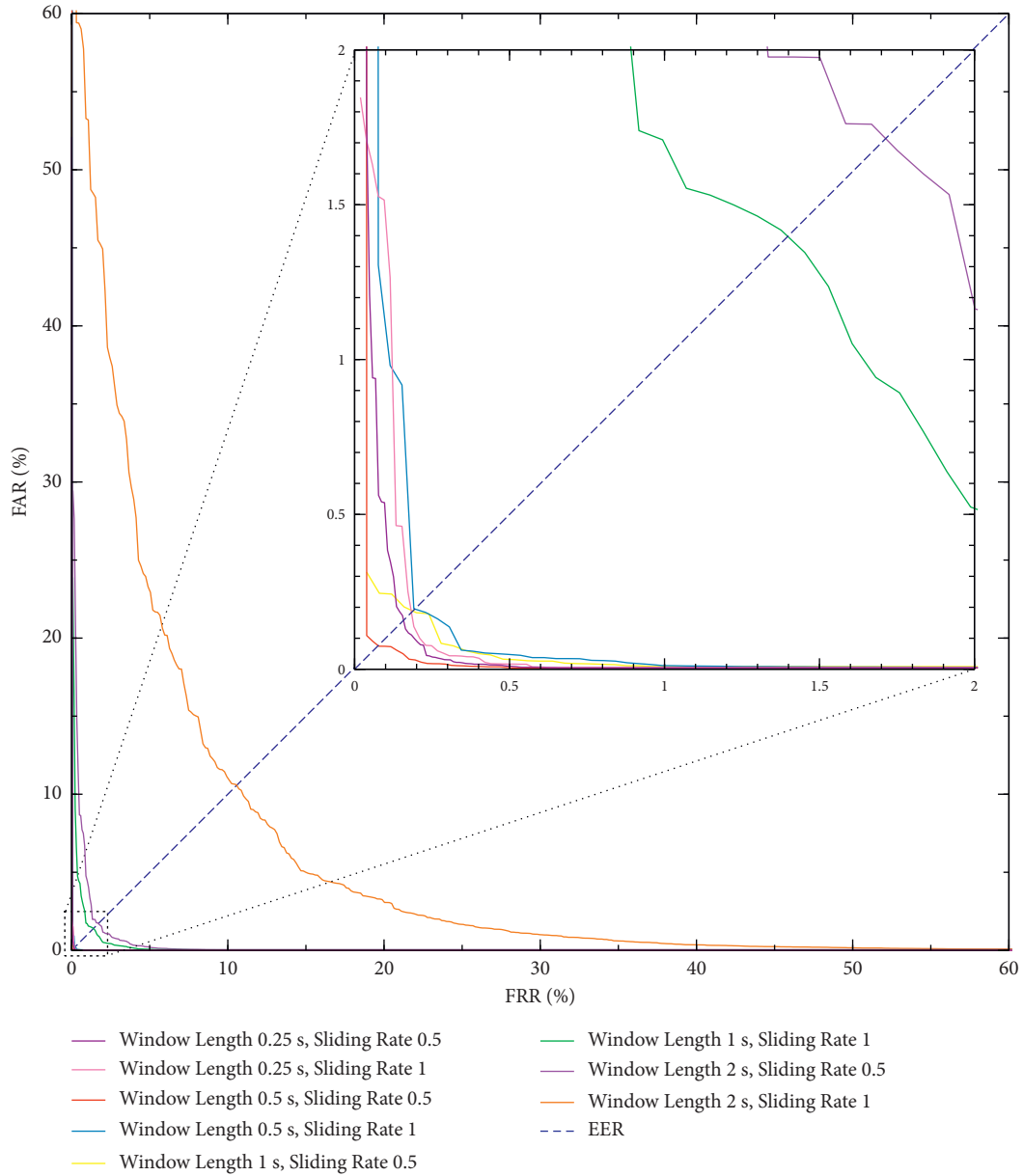


FIGURE 7: DET curves for different sliding windows.

TABLE 3: Comparison of the performance of the proposed personal identification systems with 14-, 32-, and 64-channel EEG signals (positions of the electrodes are shown in Figure 5).

Session	Channels	Rank-1 (%)	FRR (%)	FAR (%)	EER (%)
EO	14	99.04 ± 0.95	0.25 ± 0.21	0.25 ± 0.21	0.25 ± 0.21
	32	99.29 ± 0.81	0.19 ± 0.16	0.19 ± 0.16	0.19 ± 0.16
	64	99.29 ± 0.85	0.21 ± 0.22	0.21 ± 0.22	0.21 ± 0.22
EC	14	99.11 ± 0.85	0.18 ± 0.15	0.19 ± 0.16	0.19 ± 0.15
	32	99.31 ± 0.90	0.15 ± 0.23	0.17 ± 0.24	0.16 ± 0.23
	64	99.44 ± 0.75	0.16 ± 0.19	0.16 ± 0.19	0.16 ± 0.19
EO&EC	14	99.32 ± 0.60	0.18 ± 0.15	0.18 ± 0.15	0.18 ± 0.15
	32	99.64 ± 0.35	0.09 ± 0.06	0.09 ± 0.06	0.09 ± 0.06
	64	99.78 ± 0.23	0.06 ± 0.08	0.07 ± 0.08	0.07 ± 0.08

TABLE 4: Comparison with other EEG-based identification systems using PhysioNet dataset.

Reports	Approach	Session	Subjects	Channels	Sampling rate (Hz)	Window length (s)	Stride (s)	Rank-1 (%)	EER (%)
[22]	PSD and spectral coherence	EO and EC	108	56	160	10	—	100	—
[23]	Eigenvector	EO and EC	109	64	160	12	—	96.90	4.40
[24]	Eigenvector	EO and EC	109	64	160	12	—	—	1.42
[37]	Wavelet coefficients	T1-T4	108	9	160	30	15	99.00	4.50
[26]	CNN	EO and EC	109	64	160	12	0.125	—	0.19
[25]	Eigenvector	EO and EC	109	56	—	12	—	98.93	0.73
[27]	1D-Conv. LSTM	EO and EC, T1-T4	109	16	160	1	—	99.58	0.41
Proposed	CNN	EO and EC	109	14	160	0.5	0.25	99.32 ± 0.60	0.18 ± 0.15

The key parameters used in the proposed system are highlighted in bold.

TABLE 5: Model loading time (T_{model}) and averaged execution time for batch testing (T_{batch}) for 1D-Convolutional LSTM and the proposed approach.

Model	Channels	T_{model} (s)	T_{batch} (s)
1D-Convolutional LSTM [27] (TensorFlow)	16	17.852	0.065
	32	17.965	0.065
	64	18.477	0.071
Proposed (PyTorch)	14	1.106	0.002
	32	1.102	0.002
	64	1.125	0.002

high identification performance [27], as shown in Table 5. In contrast, the proposed system adopts a sliding window of 0.5 seconds and selects 14 EEG channels based on the existing low-cost EEG sensors, which still has certain advantages in Rank-1 accuracy and EER. Meanwhile, ICA-ConvNet has a shorter loading time and a faster computing speed, when the testing PC is equipped with an Intel 10700 CPU, a Nvidia 2080Ti GPU, and a Seagate 1 TB HDD. Theoretically, the proposed system has better practicability. Through a combination of ICAConvNet and optimized sliding window, our work has a better overall performance.

4. Conclusions

In this paper, a personal identification system using resting state EEG is proposed, which is designed by combining ICA and convolution computation. The number of channels in the system can be as few as 14, and a sliding window of 0.5 seconds is applied for data augmentation. Different sliding window schemes were compared on publicly accessible PhysioNet database to select the optimal data augmentation parameters. In the cross-validation of 109 subjects, Rank-1 of $99.32 \pm 0.60\%$ and EER of $0.18 \pm 0.15\%$ were achieved, respectively. Compared with related work, our system has certain advantages in the accuracy, computational complexity, and stability, which further advances the implementation of practical EEG-based identification systems.

The identification application of resting state EEG is discussed in this paper. In the future, the characteristics of nonresting state EEG can be further studied. In addition, the

challenges faced by EEG-based identification systems in practical application are also worth exploring, such as the permanence and stability of EEG.

Data Availability

PhysioNet EEG Motor Movement/Imagery Dataset can be visited at <https://physionet.org/content/eegmmidb/1.0.0/>. The code of the proposed system is publicly available at <https://github.com/hitfyd/Personal-Identification-System-using-Resting-State-EEG>.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 62071151 and 61301099).

References



- [1] Q. Gui, M. V. Ruiz-Blondet, S. Laszlo, and Z. Jin, "A survey on brain biometrics," *ACM Computing Surveys*, vol. 51, no. 6, pp. 112–121, 2019.
- [2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
- [3] J. Peng, Q. Li, A. A. Abd El-latif, N. Wang, and X. Niu, "Finger vein recognition with gabor wavelets and local binary

- patterns,” *IEICE - Transactions on Info and Systems*, vol. 8, no. 8, pp. 1886–1889, 2013.
- [4] R. Gad, M. Talha, A. A. El-Latif et al., “Iris recognition using multi-algorithmic approaches for cognitive internet of things (CIoT) framework,” *Future Generation Computer Systems*, vol. 89, pp. 178–191, 2018.
- [5] E. Marasco and A. Ross, “A survey on antispooofing schemes for fingerprint recognition systems,” *ACM Computing Surveys*, vol. 47, no. 2, pp. 1–36, 2015.
- [6] N. Wang, Q. Li, A. A. Abd El-Latif, J. Peng, and X. Niu, “An enhanced thermal face recognition method based on multi-scale complex fusion for Gabor coefficients,” *Multimedia Tools and Applications*, vol. 72, no. 3, pp. 2339–2358, 2014.
- [7] P. Campisi and D. La Rocca, “Brain waves for automatic biometric-based user recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, 2014.
- [8] G.-L. Li, J.-T. Wu, Y.-H. Xia, Q.-G. He, and H.-G. Jin, “Review of semi-dry electrodes for EEG recording,” *Journal of Neural Engineering*, vol. 17, no. 5, Article ID 051004, 2020.
- [9] E. F. M. Wijdicks, P. N. Varelas, G. S. Gronseth, and D. M. Greer, “Evidence-based guideline update: determining brain death in adults: report of the quality standards subcommittee of the American academy of neurology,” *Neurology*, vol. 74, no. 23, pp. 1911–1918, 2010.
- [10] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, “Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs,” *PeerJ*, vol. 1, p. e38, 2013.
- [11] S. Yang and F. Deravi, “On the usability of electroencephalographic signals for biometric recognition: a survey,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 958–969, 2017.
- [12] S. Sugrim, C. Liu, and J. Lindqvist, “Recruit until it fails,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 104–111, 2019.
- [13] A. Jalaly Bidgoly, H. Jalaly Bidgoly, and Z. Arezoumand, “A survey on methods and challenges in EEG based authentication,” *Computers & Security*, vol. 93, Article ID 101788, 2020.
- [14] Y. Di, X. An, F. He, S. Liu, Y. Ke, and D. Ming, “Robustness analysis of identification using resting-state EEG signals,” *IEEE Access*, vol. 7, pp. 42113–42122, 2019.
- [15] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, and O. A. Alomari, “Person identification using EEG channel selection with hybrid flower pollination algorithm,” *Pattern Recognition*, vol. 105, Article ID 107393, 2020.
- [16] Z. Yin and J. Zhang, “Cross-session classification of mental workload levels using EEG and an adaptive deep learning model,” *Biomedical Signal Processing and Control*, vol. 33, pp. 30–47, 2017.
- [17] P. Kumari and A. Vaish, “Feature-level fusion of mental task’s brain signal for an efficient identification system,” *Neural Computing & Applications*, vol. 27, no. 3, pp. 659–669, 2016.
- [18] Z. Mao, W. X. Yao, and Y. Huang, “EEG-based biometric identification with deep learning,” in *Proceedings of the 2017 8th International IEEE/EMBS Conference On Neural Engineering (NER)*, pp. 609–612, Shanghai, China, May 2017.
- [19] R. A. Ramadan and A. V. Vasilakos, “Brain computer interface: control signals review,” *Neurocomputing*, vol. 223, pp. 26–44, 2017.
- [20] Y. Zeng, Q. Wu, K. Yang et al., “EEG-based identity authentication framework using face rapid serial visual presentation with optimized channels,” *Sensors*, vol. 19, no. 1, p. 6, 2018.
- [21] B. Kaur, D. Singh, and P. P. Roy, “A Novel framework of EEG-based user identification by analyzing music-listening behavior,” *Multimedia Tools and Applications*, vol. 76, no. 24, pp. 25581–25602, 2017.
- [22] D. L. Rocca, P. Campisi, B. Vegso et al., “Human brain distinctiveness based on EEG spectral coherence connectivity,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2406–2412, 2014.
- [23] M. Frascchini, A. Hillebrand, M. Demuru, L. Didaci, and G. L. Marcialis, “An EEG-based biometric system using eigenvector centrality in resting state brain networks,” *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 666–670, 2015.
- [24] M. Garau, M. Frascchini, L. Didaci, and G. L. Marcialis, “Experimental results on multi-modal fusion of EEG-based personal verification algorithms,” in *Proceedings of the 2016 International Conference On Biometrics (ICB)*, pp. 1–6, Halmstad, Sweden, June 2016.
- [25] J.-H. Kang, Y. C. Jo, and S.-P. Kim, “Electroencephalographic feature evaluation for improving personal authentication performance,” *Neurocomputing*, vol. 287, pp. 93–101, 2018.
- [26] T. Schons, G. J. P. Moreira, P. H. L. Silva, V. N. Coelho, and E. J. S. Luz, “Convolutional network for EEG-based biometric,” in *Proceedings of the Progress In Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 601–608, Madrid, Spain, February 2018.
- [27] Y. Sun, F. P.-W. Lo, and B. Lo, “EEG-based user identification system using 1D-convolutional long short-term memory neural networks,” *Expert Systems with Applications*, vol. 125, pp. 259–267, 2019.
- [28] E. Lashgari, D. Liang, and U. Maoz, “Data augmentation for deep-learning-based electroencephalography,” *Journal of Neuroscience Methods*, vol. 346, Article ID 108885, 2020.
- [29] P. Comon, “Independent component analysis, A new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [30] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural Networks*, vol. 13, no. 4, pp. 411–430, 2000.
- [31] Y. Qiu, Q.-H. Lin, L.-D. Kuang et al., “Classification of schizophrenia patients and healthy controls using ICA of complex-valued fMRI data and convolutional neural networks,” *Advances in Neural Networks - ISNN 2019*, vol. 22, pp. 540–547, 2019.
- [32] A. Paszke, “PyTorch: an imperative style, high-performance deep learning library,” *Advances in Neural Information Processing Systems*, vol. 32, 2019, <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>.
- [33] A. Goldberger, “Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, 2000 http://www.researchgate.net/publication/306204663_physiobank_physiotoolkit_and_physionet_components_of_a_new_research_resource_for_complex_physiologic_signals.
- [34] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, “BCI2000: a general-purpose brain-computer interface (BCI) system,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

- [35] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "EEG motor movement/imagery dataset," *Physionet.org*, vol. 8, 2009.
- [36] N. S. Williams, G. M. McArthur, B. de Wit, G. Ibrahim, and N. A. Badcock, "A validation of Emotiv EPOC Flex saline for EEG and ERP research," *PeerJ*, vol. 8, Article ID e9713, 2020.
- [37] S. Yang, F. Deravi, and S. Hoque, "Task sensitivity in EEG biometric recognition," *Pattern Analysis & Applications*, vol. 21, no. 1, pp. 105–117, 2018.

Research Article

Secure Health Monitoring Communication Systems Based on IoT and Cloud Computing for Medical Emergency Applications

Ali I. Siam ¹, Mohammed Amin Almaiah ², Ali Al-Zahrani,² Atef Abou Elazm,³ Ghada M. El Banby,⁴ Walid El-Shafai,^{3,5} Fathi E. Abd El-Samie,³ and Nirmeen A. El-Bahnasawy⁶

¹Department of Embedded Network Systems Technology, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, Egypt

²College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia

³Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt

⁴Department of Industrial Electronics and Control Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt

⁵Security Engineering Laboratory, Computer Science Department, Prince Sultan University, Riyadh 11586, Saudi Arabia

⁶Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt

Correspondence should be addressed to Mohammed Amin Almaiah; malmaiah@kfu.edu.sa

Received 28 June 2021; Revised 18 September 2021; Accepted 4 October 2021; Published 13 December 2021

Academic Editor: Anastasios D. Doulamis

Copyright © 2021 Ali I. Siam 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.

Smart health surveillance technology has attracted wide attention between patients and professionals or specialists to provide early detection of critical abnormal situations without the need to be in direct contact with the patient. This paper presents a secure smart monitoring portable multivital signal system based on Internet-of-Things (IoT) technology. The implemented system is designed to measure the key health parameters: heart rate (HR), blood oxygen saturation (SpO₂), and body temperature, simultaneously. The captured physiological signals are processed and encrypted using the Advanced Encryption Standard (AES) algorithm before sending them to the cloud. An ESP8266 integrated unit is used for processing, encryption, and providing connectivity to the cloud over Wi-Fi. On the other side, trusted medical organization servers receive and decrypt the measurements and display the values on the monitoring dashboard for the authorized specialists. The proposed system measurements are compared with a number of commercial medical devices. Results demonstrate that the measurements of the proposed system are within the 95% confidence interval. Moreover, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Relative Error (MRE) for the proposed system are calculated as 1.44, 1.12, and 0.012, respectively, for HR, 1.13, 0.92, and 0.009, respectively, for SpO₂, and 0.13, 0.11, and 0.003, respectively, for body temperature. These results demonstrate the high accuracy and reliability of the proposed system.

1. Introduction

Many inventors and researchers have competed to create new systems that help specialists to diagnose and possibly treat some diseases. Diseases are usually associated with changes in some physiological parameters in the human body (e.g., heart rate, oxygen saturation, body temperature, blood pressure, etc.). The diagnosis of such diseases requires making some checks in the hospital to measure how a physiological parameter is away

from the normal rates and then determine the positive or negative presence of those diseases. More deviations from normal rates are strong markers of death for a wide range of patients [1]. However, many people cannot go to the hospital continuously because they may not have enough time to go to the hospital from time to time, they have a chronic illness, or the coordinating specialist is abroad. In addition, medical care at hospitals may cost a lot. For those people, personal health devices are reliable solutions to monitor and track vital signs at

home and also can call for medical help in case of emergency [2–6]. Personal health devices have an increasing interest and have become commercially available [7–9].

With recent advances in IoT and wireless sensor networks [10], many attempts have been made to deliver patient data remotely without going to the hospital [11]. This helps specialists to determine the appropriate action ahead or to send a specific equipped medical help. In emergency cases, the transmission of critical patient data can significantly impact patient life [12]. With cloud computing, which is a paradigm shift in computing and storage, IoT-based health monitoring systems found new ways of innovation [13, 14]. The cloud is the place where patient data is processed and stored, allowing vital signs of a patient to be monitored in real time or stored for historical reviews. Storing patient data in the cloud provides several benefits, including availability, reliability, and convenience at a relatively low cost [15, 16]. Various researchers [17–19] have addressed the opportunities and challenges of using cloud computing in the healthcare field. However, communication and storage of patient data within most cloud-based healthcare systems are in plain form, which puts the patient personality and privacy at stake [20]. Yi et al. [21] addressed some security threats regarding sensitive physiological data transmitted over the public channels and stored in the backend systems. Thus, an approach for delivering critical patient data to relevant healthcare providers without compromising patient privacy is needed.

The proposed system provides a secure and real-time solution for private health data records stored in the cloud. IoT biosensors are used to capture key biological parameters (heart rate, blood oxygen saturation (SpO_2), and body temperature) from a patient at a comfortable home. Then, an IoT-based microcontroller encrypts, processes, and delivers secure health records to the public cloud. On the other side, only specialists at trusted healthcare centers can monitor the biological parameters of the patient in real time. Also, they can review historical records to predict any unusual activities and also can assign precautions to prevent any emergency cases. The proposed health monitoring system targets several patients with medical issues, such as patients in accidents or emergency places, patients with motion disabilities, patients with chronic illness, patients whose doctors are abroad, or elders who need continuous monitoring.

Securing patient data is achieved using the AES algorithm, which is a symmetric encryption algorithm that offers an excellent compromise between encryption speed and security [22]. AES algorithm is employed in the proposed system to secure patient data before storing it in the cloud. This ensures data integrity and privacy and the secure distribution of patient data in public networks.

Although there are plenty of researches and papers on the topic of health monitoring, our research, unlike most monitoring systems, adds some key contributions in the field that are as follows:

- (i) A low cost and accurate health monitoring system is implemented to monitor the heart rate, blood oxygen saturation, and body temperature of patients without the need to be in direct contact with specialists

- (ii) Multiple medical sensors are incorporated with a compact and powerful microcontroller chip in a small-sized device with the help of IoT infrastructure. So, the implementation is simple and, at the same time, effective
- (iii) Medical measurements are encrypted before transmission to cloud storage. So, the proposed framework keeps the privacy and integrity of patient data
- (iv) End-to-end security for medical records is ensured between the patient node and the healthcare center
- (v) The proposed system relies on a Wi-Fi-based connection, which provides fast communication between the patient module and the specialists module with low power consumption compared to other technologies

The rest of this paper is organized as follows. Section 2 discusses some preliminaries in the context of the research work. Section 3 gives the previous studies related to the proposed system. Section 4 presents the proposed health monitoring system, layers, and actors of the system. System implementation is introduced in Section 5. In Section 6, the experimental results are discussed. Finally, the paper conclusion is given in Section 7.

2. Preliminaries

2.1. Blood Oxygen Saturation. Body cells and tissues need oxygen to live. Oxygen is carried from the lungs and absorbed into the Red Blood Cells (RBCs). Hemoglobin is the protein that carries oxygen in the RBCs and transports it throughout the body. The heart pumps oxygenated blood from the left ventricle to the whole body cells and tissues through the circularity system. It receives the deoxygenated blood and pumps it towards the lungs again to be oxygenated during the inhalation process. Blood oxygen saturation, termed SpO_2 , is an estimation of the amount of oxygen dissolved in the blood, which is described as the percentage of oxygenated hemoglobin to the total amount of hemoglobin, expressed as

$$SpO_2 (\%) = \frac{HbO_2}{HbO_2 + Hb} \times 100. \quad (1)$$

SpO_2 is one of the clinical vitals preferably measured by specialists to determine how much oxygen is saturated in the blood. Normal oxygen saturation for most healthy persons is 94% to 100% at sea level. SpO_2 is a key indicator for the effectiveness of the respiratory system, and it can aid in the detection of hypoxemia. Furthermore, SpO_2 level can help in the early detection of COVID-19 pneumonia [23, 24], which may cause initially unnoticeable low arterial oxygen saturation. The author in [23] reported that COVID-19 pneumonia patients have oxygen saturations as low as 50%.

The SpO_2 level is commonly measured by a pulse oximeter, which has a Light Emitting Diode (LED) to shine the light through the fingertip and a photodetector (PD) to measure the amount of the reflected light. The structure of

the pulse oximeter is depicted in Figure 1. When the light is emitted from the LED through the fingertip, some of the light is absorbed by the blood and the other amount is reflected to the PD. Figure 2 describes the resulting waveform of the output of the PD, which has a pulsatile waveform due to the periodic change of the amount of the blood underneath the sensor due to the periodic pumping of the blood from the heart, which affects the amount of the reflected light. The more the amount of blood is, the more absorbed light and less reflected light arriving at the PD. The Direct Current (DC) component in the resulting waveform is due to the reflectance of light on bones, tissues, and other stationary parts, while the Alternating Current (AC) component represents the pulsatile change of the arterial blood that forms the photoplethysmography (PPG) signal [25–27].

With two light beams with different wavelengths, typically Red (660 nm) and Infrared (IR) (880 nm), it is reported that HbO₂ and Hb absorb the two different wavelengths with different amounts (Figure 3). Hb has a higher absorption at 660 nm, while HbO₂ has a higher absorption at 880 nm. This characteristic reveals that the amount of absorbed light at 660 and 880 nm can be used to estimate the amount of dissolved oxygen in the blood (SpO₂). The two separate PPG signals determined from the Red and IR LEDs are used to find the ratio R , which is used to calculate the SpO₂ level [28].

$$R = \frac{(AC/DC)_{Red}}{(AC/DC)_{IR}} \quad (2)$$

The accurate estimation of SpO₂ is based on empirical calibration with the ratio R for the specific device. Equation (3) is often used in the literature to approximate the SpO₂ value based on R [28].

$$SpO_2 [\%] = 110 - 25(R). \quad (3)$$

Another approximation to find the value of SpO₂ using the ratio R is developed by *Maxim Integrated* based on empirical calibration for their medical products and is defined as [29]:

$$SpO_2 (\%) = 104 - 17(R). \quad (4)$$

In our study, the MAX30102 sensor, a product of *Maxim Integrated*, is adopted to measure the SpO₂ level and the heart rate.

2.2. Heart Rate. The heart rate is denoted as the frequency at which the heart pumps blood to the arteries, and it is measured by the number of contractions of the heart per minute. The heart rate is a reflection of the physical and mental state of the body. It varies conditionally according to the body physical needs, as in the case in which the oxygen saturation level is low.

Pulse oximeters can determine the frequency of the heartbeats by calculating the time between consecutive peaks in the PPG signal using a single light source (e.g., Red LED), as shown in Figure 2. The heart rate is typically measured in beats per minute (bpm). The normal heart rate of healthy adult persons is between 60 and 100 bpm, while they are at rest.

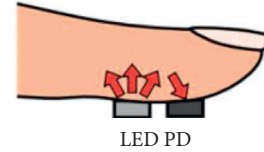


FIGURE 1: Pulse oximeter structure.

2.3. AES Algorithm. Data security is an essential target in everything in our lives on all applications. Data is required to be protected from assaults and intruders. Due to the great revolution of the Internet and its applications, there is a critical need to employ security techniques to secure the transmitted information. Authorized users can transmit and receive data from a distance with communication networks. To be reliable, data needs to be safeguarded from unapproved change (integrity), hidden from unlicensed access (confidentiality), and accessible to an approved entity when required (availability). Not only should the data be trusted when it is stored in a computer, but there should also be a way to preserve its privacy when it is transmitted over a communication network.

The AES ciphering algorithm is cost-effective, and it is based on the Rijndael procedure [22], which is an iterated block ciphering process with variable key size and variable block size. The key size and block size can be autonomously 192, 128, or 256 bits. The cipher key is a rectangular array with four rows and a number of columns equal to the key size divided by 32. In addition, the intermediate resulting ciphertext describes a state and it is in the shape of a rectangular array of four rows, and a number of columns equal to the block size divided by 32. The number of rounds performed on the intermediate state is related to the key size. For key sizes of 192, 128, and 256 bits, the numbers of rounds are 10, 12, and 14, respectively. Every round comprises a fixed sequence of transformations, except the last and the first rounds.

The AES comprises a number of rounds. Any round, except the final one, involves ShiftRows, SubBytes, AddRoundKey, and MixColumns functions. In the SubBytes step, a linear substitution for every byte is performed according to Figure 4. In the final round, no MixColumns operation is executed. Every byte in the array is updated using an 8-bit S-box, which provides nonlinearity in the cipher system. The S-box is derived from the multiplicative inverse over the finite Galois Field GF(2⁸), known to have good nonlinearity characteristics. The S-box is selected to prevent fixed-point as well as opposite-fixed-point attacks.

The step of ShiftRows operates on the rows of the state. It cyclically shifts the bytes in every row. For the AES process, the first row is left unaffected. Every byte of the second row is shifted a single byte to the left. Also, the third and fourth rows are shifted by offsets of two and three bytes, respectively. For the blocks of size 192 bits or 128 bits, the shifting patterns are the same. In this manner, every column of the output state of the ShiftRows step is composed of bytes from every column of the input state. In

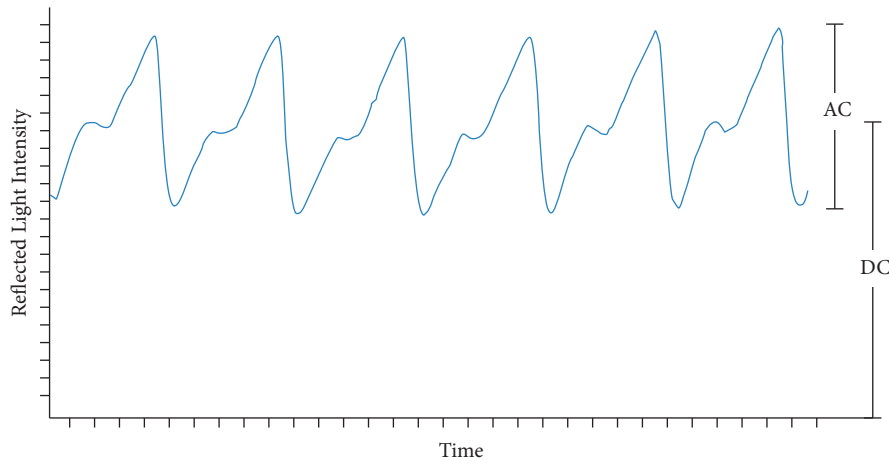


FIGURE 2: Reflected light waveform for a single light source.

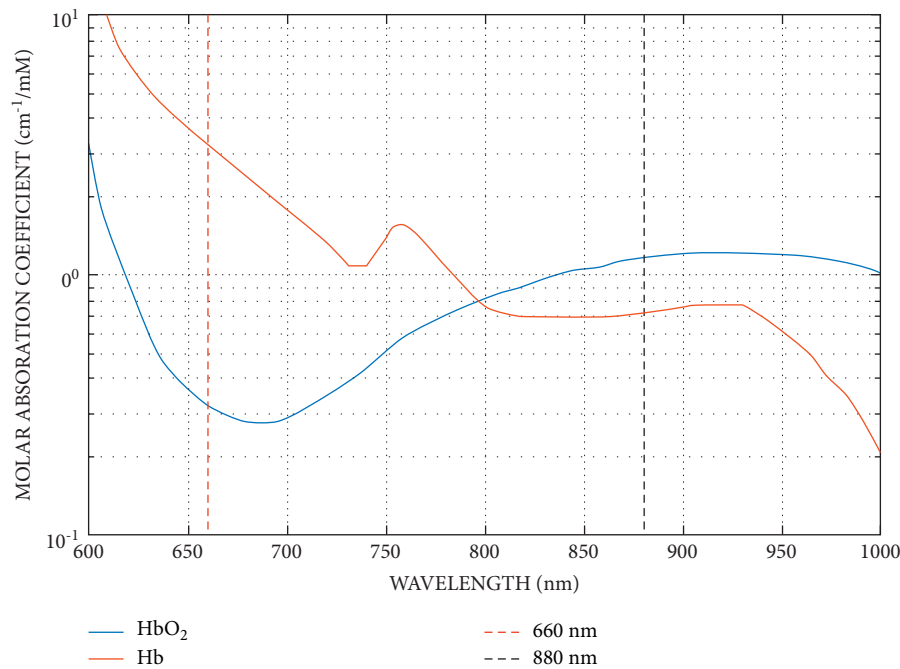


FIGURE 3: Oxygenated and deoxygenated hemoglobin absorption graph for red and infrared wavelengths [27].

the state of the 256-bit blocks, the first row is unaffected and the shiftings for the second, third, and fourth rows are 1 byte, 3 bytes, and 4 bytes, respectively, as demonstrated in Figure 5.

In the MixColumns step, the four bytes of every column of the state are merged with a linear invertible transformation. The MixColumns function requires four bytes as input and outputs four bytes, where every input byte involves all four output bytes. With ShiftRows, MixColumns delivers diffusion in the cipher system. Every column is treated as a polynomial over $GF(2^8)$ and is subsequently multiplied with a fixed polynomial $c(x) = 3x^3 + x^2 + x + 2$. The MixColumns step can also be considered as multiplication by a particular matrix, as demonstrated in Figure 6.

3. Related Studies

With recent advances in cloud computing and IoT, mobile healthcare devices were developed to provide healthcare services with more flexibility and speed at a lower cost. This helps patients receive healthcare and medical treatment anytime and helps specialists to monitor their patients in real time. From the perspective of healthcare providers, the IoT has the potential to reduce device downtime through remote provision. Besides, the IoT provides efficient scheduling of the limited resources by ensuring their best use and serves more patients [30].

In this context, several researchers have developed smart medical and healthcare surveillance and monitoring architectures. Yi et al. [21] proposed a secure health monitoring

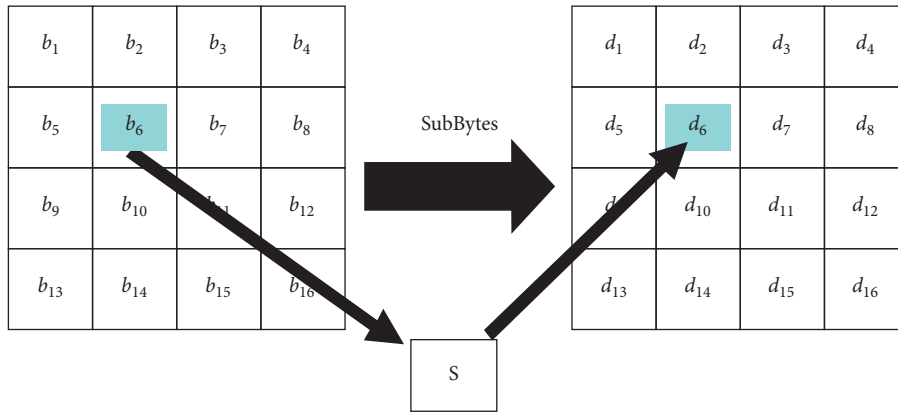


FIGURE 4: SubBytes step.

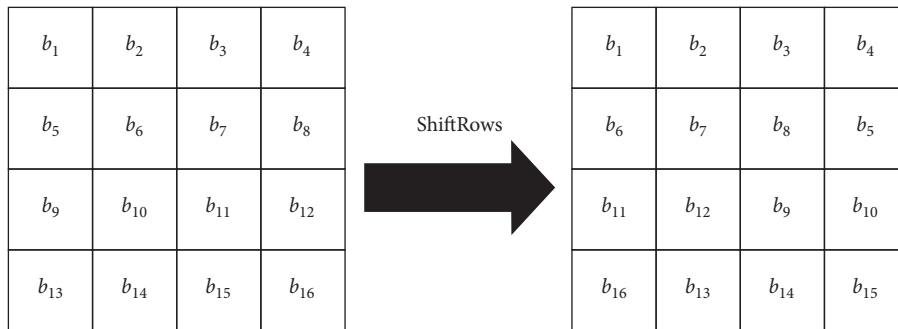


FIGURE 5: ShiftRows step.

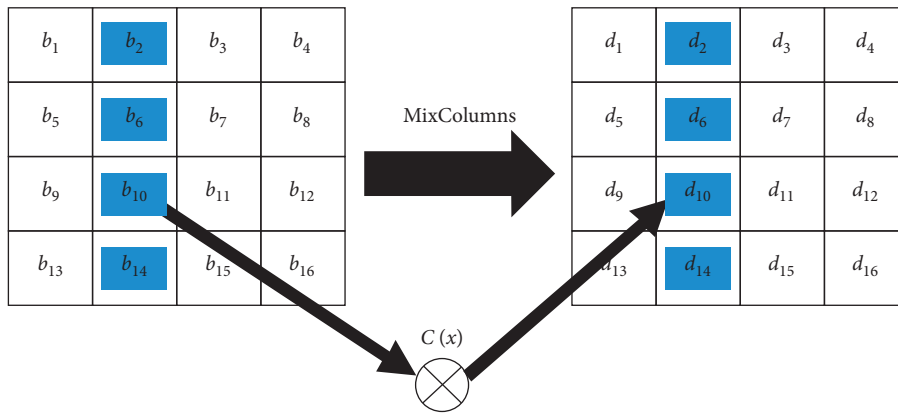


FIGURE 6: MixColumns step.

system in which private health data is encrypted using AES and split into three different servers to keep the privacy of the data even if one server is compromised. This approach defends the system against both outside and inside attacks. However, it requires more computational steps regarding generating and distributing public keys among three database servers. Ali et al. [31] implemented an IoT and Android-based health monitoring system to measure the heart rate, oxygen saturation, and body temperature of patients. Measurements are sent via Bluetooth to a mobile application and can be transmitted using Wi-Fi to the Internet. They

compared the results with those of a commercially available product and reported a maximum deviation of 2%.

Mohammed et al. [32] integrated IoT and cloud computing in building an ECG mobile application, which provides the end-users with visualization for their ECG signals and logging data uploaded to the specific medical cloud. A microcontroller board was used to capture the ECG signal from a patient and send it to the mobile device in a wireless manner using Bluetooth technology. ECG data is saved as a binary file into the SD card of the mobile phone, and the user has the ability to send this file to the cloud to become available for specialist inspection.

Al-khafajiy et al. [33] proposed a Smart Healthcare Monitoring System (SW-SHMS) to monitor elderly people in their homes in real time using a mobile application. It uses a pulse sensor connected to an Arduino Uno to track the heart rate and oxygen saturation of an elder. Data from a pulse sensor is transferred to a mobile device through Bluetooth. The mobile application collects vital data from wearable sensors and sends the data to the cloud for processing and storage to become available for relevant hospitals or specialists.

Gupta et al. [34] adopted IoT and a microcontroller to monitor the vital signs of the patient. They considered only one perspective, which is the ECG signal. Raspberry Pi was used to collect data from wearable sensors and send it to a MySQL database. The authors also employed the GSM wireless network to send alert messages to healthcare centers in emergency cases. Ghosh et al. [35] implemented an IoT sensing module to measure various vital signs (ECG, body temperature, and patient position). This module is connected to a local web server via a COM connection for local monitoring and it can send measurements to cloud storage for remote monitoring.

Lloret et al. [36] presented a framework to improve the life of elders that depends on several types of communication to ease their daily affairs. They proposed Ambient Assisted Living (AAL) based wireless communication sensors, which help elders to avoid dangerous situations. Elsts et al. [37] proposed a Sensing Platform for HHealthcare in a Residential Environment (SPHERE) based on IoT technology. They presented SPHERE IoT network infrastructure for healthcare in a home environment with low power wireless network performance. Moustafa et al. [38] introduced a remote monitoring solution for developing real-time control of medical devices in eHealth applications. They presented a secure, scalable, unified, and real-time infrastructure based on sensors and IoT to remotely monitor patients. Park et al. [39] presented an emergency alert and an elderly health monitoring system that encompasses active capturing of brain and body movement signals, communication signal analysis, warning, and detection processes.

Khan et al. [40] presented a healthcare model to employ IoT technology within the field of crafty wellness care. They introduced a complete and effective healthcare monitoring framework planned based on the IoT and RFID tags. Mighali et al. [41] described a reliable and smart remote monitoring system with low cost for controlling the body motility and the position of elderly people. Tuli et al. [42] suggested a novel model named HealthFog for incorporating deep learning in edge computing devices and implementing it for automatic heart disease analysis. This model delivers healthcare as a fog service through IoT devices and proficiently manages the data of heart patients. The presented model is adaptive to a variety of operation modes of quality of service (QoS) and prediction accuracy based on user demands. Sodhro et al. [43] presented an efficient and intelligent monitoring and measurement approach for medical healthcare applications by transmitting critical patient data with good QoS through wireless networks. Alabdulatif et al. [44] discussed the main concept of a smart health IoT surveillance system in real time for cloud medium.

Table 1 summarizes the different features adopted with different healthcare-monitoring-related studies in the

literature. Generally, there are few contributions in the literature on medical emergency applications adopting IoT and cloud computing technologies. Some of those introduced techniques have critical problems with medical data security and real-time communication. The traditional health monitoring systems are believed not to achieve adequate security, and they are not recommended for real-time communication. In addition, they have low robustness and require more computations in medical data processing and transmission. Hence, they increase the computational overhead. Taking into account the limitations of the state-of-the-art works, an efficient IoT and cloud-computing-based secure and real-time health monitoring communication system for medical emergency applications is the main contribution of this paper. This framework consists of IoT biosensors, an IoT-based microcontroller, an AES mechanism, and cloud storage to efficiently monitor, process, protect, store, and transmit patient medical data. Moreover, the proposed system achieves real-time communication of transmitted medical data with high quality, high robustness, and low computational complexity compared to the traditional related systems.

4. Proposed IoT and Cloud-Computing-Based Secure Health Monitoring System

The proposed health monitoring system aims to monitor vital data from patients or elderly people, secure it, transmit it to a public cloud database, and provide a real-time monitoring dashboard for authorized caregivers or healthcare centers at any time and anywhere. The implementation of the proposed system involves a three-layer structure of different technologies. The layers of the proposed system are the patient layer, the cloud layer, and the doctor layer. The system architecture of the proposed model with the three layers is shown in Figure 7 and described as follows.

4.1. Patient Layer. The patient layer consists of the patient and an IoT module. The IoT module acquires vital data from medical sensors attached to the patient body, encrypts that data, and sends ciphered data to the cloud database (i.e., second layer). The IoT module consists of a number of biomedical sensors that measure the key vital data, (heart rate, blood oxygen saturation, and body temperature), and a Wi-Fi-based microcontroller that processes this vital data, encrypts it using AES algorithm and sends it directly to the cloud database over Wi-Fi without the need to a local server or an intermediate mobile application. This procedure is performed automatically without patient interaction, making it more convenient for patients with motion disabilities and elderly people.

MAX30102 [53], shown in Figure 8(a), is a high sensitivity pulse oximeter employed to measure the heart rate and blood oxygen saturation of a patient through his fingertip. DS18B20 sensor [54], shown in Figure 8(b), is used to measure body temperature. These sensors are connected to the ESP8266 NodeMCU [55] microcontroller, which controls the whole system and provides the processing and transmission functionalities (Figure 8(c)). ESP8266

TABLE 1: Summary of different features in healthcare monitoring studies.

Feature	Type	Sample studies
Monitoring mode	Local	[31–33, 35]
	Remote	[12, 21, 34]
Transmission type	Cloud-based	[21, 31, 33, 34]
	Device-to-Device	[12, 32, 35]
Communication protocol	Wi-Fi	[31, 34]
	Bluetooth	[31–33]
	Mobile cellular network	[34]
	Zigbee	[12]
Is secured	Yes	[12, 21, 45]
	No	[32–35]
Monitored sign	Respiration	[46–50]
	Heart rate & SpO ₂	[31, 33]
	Body temperature	[31, 35]
	ECG	[12, 32, 34, 35]
	Blood pressure	[51, 52]
	Patient position	[35, 41]

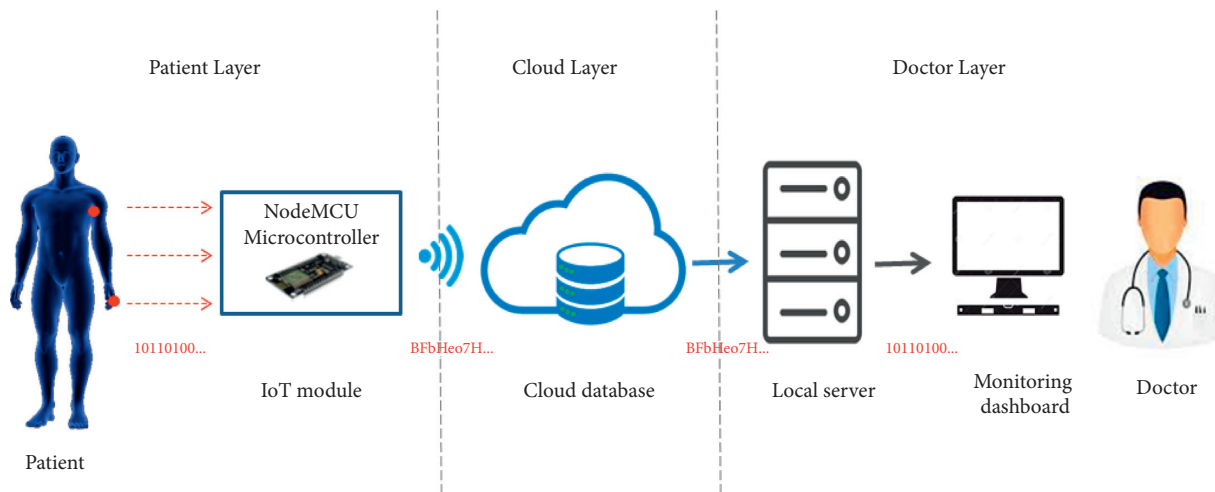


FIGURE 7: Architecture of the proposed model.

NodeMCU is an emerging IoT chip with a small-size, low-cost, self-contained Wi-Fi module, high processing speed, and capability of running self-contained applications.

The ESP8266 Crypto library [56] is adopted to provide the AES implementation for the ESP8266 module. Vital data is encrypted with a 128-bit key using Cipher Block Chaining (CBC) mode, and then encoded with Base64 format. After that, it is sent to the cloud.

AES algorithm is selected to encrypt the sensor readings, because it is simple to be implemented within the hardware using the appropriate software library, unlike other encryption algorithms, which may not be supported to be implemented in the hardware devices. In addition, it provides a good compromise between the speed of computations and the complexity.

4.2. Cloud Layer. The cloud layer is responsible for providing a safe place for private health data. Cloud receives sensitive data from the patient layer and stores it in a ciphered form, which makes the system more robust against not only external attacks but also internal attacks that can be initiated by the cloud

service provider [57]. The cloud layer is not charged in any processing of data, but it delivers data as it is to the next layer.

Firestore [58] is employed in this work. It is a real-time cloud database, acquired by Google, and intended for IoT solutions. Figure 9 shows a screenshot of the created real-time database on Firestore, showing encoded values for heart rate, SpO₂, and body temperature.

4.3. Doctor Layer. This layer enables doctors at trusted healthcare centers to monitor and track vital data in real time. This enables doctors to predict any unusual activity and it can assign precautions to prevent any emergency case. This layer is synchronized with the cloud layer to receive updates of patient data in real time, which is in a ciphered form. A backend mechanism is used to fetch and decrypt received data and deliver it to the monitoring dashboard. First, doctors should log in via a web interface to be authenticated to prevent fraud access; and then, they are directed to the monitoring dashboard. The web interface is developed using HTML5, JavaScript, Bootstrap, and ASP.NET.

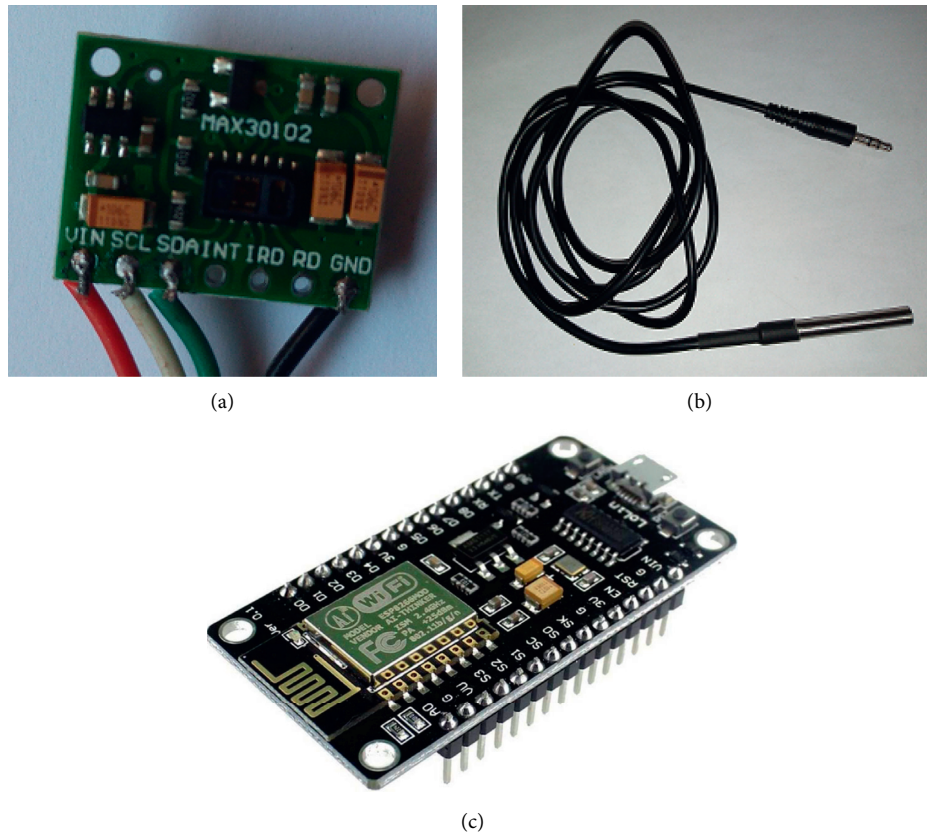


FIGURE 8: (a) MAX30102 sensor, (b) DS18B20 sensor, and (c) ESP8266 NodeMCU WiFi Devkit.

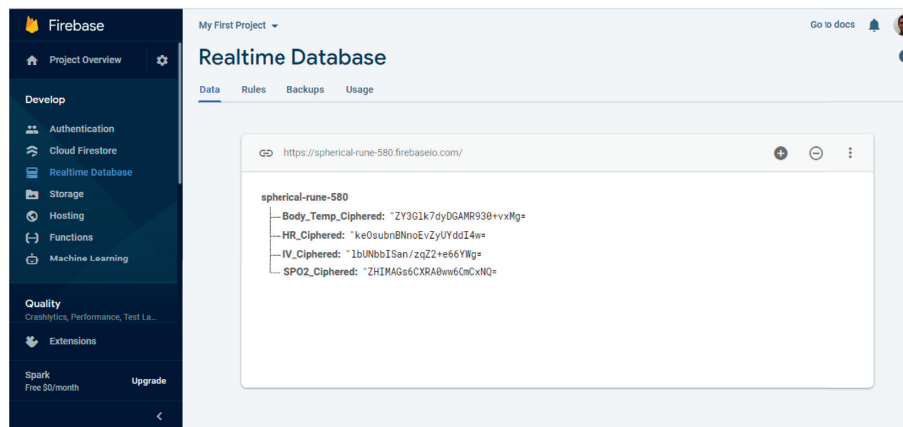


FIGURE 9: Screenshot of the real-time cloud-based database.

5. System Design and Implementation

The proposed system uses the MAX30102 pulse oximeter to measure the heart rate and the blood oxygen saturation by calculating the ratio of oxygenated hemoglobin to deoxygenated hemoglobin, which is then used to calculate the percentage of oxygenated blood levels (SpO_2), as discussed in Sections 2.1 and 2.2. For the heart rate and SpO_2 measurements, the patient is asked to put his fingertip on the finger probe shown in Figure 10. The finger probe consists of a plastic holder with a soft contact surface, which is used to fit the fingertip on the sensor.

The other end of the finger probe is connected to the device circuit board. The temperature sensor is placed under the armpit of the patient, whereas this position is recommended by specialists to measure the body temperature. Also, the other end of the sensor is connected to the specified socket in the circuit board. The block diagrams describing the procedures for measuring the heart rate, SpO_2 , and body temperature are shown in Figures 11 and 12. The device is powered by a 3.7 V rechargeable battery, which is a good choice for small size and long-time operation. The proposed device sends the measured medical data every five seconds, periodically. So, this is an



FIGURE 10: Finger probe used to fit the fingertip on the MAX30102 sensor. The sensor is placed inside the plastic holder.

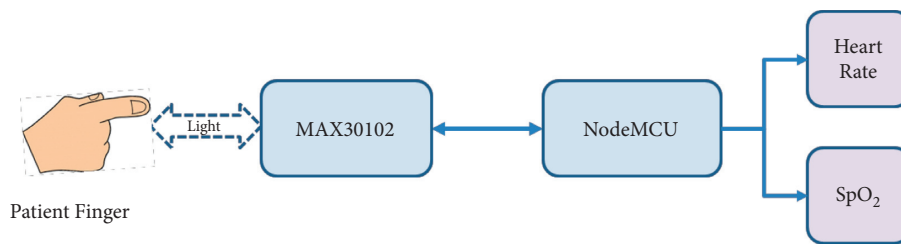


FIGURE 11: Heart rate and SpO₂ monitoring block diagram.

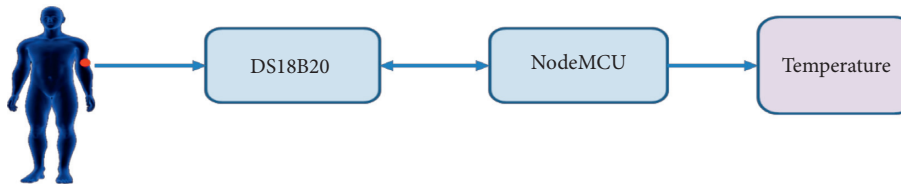


FIGURE 12: Block diagram of body temperature monitoring.

important factor that guarantees the achievement of acceptable QoS of delivering the proposed device measurements. Figure 13 shows the complete hardware implementation of the IoT module with relevant sensors and the microcontroller being connected. The complete system flowchart is depicted in Figure 14, indicating basic actors and their roles, whereas each actor has its own functionality to achieve the system goal, as follows.

5.1. Sensor Module. This module involves capturing raw physiological data from the patient's body, and sends this data to a Microcontroller Unit (MCU) for processing. This module comprises two sensor types: pulse oximeter sensor and body temperature sensor. The MAX30102 pulse oximeter sensor, shown in Figure 8(a), is involved in this study to measure the heart rate and the blood oxygen saturation. The MAX30102 is a reflective pulse oximeter that includes internal LEDs, photodetectors, optical elements, and low-noise electronics with ambient light cancellation. The

communication between the MAX30102 sensor and the MCU is through the I²C interface. DS18B20 temperature sensor, shown in Figure 8(b), is used to sense the patient skin temperature. The DS18B20 digital thermometer provides 9-bit to 12-bit Celsius temperature measurements and communicates with the MCU through the 1-Wire interface. Table 2 shows a summary of the technical specifications for the utilized sensors.

5.2. IoT Module. The IoT module is the coordinator of the whole patient layer. The process flow along this module includes the following steps:

- (1) Receive raw physiological data from sensors through an appropriate interface (I²C or 1-Wire).
- (2) Process the received data and convert it into numerical values (heart rate, blood oxygen saturation, and body temperature).
- (3) Encrypt vital signs using the AES algorithm with a 128-bit key.

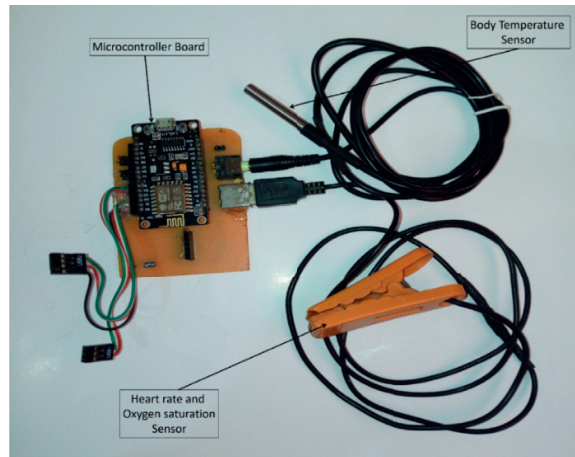


FIGURE 13: Hardware implementation of the IoT module.

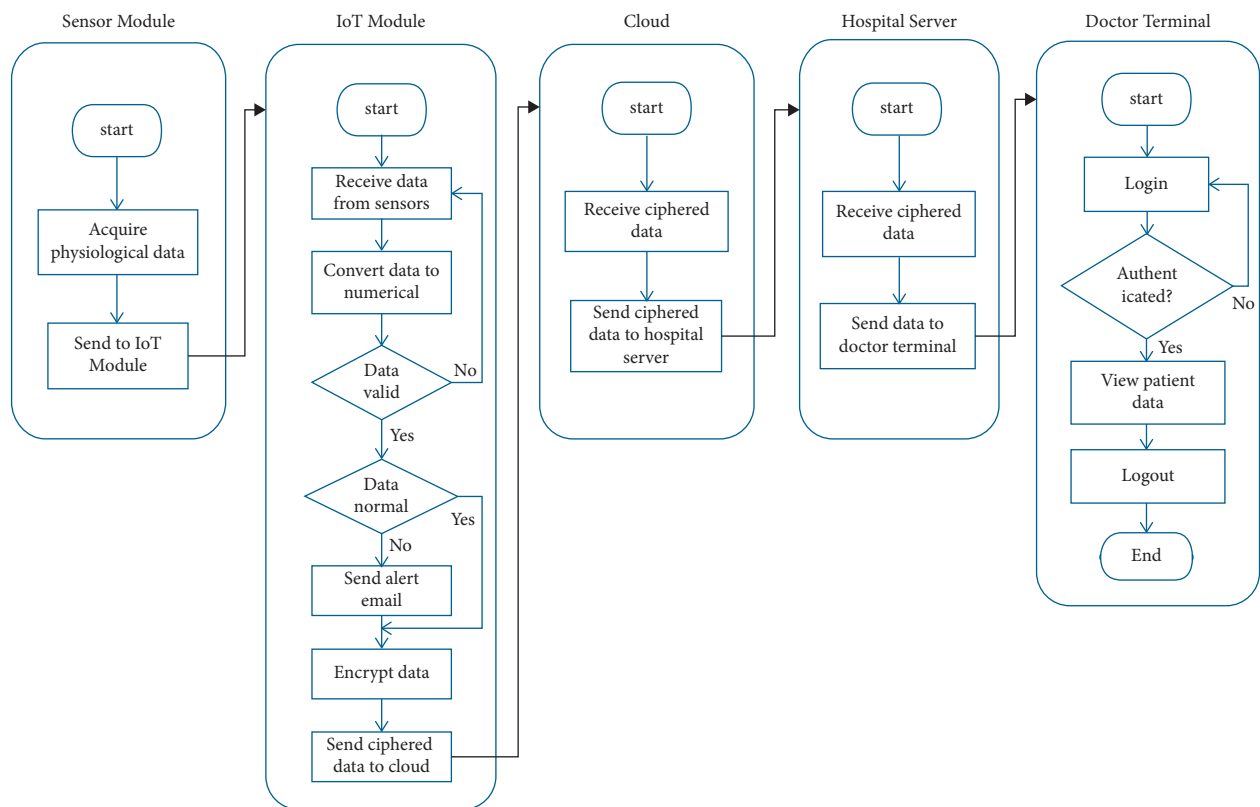


FIGURE 14: Complete system implementation flowchart.

TABLE 2: Summary of technical specifications for the utilized sensors [53, 54].

Sensor	Accuracy	Resolution	Current
MAX30102	—	18 bits	Standby: 600 μ A IR LED: 20 mA RED LED: 20 mA
DS18B20	$\pm 0.5^{\circ}\text{C}$	0.0625°C	1 mA

- (4) Establish a connection to the cloud database over a Wi-Fi link.
- (5) Send ciphered data to cloud storage.

These tasks are accomplished using the ESP8266 NodeMCU developing kit, shown in Figure 8(c). ESP8266 is an Arduino-like board with extra beneficial features, such as 802.11 b/g/n Wi-Fi support, integrated TCP/IP protocol stack, 3.3 V operating voltage, low current consumption (10 μ A~170 mA), attachable flash memory (16 MB), and high processor speed (80~160 MHz). ESP8266 is programmed using the open-source Arduino IDE in order to accomplish its commissioned tasks.

5.3. Cloud. The cloud is the place where patient data is stored. Firebase cloud database server is adopted in this work to store patient data, so that the IoT module can communicate with the medical organization to allow the specialist to access and diagnose patient vital signs from anywhere at any time.

5.4. Hospital Local Server. This entity is responsible for receiving data from cloud storage, decrypting it with the appropriate decryption key, and then delivering it to the doctor's terminal. It also holds a SQL database comprising a table for patient information and another table for login credentials in order to control access to the system and provide authorization for users according to granted permissions.

5.5. Doctor Terminal. It is the last destination of patient data, where vital data of the patient is examined by a specialist to determine any health issues associated with this data and assign precautions to prevent any emergency cases. First, the specialist is asked to provide his credentials to determine his roles, after which he can proceed to the monitoring dashboard to view and interact with patient data in real time. The monitoring dashboard is updated automatically with every update in the cloud database.

6. Experimental Results

The proposed system provides a way to keep an eye on key biological indicators of a patient on a secure and real-time basis. With the proposed system, securing patient data is assured by encrypting the data to ensure data privacy and secure distribution of patient data in public networks. The proposed system initiates the encryption process on the IoT module, as illustrated in Figure 14, and then sends the ciphered data to the cloud. The server at the trusted healthcare center is synchronized with the cloud storage, and it is notified when the cloud storage is updated. After that, the healthcare center server fetches the new data from the cloud, which is in ciphered form. Then, the healthcare center server decipheres the data using the decryption key, which is kept secret between the system and the healthcare center. Hence, if a non-trusted user tries to sniff the outgoing data or gain access to the cloud storage, he will get ciphered data that cannot be

deciphered except by using the correct decryption key. Moreover, the decryption key is unique for each module, and it is hard-coded on the microprocessor program and it cannot be inferred by an attacker.

The monitoring dashboard is shown in Figure 15. It displays the received patient data in cipher form and the decrypted values.

To evaluate the accuracy and effectiveness of the proposed health monitoring system, the measurements are compared to those of a number of commercial devices: *High Care* heart rate monitor, pulse oximeter, and a medical thermometer to measure the heart rate, SpO₂, and the body temperature. The reference devices used in the comparison are shown in Figure 16.

The two statistical analysis tools, namely, linear correlation and Bland-Altman plot, are adopted to validate the proposed system accuracy. The measurement setup is shown in Figure 17, indicating that the proposed system values appear on the laptop screen, and the reference measurements are shown in the reference device.

A number of measurement points (50 heart rate points, 50 oxygen saturation level points, and 40 body temperature points) are taken from 20 different individuals (8 males and 12 females) of different ages (4–60 years) at different times. The data points are collected and compared against the reference measurements.

The experimental and actual measurements with error for heart rate sensor are shown in Table 3. The results reveal high agreement with the reference measurements, as shown in Figure 18, demonstrating that the proposed device is highly accurate.

Similarly, the results for the SpO₂ and body temperature sensors are shown in Table 4, Figure 19, Table 5, and Figure 20.

Moreover, the RMSE, MAE, and MRE are computed for the proposed system as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^K (\text{HR}_{ref_i} - \text{HR}_{mes_i})^2}{K}}, \quad (5)$$

$$\text{MAE} = \frac{1}{K} \sum_{i=1}^K |\text{HR}_{ref_i} - \text{HR}_{mes_i}|, \quad (6)$$

$$\text{MRE} = \frac{1}{K} \sum_{i=1}^K \frac{|\text{HR}_{ref_i} - \text{HR}_{mes_i}|}{\text{HR}_{ref_i}}, \quad (7)$$

where HR_{ref} is the reference measurement from the commercial device, HR_{mes} is the measurement from the proposed device, and K is the number of measurements.

The coefficient of determination, denoted as R^2 , is a measure of the correlation between two variables. It ranges from 0, which indicates no correlation, to 1, which indicates a perfect match. Table 6 summarizes the results of the proposed system for the three monitored health parameters.

In addition, Tables 7–9 compare the error rates for the proposed system against those of a number of solutions in

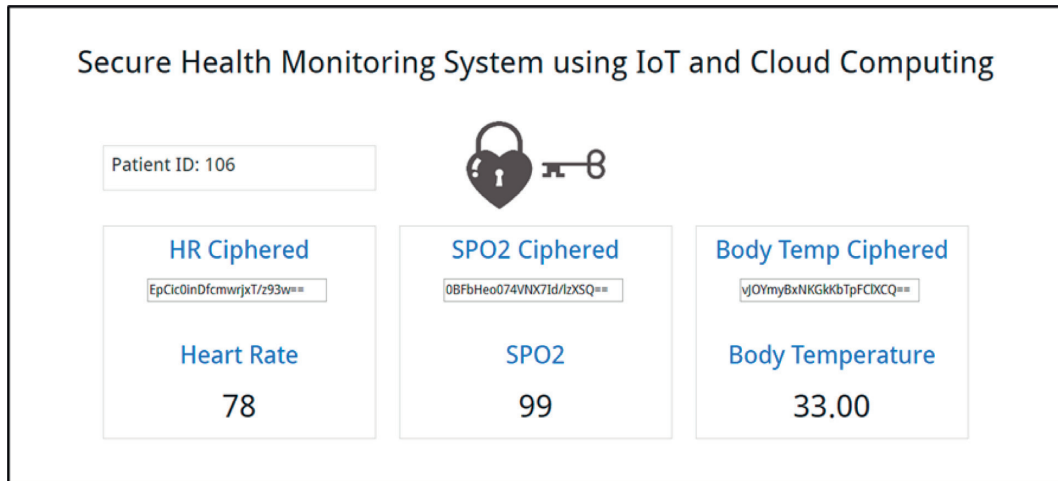


FIGURE 15: Monitoring dashboard, indicating the ciphred and decrypted values for heart rate, SpO₂, and body temperature readings.

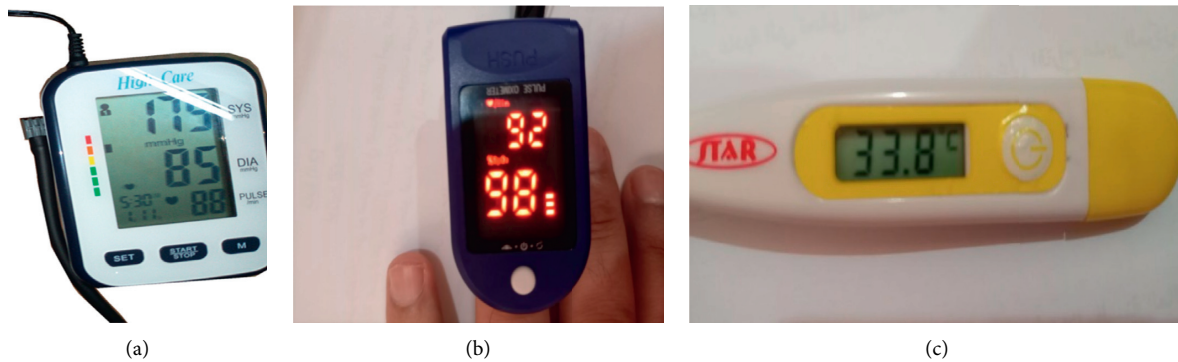


FIGURE 16: Reference devices. (a) Heart rate measuring device, (b) SpO₂ measuring device, and (c) temperature measuring device.

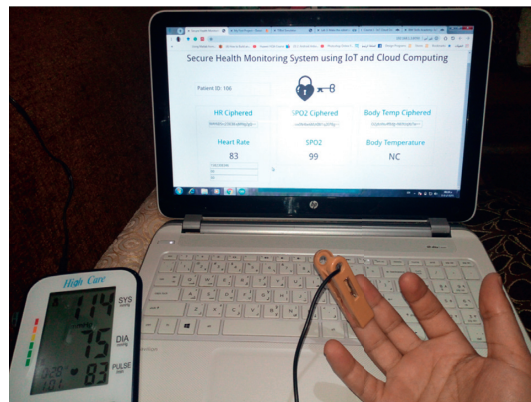


FIGURE 17: Reading from the proposed system versus high care reading.

the literature. This demonstrates the high accuracy and reliability of the proposed system against the solutions in the literature and the feasibility of applying the proposed device for clinical use.

The linear correlation analysis measures the degree of the linear relationship between two variables. The linear relationship between two variables x and y is defined as;

$$y = ax + b, \tag{8}$$

where a and b are the slope and the intercept of the line, respectively.

The line of the perfect match has slope = 1 and intercept = 0, i.e.,

$$y = x. \tag{9}$$

TABLE 3: Proposed system readings versus commercial device (High Care) readings for HR.

Reading no.	Proposed system reading	Reference reading	Error (%)
1	88	88	0.00
2	107	104	2.88
3	78	78	0.00
4	100	99	1.01
5	68	67	1.49
6	107	108	0.93
7	83	82	1.22
8	88	85	3.53
9	93	91	2.20
10	78	78	0.00
11	75	76	1.32
12	78	76	2.63
13	83	81	2.47
14	65	66	1.52
15	83	84	1.19
16	78	79	1.27
17	75	75	0.00
18	83	83	0.00
19	93	95	2.11
20	65	65	0.00
21	78	77	1.30
22	125	124	0.81
23	125	127	1.57
24	88	85	3.53
25	83	85	2.35
26	88	89	1.12
27	75	75	0.00
28	107	109	1.83
29	71	71	0.00
30	75	75	0.00
31	78	77	1.30
32	83	82	1.22
33	115	117	1.71
34	116	115	0.87
35	107	107	0.00
36	115	112	2.68
37	93	92	1.09
38	100	101	0.99
39	93	93	0.00
40	115	117	1.71
41	93	91	2.20
42	78	79	1.27
43	88	88	0.00
44	109	107	1.87
45	100	100	0.00
46	107	107	0.00
47	100	99	1.01
48	83	82	1.22
49	93	92	1.09
50	107	109	1.83

Figures 21–23 show the linear correlation plots for heart rate, SpO₂, and body temperature results, respectively. As shown in the figures, most of the measurements are close to the line of the perfect match. The statistical analysis results indicate that the fit line for measurement points closely coincides with the line of the perfect match.

Figures 24–26 show the corresponding Bland-Altman plots of the difference between measurements versus the average for the three health parameters, respectively. The plots indicate that

all difference points are within the 95% limits of agreement, where the upper 95% limit of the agreement is defined as:

$$95\%_{\text{upper}} = \text{mean} + 1.96 \times \text{SD}, \quad (10)$$

and the lower 95% limit of the agreement is defined as:

$$95\%_{\text{lower}} = \text{mean} - 1.96 \times \text{SD}, \quad (11)$$

where SD is the standard deviation for the differences.

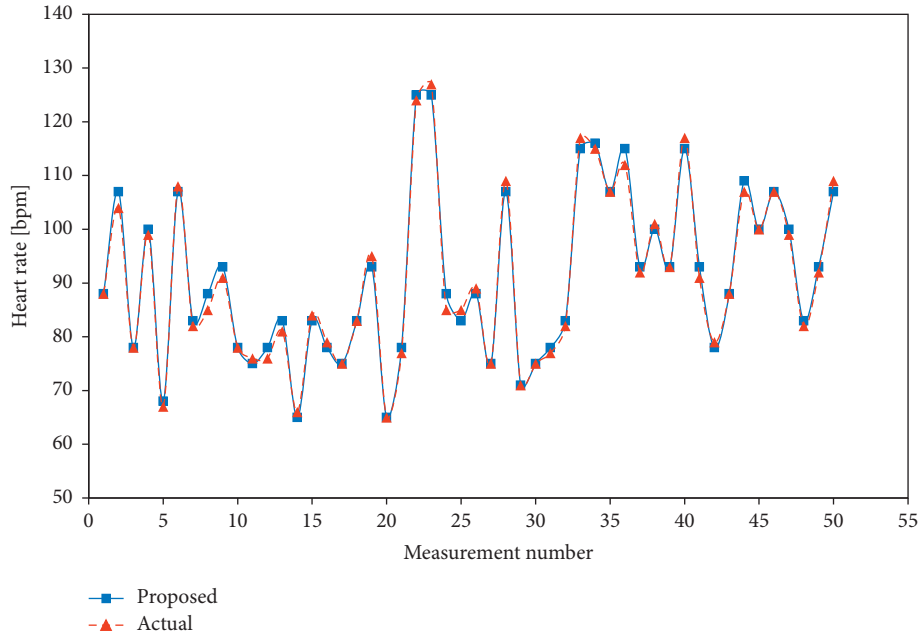


FIGURE 18: Heart rate comparison.

TABLE 4: Proposed system readings versus commercial device (Oximeter) readings for SpO₂ levels.

Reading no.	Proposed system reading	Reference reading	Error (%)
1	99	98	1.02
2	98	97	1.03
3	100	98	2.04
4	99	99	0.00
5	100	99	1.01
6	97	98	1.02
7	98	99	1.01
8	99	98	1.02
9	98	98	0.00
10	99	99	0.00
11	99	99	0.00
12	99	99	0.00
13	99	100	1.00
14	100	98	2.04
15	99	100	1.00
16	99	97	2.06
17	100	99	1.01
18	99	98	1.02
19	100	99	1.01
20	99	98	1.02
21	98	97	1.03
22	100	100	0.00
23	99	99	0.00
24	100	98	2.04
25	98	99	1.01
26	99	97	2.06
27	99	100	1.00
28	98	98	0.00
29	97	99	2.02
30	98	99	1.01
31	99	98	1.02
32	98	98	0.00
33	99	97	2.06
34	100	99	1.01

TABLE 4: Continued.

Reading no.	Proposed system reading	Reference reading	Error (%)
35	99	98	1.02
36	98	97	1.03
37	98	100	2.00
38	98	99	1.01
39	98	98	0.00
40	99	98	1.02
41	97	98	1.02
42	100	100	0.00
43	100	99	1.01
44	97	99	2.02
45	97	97	0.00
46	98	99	1.01
47	100	99	1.01
48	99	99	0.00
49	98	97	1.03
50	99	100	1.00

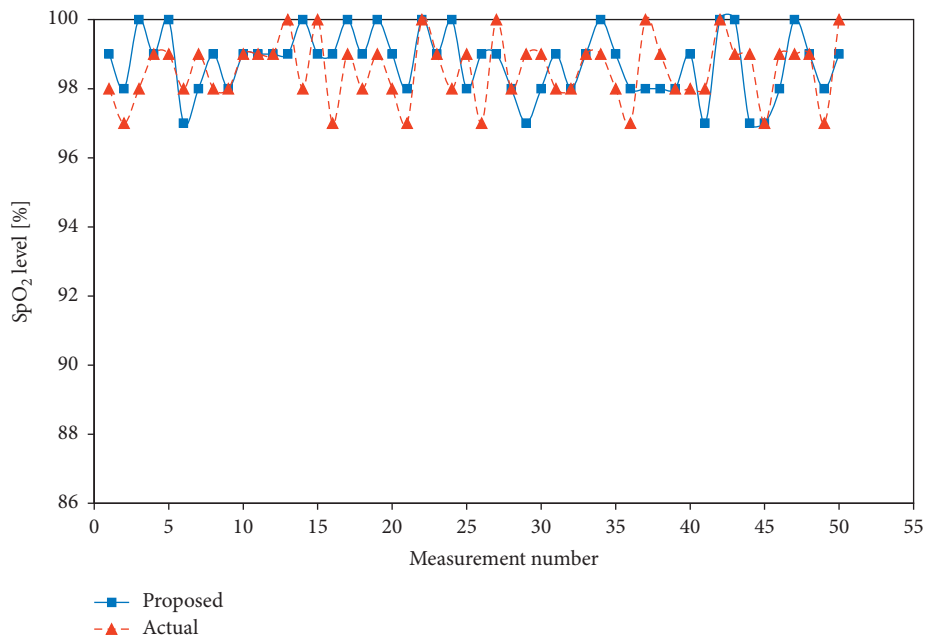


FIGURE 19: SpO₂ level comparison.

TABLE 5: Proposed system readings versus commercial device (medical thermometer) readings for body temperature.

Reading no.	Proposed system reading	Reference reading	Error (%)
1	34.75	34.6	0.43
2	35.5	35.4	0.28
3	33	32.9	0.30
4	35.13	35.1	0.09
5	33.5	33.7	0.59
6	35.63	35.6	0.08
7	36.63	36.5	0.36
8	35.5	35.6	0.28
9	36.5	36.4	0.27
10	35.75	35.9	0.42
11	35.25	35.4	0.42
12	35.63	35.8	0.47
13	34.75	34.8	0.14
14	35.75	35.9	0.42
15	34.5	34.5	0.00

TABLE 5: Continued.

Reading no.	Proposed system reading	Reference reading	Error (%)
16	32.9	32.8	0.30
17	34.75	34.7	0.14
18	33.63	33.6	0.09
19	36	36.1	0.28
20	35.5	35.5	0.00
21	33.5	33.3	0.60
22	35.5	35.4	0.28
23	34.33	34.2	0.38
24	33.25	33.2	0.15
25	35.13	35	0.37
26	35.5	35.5	0.00
27	33.75	33.9	0.44
28	36.13	36.2	0.19
29	35.63	35.6	0.08
30	34.88	34.8	0.23
31	33.88	34.1	0.65
32	35.8	36.1	0.83
33	34.75	34.8	0.14
34	35.75	36	0.69
35	35.63	35.7	0.20
36	36.13	36.1	0.08
37	35.25	35	0.71
38	35.38	35.3	0.23
39	35.25	35.1	0.43
40	35.88	36.1	0.61

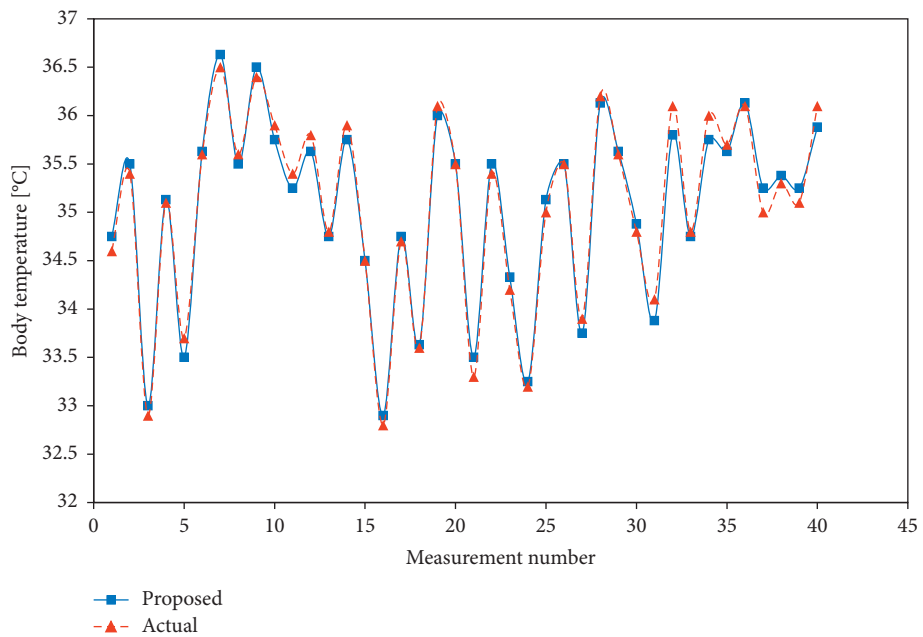


FIGURE 20: Body temperature comparison.

TABLE 6: Summary of the proposed system results for different health parameters.

Parameter	RMSE	MAE	MRE	R^2
HR	1.44	1.12	0.012	0.992
SpO ₂	1.13	0.92	0.009	0.074
Body temperature	0.13	0.11	0.003	0.982

TABLE 7: Comparison of HR error rates for the proposed and other solutions.

Work	RMSE	MAE	MRE (%)
[59]	2.34	2.17	2.93
[31]	3.87	3.4	4.93
Proposed	1.44	1.12	1.20

TABLE 8: Comparison of SpO₂ error rates for the proposed and other solutions.

Work	RMSE	MAE	MRE (%)
[31]	1.41	1.2	1.24
Proposed	1.13	0.92	0.93

TABLE 9: Comparison of body temperature error rates for the proposed and other solutions.

Work	RMSE	MAE	MRE (%)
[59]	0.70	0.65	0.66
[31]	0.61	0.50	1.66
Proposed	0.13	0.11	0.31

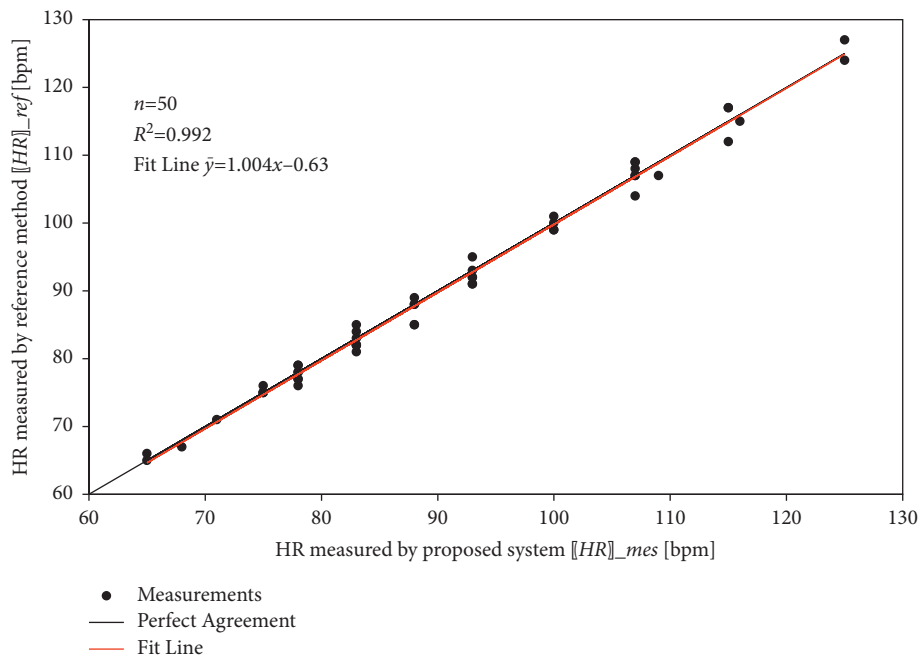


FIGURE 21: Linear relationship between measured and reference HR measurements.

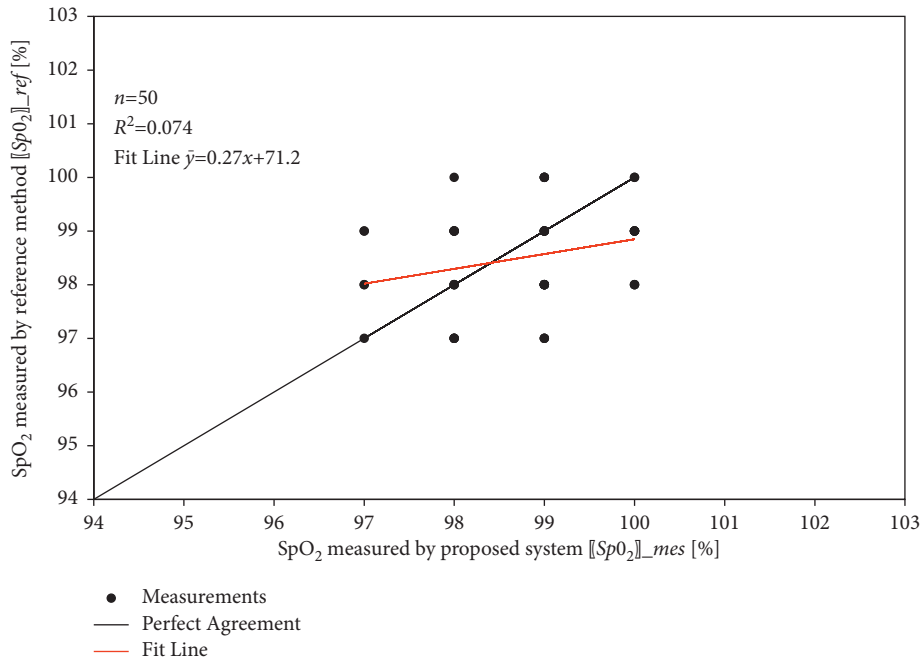


FIGURE 22: Linear relationship between measured and reference SpO₂ measurements.

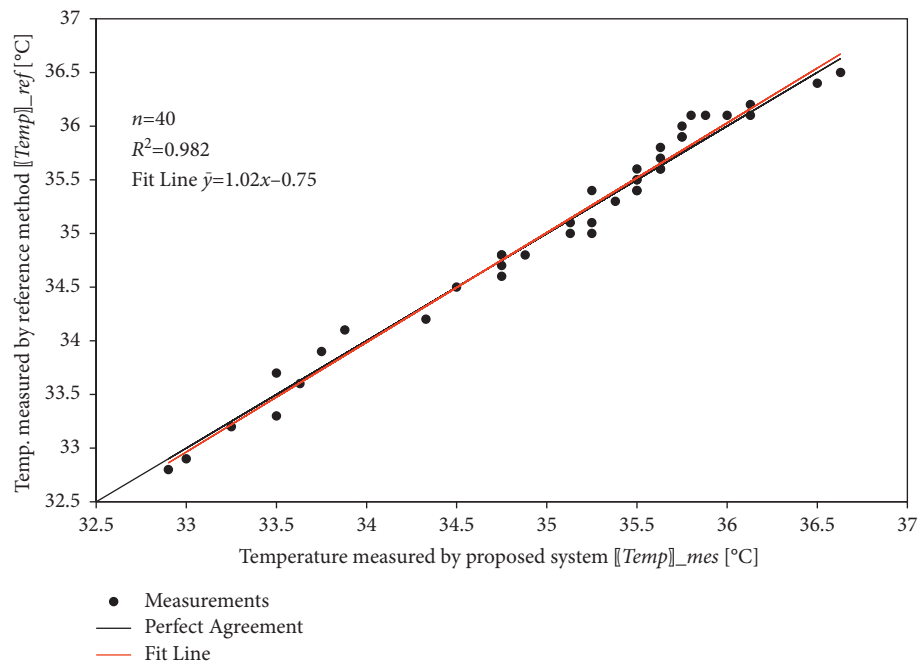


FIGURE 23: Linear relationship between measured and reference body temperature measurements.

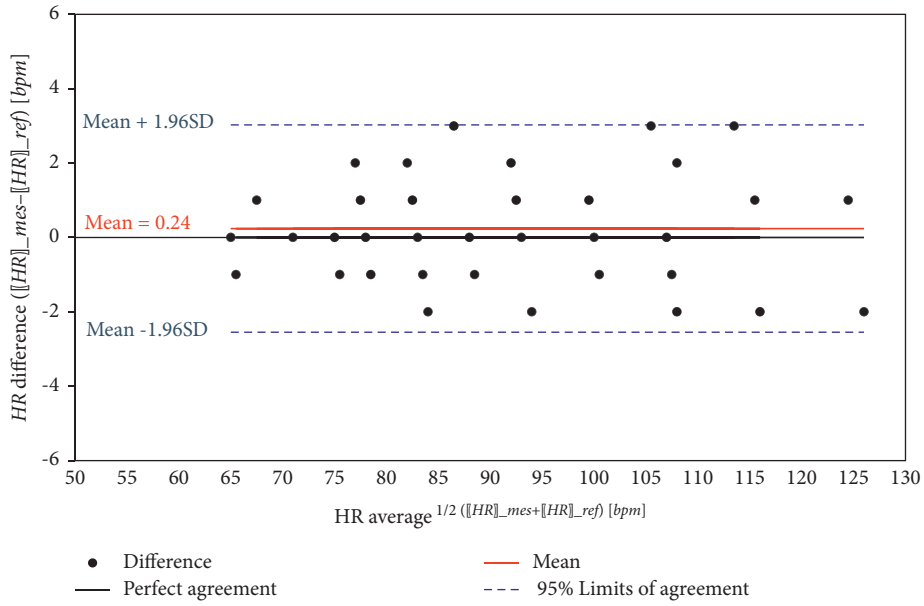


FIGURE 24: Bland-Altman plot of HR.

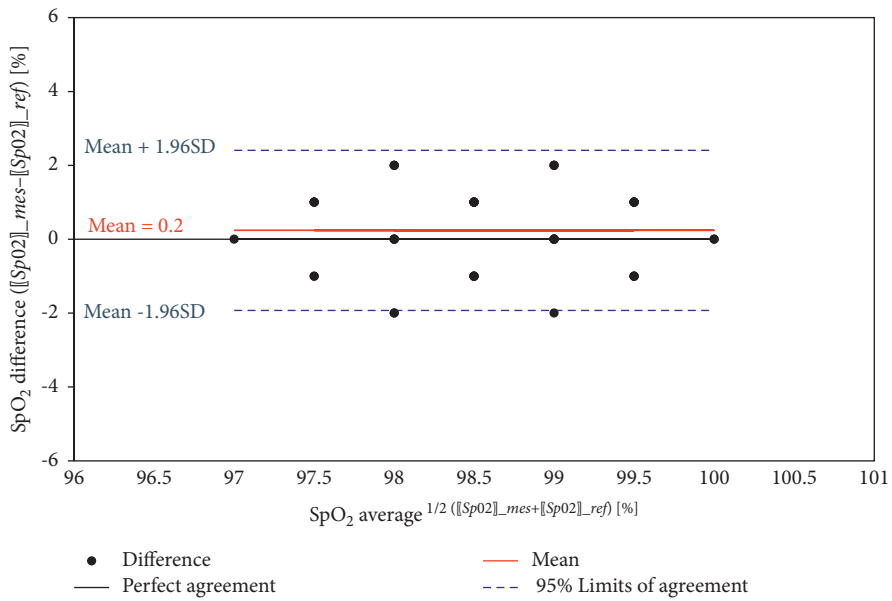


FIGURE 25: Bland-Altman plot of SpO₂.

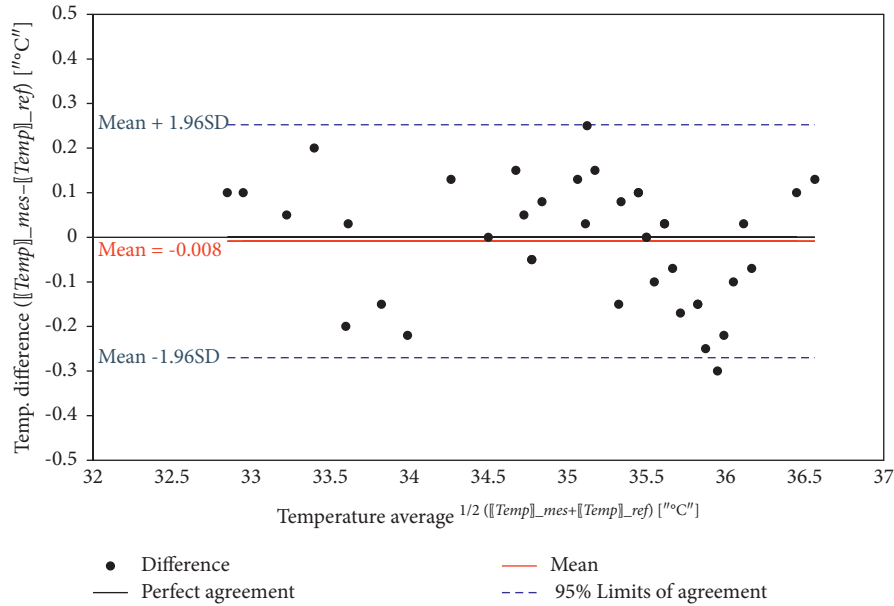


FIGURE 26: Bland–Altman plot of body temperature.

The above results indicate that the measurements of the proposed system closely coincide with the reference measurements of the commercial products.

7. Conclusions and Future Works

Health monitoring systems play a crucial role in the field of health care, diagnosis and early predicting issues regarding one's health. In addition, these systems are a means of cutting medical costs regarding periodical hospital checks and doctor visits. Thus, developing a system that delivers health data from the patient place to a relative or a medical specialist became a necessity with the increasing demand.

The main outcomes of this paper are as follows:

- (i) This paper presented a secure, low-cost, real-time, and trustable health monitoring system that provides a real-time monitoring dashboard for biological indicators within a secure environment using IoT and cloud computing.
- (ii) The proposed system adopts the AES algorithm to encrypt vital signals captured from sensors before sending them to the cloud for storage.
- (iii) An ESP8266 microcontroller is utilized to carry out the processing and encryption functions and connectivity to the cloud over Wi-Fi.
- (iv) The proposed system measurements are compared with those of a commercially available *High Care* medical device.
- (v) The results have revealed high agreement with the reference measurements.
- (vi) The RMSE, MAE, and MRE between the reference and the measured readings are computed as 1.44, 1.12, and 0.012, respectively, for HR, 1.13, 0.92, and

0.009, respectively, for SpO₂, and 0.13, 0.11, and 0.003, respectively, for body temperature. This indicates the high accuracy of the proposed system and its reliability to monitor the health and vital signs of patients and elders at home.

We have tried to guarantee an acceptable computational complexity for the proposed system by adopting the following approaches:

- (i) AES algorithm is selected to encrypt the sensor data, because it is simple to be implemented within the hardware using the appropriate software library, unlike other encryption algorithms, which may not be supported to be implemented in the hardware devices. In addition, it provides a good compromise between the speed of computations and the complexity.
- (ii) Wi-Fi technology is adopted in this solution, because it is faster and more reliable.
- (iii) We rely on the cloud servers as a backend for our solution as they are characterized by their super computational power, unlimited storage, high resource utilization, and low cost.
- (iv) Messages are sent from the device every 2 seconds. Each message contains a single read. Moreover, data transmission is based on TCP. So, there is no need for a retransmission mechanism, because the packets are automatically retransmitted if the transmission fails.
- (v) The solution has three layers: patient layer, cloud layer, and doctor layer. In real cases, where many patients are enrolled into the system, each patient will have his own IoT module to connect to the cloud server. Each patient will be located inside a different patient layer. In this case, the architecture

involves multiple instances of the patient layer, while the cloud layer and the doctor layer remain as single instances. The cloud and the doctor layers are constructed with high processing and large storage capabilities to support processing of a huge amount of data that could be received from the patient layers.

- (vi) In this work, we employed the Firebase cloud server, and a real-time cloud database acquired by Google and intended for IoT solutions.

However, some limitations of the proposed solution may be encountered, which may make the device fail. The failure cases include the following:

- (i) Loss of Internet connectivity
- (ii) Loss of the direct Wi-Fi link between the node and the local access point (e.g., wrong credentials)
- (iii) Loss or drop of the power source, such that the nodes or the sensors cannot be powered up
- (iv) Misconfiguration or utilization of the sensors in a wrong way
- (v) Operation at exceeded limits for sensors that are defined in the datasheet

Future research directions may include further development of the proposed system to monitor more biomedical aspects such as heart activity, blood pressure, and blood glucose by integrating appropriate sensors. In addition, automated diagnosis for common diseases may be integrated with the proposed device. Moreover, a framework to process encrypted data may be developed to provide decision-making about the status of individuals, while data is encrypted.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Sajadieh, O. W. Nielsen, V. Rasmussen, H. O. Hein, S. Abedini, and J. F. Hansen, "Increased heart rate and reduced heart-rate variability are associated with subclinical inflammation in middle-aged and elderly subjects with no apparent heart disease," *European Heart Journal*, vol. 25, no. 5, pp. 363–370, 2004.
- [2] M. Adil, M. A. Almaiah, A. Omar Alsayed, and O. Almomani, "An anonymous channel categorization scheme of edge nodes to detect jamming attacks in wireless sensor networks," *Sensors*, vol. 20, no. 8, p. 2311, 2020.
- [3] A. Doulamis, N. Doulamis, A. Angeli et al., "A non-invasive photonics-based device for monitoring of diabetic foot ulcers: architectural/sensorial components & technical specifications," *Inventions*, vol. 6, no. 2, p. 27, 2021.
- [4] A. Sedik, M. Hammad, F. E. Abd El-Samie, B. B. Gupta, and A. A. Abd El-Latif, "Efficient deep learning approach for augmented detection of coronavirus disease," *Neural Computing and Applications*, pp. 1–18, 2021.
- [5] M. Hammad, A. M. Ilyasu, A. Subasi, E. S. L. Ho, and A. A. Abd El-Latif, "A multitier deep learning model for arrhythmia detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2020.
- [6] A. Sedik, A. M. Ilyasu, B. Abd El-Rahiem et al., "Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections," *Viruses*, vol. 12, no. 7, p. 769, 2020.
- [7] Beurer, Heart Rate Monitors-eurer, 2021, <https://www.beurer.com/web/gb/products/active/sport-and-activity/pulsuhren/>.
- [8] "A complete guide to the apple watch heart rate monitor," 2021, <https://www.wareable.com/apple/apple-watch-heart-rate-monitor-guide-340>.
- [9] Alivecor, "kardiamobile," 2021, <https://www.alivecor.com/>.
- [10] M. Adil, R. Khan, J. Ali, B.-H. Roh, Q. T. H. Ta, and M. A. Almaiah, "An energy proficient load balancing routing scheme for wireless sensor networks to maximize their lifespan in an operational environment," *IEEE Access*, vol. 8, pp. 163209–163224, 2020.
- [11] A. I. Siam, A. Abou Elazm, N. A. El-Bahnasawy, G. El Banby, and F. E. A. E.-S. El-Samie, "Smart health monitoring system based on IoT and cloud computing," *Menoufia Journal of Electronic Engineering Research*, vol. 28, no. 1, pp. 37–42, 2019.
- [12] S. Dağtas, G. Pekhteryev, Z. Sahinoğlu, H. Cam, and N. Challa, "Real-time and secure wireless health monitoring," *International Journal of Telemedicine and Applications*, vol. 2008, pp. 1–10, Article ID 135808, 2008.
- [13] M. A. Almaiah and A. Al-Khasawneh, "Investigating the main determinants of mobile cloud computing adoption in university campus," *Education and Information Technologies*, vol. 25, no. 4, pp. 3087–3107, 2020.
- [14] Y. A. Qadri, A. Nauman, Y. B. Zikria, A. V. Vasilakos, and S. W. Kim, "The future of healthcare internet of things: a survey of emerging technologies," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1121–1167, 2020.
- [15] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K.-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.
- [16] A. I. Siam, H. A. El-khobby, H. S. Abd Elkader, M. M. Abdelnaby, H. S. A. Elkader, and M. M. Abdelnaby, "Enhanced data security model for cloud computing platform," *International Journal of Scientific Research in Science, Engineering and Technology*, vol. 1, no. 4, pp. 450–460, 2015.
- [17] R. A. Nafea and M. Amin Almaiah, "Cyber security threats in cloud: literature review," in *Proceedings of the 2021 International Conference on Information Technology (ICIT)*, pp. 779–786, Amman, Jordan, July 2021.
- [18] V. Casola, A. Castiglione, K.-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.
- [19] O. Ali, A. Shrestha, J. Soar, and S. F. Wamba, "Cloud computing-enabled healthcare opportunities, issues, and applications: a systematic review," *International Journal of Information Management*, vol. 43, pp. 146–158, 2018.
- [20] M. A. Almaiah, A. Al-Zahrani, O. Almomani, and A. K. Alhwaitat, "Classification of cyber security threats on mobile devices and applications," *Studies in Big Data*, pp. 107–123, 2021.

- [21] X. Yi, A. Bouguettaya, D. Georgakopoulos, A. Song, and J. Willemsen, "Privacy protection for wireless medical sensor data," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 3, pp. 369–380, 2016.
- [22] J. Daemen and V. Rijmen, *The Design of Rijndael: AES-The Advanced Encryption Standard*, Springer Science & Business Media, Berlin, Germany, 2013.
- [23] R. Levitan, "The Infection That's Silently Killing Coronavirus Patients," *The New York Times*, New York, NY, USA, 2020, <https://web.archive.org/web/20200421002343/https://www.nytimes.com/2020/04/20/opinion/coronavirus-testing-pneumonia.html>.
- [24] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, evaluation and treatment coronavirus (COVID-19)," in *Statpearls [Internet]* StatPearls Publishing, Treasure Island, FL, USA, 2020.
- [25] A. I. Siam, A. A. Elazm, N. A. El-Bahnasawy, G. M. El Banby, and F. E. Abd El-Samie, "PPG-based human identification using Mel-frequency cepstral coefficients and neural networks," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26001–26019, 2021.
- [26] A. I. Siam, A. Sedik, W. El-Shafai et al., "Biosignal classification for human identification based on convolutional neural networks," *International Journal of Communication Systems*, vol. 34, no. 7, 2021.
- [27] M. W. Wukitsch, M. T. Petterson, D. R. Tobler, and J. A. Pologe, "Pulse oximetry: analysis of theory, technology, and practice," *Journal of Clinical Monitoring*, vol. 4, no. 4, pp. 290–301, 1988.
- [28] S.-S. Oak and P. Aroul, *How to Design Peripheral Oxygen Saturation (Spo2) and Optical Heart Rate Monitoring (OHRM) Systems Using the Afe4403*, Texas Instruments, Dallas, Texas, USA, 2015.
- [29] M. Integrated, *Recommended Configurations and Operating Profiles for max30101/max30102 EV Kits*, Maxim Integrated, San Jose, CA, USA, 2018.
- [30] S. M. Riazul Islam, D. Kwak, M. Humaun Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE access*, vol. 3, pp. 678–708, 2015.
- [31] M. M. Ali, S. Haxha, M. M. Alam, C. Nwibor, and M. Sakel, "Design of internet of things (IoT) and android based low cost health monitoring embedded system wearable sensor for measuring SpO₂, heart rate and body temperature simultaneously," *Wireless Personal Communications*, vol. 111, no. 4, pp. 2449–2463, 2020.
- [32] J. Mohammed, C.-H. Lung, A. Ocneanu, A. Thakral, C. Jones, and A. Adler, "Internet of things: remote patient monitoring using web services and cloud computing," in *Proceedings of the 2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)*, pp. 256–263, Taipei, Taiwan, September 2014.
- [33] M. Al-khafajiy, T. Baker, C. Chalmers et al., "Remote health monitoring of elderly through wearable sensors," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24681–24706, 2019.
- [34] M. Gupta, V. Patchava, and V. Menezes, "Healthcare based on IoT using Raspberry Pi," in *Proceedings of the 2015 IEEE International Conference on Green Computing and Internet of Things (ICGCIoT)*, pp. 796–799, Delhi, India, October 2015.
- [35] A. M. Ghosh, D. Halder, and S. A. Hossain, "Remote health monitoring system through IoT," in *Proceedings of the 2016 IEEE 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 921–926, Dhaka, Bangladesh, May 2016.
- [36] J. Lloret, A. Canovas, S. Sendra, and L. Parra, "A smart communication architecture for ambient assisted living," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 26–33, 2015.
- [37] A. Elsts, X. Fafoutis, P. Woznowski et al., "Enabling healthcare in smart homes: the SPHERE IoT network infrastructure," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 164–170, 2018.
- [38] H. Moustafa, E. M. Schooler, G. Shen, and S. Kamath, "Remote monitoring and medical devices control in eHealth," in *Proceedings of the 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–8, New York, NY, USA, October 2016.
- [39] S. J. Park, M. Subramaniyam, S. E. Kim et al., "Development of the elderly healthcare monitoring system with IoT," in *Advances in Human Factors and Ergonomics in Healthcare*, pp. 309–315, Springer, Berlin, Germany, 2017.
- [40] S. F. Khan, "Health care monitoring system in Internet of Things (IoT) by using RFID," in *Proceedings of the 2017 6th International Conference on Industrial Technology and Management (ICITM)*, pp. 198–204, Cambridge, UK, March 2017.
- [41] V. Mighali, L. Patrono, M. L. Stefanizzi, J. J. P. C. Rodrigues, and P. Solic, "A smart remote elderly monitoring system based on IoT technologies," in *Proceedings of the 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 43–48, Milan, Italy, July 2017.
- [42] S. Tuli, N. Basumatary, S. S. Gill et al., "Healthfog: an ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated Iot and fog computing environments," *Future Generation Computer Systems*, vol. 104, pp. 187–200, 2020.
- [43] A. H. Sodhro, A. S. Malokani, G. H. Sodhro, M. Muzammal, and L. Zongwei, "An adaptive QoS computation for medical data processing in intelligent healthcare applications," *Neural Computing and Applications*, vol. 32, no. 3, pp. 723–734, 2020.
- [44] A. Alabdulatif, I. Khalil, A. R. M. Forkan, and M. Atiqzaman, "Real-time secure health surveillance for smarter health communities," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 122–129, 2019.
- [45] X. Yi, J. Willemsen, and F. Nat-Abdesselam, "Privacy-preserving wireless medical sensor network," in *Proceedings of the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 118–125, Melbourne, Australia, July 2013.
- [46] A. I. Siam, N. A. El-Bahnasawy, G. M. El Banby, A. Abou Elazm, and F. E. Abd El-Samie, "Efficient video-based breathing pattern and respiration rate monitoring for remote health monitoring," *Journal of the Optical Society of America A*, vol. 37, no. 11, p. C118, 2020.
- [47] C. Massaroni, D. S. Lopes, D. Lo Presti, E. Schena, and S. Silvestri, "Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: a single camera approach," *Journal of Sensors*, vol. 2018, Article ID 4567213, 13 pages, 2018.
- [48] G. O. Ganfure, "Using video stream for continuous monitoring of breathing rate for general setting," *Signal, Image and Video Processing*, vol. 13, no. 7, pp. 1395–1403, 2019.
- [49] M. Hu, G. Zhai, D. Li et al., "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under

- sleep situation,” *PLoS One*, vol. 13, no. 1, Article ID e0190466, 2018.
- [50] S. Sanyal and K. K. Nundy, “Algorithms for monitoring heart rate and respiratory rate from the video of a user’s face,” *IEEE Journal of translational engineering in health and medicine*, vol. 6, pp. 1–11, 2018.
- [51] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, “Cuffless blood pressure estimation algorithms for continuous health-care monitoring,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, 2017.
- [52] J. Sola, M. Proenca, D. Ferrario et al., “Noninvasive and nonocclusive blood pressure estimation via a chest sensor,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3505–3513, 2013.
- [53] Maxim Integrated, “MAX30102 datasheet,” 2021, <https://datasheets.maximintegrated.com/en/ds/MAX30102.pdf>.
- [54] Maxim Integrated, “DS18B20 datasheet,” 2021, <https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf>.
- [55] NodeMCU, “NodeMCU documentation,” 2021, <https://nodemcu.readthedocs.io/>.
- [56] ESP8266 Crypto Library, 2021, <https://github.com/intrbiz/arduino-crypto>.
- [57] M. H. Qasem, N. Obeid, A. Hudaib, M. A. Almaiah, A. Al-Zahrani, and A. Al-Khasawneh, “Multi-agent system combined with distributed data mining for mutual collaboration classification,” *IEEE Access*, vol. 9, pp. 70531–70547, 2021.
- [58] Firebase Cloud Database, 2021, <http://www.firebaseio.com>.
- [59] M. M. Islam, A. Rahaman, and M. R. Islam, “Development of smart healthcare monitoring system in IoT environment,” *SN Computer Science*, vol. 1, no. 3, p. 185, 2020.

Research Article

A Computer-Aided Diagnosis System Using Deep Learning for Multiclass Skin Lesion Classification

Mehak Arshad,¹ Muhammad Attique Khan ,¹ Usman Tariq ,² Ammar Armghan,³ Fayadh Alenezi ,³ Muhammad Younus Javed,¹ Shabnam Mohamed Aslam ,⁴ and Seifedine Kadry⁵

¹Department of Computer Science, HITEC University Taxila, Taxila, Pakistan

²College of Computer Engineering and Science, Prince Sattam Bin Abdulaziz University, Al-Kharaj, Saudi Arabia

³Department of Electrical Engineering, Jouf University, Sakaka 75471, Saudi Arabia

⁴Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia

⁵Faculty of Applied Computing and Technology, Noroff University College, Kristiansand, Norway

Correspondence should be addressed to Muhammad Attique Khan; attique@ciitwah.edu.pk and Shabnam Mohamed Aslam; s.aslam@mu.edu.sa

Received 7 August 2021; Revised 28 October 2021; Accepted 10 November 2021; Published 6 December 2021

Academic Editor: Ahmed A. Abd El-Latif

Copyright © 2021 Mehak Arshad 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 USA, each year, almost 5.4 million people are diagnosed with skin cancer. Melanoma is one of the most dangerous types of skin cancer, and its survival rate is 5%. The development of skin cancer has risen over the last couple of years. Early identification of skin cancer can help reduce the human mortality rate. Dermoscopy is a technology used for the acquisition of skin images. However, the manual inspection process consumes more time and required much cost. The recent development in the area of deep learning showed significant performance for classification tasks. In this research work, a new automated framework is proposed for multiclass skin lesion classification. The proposed framework consists of a series of steps. In the first step, augmentation is performed. For the augmentation process, three operations are performed: rotate 90, right-left flip, and up and down flip. In the second step, deep models are fine-tuned. Two models are opted, such as ResNet-50 and ResNet-101, and updated their layers. In the third step, transfer learning is applied to train both fine-tuned deep models on augmented datasets. In the succeeding stage, features are extracted and performed fusion using a modified serial-based approach. Finally, the fused vector is further enhanced by selecting the best features using the skewness-controlled SVR approach. The final selected features are classified using several machine learning algorithms and selected based on the accuracy value. In the experimental process, the augmented HAM10000 dataset is used and achieved an accuracy of 91.7%. Moreover, the performance of the augmented dataset is better as compared to the original imbalanced dataset. In addition, the proposed method is compared with some recent studies and shows improved performance.

1. Introduction

The development of skin cancer has risen throughout the previous decade [1]. Ultraviolet rays in the sun damage the skin over time and cause cancer cells to develop [2]. Usually, such conditions have hidden risks that lead to a lack of confidence and psychological distress in humans and to skin cancer risks. Several types of skin cancer exist,

including basal cells, melanoma, actinic keratosis, and squamous cell carcinoma [3]. The squamous cell carcinoma is contrasted against actinic keratosis (solar keratosis) [4]. Each year, the incidence rate of both melanoma and nonmelanoma continues to grow [2]. The deadliest form of skin cancer is melanoma and quickly spread to other body parts due to the malignancy of neural crest neoplasia of melanocytes [5].

In the United States, almost 5.4 million new cases of skin cancer are detected each year. Due to melanoma, more than 10,000 deaths are registered every year in the USA [6]. In the USA, 104,350 new cases of skin cancers were diagnosed during the year 2019, where the numbers of deaths were 7230. In the year 2020, 196,060 Americans are diagnosed with melanoma. According to these facts, melanoma cases are increasing approximately 2% [7]. Recently, in the year 2021, 207.39 K peoples are diagnosed with skin cancer whereas the numbers of deaths are 70.18 K. According to the facts, when the lesion is detected earlier, the survival rate increases approximately 98% [7]. The summary of diagnoses and deaths due to skin cancer is illustrated in Figure 1.

Dermatologists diagnose malignant lesions via a dermoscopic visual examination technique [8]. Diagnosis of skin cancer using dermoscopy is challenging due to various textures and wounds [9]. However, the manual inspection of dermoscopic images makes it difficult to diagnose skin cancer with better accuracy. The accuracy of the lesion diagnosis depends on the dermatologist's experience [9]. Few other techniques are available for diagnosing skin cancer, such as biopsy [7] and macroscopic [10]. Due to the complex nature of skin lesions, the clinical methods need more attention and time [11, 12].

The computer-based detection (CAD) techniques are introduced by several researchers in medical imaging [7, 13]. They introduced CAD techniques for several cancers such as skin cancer [14], brain tumor [15, 16], lung cancer [17, 18], COVID-19 [19, 20], and more [21–23]. A simple CAD technique consists of four key steps such as preprocessing of input images, detection of infected parts, features extraction, and classification. A computerized method can be helpful as a second opinion for dermatologists to verify the manual diagnosis results [8]. The advancement in machine learning, like deep learning, has shown much achievement in medical imaging in the last couple of years. Convolutional Neural Network (CNN) is a form of deep learning used for automated features extraction [6]. A convolutional neural network is a computer vision technique that automatically distinguishes and recognizes images' features [24]. Due to its high accuracy, it has attracted interest in medical image processing, agriculture, biometric, and surveillance, to name a few. A simple CNN typically entails a series of layers such as a convolutional layer, ReLU layer [25], normalization layer, pooling layer [26], fully connected layer, and Softmax layer [27]. In many techniques, researchers used some pretrained deep learning models for the classification tasks. A few publically available pretrained deep learning models are AlexNet, VGG, GoogleNet, InceptionV3, and ResNet to name a few [28]. They used these models through transfer learning [7]. Few researchers used feature selection and fusion techniques to improve recognition accuracy [29, 30].

The computer-aided diagnostic systems can allow dermatologists and physicians to make decisions, decrease diagnostic costs, and increase diagnostics reliability [31]. An automated skin lesion identification mechanism is challenging due to several challenges such as changing appearance and imbalanced datasets to name a few [32]. Chaturvedi et al. [6] presented an automated framework for

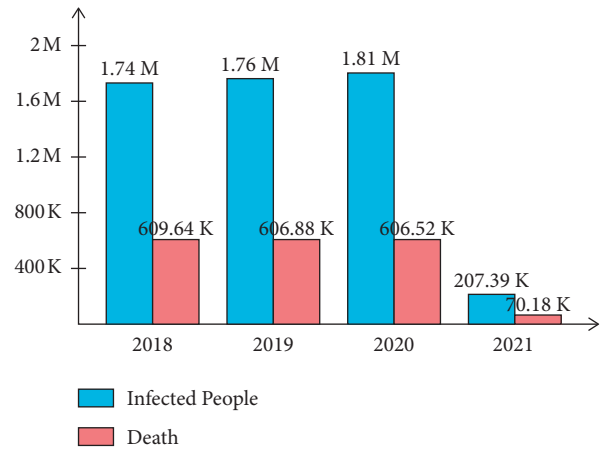


FIGURE 1: Graph of infected and death cases of skin lesion.

multiclass skin cancer classification. Five steps were involved in the presented method: dataset preprocessing, classification models (pretrained deep learning), fine-tuning, feature extraction, and performance evaluation. During the evaluation process, it is noted that the maximum accuracy of 93.20% was achieved for an individual model (ResNet-101), whereas a complete precision of 92.83% was performed on the ensemble model (InceptionResNetV2 + ResNet-101). In the end, they concluded that the training of deep learning models with the best setup of hyperparameters could be performed better than even ensemble models. Hsin et al. [33] presented the automatic lightweight diagnostic algorithm for skin lesion diagnosis. The presented algorithm was more reliable, feasible, and easy to use. For the experimental process, the HAM10000 dataset was used and achieved an accuracy of 85.8%. Besides, this method was tested on a five-class KCGMH dataset and achieved an accuracy of 89.5%. Kumar et al. [9] presented an automated electronic device. They considered numerous challenges such as skin cancer injuries, skin colors, asymmetric skin, and the shape of the area affected. They used fuzzy C-means to divide homogeneous image regions. Then, some texture features are extracted and trained with the Differential Evolution (DE) algorithm. The experimental process was conducted on HAM10000 and achieved an accuracy of 97.4%.

Afshar et al. [8] presented a computerized method for lesion localization and identification. For the lesion localization, they used RCNN architecture and extract deep features. Later, the best features are selected using Newton-Raphson (IcNR) and artificial bee colony (ABC) optimization. Daghrir et al. [5] developed a hybrid approach for diagnosing suspect lesions that may be checked for melanoma skin cancer. They used a coevolutionary neural network and two classical classifiers in three different methods. Shayini [2] presented a classification framework using geometric and textural information. They used ANN for the final features classification. Results showed improved accuracy as compared to the existing techniques. Akram et al. [7] presented deep learning-based lesion segmentation and classification process. They used Mask RCNN architecture

for lesion segmentation. Later, a 24-layered CNN architecture was designed for the multiclass skin lesion classification.

Moreover, many other techniques are introduced such as deep learning and improved moth-flame optimization [34], teledermatology-based architecture [35], hierarchical three-step deep framework [35], and more [36, 37].

1.1. Challenges. Several challenges affect the multiclass lesion classification accuracy. As compared to binary class classification, the multiclass problem is a complex and challenging recognition process. The following challenges are considered in this research work:

- (i) Classifying multiple skin lesions into a correct class is challenging due to the high similarity among different lesions.
- (ii) The imbalanced dataset classes increase the probability of a higher sample class.
- (iii) Multiclass skin lesion types have similar shapes, colors, and textures, which also extract similar features. In the later stage, those features are classified into an incorrect skin class.
- (iv) In the fusion step, multiproperties features are fused in one matrix for better accuracy, but it is a high chance that several redundant features are also added. This kind of problem later increases the computational time.
- (v) In the feature extraction step, several essential features are also removed, which may cause a problem of misclassification. Therefore, a good feature optimization technique is required [38].

1.2. Major Contributions. In this work, an automated technique has been proposed for multiclass skin lesion classification. The significant contributions in this work are as follows:

- (i) Intra-class pixel change operations are implemented for data augmentation based on the left to right flip, up-to-down flip, and rotation at 90 degrees. This step shifts entire image pixels for differentiating the images from each other for a fair training of a deep model.
- (ii) A modified serial-based approach is proposed for the fusion of extracted deep features.
- (iii) A novel skewness-controlled SVR approach is proposed for the best feature selection. The best-selected features are finally classified using supervised learning algorithms.

The rest of the manuscript is organized in the following order. Section 2 presented the proposed methodology including deep feature, selection of best features, and fusion process. Results and comparisons with existing techniques are presented in Section 3. Finally, the manuscript is concluded in Section 4.

2. Proposed Methodology

For the multiclass skin lesion classification, a new framework was proposed using deep learning and features selection. The proposed framework consists of a series of steps such as data augmentation, model fine-tuning, transfer learning, feature extraction, the fusion of extracted features, and selection of best features. In the augmentation phase, three operations are performed: rotate 90, right-left flip, and up and down flip. In the fine-tuning model step, two models are opted, such as ResNet-50 and ResNet-101, and updated their layers. Later, transfer learning is applied to train both fine-tuned deep models on augmented datasets. In the subsequent step, features are extracted and performed fusion using a modified serial-based approach. Finally, the fused vector is further enhanced by selecting the best features using the skewness-controlled SVR approach. The main architecture diagram of the proposed framework is illustrated in Figure 2.

2.1. Data Augmentation. Data augmentation is a vital information extension approach in machine learning (ML). Data augmentation showed much importance in deep learning due to a massive amount of data for training a model. In this article, the HAM10000 dataset is selected for the experimental process. This dataset consists of seven highly imbalanced classes. Initially, the HAM10000 dataset includes more than 10,000 images of seven skin classes such as 6705 images of melanocytic nevi, 1113 images in melanomas, 1099 images in benign keratoses, 514 images in basal cell carcinomas, 327 images of actinic keratoses, 142 images in vascular lesions, and 115 images in dermatofibromas [39]. From this information, it is noted that few classes are highly imbalanced; therefore, it is essential to balance this dataset. On imbalanced datasets, the deep learning models are not trained for better performance. A few sample images are shown in Figure 3.

Three operations are performed in the data augmentation phase: rotate 90, right-left flip (LR), and up and down flip (UD). These operations are applied multiple times until the number of images in each class reached 6000. In the end, the numbers of images in the newly updated dataset are 42,000, which are previously 10,000. Mathematically, these operations are performed as follows.

Consider an image dataset $\rho = \{a_1, \dots, a_k\}$ [40], where $a_k \in U$ is an example image from the dataset. Let a_k have fully N pixels; then, the homogeneous pixel matrix coordinates C_k or a_k is defined as follows:

$$C_k = \begin{bmatrix} Y_1 & Z_1 & 1 \\ Y_2 & Z_2 & 1 \\ \vdots & \vdots & 1 \\ Y_n & Z_n & 1 \end{bmatrix}, \quad (1)$$

where each row of single-pixel indicates the exact coordinates. Consider that the size of an input image is $256 \times 256 \times 3$, represented by $U_{i,j,k}$ having i th rows, j th columns, and k th channels, where $U_{i,j} \in R^{i \times j}$. The flip-up (UD) operation is formulated as follows [41]:

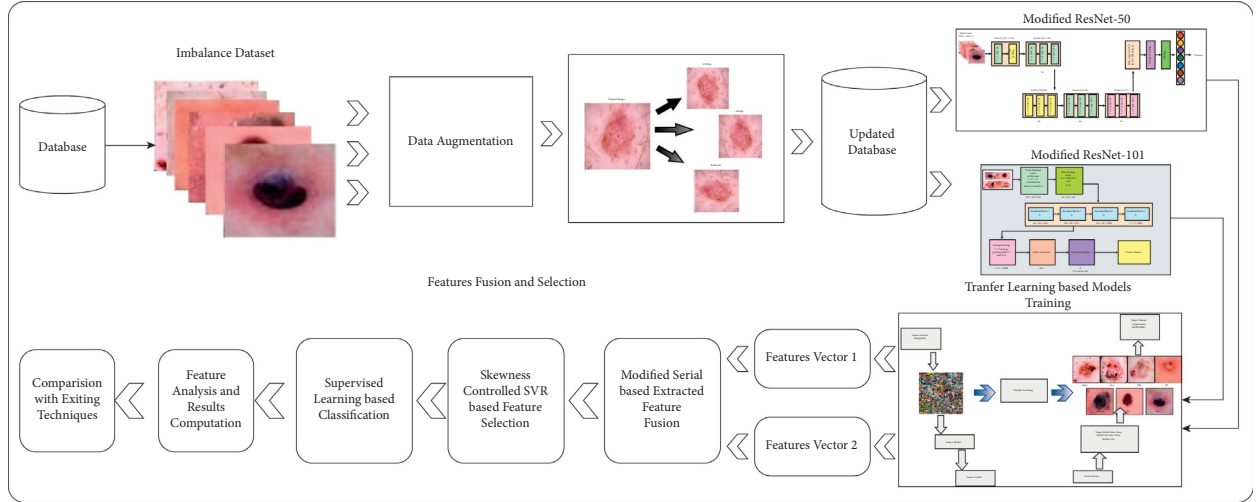


FIGURE 2: Architecture of proposed methodology.

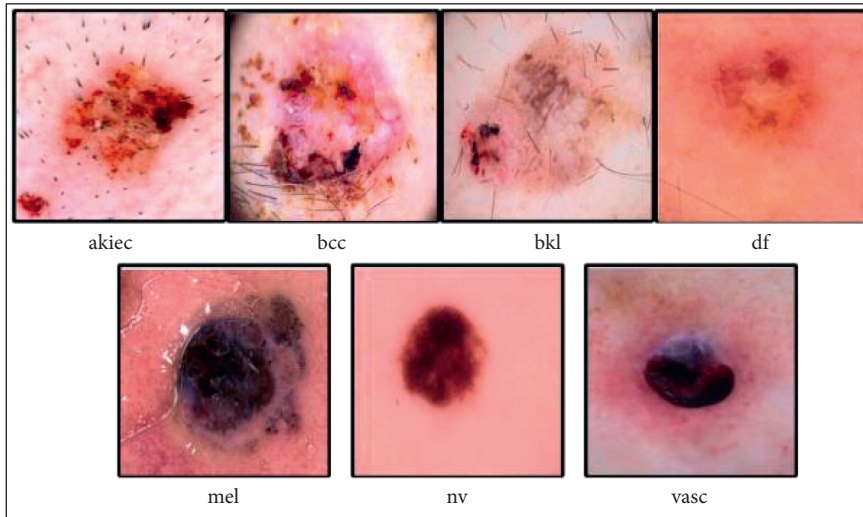


FIGURE 3: Sample skin lesion images of the HAM10000 dataset [7].

$$U^t = U_{j,i}^t, \quad (2)$$

where U^t denotes the transposition of the original image. This image is further updated as follows:

$$U^V = U_{(m+1-i)j}, \quad (3)$$

where U^V denotes the vertical flip image. The horizontal flip (LR) operation is performed as follows:

$$U^H = U_{i(n+1-j)}, \quad (4)$$

where U^H denotes the horizontal flip image. The third operation, named rotate 90, is formulated as follows:

$$\text{Rot} = \begin{bmatrix} \cos \beta & -\sin \beta & 0 \\ \sin \beta & \cos \beta & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5)$$

where Rot denotes the rotation matrix of the image. Visually, these operations are illustrated in Figure 4. This figure shows that three operations are performed on each original image: vertical flip (UD), horizontal flip (LR), and rotate 90.

2.2. Convolutional Neural Networks. A convolutional neural network (CNN) is a computer vision technique that automatically distinguishes and recognizes images' features [24]. A simple CNN architecture for image classification is illustrated in Figure 5. In this figure, skin lesion images are considered as input, passed to the convolutional layer. In this layer, weights are transformed into features that are further refined into the pooling layer. Later, the features are transformed into 1D in a fully connected layer. The features of this layer are finally classified through the Softmax layer.

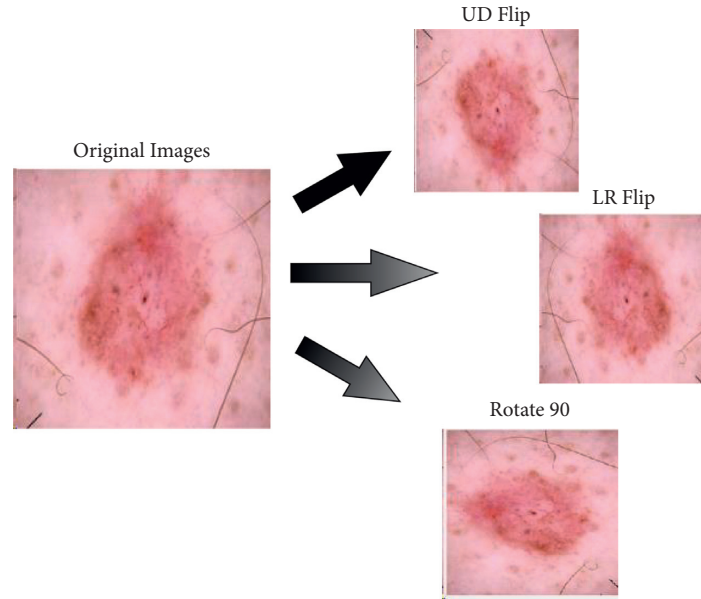


FIGURE 4: Flip operations in data augmentation.

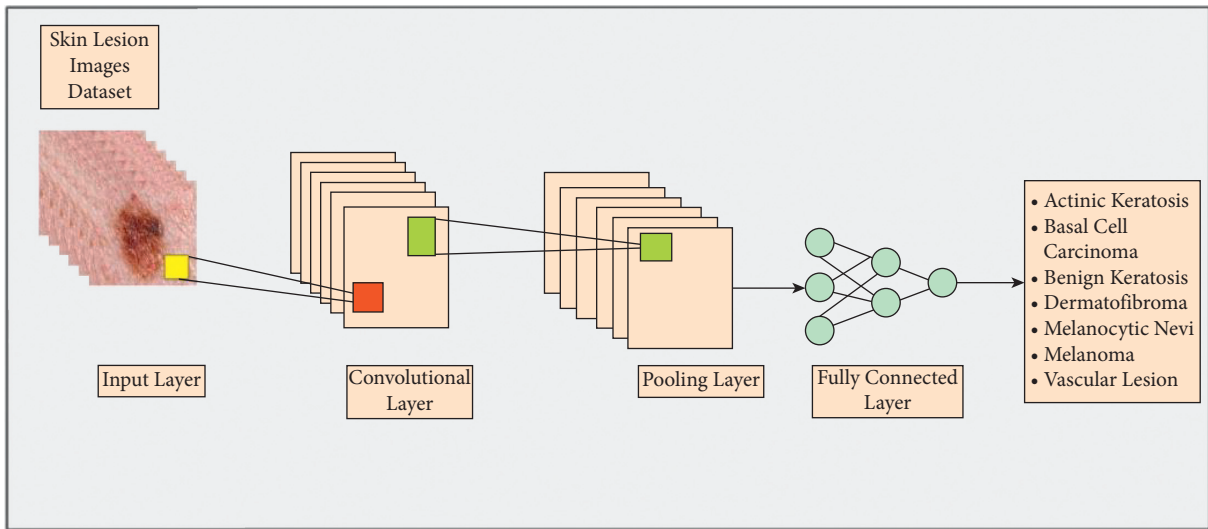


FIGURE 5: A general structure of a CNN model for skin lesion classification.

2.3. *Transfer Learning.* Transfer learning is a technique to define applied knowledge based on one or more source activities. Consider a domain M consisting of two parts:

$$M = \{z, Q(Z)\}, \quad (6)$$

where y is a feature space, and the distribution is marginal:

$$q(Z), \quad Z = \{z_1, \dots, z_n\}, z_i \in Z. \quad (7)$$

Given a two-component task U and X ,

$$\begin{aligned} U &= \{\varphi, Q(X|Z)\} = \{\varphi, \delta\}; \\ X &= \{x_1, \dots, x_n\}, \quad x_i \in \varphi, \end{aligned} \quad (8)$$

where φ is label space containing a prediction function; then, δ is trained as

$$(x_i, z_i), \quad z_i \in Z, x_i \in \varphi. \quad (9)$$

Each vector of features in the M domain and δ represents an appropriate label.

$$\delta(x_i) = (z_i). \quad (10)$$

Suppose the source domain M_S and an objective domain M_T , where $M = \{z, Q(Z)\}$ and the task is U_S and U_T , where $U = \{\varphi, Q(\varphi|Z)\}$. Hence, TL is defined as follows:

- (i) $y_S \neq y_T$: different feature space
- (ii) $Q(Z_S) \neq Q(Z_T)$: different marginal possibilities
- (iii) $\varphi_S \neq \varphi_T$: different label spaces
- (iv) $Q(Z_S|Y_S) \neq Q(Z_T|Y_T)$: different conditional probabilities

Visually, this process is illustrated in Figure 6. This figure describes that the ImageNet dataset used as source data has 1000 object classes. After transferring knowledge of the source model to the target model, the weights and labels are updated according to the target dataset. The HAM10000 skin cancer dataset is utilized as a target dataset with seven skin classes in this work.

2.4. Fine-Tuned ResNet-50 Deep Features. Residual Network (ResNet) is a traditional neural network model for many computer vision tasks utilized as an integrated network element. The network has a depth of 50 layers and a size of 224×224 pixels in the input [42]. When it comes to residual learning functions, ResNet may reformulate network layers given an input mapping reference. The layers are stacked directly within ResNet. The basic idea of ResNet-50 is to use identity mapping to anticipate what is required to obtain the final prediction of previous layer output [43]. ResNet-50 reduces the disappearing gradient effect by applying an alternative bypass shortcut. It may help the model overcome the overfitting training problem. Visually, it is shown in Figure 7.

Moreover, a complete architecture is also given in Figure 8. This figure describes that five residual blocks are used in this network, and in each residual block, multiple layers are added to convolve hidden layer features. Overall, this network includes 50 deep layers with a 7×7 input layer receptive field, followed by a max-pooling layer of 3×3 kernel size.

The last fully connected (FC) layer is removed, and a new FC layer is added in the fine-tuning process. Then, the new FC layer is connected with the Softmax layer and final classification output layer. The fine-tuned architecture is shown in Figure 9. This figure describes that the augmented skin lesion dataset is considered an input to this network, and in the output, seven classes of different skin cancer types are gotten. After this, the TL technique is employed to train this network, and a new modified network is obtained. In the training process, the following parameters are initialized; for example, the learning rate is 0.0001, the epochs are 100, the minibatch size is 64, and the learning method is Stochastic Gradient Descent (SGD). Features are extracted from the global average pooling layer, which is later utilized for the classification process. The dimension of an extracted feature on this layer is $N \times 2048$, where N denotes the dermoscopy images.

2.5. Fine-Tuned ResNet-101 Deep Features. ResNet-101 consists of 104 layers composed of 33 squares, of which the previous blocks use 29 squares directly [44]. Figure 10 shows a brief description of the ResNet-101 CNN model. In this figure, it is described that the output of the first residual block is 112×112 . After the first convolutional layer, a max-pooling layer is added of filter size 3×3 and stride 2. Using the same sequence, four more residual blocks are added, and each block consists of several layers, as given in Figure 11. This model was initially trained on the ImageNet dataset; therefore, the output was 1000D.

In this work, this model is fine-tuned according to the target dataset named HAM10000 having seven skin classes. The FC layer is removed in the fine-tuning process and a new FC layer is added with seven outputs. Later, the FC layer is connected with the Softmax layer and output layer and trained using TL. The following parameters are initialized in the training process: the learning rate is 0.0001, epochs are 100, the minibatch size is 64, and the learning method is Stochastic Gradient Descent (SGD). Features are extracted from the average pooling layer, which is later utilized for the classification process. On this layer, the dimension of extracted features is $N \times 2048$.

2.6. Feature Fusion. Feature fusion is an essential topic in pattern recognition, where multisource features are fused in one vector. The main purpose of feature fusion is to increase the object information for accurate classification. In this work, we consider the idea of a serial-based approach named modified serial-based feature fusion. The proposed fusion approach works in two sequential steps. In the first step, all features of vectors are fused in one matrix, and later on, a standard error mean- (SEM-) based threshold function is proposed.

Assume that P and Q are two function rooms on the sample size pattern Δ . The corresponding two characteristic vectors $\delta \in P$ and $\gamma \in Q$ for an arbitrary sample are $f \in \Delta$. The serial-based feature combination of f is defined as $\omega = \begin{pmatrix} \delta \\ \gamma \end{pmatrix}$. Of course, if the vector feature δ is n -dimensional and γ is m -dimensional, then the combined serial feature ω is $(n + m)$ -dimensions [45]. A serial combined feature space is created by combining all serially merged feature vectors of pattern samples of $(n + m)$ -dimensions. The resultant ω vector has dimension $N \times 4096$. After this step, SEM is computed of ω using the following formulation:

$$\text{SEM} = \frac{s}{\sqrt{n}},$$

$$s = \sqrt{\frac{\sum_{i=1}^n (\omega_i - \text{Mean})^2}{n - 1}}, \quad (11)$$

$$\text{Thr} = \begin{cases} \text{Fus}(i), & \text{for } \omega_i \geq \text{SEM}, \\ \text{Nfus}(j), & \text{elsewhere,} \end{cases}$$

where Thr denotes the threshold function, $\text{Fus}(i)$ is fused feature vector of dimension $N \times 2506$, $\text{Nfus}(j)$ is a feature that is not considered in the fused vector, and s is a standard deviation value. The output of this step is further refined in the feature selection step, as given below.

2.7. Feature Selection. The goal of feature selection is to reduce input variables when a predictive model is developed. This process minimizes the computational time of a proposed system and improves classification accuracy. In this work, a new heuristic search-based feature selection method is proposed named skewness-controlled SVR. In the first step, a skewness feature vector is extracted from the fused

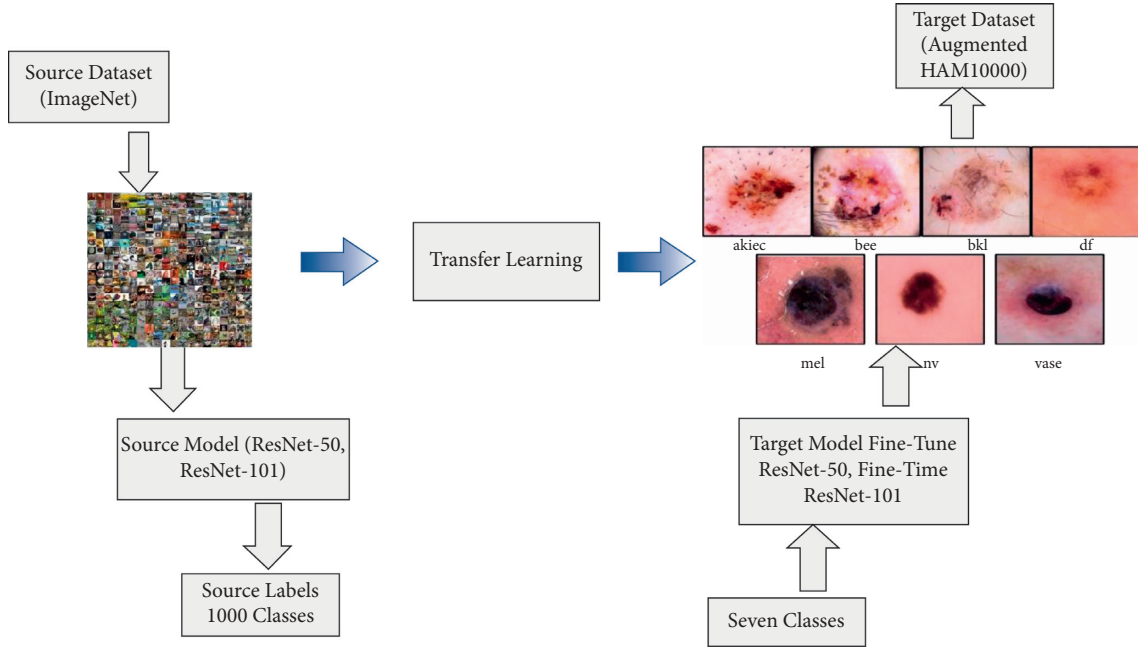


FIGURE 6: Transfer learning-based training a new target model for the classification of skin lesions.

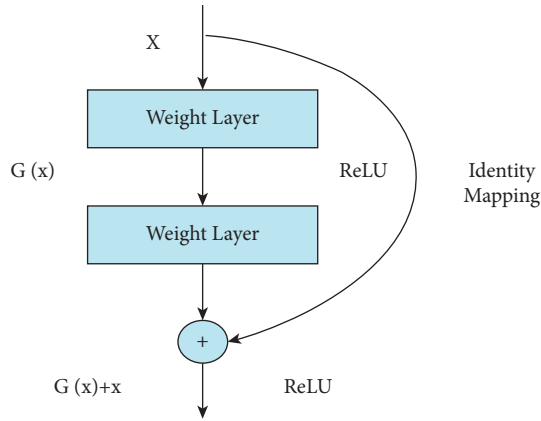


FIGURE 7: Residual identity mapping.

vector $Fus(i)$. This step aims to find the likelihood of the features falling in the specific probability distribution. Mathematically, skewness is computed as follows:

$$Skew = \frac{3\overline{Fus(i)} - Median}{s}, \quad (12)$$

where Skew is the skewness feature vector, $\overline{Fus(i)}$ is the mean value of the fused feature vector, and s is the standard deviation. Using this skewness value, a threshold function is defined to select features at the first stage.

$$Thr 1 = \begin{cases} Sel(i), & \text{for } Fus(i) \geq Skew, \\ Ignore, & \text{elsewhere.} \end{cases} \quad (13)$$

Using this threshold function, features are selected at the initial phase. The selected features of this phase are later validated using a fitness function Support Vector Regression (SVR). The SVR is formulated as follows.

Layer Name	Output Size	50-Layer
Conv1	112×112	7×7,64, stride2
Conv2	56×56	3×3 max pool, stride2
		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
		$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 512 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv4	14×14	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2,048 \end{bmatrix} \times 3$
		$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2,048 \end{bmatrix} \times 3$
Conv5	7×7	Average pool, 7D fc, Softmax
		1×1
FLOPs		3.8×10 ⁹

FIGURE 8: Architecture of ResNet-50.

Assume that the dataset for training Q comprises the instances q , each having an attribute u_i , an associated class, and v_i . $u_i \in Sel(i)$ is a selected feature and v_i represents labels; i.e., $\{(u_1, v_1), (u_2, v_2), \dots, (u_q, v_q)\}$. On the dataset D , b is a bias, and the linear function $f(x)$ may be defined as follows:

$$f(u) = \delta_1 u_1 + \delta_2 u_2 + \dots + \delta_d u_d + b, \quad (14)$$

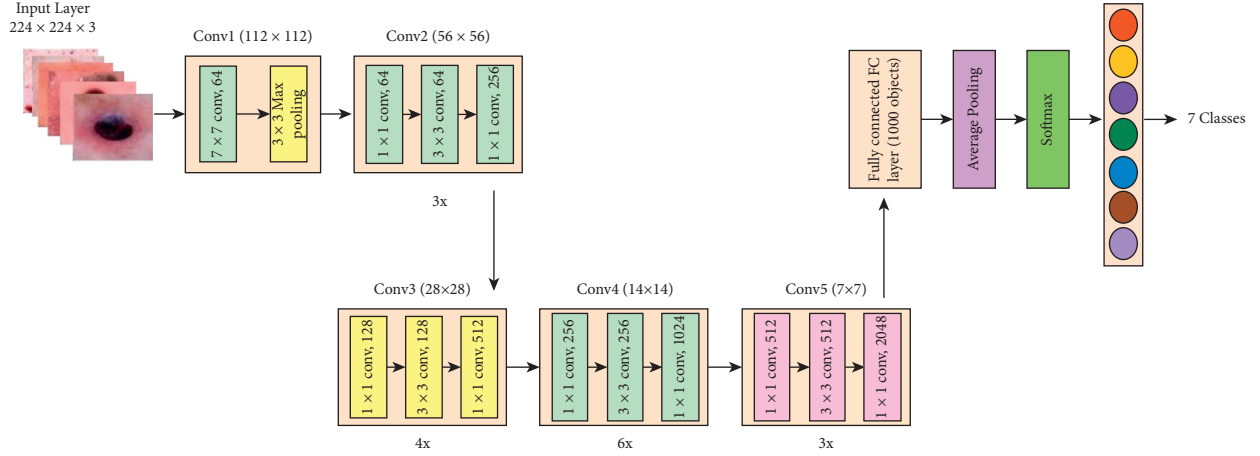


FIGURE 9: Block diagram of fine-tuned ResNet-50.

Layer Name	Output Size	100-Layer
Conv1	112x112	7x7,64,stride2
Conv2	56x56	3x3max pool,stride2 $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3	28x28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4	14x14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 512 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
Conv5	7x7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1x1	Average pool,7D fc,Softmax
FLOPs		7.6×10^9

FIGURE 10: Fine-tuned architecture of ResNet-101.

where the weight δ_i is defined as input space S^d ; i.e., $\delta_i \in S^d$. The maximum margin size is determined by the Euclidean weight ($\|Y\|$). The flatness, therefore, requires a minimum weight standard in the case of the following equation. Here, the definition of ($\|Y\|$) is

$$\|Y\|^2 = Y_1^2 + Y_2^2 + \dots + Y_d^2. \quad (15)$$

Each training data error may be represented as $\langle u_i, v_i \rangle$.

$$\text{Err}_i(u_i) = v_i - (\delta_i u_i + b). \quad (16)$$

If there is error $\text{Err}_i(u_i)$, the deviation is permitted to be within it, and the previous equation may be expressed as η .

$$\begin{aligned} v_i - (\delta_i u_i + b) &\leq \eta, \\ (\delta_i u_i + b) - v_i &\leq \eta. \end{aligned} \quad (17)$$

Using these two equations, the minimization issue for δ can be formulated as follows:

$$\text{minimized: } \frac{1}{2} \|Y\|^2, \quad (18)$$

subject to

$$\begin{aligned} v_i - (\delta_i u_i + b) &\leq \eta, \\ (\delta_i u_i + b) - v_i &\leq \eta. \end{aligned} \quad (19)$$

The restrictions of the above equation imply that the function f corresponds to all pairings (u_i, v_i) with a deviation of η . However, the assumption is not accepted in all instances when the slack variables δ_i, δ_i^* are neither required nor necessary in case of violation of the assumption. The optimization problem may be reformulated using slack variables as follows:

$$\text{minimize: } \frac{1}{2} \|Y\|^2 + C \sum_{i=0}^d (\delta_i + \delta_i^*), \quad (20)$$

subject to

$$\begin{aligned} \forall_i: v_i - (\delta_i u_i + b) &\leq \eta + \delta_i, \\ \forall_i: (\delta_i u_i + b) - v_i &\leq \eta + \delta_i^*, \\ \forall_i: \delta_i &\geq 0, \\ \forall_i: \delta_i^* &\leq 0, \end{aligned} \quad (21)$$

where C is the penalty constant, which does not meet the constraints. It also helps in reducing overfitting. The Kernel is defined by the input data $K(u_i, u_j)$ and can substitute the occurrence of the dot product between the tuples to avoid the dot product on a data tuple changed. All computations are therefore done in the original input areas. In this work, a radial basis Kernel/Gaussian function is utilized:

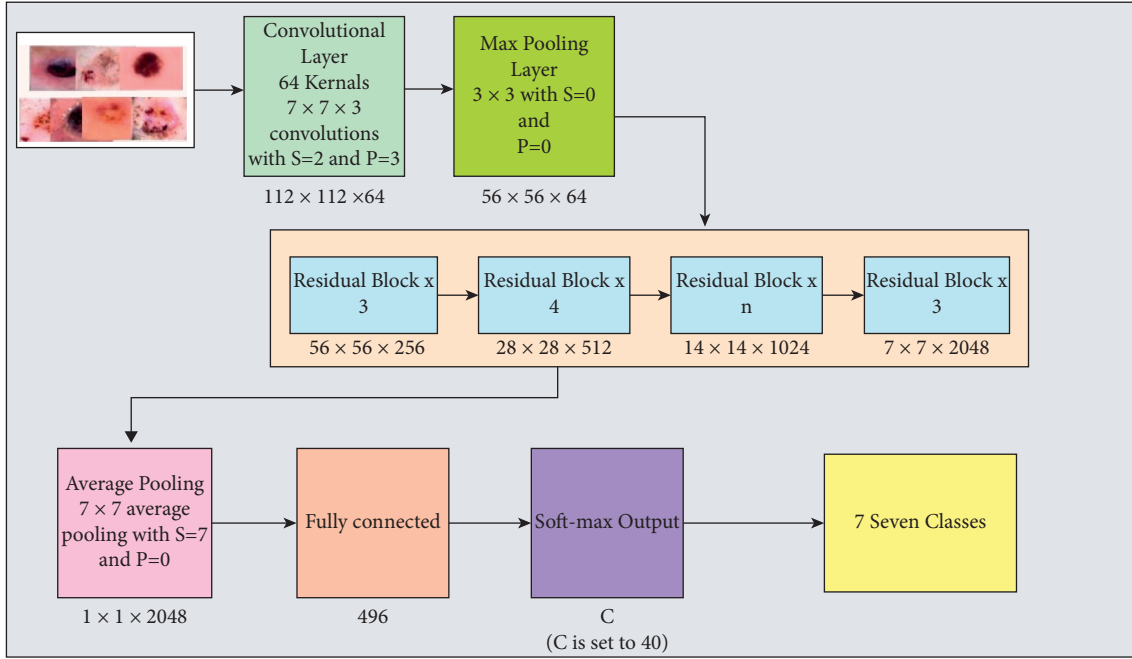


FIGURE 11: Transfer learning architecture with modified ResNet-101 model ($n=23$ for ResNet-101; stride * S and padding * P).

$$K(u_i, u_j) = \exp - \frac{\left(\|u_i, u_j + 1\|^2 \right)}{2\rho^2}. \quad (22)$$

The accuracy is computed using SVR, and if accuracy is less than the target accuracy value, then $Sel(i)$ is again updated. This process is continued until the maximum number of iterations is performed. In this work, the target accuracy is 90%, and the numbers of iterations are 5. Following this process, a feature vector is obtained called the best-selected feature vector of dimension $N \times 1456$ and further fed to supervised learning algorithms for final classification.

3. Experimental Results and Discussion

The proposed method is evaluated on the augmented HAM10000 dataset. Dataset is divided into 70 : 30, where the 70% data is used for the training of a model, and the rest of the 30% is utilized for the testing process. The other training hyperparameters; for example, epochs are 100, the mini-batch size is 64, and the learning rate is 0.0001. The 10-fold method was carried out for cross-validation [46]. Seven performance measures are used for the experimental process: recall rate, precision rate, false-negative rate (FNR), Area under Curve (AUC), accuracy, time, and F1-score. The proposed method is implemented in MATLAB 2020b, Corei7, with a RAM 16GB and 8GB graphics card.

3.1. Results. In this section, the proposed method results are described in numerical values (Tables) and confusion matrixes. Total ten classifiers are utilized for the experimental process, such as Linear Support Vector Machine (LSVM), Quadratic SVM (QSVM), Cubic SVM (CSVM), Medium

Gaussian SVM (MGSVM), Cosine K -Nearest Neighbor (CKNN), Weighted KNN (WKNN), Coarse KNN (CKNN), Ensemble Subspace Discriminative (ESD), Ensemble Boosted Tree (EBT), and Ensemble Subspace KNN (ESKNN). Five experiments are performed for the validation of the proposed framework such as (i) Experiment # 1: classification using fine-tuned ResNet-50 CNN model, (ii) Experiment # 2: classification using Fine-Tuned ResNet-101CNN model, (iii) Experiment # 3: perform features fusion of Fine-Tuned ResNet-50 and ResNet-101 CNN models, and (iv) Best Features (BF) selection.

3.1.1. Experiment # 1. In the first experiment, features are extracted using fine-tuned ResNet-50 CNN model, and results are computed. The augmented dataset was used for the experimental process. The results of this experiment are given in Table 1. CSVM has the highest accuracy of 92.7% in this table, with computational time 1190.3 (sec). Figure 12 shows the confusion matrix of CSVM for this experiment. In this figure, the diagonal values represent the correct predicted values such as AKIEC (96%), BCC (93%), BKL (87%), DF (97%), MEL (86%), NV (94%), and VASC (99%), respectively. Moreover, the recall rate is 93.14, the precision rate is 93.14, and F1-score is 93.14%, respectively. Compared with the rest of the classifiers, it is noticed that the CSVM showed better classification accuracy. Moreover, the computational time of each classifier is also noted and plotted in Figure 13. This figure shows that the CKNN has the lowest computational time of 274.55 (sec).

3.1.2. Experiment # 2. Table 2 presents the results of fine-tuned ResNet-101 CNN features using the augmented HAM10000 dataset. This table shows that the best accuracy

TABLE 1: Classification accuracy of fine-tuned ResNet-50 deep features using augmented HAM10000 dataset.

Classifier	Recall rate (%)	Precision rate (%)	FNR (%)	AUC	Accuracy (%)	Time (sec)	F1-score (%)
LSVM	86.42	86.85	13.57	0.988	86.5	742.5	86.63
QSVM	92.00	92.14	8.00	0.992	91.7	1046.1	92.07
CSVM	93.14	93.14	6.858	0.994	92.7	1190.3	93.14
MG SVM	89.57	90.00	10.42	0.988	89.3	1906.8	89.78
CKNN	53.25	65.28	46.75	0.898	60.8	274.5	58.65
CKNN	80.42	79.00	19.57	0.967	78.7	287.6	79.70
WKNN	85.14	84.42	14.85	0.98	83.6	262.3	84.78
ESKNN	93.14	92.57	6.858	0.99	92.3	4514.7	92.85
EBT	55.28	86.71	44.71	0.974	57.1	1546.0	86.49
ESD	86.28	57.00	13.71	0.85	86.1	839.8	56.12

The bold value represents best ones.

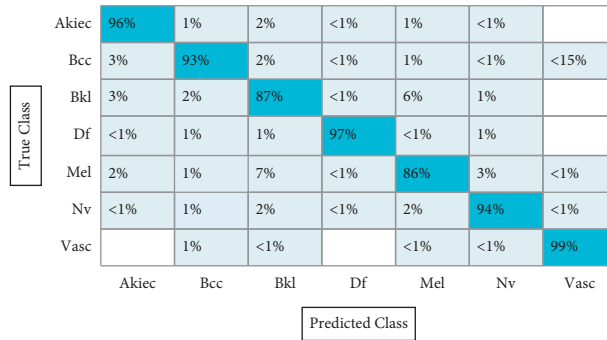


FIGURE 12: Confusion matrix of CSVM using ResNet-50 model for HAM10000 dataset.

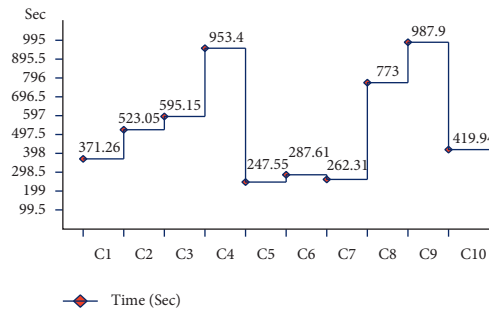


FIGURE 13: Time plot for fine-tuned ResNet-50 CNN model using augmented HAM10000 dataset.

TABLE 2: Classification accuracy of fine-tuned ResNet-101 deep features using augmented HAM10000 dataset.

Classifier	Recall rate (%)	Precision rate (%)	FNR (%)	AUC	Accuracy (%)	Time (sec)	F1-score (%)
LSVM	86.00	86.28	14.00	0.98	85.5	746.2	86.14
QSVM	91.57	91.71	8.428	0.992	91.1	1010.2	91.64
CSVM	92.71	92.42	7.285	0.992	92.1	11321.1	92.56
MG SVM	89.42	89.42	10.57	0.988	88.9	1919.4	89.42
CKNN	78.87	77.85	21.14	0.961	77.2	268.2	78.36
CKNN	58.42	63.00	41.27	0.887	58.9	263.6	60.62
WKNN	84.85	84.00	15.14	0.977	83.3	260.5	84.42
EBT	56.57	56.57	43.42	0.855	57.2	1544.5	56.57
ESKNN	80.28	92.28	19.71	0.99	92.1	4590.9	85.86
ESD	85.85	86.28	14.14	0.98	85.6	821.79	86.06

The bold value represents best ones.

achieved by CSVM is 92.1%, with a computational time of 11321.1 (sec), recall rate is 92.7, the precision rate is 92.42, and F1-score is 92.56%, respectively. Figure 14 shows the

confusion matrix of CSVM. In this figure, the diagonal values represent the correct predicted values such as AKIEC (96%), BCC (92%), BKL (85%), DF (98%), MEL (86%), NV

True Class	Akiec	96%	1%	1%		1%	<1%	
	Bcc	3%	92%	2%	1%	2%	1%	<1%
	Bkl	2%	3%	85%	<1%	7%	2%	<1%
	Df	1%	1%	<1%	98%		<1%	
	Mel	2%	1%	7%	<1%	86%	3%	<1%
	Nv	<1%	1%	3%	<1%	2%	93%	<1%
	Vasc		1%	<1%			<1%	99%
			Akiec	Bcc	Bkl	Df	Mel	Nv
		Predicted Class						

FIGURE 14: Confusion matrix of CSVM using ResNet-101 model for HAM10000 dataset.

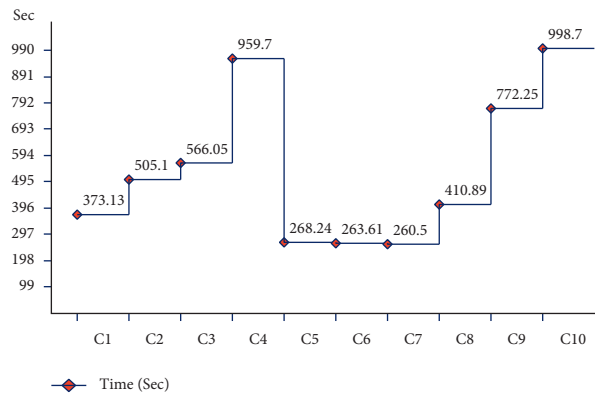


FIGURE 15: Time plot for fine-tuned ResNet-101 CNN model using augmented HAM10000 dataset.

(93%), and VASC (99%), respectively. As given in this table, a few other classifiers are also implemented and show that the CSVM gives better accuracy. Moreover, the computational time is computed for each classifier, and the minimum noted time is 260.5 (sec) for the W-KNN classifier. The noted time is also plotted in Figure 15.

3.1.3. Experiment # 3. In the next experiment, features are fused using the serial-based extended (SbE) approach. Results are given in Table 3. This table represents the best accuracy achieved by the ESD classifier of 95%, further demonstrating in a confusion matrix, given in Figure 16. This figure represents the correct predicted values such as AKIEC (97%), BCC (94%), BKL (89%), DF (98%), MEL (89%), NV (99%), and VASC (99%), respectively. The other computed measures are recall rate, precision rate, FNR, AUC, and F1-score of 95.0, 95.0, 5.00, 0.99, and 95.0%, respectively. The CSVM achieved the second-best accuracy of 94.9%, whereas the recall rate and precision rates are 95.0%. Comparison with the rest of the classifiers shows the superiority of the ESD classifier. Moreover, the computational time is also noted, as illustrated in Figure 17.

Compared with the results of this experiment with Tables 1 and 2, it is noticed that the fusion using the SbE approach significantly improves the classification accuracy. The limitation of this step increases computational time, which needs to be minimized.

3.1.4. Experiment # 4. Finally, the proposed feature selection algorithm is applied on the fused feature vector and achieved an accuracy of 91.7% on the ESD classifier, where the computational time is 1367 (sec), given in Table 4. The recent time was 4118 (sec), which is significantly minimized after the selection algorithm. This table also showed that the proposed accuracy decreases, but on the other side, it helps to minimize the computational time. The accuracy of the ESD classifier is further verified using a confusion matrix given in Figure 18. In this figure, the diagonal values represent the correct predicted values such as AKIEC (94%), BCC (91%), BKL (85%), DF (93%), MEL (83%), NV (97%), and VASC (99%), respectively.

The F1-score-based analysis is also conducted and plotted in Figure 19. In this figure, it is illustrated that the value of the F1-score is improved after the feature fusion process except the CKNN and EBT classifier. Moreover, the feature selection approach reduced the computational time but accuracy is degraded. Overall, the proposed framework performed well on the selected dataset. In the last, the proposed method accuracy is compared with some recent techniques, as given in Table 5. In this table, Khan et al. [7] presented a deep learning method for skin lesion classification. They used the HAM10000 dataset and achieved an accuracy of 88.5%. The recent best-reported accuracy was 91.5%, achieved by Sevli [47]. The proposed accuracy is 91.7% and 95% for the best feature selection approach and fusion approach. Based on this accuracy, it is noted that the proposed method showed improved accuracy.

TABLE 3: Classification results using SbE approach-based deep features fusion on augmented HAM10000 dataset.

Classifier	Recall rate (%)	Precision rate (%)	FNR (%)	AUC	Accuracy (%)	Time (sec)	F1-score (%)
LSVM	92.71	92.85	7.285	0.992	92.50	1303.8	92.78
QSVM	94.85	94.85	5.142	0.997	94.80	2400.4	94.75
CSVM	95.00	95.00	5.00	0.854	94.90	2868.5	95.00
MG SVM	92.85	93.14	7.142	0.995	92.60	4501.6	92.99
CKNN	61.71	73.14	38.28	0.910	62.20	562.0	66.94
CKNN	84.14	83.57	15.85	0.975	82.60	550.5	83.85
WKNN	83.42	84.57	16.27	0.971	82.10	530.7	83.99
ESD	95.00	95.00	5.00	0.997	95.00	4118.2	95.00
FKNN	88.14	87.57	11.85	0.931	87.00	543.2	87.85
EBT	62.00	62.22	38.00	0.855	62.80	69886.0	62.11

The bold value represents best ones.

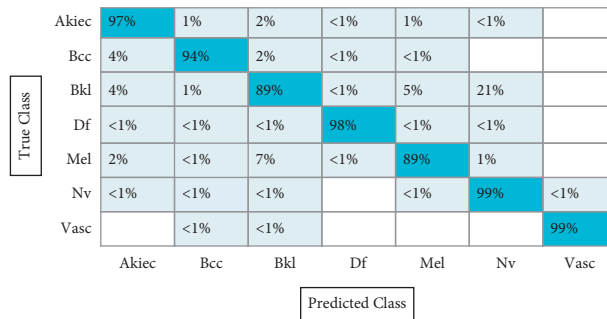


FIGURE 16: Confusion matrix of ESD using ResNet-50 and ResNet-101 model for HAM10000 dataset.

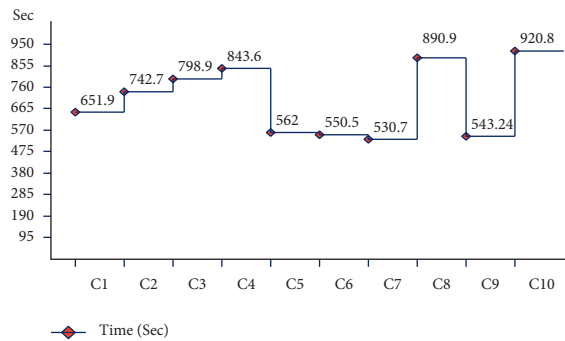


FIGURE 17: Time plot for the fusion of ResNet-50 and ResNet-101 using augmented dataset (HAM10000).

TABLE 4: Classification results using proposed feature selection algorithm on augmented HAM10000 dataset.

Classifier (ESD)	Recall rate (%)	Precision rate (%)	FNR (%)	AUC	Accuracy (%)	Time (sec)	F1-score (%)
500	78.42	79.28	21.57	0.954	78.3	102.4	78.85
1000	64.85	66.71	35.14	0.902	65.1	132.5	65.77
1500	87.14	87.57	12.85	0.978	86.8	406.0	87.35
2000	83.57	84.42	16.42	0.97	83.3	748.4	83.99
2500	89.71	90.14	10.28	0.984	89.5	1293.5	89.92
3000	91.85	92.00	8.142	0.99	91.7	1367.6	91.92

The bold value represents best ones.

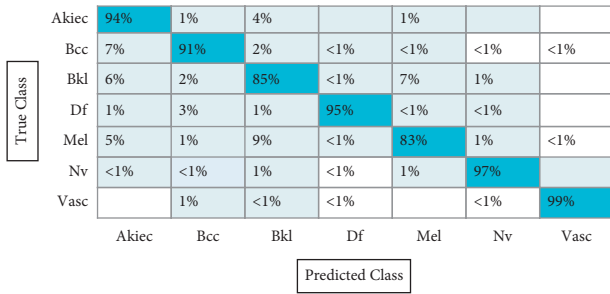


FIGURE 18: Confusion matrix of ESD classifier using a proposed feature selection algorithm.

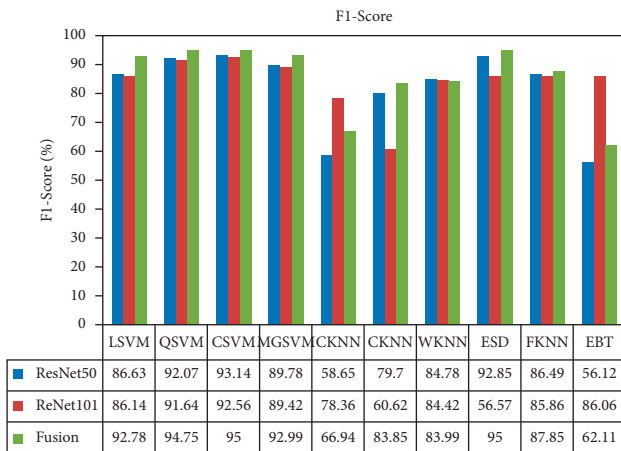


FIGURE 19: F1-score-based analysis of middle steps like ResNet-50, ResNet-101, and fusion.

TABLE 5: Comparison with existing techniques.

Reference	Year	Dataset	Accuracy (%)
[7]	2020	HAM10000	88.5
[48]	2020	HAM10000	86.1
[49]	2020	HAM10000	83.1
[30]	2021	HAM10000	85.50
[47]	2020	HAM10000	91.5
Proposed (fusion)	2021	HAM10000	95.0
Proposed (feature selection)	2021	HAM10000	91.7

The bold value represents best ones.

4. Conclusion

In this work, a new framework is presented for multiclass skin lesion classification using deep learning. The proposed method consisted of a series of steplike data augmentation, feature extraction using deep learning models, the fusion of features, selection of parts, and classification. The experiment was performed on an augmented HAM10000 dataset. The number of experiments was performed, such as non-augmented and augmented datasets, and achieved accuracy with a nonaugmented dataset of 64.36% using ResNet-50 and 49.98% using ResNet-101. The augmented dataset achieved an accuracy of 95.0% for feature fusion and 91.7% for feature selection. The results show that the augmentation process helps improve the classification accuracy for a complex dataset.

Moreover, the fusion process increases the performance but also increases the computational time. This process can be further refined through a feature selection process. However, according to the results, the feature selection process decreases the computational time and reduces accuracy. But from the overall comparison with recent techniques, feature fusion and feature selection technique both perform better than previous techniques. The new datasets ISBI 2020 and ISIC 2020 can be used for the experimental process in future work. Latest deep learning models can be used as feature extraction. Fusion can be performed using parallel approaches. The selection process can be refined, which not only reduces the time but also increases accuracy.

Data Availability

The HAM10000 dataset is utilized in this work for the experimental process. The dataset is publically available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>.

Conflicts of Interest

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

Acknowledgments

The work of Shabnam M. Aslam was supported by the Majmaah University’s Deanship of Scientific Research under Project 155/46683.

References

- [1] M. H. Trager, L. J. Geskin, F. H. Samie, and L. Liu, “Bio-markers in melanoma and non-melanoma skin cancer prevention and risk stratification,” *Experimental Dermatology*, 2020.
- [2] R. Shayini, “Classification of skin lesions in digital images for the diagnosis of skin cancer,” in *Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 162–166, Trichy, India, 2020.
- [3] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, “Discriminative feature learning for skin disease classification using deep convolutional neural network,” *IEEE Access*, vol. 8, pp. 39025–39033, 2020.
- [4] M. Fernandez-Figueras, C. Carrato, X. Sáenz, L. Puig, E. Musulen, and C. Ferrándiz, “Actinic keratosis with atypical basal cells (AK I) is the most common lesion associated with invasive squamous cell carcinoma of the skin,” *Journal of the European Academy of Dermatology and Venereology*, vol. 29, pp. 991–997, 2015.
- [5] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, “Melanoma skin cancer detection using deep learning and classical machine learning techniques: a hybrid approach,” in *Proceedings of the 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–5, Sousse, Tunisia, 2020.
- [6] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, “A multi-class skin cancer classification using deep convolutional

- neural networks,” *Multimedia Tools and Applications*, vol. 79, no. 39-40, pp. 28477–28498, 2020.
- [7] T. Akram, Y.-D. Zhang, and M. Sharif, “Attributes based skin lesion detection and recognition: a mask RCNN and transfer learning-based deep learning framework,” *Pattern Recognition Letters*, vol. 143, pp. 58–66, 2021.
 - [8] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, “Covid-caps: a capsule network-based framework for identification of covid-19 cases from x-ray images,” *Pattern Recognition Letters*, vol. 138, pp. 638–643, 2020.
 - [9] M. Kumar, M. Alshehri, R. AlGhamdi, P. Sharma, and V. Deep, “A de-ann inspired skin cancer detection approach using fuzzy c-means clustering,” *Mobile Networks and Applications*, vol. 25, no. 4, pp. 1319–1329, 2020.
 - [10] A. R. Lopez, X. Giro-i-Nieto, J. Burdick, and O. Marques, “Skin lesion classification from dermoscopic images using deep learning techniques,” in *Proceedings of the 2017 13th IASTED international conference on biomedical engineering (BioMed)*, pp. 49–54, Innsbruck, Austria, 2017.
 - [11] S. Roy, T. D. Whitehead, S. Li, F. O. Ademuyiwa, R. L. Wahl, and F. Dehdashti, “Co-clinical FDG-PET radiomic signature in predicting response to neoadjuvant chemotherapy in triple negative breast cancer,” *European Journal of Nuclear Medicine and Molecular Imaging*, 2021.
 - [12] K. Dutta, S. Roy, T. D. Whitehead et al., “Deep learning segmentation of triple-negative breast cancer (TNBC) patient derived tumor xenograft (PDX) and sensitivity of radiomic pipeline to tumor probability boundary,” *Cancers*, vol. 13, no. 15, p. 3795, 2021.
 - [13] A. Sedik, A. Ilyasu, M. E El-Rahiem et al., “Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections,” *Viruses*, vol. 12, no. 7, p. 769, 2020.
 - [14] T. Saba, M. A. Khan, A. Rehman, and S. L Marie-Sainte, “Region extraction and classification of skin cancer: a heterogeneous framework of deep CNN features fusion and reduction,” *Journal of Medical Systems*, vol. 43, pp. 289–319, 2019.
 - [15] I. Ashraf, M. Alhaisoni, R. Damaševičius, R. Scherer, and A. Rehman, “Multimodal brain tumor classification using deep learning and robust feature selection: a machine learning application for radiologists,” *Diagnostics*, vol. 10, p. 565, 2020.
 - [16] U. Nazar, M. A. Khan, I. U. Lali et al., “Review of automated computerized methods for brain tumor segmentation and classification,” *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 16, no. 7, pp. 823–834, 2020.
 - [17] S. A. Khan, M. Nazir, M. A. Khan et al., “Lungs nodule detection framework from computed tomography images using support vector machine,” *Microscopy Research and Technique*, vol. 82, no. 8, pp. 1256–1266, 2019.
 - [18] T. Meraj, H. T. Rauf, S. Zahoor, A. Hassan, M. I. Lali, and L. Ali, “Lung nodules detection using semantic segmentation and classification with optimal features,” *Neural Computing & Applications*, vol. 33, pp. 10737–10750, 2021.
 - [19] A. Sedik, M. Hammad, F. E. Abd El-Samie, B. B. Gupta, and A. A. Abd El-Latif, “Efficient deep learning approach for augmented detection of Coronavirus disease,” *Neural Computing and Applications*, Springer, Berlin, Germany, pp. 1–18, 2021.
 - [20] H. T. Rauf, M. I. U. Lali, S. Kadry, H. Alolaiyan, and A. Razaq, “Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks,” *Personal and Ubiquitous Computing*, pp. 1–18, 2021.
 - [21] M. Hammad, M. H. Alkinani, B. Gupta, and A. A. Abd El-Latif, “Myocardial infarction detection based on deep neural network on imbalanced data,” *Multimedia Systems*, pp. 1–13, 2021.
 - [22] M. Hammad, A. M. Ilyasu, A. Subasi, E. S. Ho, and A. A. Abd El-Latif, “A multitier deep learning model for arrhythmia detection,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2020.
 - [23] A. Alghamdi, M. Hammad, H. Ugail, A. Abdel-Raheem, K. Muhammad, and H. S. Khalifa, “Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities,” 2019, <https://arxiv.org/abs/1906.09358>.
 - [24] X. Yu, X. Wu, C. Luo, and P. Ren, “Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework,” *GIScience and Remote Sensing*, vol. 54, no. 5, pp. 741–758, 2017.
 - [25] A. F. Agarap, “Deep learning using rectified linear units (relu),” 2018, <https://arxiv.org/abs/1803.08375>.
 - [26] L. Liu, C. Shen, and A. Van Den Hengel, “The treasure beneath convolutional layers: cross-convolutional-layer pooling for image classification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4749–4757, Boston, MA, USA, 2015.
 - [27] S. Sharma and S. Sharma, “Activation functions in neural networks,” *Data Science*, vol. 6, pp. 310–316, 2017.
 - [28] P. Marcelino, “Transfer learning from pre-trained models,” *Data Science*, 2018.
 - [29] A. Majid, N. Hussain, M. Alhaisoni, Y.-D. Zhang, and S. Kadry, “Multiclass stomach diseases classification using deep learning features optimization,” *Computers, Materials and Continua*, vol. 67, no. 3, 2021.
 - [30] F. Afza, M. Sharif, M. Mittal, and D. J. Hemanth, “A hierarchical three-step superpixels and deep learning framework for skin lesion classification,” *Methods*, 2021.
 - [31] A. K. Verma, S. Pal, and S. Kumar, “Comparison of skin disease prediction by feature selection using ensemble data mining techniques,” *Informatics in Medicine Unlocked*, vol. 16, Article ID 100202, 2019.
 - [32] T. Akram, H. M. J. Lodhi, S. R. Naqvi, S. Naeem, M. Alhaisoni, and M. Ali, “A multilevel features selection framework for skin lesion classification,” *Human-Centric Computing and Information Sciences*, vol. 10, pp. 1–26, 2020.
 - [33] H. W. Huang, B. W. Y. Hsu, C. H. Lee, and V. S. Tseng, “Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers,” *The Journal of Dermatology*, vol. 48, no. 3, 2020.
 - [34] M. Sharif, T. Akram, R. Damaševičius, and R. Maskeliūnas, “Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization,” *Diagnostics*, vol. 11, p. 811, 2021.
 - [35] K. Muhammad, M. Sharif, T. Akram, and V. H. C. de Albuquerque, “Multi-class skin lesion detection and classification via teledermatology,” *IEEE Journal of Biomedical and Health Informatics*, 2021.
 - [36] K. Melbin and Y. J. V. Raj, “Integration of modified ABCD features and support vector machine for skin lesion types classification,” *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 8909–8929, 2021.
 - [37] H. Zanddzari, N. Nguyen, B. Zeinali, and J. M. Chang, “A new preprocessing approach to improve the performance of CNN-based skin lesion classification,” *Medical, & Biological Engineering & Computing*, vol. 59, no. 5, pp. 1123–1131, 2021.
 - [38] H. T. Rauf, U. Shoaib, M. I. Lali, M. Alhaisoni, M. N. Irfan, and M. A. Khan, “Particle swarm optimization with

- probability sequence for global optimization,” *IEEE Access*, vol. 8, pp. 110535–110549, 2020.
- [39] H. Younis, M. H. Bhatti, and M. Azeem, “Classification of skin cancer dermoscopy images using transfer learning,” in *Proceedings of the 2019 15th International Conference on Emerging Technologies (ICET)*, pp. 1–4, Peshawar, Pakistan, 2019.
- [40] N. Hussain, M. Sharif, S. A. Khan, A. A. Albeshier, and T. Saba, “A deep neural network and classical features based scheme for objects recognition: an application for machine inspection,” *Multimedia Tools and Applications*, Springer, Berlin, Germany, pp. 1–23, 2020.
- [41] O. A. Shawky, A. Hagag, E.-S. A. El-Dahshan, and M. A. Ismail, “Remote sensing image scene classification using CNN-MLP with data augmentation,” *Optik*, vol. 221, Article ID 165356, 2020.
- [42] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou, “Fixing the train-test resolution discrepancy,” 2019, <https://arxiv.org/abs/1906.06423>.
- [43] J.-A. Almaraz-Damian, V. Ponomaryov, S. Sadovnychiy, and H. Castillejos-Fernandez, “Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures,” *Entropy*, vol. 22, no. 4, p. 484, 2020.
- [44] A. Demir, F. Yilmaz, and O. Kose, “Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3,” in *Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO)*, pp. 1–4, Izmir, Turkey, 2019.
- [45] S. Mustafa, “Feature selection using sequential backward method in melanoma recognition,” in *Proceedings of the 2017 13th International Conference on Electronics, Computer and Computation (ICECCO)*, pp. 1–4, Abuja, Nigeria, 2017.
- [46] D. Zhao, G. Yu, P. Xu, and M. Luo, “Equivalence between dropout and data augmentation: a mathematical check,” *Neural Networks*, vol. 115, pp. 82–89, 2019.
- [47] O. Sevli, “A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation,” *Neural Computing and Applications*, Springer, Berlin, Germany, pp. 1–12, 2021.
- [48] M. A. Khan, Y.-D. Zhang, M. Sharif, and T. Akram, “Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification,” *Computers & Electrical Engineering*, vol. 90, Article ID 106956, 2021.
- [49] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, “Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet,” in *Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications*, pp. 165–176, Cairo, Egypt, 2020.

Research Article

An IoMT-Enabled Smart Healthcare Model to Monitor Elderly People Using Machine Learning Technique

Muhammad Farrukh Khan,^{1,2} Taher M. Ghazal,^{3,4} Raed A. Said,⁵ Areej Fatima,⁶ Sagheer Abbas ¹, M.A. Khan ⁷, Ghassan F. Issa,⁴ Munir Ahmad ¹, and Muhammad Adnan Khan ⁸

¹School of Computer Science, National College of Business Administration and Economics, Lahore 54000, Pakistan

²Lahore Institute of Science and Technology, Lahore 54792, Pakistan

³Center for Cyber Security, Faculty of Information Science and Technology, University Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

⁴School of Information Technology, Skyline University College, University City Sharjah, 1797 Sharjah, UAE

⁵Canadian University Dubai, Dubai, UAE

⁶Department of Computer Science, Lahore Garrison University, Lahore 54792, Pakistan

⁷Riphah School of Computing & Innovation, Faculty of Computing, Riphah International University Lahore Campus, Lahore 54000, Pakistan

⁸Pattern Recognition and Machine Learning Lab, Department of Software, Gachon University, Seongnam Gyeonggi-do 13120, Republic of Korea

Correspondence should be addressed to Munir Ahmad; munir@ncbae.edu.pk and Muhammad Adnan Khan; adnan@gachon.ac.kr

Received 27 August 2021; Revised 3 November 2021; Accepted 12 November 2021; Published 25 November 2021

Academic Editor: Lo'ai Tawalbeh

Copyright © 2021 Muhammad Farrukh Khan 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 Internet of Medical Things (IoMT) enables digital devices to gather, infer, and broadcast health data via the cloud platform. The phenomenal growth of the IoMT is fueled by many factors, including the widespread and growing availability of wearables and the ever-decreasing cost of sensor-based technology. The cost of related healthcare will rise as the global population of elderly people grows in parallel with an overall life expectancy that demands affordable healthcare services, solutions, and developments. IoMT may bring revolution in the medical sciences in terms of the quality of healthcare of elderly people while entangled with machine learning (ML) algorithms. The effectiveness of the smart healthcare (SHC) model to monitor elderly people was observed by performing tests on IoMT datasets. For evaluation, the precision, recall, fscore, accuracy, and ROC values are computed. The authors also compare the results of the SHC model with different conventional popular ML techniques, e.g., support vector machine (SVM), K-nearest neighbor (KNN), and decision tree (DT), to analyze the effectiveness of the result.

1. Introduction

The Internet of things (IoT) is a system of interrelated computing devices that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. The practical application of IoT devices with medical technology used in healthcare is the IoMT. IoT helps to transfer the data of healthcare devices and

applications on medical IT servers for remote analysis. IoMT allows medical staff to access patients' healthcare data remotely through a web platform or any mobile application in real time to deal with patients' medical issues and help them avoid any future severe circumstances. This technology of interconnected medical devices allows the patients to monitor their health conditions following the treatment suggestions of the doctors by engaging in smart devices and applications while creating ease for the doctors to know the

medical history of the patients before the checkup through the collection of real-time data using IoMT [1].

In short, healthcare coupled with IoMT improves the quality of life, gives superior care administrations, and can make more cost-effective frameworks. In IoMT, a supportive mechanism between the sensors, communication modules, and users is required to efficiently and securely provide health services. IoMT technology is considered helpful in strengthening healthcare by providing self-care and early diagnosis features using a remote monitoring system. People are becoming more engaged and aware of their well-being as healthcare innovations advance. In this situation, the need for remote treatment is higher than ever. However, existing healthcare systems require technology to transform patient care by providing real-time patient information and encouraging doctors to take practical treatment steps [2].

Health facilities are not accessible or affordable to all, despite having excellent infrastructure and cutting-edge technology. Smart healthcare (SHC) aims to assist users by informing them about their medical conditions and keeping them informed about their health. SHC allows people to handle specific emergencies on their own. SHC employs modern information technology, e.g., IoT, big data, cloud computing, and artificial intelligence (AI), to completely change the current healthcare system into a more efficient and convenient one [3].

SHC promotes interaction among all stakeholders in the medical industry. It ensures that users receive the services they require, assisting parties in making informed decisions and facilitating resource distribution. SHC technology improves disease diagnosis, patient treatment, and overall quality of life. SHC systems based on IoT and big data can link patients with providers across various healthcare systems efficiently. SHC systems are also becoming more linked to different wearable devices used for real-time healthcare surveillance through the Internet. In SHC, wearable health devices (such as blood pressure monitors, glucometers, smartwatches, and so on) combined with IoT gadgets allow for continuous patient surveillance and treatment even when they are at home. The World Health Organization (WHO) predicts that “by 2050, the number of senior citizens will have risen to about 1.5 billion” [4]. The elderly population (including persons older than 60 years) requires the most significant healthcare needs. Elderly people are more vulnerable to chronic diseases because of a decreasing immune system and require regular visits to healthcare facilities and more hospitalizations.

Elderly people move from one level of treatment to another as they age. These patients, on the other hand, have limited access to specialized senior care facilities. A multifaceted strategy is essential to avoid health issues in the elderly population. An excellent preventive system that includes routine medical checkups allows for early disease detection and optimal care. Furthermore, older adults and their families should be aware of potential diseases to recognize them and start treatment as soon as possible [5].

Many countries built technologies and communication networks to help people live their lives more efficiently and simply. Many industries drive technology development, which transforms people’s lifestyles. A related perspective on the link between technology and aging is technological transition and consumption, emphasizing the aged as active users and co-creators. Smart homes (SHs), smart cities, and mobile apps are examples of developments and innovations to assist the senior population’s well-being through generic design. In the case of elderly people, IoMT-enabled SHC is an emerging solution for providing constant and holistic monitoring, reducing human caregiver effort, and assisting in clinical decision making. Rather than being hospitalized, elderly people can be supported using various “smart” devices in their own homes [6]. The SH idea is a viable and cost-effective approach to improve non-intrusive home treatment for seniors, enabling greater independence, ensuring good health, and avoiding social isolation. According to [7], SH solutions are regarded as information-based technology that gathers and disseminates user data with the resident, family members, and primary care physicians in a passive manner. SH solution also refers to using essential and assistive gadgets to create an environment where many house features are automated and devices may connect. SHs assist elderly people in their homes. Sensors and actuators integrated into the housing infrastructure track the occupants’ bodily signals, ambient conditions, daily behavior patterns, and sleep patterns, among other things. SHs also have a role in improving people’s quality of life. Health and wellness tracking technologies include wearable activity trackers using accelerometers and sensors. They also include non-wearable, embedded sensor activity monitors to track everyday activities.

An emergency medical service system (EMSS) is a complete system that organizes individuals, facilities, and equipment to offer health and safety services to sufferers of unexpected sickness or injury in a quick, effective, and coordinated way. EMSS aims to provide prompt treatment to sufferers of unexpected and life-threatening accidents to avoid unnecessary fatality or long-term morbidity. With advanced information and communication technologies, EMSS can deliver services that address the requirements of the elderly [8].

Elderly people having a weak immune system require daily checkups to maintain their health. For this purpose, they need to go to the hospitals or clinics, which is the main issue because of the mobility problems faced by elderly people [9]. In this case, SHC can provide the facility to the elderly by continuously monitoring their health without going to the hospital and helping medical experts make efficient decisions about patient health. Despite having this technology there exist some other challenges that restrict the usage of SHC for elderly people’s health monitoring. Many senior citizens are reluctant to adopt this technology.

Many senior citizens are unfamiliar with current technology [10], its advantages, and how to use the gadgets and applications. Furthermore, many individuals lack access to training and technical assistance, and their capacity to utilize and manage technological systems is a worry. Other

concerns include issues of privacy, protection, and reliability. The cost and use of communication technologies are critical elements in expanding medical facilities for the elderly to address technological challenges. Device design and usability are important requirements that should be prioritized to make the software simple to understand and use.

ML, an application of AI, gives systems the ability to learn from data automatically and make decisions without human assistance. ML is the study of allowing machines to learn and create their programs to make them more human-like in their actions and decisions. ML enables the systems to improve them from experience without being explicitly programmed. It also enables the machines or software to analyze, predict, and sort huge amounts of data. To make better decisions in the future, the learning process begins with data, instructions, and observations. ML algorithms use statistics to identify the patterns in massive quantities of data, e.g., numbers, words, images, etc. ML is subdivided into supervised, unsupervised, semisupervised, and reinforcement ML that uses different types of data and produces specific results [11].

2. Literature Review

Many studies have been done on the emerging technology named IoMT. Researchers have published many papers from a different perspective, focusing on specific issues and challenges.

Iyer [12] proposed a framework and protocol related to an IoT-based patient monitoring system and suggested that patients' health can be monitored using IoT devices and sensors connected to the Internet. The medical nursing system (MNS) based on IoT has been designed that uses different communication methods to transfer the data, e.g., sensors, Wireless Fidelity (Wi-Fi), radio-frequency identification (RFID), and Bluetooth [13]. In [14], the authors proposed the patient monitoring system with the help of patient body weight. Another IoT-based system was proposed in [15] to monitor and track autism patients with the help of sensors by collecting the signals from the brain.

A new scheme based on IoT was proposed in [16] to convert the old hospitals into smart hospitals that will help manage the information in an advanced way. Another design using IoT as a back-end platform was proposed in [17] to monitor the health of aged patients by a method of the end-to-end medical healthcare system. An Indian researcher [18] presented a model based on IoT for the electronic healthcare unit using RFID technology and an experienced healthcare system Mycin (an AI-based system to identify bacteria causing different infections).

Pinto [19] presented a novel idea in cardio signals and proposed a technique for cardiac patients named Inexpensive Cardiac Arrhythmia Management (iCarMa) that will indicate the severity of the cardiac patient and its timely detection and diagnosis. Research has been conducted to continuously monitor the patient's health using an accurate algorithm for sensing the patient's events, e.g., steps counting, immobility and fall, etc.

A fascinating idea presented in [20] was to help non-professionals know about the disease with the help of a bot.

This bot can be linked with various sensors on a smartphone to give a more flexible service utilizing IoMT. An emphasis on its use in [21] is that it will be very beneficial for the patients if major diseases can be predicted early on. This can be done with the help of IoT that facilitates the patients in the domain of remote healthcare systems.

Papageorgiou [22] proposed cost-efficient IoT-based living assistance for elderly people. It tracks and stores the critical details of patients employing a cloud-connected wristband. This proposed scheme triggered an alarm during emergency conditions to assist patients by alerting the healthcare experts to take the appropriate steps and decisions. Darshan [23] presented a healthcare monitoring system that offers emergency assistance to the patients through assessing their emergency condition based on their movement monitoring.

This paper reports a practical cryptosystem to secure the transmission of medical images in an Internet of Healthcare Things (IoHT) environment. The dynamics of a 2D trigonometric map created utilizing various well-known maps, such as logistic-sine-cosine maps, are investigated in this research. The map has an endless number of solutions, according to a stability study. The map's complicated dynamics are demonstrated using the Lyapunov exponent, bifurcation figure, and phase portrait. A strong cryptosystem is built using the map's sequences. First, the newly developed trigonometric map generates three key streams combined with the picture mechanisms (R, G, and B) to calculate the Hamming distance. The output distance vector, conforming to each section, is then Bit-XORed with each of the critical streams. The subsequent shuffled vectors are then Bit-XORed (diffusion) by the saved outputs as of the early stage, and finally, the image vectors are joined to create the encrypted image. The data stored in the system must be encrypted or anonymized using cryptography and data anonymization techniques [24].

Tamper-proof steganography includes efficient procedures to encrypt the image or concealed message before implanting it. Quantum inspired variants of controlled alternative quantum walks (CAQWs) which are used to determine the pixels for secret/hidden bits in the carrier image. The design employed in our method prevents the need for pre or post-encryption of the carrier and secret images. Also, our design shortens removing hidden material since only the stego image and primary conditions to run the CAQWs are required. The designed protocol showed extraordinary outcomes in terms of their security, good visual quality, high resistance to data loss bouts, and high embedding capacity [25].

In quantum computing, a quantum algorithm is an algorithm which runs on a realistic model of quantum computation, the most commonly used model being the quantum circuit model of computation. Quantum walks establish a universal quantum computational model extensively utilized in cryptography. This research designed a new encryption appliance for privacy-preserving IoT-based healthcare schemes to defend patients' privacy. The encryption/decryption procedures are based on measured alternate quantum walks. The simulation results show that image encryption protocol is healthy and well organized for protecting patients' privacy protection [26]. Another

healthcare monitoring system was proposed in [27] that is based on lightweight sensor-enabled wearable devices. These devices collect, analyze, and share real-time patient healthcare information. In this model, data are collected using an “Arduino-based wearable system” with body sensor networks. Also, “LabVIEW” is integrated with this system to allow for remote patient surveillance.

In [28], a proposed system includes feasible IoT wearable devices to collect the patient’s health-related information. It used different ML-based classification methods such as “DT, logistic regression (LR), and library for support vector machine (LibSVM)” to forecast the occurrence of illnesses. Lastly, a “mathematical model” is utilized to propose a personalized IoT solution for each scenario.

Ghose [29] proposed an ML algorithm that uses the Naive Bayes (NB) algorithm and the SVM to detect and analyze heart disease. To predict coronary heart disease, SVM and Bayes Net algorithms were used [30, 31].

Various ML methods are used to improve the accuracy of diabetic disease inspection. The authors suggested an ML algorithm that uses NB [32] and DT [33] to predict diabetic disorders. Ootom [34] used ML algorithm classification and regression tree (CART) to help diagnose diabetes.

ML methods are also used to forecast thyroid diseases. SVM and DT are employed as classification algorithms, with the dataset coming from the UCI repository. Advanced methods for thyroid disease diagnosis proposed in [35] use fuzzy maps and data mining algorithms.

A wireless body sensor network (WBSN)-enabled IoT healthcare solution was proposed in [36, 37]. It uses a wireless body network made up of small, lightweight sensor nodes to keep track of the patient. This solution employs a variety of ML techniques to improve security by protecting WBSN from hackers.

3. Effectiveness Comparison of IoMT-Enabled Smart Healthcare Model

In this research work, a model is developed to monitor elderly people’s activities and provide them automated assistance when required. The IoMT has a vital role to play in monitoring elderly people. In this model, an artificial neural network (ANN) approach is used to monitor elderly people intelligently and efficiently.

Figure 1 demonstrates the IoMT-enabled SHC model, which depends on two phases: training and validation. The cloud is used for communication between the two phases. The training phase is composed of the three layers: the sensory layer, preprocessing layer, and application layer. The sensory layer contains various input parameters that get the human body’s values and pass these values through IoMT to store in a database. The data received through IoMT might hold missing or noisy data. So, they are known as raw data.

The next layer is the preprocessing layer. It is an important layer that handles the missing values by moving average and normalization to remove the noisy data. After this process, the preprocessing layer’s output is sent to the application layer. The application layer is further divided into two layers, namely, prediction layer and performance layer.

In the prediction layer, ANN is utilized to predict the disease further. It consists of three layers named as input, hidden, and output layers. The backpropagation algorithm involves several steps: weight initialization, feedforward, feedback error propagation, and weight and bias updates. In the hidden layer, each neuron has a function of activation like $f(j) = \text{sigmoid}(j)$.

The sigmoid function for input and the model’s hidden layer can be written as

$$\psi_j = b_1 + \sum_{i=1}^m (\omega_{ij} * r_i), \quad (1)$$

$$\varphi_j = \frac{1}{1 + e^{-\psi_j}}, \quad \text{where } j = 1, 2, 3, \dots, n.$$

Input is taken from the output layer as shown in the following equation:

$$\psi_k = b_2 + \sum_{j=1}^n (v_{jk} * \varphi_j). \quad (2)$$

Output layer activation function is shown in the following equation

$$\varphi_k = \frac{1}{1 + e^{-\psi_k}}, \quad \text{where } k = 1, 2, 3, \dots, r. \quad (3)$$

Error in backpropagation can be written as

$$E = \frac{1}{2} \sum_k (\tau_k - \varphi_k)^2, \quad (4)$$

where τ_k represents the desired output and φ_k is the calculated output.

In equation (5), the layer is written as the rate of change in weight for the output.

$$\begin{aligned} \varphi W &\propto -\frac{\partial E}{\partial W}, \\ \varphi v_{j,k} &= -\epsilon \frac{\partial E}{\partial v_{j,k}}. \end{aligned} \quad (5)$$

The chain rule method is written as follows:

$$\varphi v_{j,k} = -\epsilon \frac{\partial E}{\partial \varphi_k} \times \frac{\partial \varphi_k}{\partial \psi_k} \times \frac{\partial \psi_k}{\partial v_{j,k}}. \quad (6)$$

The value of weight change can be found by inserting the values in equation (6), and the results are shown in the following equation:

$$\begin{aligned}\varphi v_{j,k} &= \epsilon (\tau_k - \varphi_k) \times \varphi_k (1 - \varphi_k) \times (\varphi_j), \\ \varphi v_{j,k} &= \epsilon \xi_k \varphi_j,\end{aligned}\quad (7)$$

where

$$\begin{aligned}\xi_k &= (\tau_k - \varphi_k) \times \varphi_k (1 - \varphi_k), \\ \varphi \omega_{i,j} &\propto - \left[\sum_k \frac{\partial E}{\partial \varphi_k} \times \frac{\partial \varphi_k}{\partial \psi_k} \times \frac{\partial \psi_k}{\partial \varphi_j} \right] \times \frac{\partial \varphi_j}{\partial \psi_j} \times \frac{\partial \psi_j}{\partial \omega_{i,j}}, \\ \varphi \omega_{i,j} &= -\epsilon \left[\sum_k \frac{\partial E}{\partial \varphi_k} \times \frac{\partial \varphi_k}{\partial \psi_k} \times \frac{\partial \psi_k}{\partial \varphi_j} \right] \times \frac{\partial \varphi_j}{\partial \psi_j} \times \frac{\partial \psi_j}{\partial \omega_{i,j}}, \\ \varphi \omega_{i,j} &= \epsilon \left[\sum_k (\tau_k - \varphi_k) \times \varphi_k (1 - \varphi_k) \times (v_{j,k}) \right] \times \varphi_k (1 - \varphi_k) \times r_i, \\ \varphi \omega_{i,j} &= \epsilon \left[\sum_k (\tau_k - \varphi_k) \times \varphi_k (1 - \varphi_k) \times (v_{j,k}) \right] \times \varphi_j (1 - \varphi_j) \times r_i, \\ \varphi \omega_{i,j} &= \epsilon \left[\sum_k \xi_k (v_{j,k}) \right] \times \varphi_j (1 - \varphi_j) \times r_i, \\ \varphi \omega_{i,j} &= \epsilon \xi_j r_i,\end{aligned}\quad (8)$$

where

$$\xi_j = \left[\sum_k \xi_k (v_{j,k}) \right] \times \varphi_j (1 - \varphi_j). \quad (9)$$

Output and hidden layers are shown in equation (10) in which the weight and bias between them are updated:

$$v_{j,k}^+ = v_{j,k} + \lambda_F \varphi v_{j,k}. \quad (10)$$

The process of updating weight and bias between the input layer and the hidden layer is shown in the following equation:

$$\omega_{i,j}^+ = \omega_{i,j} + \lambda_F \varphi \omega_{i,j}, \quad (11)$$

where λ_F is the learning rate of the IoMT-enabled smart model. The convergence of the model depends upon the careful selection of λ_F . The equation above is used to update the weight of the hidden and input layers.

After the prediction layer, the output of the prediction layer will be sent to the performance layer to predict the healthcare issue based on accuracy and miss rate whether the learning criteria meet or not. In the case of “no,” the

prediction layer will be updated, but in the case of “yes,” the output will be stored on the cloud database. Now, in the validation phase, the input will be sensed from input layer parameters and sent to ANN to determine whether the healthcare issue is found or not. In the case of “no,” the process will be discarded, and in the case of “yes,” the message will display that healthcare issue found.

4. Simulation Results

This paper presents a SHC system that can enable remote healthcare consultations, elderly management, and home-care over global wide area networks and heterogeneous platforms. The simulation results of the research are described below.

Tables 1 and 2 show the training and validation concerning accuracy and miss rate. ANN algorithm has been implemented to the dataset of 4848 sets of records; furthermore, 3393 samples (70%) and 1455 samples (30%) are used for training and validation purpose, respectively. In performance evaluation layer, various statistical metrics are used to measure the proposed system performance as shown in equation (12). The parameters are derived by the following formulas:

$$\begin{aligned}
\text{sensitivity} &= \frac{\sum \text{true positive}}{\sum \text{condition positive}}, \\
\text{specificity} &= \frac{\sum \text{true negative}}{\sum \text{condition negative}}, \\
\text{accuracy} &= \frac{\sum \text{true positive} + \sum \text{true negative}}{\sum \text{total population}}, \\
\text{miss rate} &= \frac{\sum \text{false negative}}{\sum \text{condition positive}}, \\
\text{fallout} &= \frac{\sum \text{false positive}}{\sum \text{condition negative}}, \\
\text{likelihood positive ratio} &= \frac{\sum \text{true positive ratio}}{\sum \text{false positive ratio}}, \\
\text{likelihood negative ratio} &= \frac{\sum \text{true positive ratio}}{\sum \text{false positive ratio}}, \\
\text{positive predictive value} &= \frac{\sum \text{true positive}}{\sum \text{predicted condition positive}}, \\
\text{negative predictive value} &= \frac{\sum \text{true negative}}{\sum \text{predicted condition negative}}.
\end{aligned} \tag{12}$$

The IoMT-enabled SHC model estimates anticipated output as negative (−1) and positive (1). The consequent output of value negative (−1) indicates that a health issue is found, whereas positive (1) indicates no health issue.

Table 1 shows IoMT-enabled SHC model monitoring of elderly people during the training phase. During training, 1522 positive (healthy) and 1871 negative (unhealthy) samples are used. It clearly observed that 1422 samples are correctly predicted in the case of healthy samples., while 100 records are wrongly predicted as negative, implying that there is a healthcare issue. Similarly, a total of 1871 samples were picked, with negative indicating the presence of a healthcare issue. 1754 samples are correctly predicted as negative, indicating the presence of a healthcare issue. 117 samples are incorrectly predicted as positive, indicating the absence of a healthcare issue, although a healthcare issue exists.

Table 2 displays the model's prediction of health status during the validation phase. During validation, a total of 1455 samples are utilized in which 686 are positive (healthy) and 769 are negative (unhealthy) samples are used. It is discovered that 626 samples contain true positives that are accurately forecasted, and no healthcare issue is discovered, while 60 records are mistakenly predicted as negative, implying that a healthcare issue is discovered. Similarly, 769

samples were picked, with negative indicating the presence of a health issue. 710 samples are accurately predicted as negative, indicating the presence of a healthcare issue. 59 samples are incorrectly predicted as positive, indicating the absence of a healthcare issue, although healthcare issues exist.

Table 3 shows the model's accuracy, sensitivity, specificity, miss rate, and precision during the training and validation phase. The model during training gives accuracy, sensitivity, specificity, miss rate, and precision of 0.936, 0.934, 0.937, 0.064, and 0.924, respectively. During validation, the model gives accuracy, sensitivity, specificity, miss rate, and precision of 0.918, 0.912, 0.923, 0.082, and 0.913, respectively.

Furthermore, specific statistical measures of the model are included to forecast values during training, e.g., fallout, LR+, LR−, and NPV give the result of 0.062, 15.064, 0.068, and 0.946, and during validation, they are 0.076, 12, 0.088, and 0.922, respectively.

According to the results shown in Table 4, the IoMT-enabled SHC model is more effective than the conventional ML classification-based methods like SVM, KNN, and DT while analyzing the IoMT data.

The comparison of the previously published approach and the proposed model to monitor elderly people is

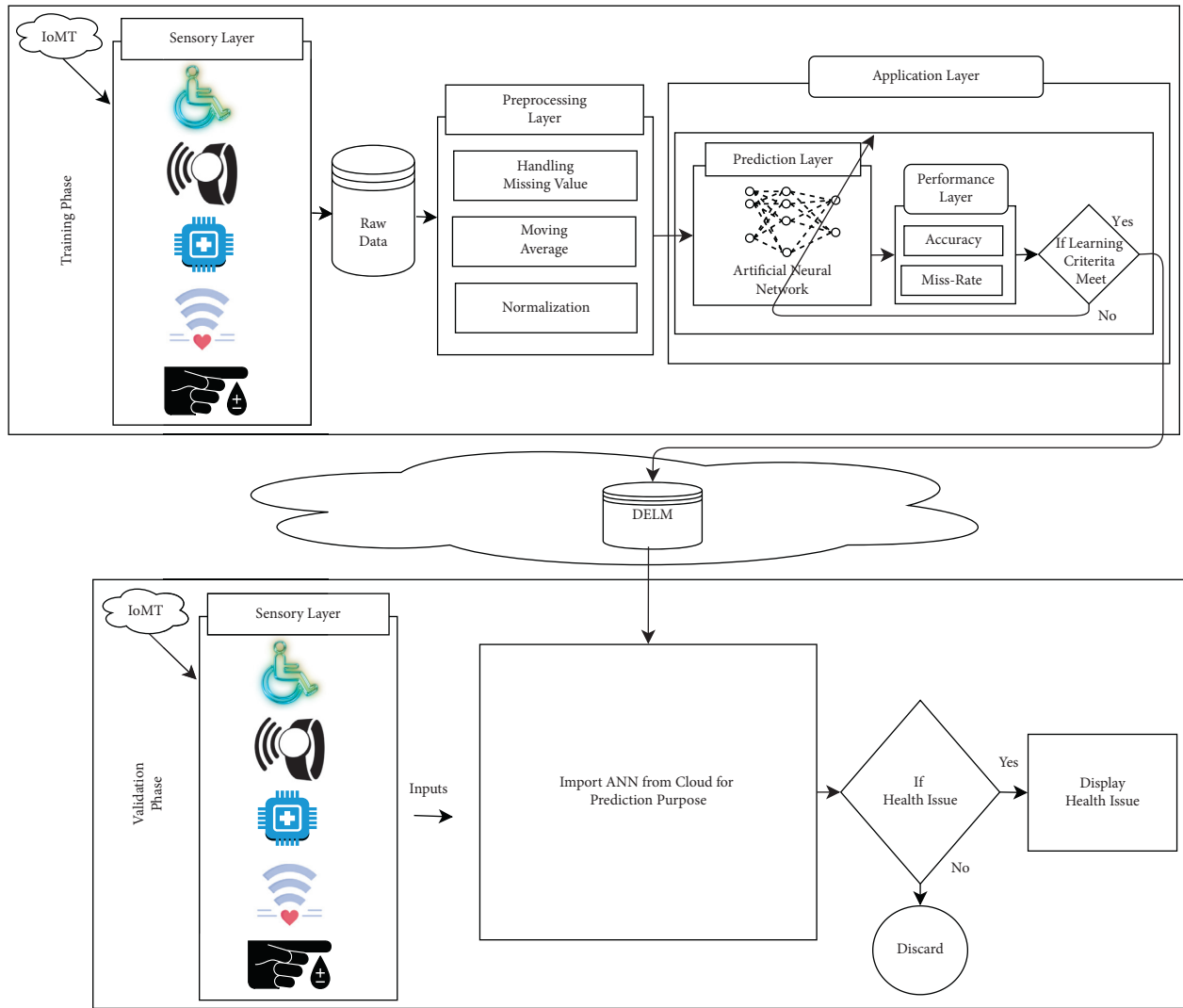


FIGURE 1: IoMT-enabled smart model to monitor elderly people’s health using ML.

TABLE 1: Training of the IoMT-enabled SHC model during the monitoring of elderly people’s health.

Training results			
	Total number of samples (3393)		Result (output)
Input	Expected output	Predicted positive	Predicted negative
	1522 (positive)	True positive (TP) 1422	False negative (FN) 100
	1871 (negative)	False positive (FP) 117	True negative (TN) 1754

TABLE 2: Validation of the IoMT-enabled SHC model during the monitoring of elderly people’s health.

Validation results			
	Total number of samples (1455)		Result (output)
Input	Expected output	Predicted positive	Predicted negative
	686 (positive)	True positive (TP) 626	False negative (FN) 60
	769 (negative)	False positive (FP) 59	True negative (TN) 710

TABLE 3: The training and validation of the IoMT-enabled SHC model using different statistical measures.

	Accuracy	Sensitivity (TPR)	Specificity (TNR)	Miss rate (FNR)	Fallout (FPR)	LR+	LR-	PPV (precision)	NPV
Training	0.936	0.934	0.937	0.064	0.062	15.064	0.068	0.924	0.946
Validation	0.918	0.912	0.923	0.082	0.076	12	0.088	0.913	0.922

TABLE 4: Effectiveness comparison of precision, recall, fscore, and accuracy of different ML-based IoMT-enabled SHC models.

	Accuracy	Sensitivity (TPR)	Specificity (TNR)	Miss rate (FNR)	Fallout (FPR)	LR+	LR-	PPV (precision)	NPV
ANN	0.936	0.934	0.937	0.064	0.062	15.064	0.068	0.924	0.946
SVM	0.89233	0.89213	0.90113	0.08523	0.07423	9.3443	0.09143	0.65432	0.90501
KNN	0.84828	0.84765	0.8513	0.08012	0.06923	8.2311	0.87213	0.84612	0.84621
Decision tree	0.83886	0.83776	0.8432	0.07933	0.06831	8.1432	0.85231	0.83531	0.83632

TABLE 5: Comparison of the proposed system with previously published approaches.

Authors	Approach	Accuracy	Miss rate
Qin et al. [38]	Logistic regression	0.672	0.328
Tang et al. [39]	XGBoost classifier	0.892	0.108
Abro et al. [40]	Random forest classifier	0.606	0.394
Dimitriadis et al. [41]	RF	0.619	0.381
Liu et al. [42]	MSDNN	0.754	0.246
Lu et al. [43]	3D ResNet	0.830	0.170
The proposed model to monitor elderly people	Artificial neural network	0.936	0.064

shown in Table 5. The proposed model achieved 0.936 accuracy for monitoring the healthcare of elderly people in an efficient way which is better than the previous approaches.

5. Conclusion

The research is responsible for overcoming the challenges of elderly care services. The research realizes the needs of the elderly healthcare system. In this research work, innovative medical services for the elderly are compared based on the real needs and challenges of the elderly and caregivers. To meet the basic needs of elderly healthcare, the researchers used ML techniques for getting better outcomes. After simulation results, the research conclusions are summarized: the elderly healthcare service interface of the IoMT has a higher accuracy during validation, which gives accuracy, sensitivity, specificity, miss rate, and precision of 0.918, 0.912, 0.923, 0.082, and 0.913, respectively. The system of the proposed approach may be improved in the future by using a fusion-based machine learning approach and federated learning approach as well.

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 regarding the publication of this paper.

References

- [1] T. Abbas Khan, S. Abbas, A. Ditta et al., "IoMT-based smart monitoring hierarchical fuzzy inference system for diagnosis of covid-19," *Computers, Materials & Continua*, vol. 65, no. 3, pp. 2591–2605, 2020.
- [2] K. Sultan, I. Naseer, R. Majeed et al., "Supervised machine learning-based prediction of covid-19," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 21–34, 2021.
- [3] Q.-T.-A. Khan, S. Abbas, M. Adnan Khan, A. Fatima, S. Alanazi, and N. Sabri Elmitwally, "Modelling intelligent driving behaviour using machine learning," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 3061–3077, 2021.
- [4] WHO, *Global Health and Aging (NIH Publication No. 11-7737)*, World Health Organization, Geneva, 2011.
- [5] A. Ghaffar, S. Alanazi, M. Alruwaili et al., "Multi-stage intelligent smart lockdown using sir model to control covid 19," *Intelligent Automation & Soft Computing*, vol. 28, no. 2, pp. 429–445, 2021.
- [6] S. Aftab, S. Alanazi, M. Ahmad, M. Adnan Khan, A. Fatima, and N. Sabri Elmitwally, "Cloud-based diabetes decision support system using machine learning fusion," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 1341–1357, 2021.
- [7] A. Inam, Zhuli, A. Sarwar et al., "Detection of covid-19 enhanced by a deep extreme learning machine," *Intelligent Automation & Soft Computing*, vol. 27, no. 3, pp. 701–712, 2021.
- [8] S. Majed Alotaibi, A. U. Atta-ur-Rahman, M. Adnan Khan, and M. A. Khan, "Ensemble machine learning based identification of pediatric epilepsy," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 149–165, 2021.
- [9] A. Hannan Khan, M. Adnan Khan, S. Abbas et al., "Simulation, modeling, and optimization of intelligent kidney disease predication empowered with computational

- intelligence approaches,” *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1399–1412, 2021.
- [10] N. Tabassum, A. Ditta, T. Alyas et al., “Prediction of cloud ranking in a hyperconverged cloud ecosystem using machine learning,” *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3129–3141, 2021.
- [11] K. L. Courtney, *Needing Smart Home Technologies: The Perspectives of Older Adults in Continuing Care Retirement Communities*, Radcliffe Medical, Radcliffe, England, 2008.
- [12] A. S. Iyer, “Diagnosis of diabetes using classification mining techniques,” pp. 1–8, 2015, <https://arxiv.org/abs/1502.03774>.
- [13] S. A. Asthana, “A recommendation system for proactive health monitoring using IoT and wearable technologies,” in *Proceedings of the IEEE International Conference on AI & Mobile Services (AIMS)*, pp. 1–10, Honolulu, HI, USA, June 2017.
- [14] S.-H. Kim and K. Chung, “Emergency situation monitoring service using context motion tracking of chronic disease patients,” *Cluster Computing*, vol. 18, no. 2, pp. 747–759, 2015.
- [15] A. J. Jara, M. A. Zamora-Izquierdo, and A. F. Skarmeta, “Interconnection framework for mHealth and remote monitoring based on the internet of things,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 47–65, 2013.
- [16] M. Chan, E. Campo, D. Estève, and J.-Y. Fourniols, “Smart homes-current features and future perspectives,” *Maturitas*, vol. 64, no. 2, pp. 90–97, 2009.
- [17] V. Chandel, A. Sinharay, N. Ahmed, and A. Ghose, “Exploiting IMU sensors for IoT enabled health monitoring,” in *Proceedings of the First Workshop on IoT-Enabled Healthcare and Wellness Technologies and Systems*, pp. 21–22, Singapore, June 2016.
- [18] R. Singh, “A proposal for mobile e-care health service system using IoT for Indian scenario,” *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 6, no. 1, 2014.
- [19] S. J. Pinto, “We-care: an IoT-based health care system for elderly people,” in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT)*, pp. 1–7, Toronto, ON, Canada, March 2017.
- [20] C. H. Huang and K. W. Cheng, “RFID technology combined with IoT application in the medical nursing system,” *Bulletin of Networking, Computing, Systems, and Software*, vol. 3, no. 1, pp. 20–24, 2014.
- [21] J. Cocco, “Smart home technology for the elderly and the need for regulation,” *Pittsburgh Journal of Environmental & Public Health Law*, vol. 6, no. 85, pp. 1–14, 2011.
- [22] E. I. Papageorgiou, “Fuzzy cognitive map-based decision support system for thyroid diagnosis management,” in *Proceedings of the IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, pp. 10–14, Hong Kong, China, September 2008.
- [23] K. R. Darshan, “A comprehensive review on the usage of Internet of Things (IoT) in healthcare system,” in *Proceedings of the 2015 IEEE International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, pp. 8–13, Mandya, India, December 2015.
- [24] N. Tsafack, S. Sankar, B. Abd-El-Atty et al., “A new chaotic map with dynamic analysis and encryption application in internet of health things,” *IEEE Access*, vol. 8, pp. 137731–137744, 2020.
- [25] B. Abd-El-Atty, A. M. Iliyasa, H. Alaskar, A. A. Abd El-Latif, and A. Ahmed, “A robust quasi-quantum walks-based steganography protocol for secure transmission of images on cloud-based E-healthcare platforms,” *Sensors*, vol. 20, no. 11, pp. 3108–3121, 2020.
- [26] A. A. Abd EL-Latif, B. Abd-El-Atty, E. M. Venegas-Andraca, and S. E. V. Andraca, “Controlled alternate quantum walks based privacy preserving healthcare images in Internet of Things,” *Optics & Laser Technology*, vol. 124, Article ID 105942, 2020.
- [27] V. A. Vippalapalli, “Internet of things (IoT) based smart health care system,” in *Proceedings of the IEEE International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)*, pp. 21–26, Paralakhemundi, India, October 2016.
- [28] M. A. Fischer, “From books to bots: using medical literature to create a chatbot,” in *Proceedings of the First Workshop on IoT-Enabled Healthcare and Wellness Technologies and Systems*, pp. 43–46, Singapore, June 2016.
- [29] A. E. Ghose, “UbiHeld: ubiquitous healthcare monitoring system for elderly and chronic patients,” in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pp. 23–27, Switzerland, September 2013.
- [30] P. A. Gope, “BSN-Care: a secure IoT-based modern healthcare system using body sensor network,” *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1368–1376, 2015.
- [31] K. S. Kumar, “IoT based health monitoring system for autistic patients,” in *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC-16’)*, pp. 25–32, Springer, Cham, 2016.
- [32] G. E. Matar, “Internet of Things in sleep monitoring: an application for posture recognition using supervised learning,” in *Proceedings of the IEEE 18th International Conference on E-Health Networking, Applications and Services (Healthcom)*, pp. 41–46, Munich, Germany, September 2016.
- [33] L. Moore, “Measuring quality and effectiveness of prehospital EMS,” *Prehospital Emergency Care*, vol. 3, no. 4, pp. 325–331, 1999.
- [34] A. F. Otoom, “Effective diagnosis and monitoring of heart disease,” *International Journal of Software Engineering and Its Applications*, vol. 9, no. 1, pp. 143–156, 2015.
- [35] G. Parthiban and K. Srivatsa, “Applying machine learning methods in diagnosing heart disease for diabetic patients,” *International Journal of Applied Information Systems*, vol. 3, no. 7, pp. 25–30, 2012.
- [36] C. E. Puri, “ICarMa: inexpensive cardiac Arrhythmia management—an IoT healthcare analytics solution,” in *Proceedings of the First Workshop on IoT-Enabled Healthcare and Wellness Technologies and Systems*, pp. 1–6, Singapore, June 2016.
- [37] S. K. Sen, “Application of meta-learning algorithms for the prediction of diabetes disease,” *International Journal of Advanced Research in Computer Science and Management Studies*, vol. 2, no. 12, pp. 123–132, 2014.
- [38] F. Y. Qin, Z. Q. Lv, D. N. Wang, B. Hu, and C. Wu, “Health status prediction for the elderly based on machine learning,” *Archives of Gerontology and Geriatrics*, vol. 90, pp. 104121–104211, 2020.
- [39] H. Tang, E. Yao, G. Tan, and X. Guo, “A fast and accurate 3D fine-tuning convolutional neural network for alzheimer’s disease diagnosis,” in *Proceedings of the International CCF Conference on Artificial Intelligence*, pp. 115–126, Springer, Singapore, 2018.
- [40] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, “Deep residual learning for neuroimaging: an application to predict progression to Alzheimer’s disease,”

- Journal of Neuroscience Methods*, vol. 339, no. 9, pp. 108701–108713, 2020.
- [41] S. I. Dimitriadis, D. Liparas, and M. N. Tsolaki, “Random forest feature selection, fusion and ensemble strategy: combining multiple morphological MRI measures to discriminate among healthy elderly, MCI, cMCI and alzheimer’s disease patients: from the alzheimer’s disease neuroimaging initiative (ADNI) database,” *Journal of Neuroscience Methods*, vol. 302, no. 2, pp. 14–23, 2018.
- [42] S. Liu, S. Liu, W. Cai, K. Pujol, R. Kikinis, and D. Feng, “Early diagnosis of Alzheimer’s disease with deep learning,” in *Proceedings of the IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014*, pp. 1015–1018, Beijing, China, January 2014.
- [43] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, “Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer’s disease using structural MR and FDG-PET images,” *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018.

Research Article

A Blockchain-Based Federated Learning Method for Smart Healthcare

Yuxia Chang ^{1,2,3}, Chen Fang ⁴, and Wenzhuo Sun ^{1,2,3}

¹Department of Emergency Medicine, Henan Provincial People's Hospital, Zhengzhou 450001, China

²Key Laboratory of Nursing Medicine of Henan Province, Zhengzhou 450001, China

³People's Hospital of Zhengzhou University, Zhengzhou 450001, China

⁴Information Engineering University, Zhengzhou 450001, China

Correspondence should be addressed to Yuxia Chang; cyx9703@163.com and Chen Fang; 17756230629@163.com

Received 1 September 2021; Revised 7 November 2021; Accepted 10 November 2021; Published 24 November 2021

Academic Editor: Ahmed A. Abd El-Latif

Copyright © 2021 Yuxia Chang 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 of artificial intelligence and worldwide epidemic events has promoted the implementation of smart healthcare while bringing issues of data privacy, malicious attack, and service quality. The Medical Internet of Things (MIoT), along with the technologies of federated learning and blockchain, has become a feasible solution for these issues. In this paper, we present a blockchain-based federated learning method for smart healthcare in which the edge nodes maintain the blockchain to resist a single point of failure and MIoT devices implement the federated learning to make full of the distributed clinical data. In particular, we design an adaptive differential privacy algorithm to protect data privacy and gradient verification-based consensus protocol to detect poisoning attacks. We compare our method with two similar methods on a real-world diabetes dataset. Promising experimental results show that our method can achieve high model accuracy in acceptable running time while also showing good performance in reducing the privacy budget consumption and resisting poisoning attacks.

1. Introduction

With the growth in volume and types of clinical data, there is an urgent need for efficient mining models to analyze these data so as to help disease diagnosis, provide medical solutions, and improve the medical care for patients. Machine learning is an effective tool with powerful computation capabilities, which has been used in many fields, such as image recognition, natural language processing, and healthcare. However, machine learning models can only reach high accuracy with abundant training data, which is especially important in healthcare that sometimes decides whether a patient's life can be saved. Traditional centralized training methods usually require collecting a large amount of data from a powerful cloud machine, which may cause serious user privacy leakage, especially in the medical domain. Many governments have issued laws prohibiting collecting data relevant to user privacy, such as the European Union's General Data Protection Regulation (GDPR). The

emergence of the Medical Internet of Things (MIoT) empowers traditional fields such as healthcare, medical care, public health, and community services, where large numbers of MIoT devices such as wearable sensors are distributed at the edge of the network to collect patient data. Federated learning (FL) [1], as a distributed machine learning framework, can allow multiple devices to train machine learning models collaboratively without sharing their raw data, which just contributes to realizing smart healthcare in the MIoT while reducing the privacy leakage of patients.

A typical FL-based smart healthcare application is shown in Figure 1, where onboard sensors collect clinical data from patients, multiple edge devices perform FL algorithm collaboratively, and the final machine learning models evaluate the patient's physical health and even request the emergency service in the cloud if necessary. However, one of the drawbacks of vanilla FL is that it needs a trustworthy central server to aggregate the model parameters uploaded by devices and distribute the global model to all devices. Once the

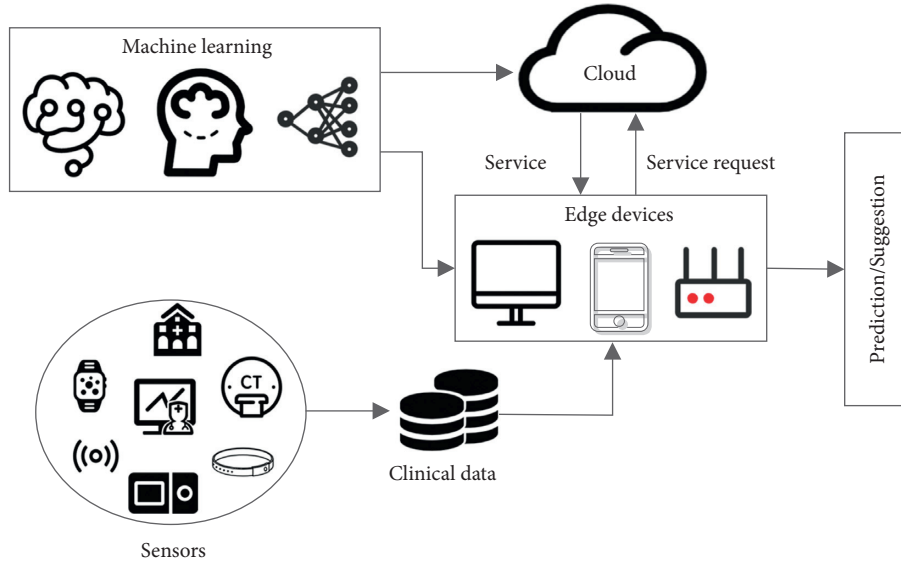


FIGURE 1: A typical FL-based smart healthcare application.

central server is crashed by attackers, the FL training will stop. As a ledger with properties of tamper-proof, collective maintenance, and traceability, blockchain can replace the central server to decentralize the coordination process in FL, thus resisting single points of failure and illegal tampering attacks. In this way, the traditional elements in blockchain can be mapped into the training stages of FL as follows: each block represents a single training round, where the stored transactions represent model parameters uploaded by devices in that round. Then all devices can look up the model parameters from the latest block and update their local models. In view of these advantages, lots of blockchain-based FL methods have been proposed to be applied in many fields, such as smart home [2], Industrial Internet of Things (IoT) [3], and smart healthcare [4]. But with more and more advanced privacy attacks, there are still several challenges that need to be addressed while applying blockchain-based FL to healthcare: (1) the model parameters stored in the blockchain may still be stolen by attackers to infer the original private clinical data; (2) clinical data of some edge devices may be poisoned to mislead the FL process; (3) edge devices have no incentive to contribute dataset and computing power to FL.

Aimed at the above challenges, this paper integrates FL with blockchain and advanced cryptography to realize a smart healthcare model in a secure and privacy-preserving manner. The main contributions of the paper are mainly as follows:

- (1) We propose a blockchain-based FL framework for smart healthcare, which not only builds an accurate collaborative model based on multiple edge devices but also provides governance of the whole training process.
- (2) To add an extra layer of security of blockchain-based FL, we propose adaptive differential privacy (DP) algorithm that adapts noise according to the training process, balancing privacy, and model accuracy.

- (3) We design an efficient consensus protocol based on gradient verification, which encourages reliable MIoT devices and edge nodes to contribute their data and computing power to federated learning.

The rest of the paper is organized as follows. We introduce related works in Section 2 and give our method in Section 3. Section 4 shows the experimental results of our method. We draw conclusions and outline future work in Section 5.

2. Related Work

With the development of artificial intelligence (AI), it is a common practice to deploy AI applications to assist medical diagnosis, which can improve the diagnosis rate of diseases and reduce the waiting time of patients. Dai et al. [5] turned the prediction of hospitalization task into a supervised classification problem, resulting in a considerable amount of potential saving in medical resources. Son et al. [6] developed a Support Vector Machine (SVM) model to identify predictors of medication adherence in heart failure patients. Tariq et al. [7] developed a multimodal fusion AI model from past medical data to predict the severity of COVID-19. In order to solve the problem of the absence of reliable data, Sedik et al. [8] presented a data augmentation framework to expand the limited dataset and used convolutional neural network and convolutional long short-term memory models to detect the COVID-19. However, the above-centralized training methods [5–8] need to collect sensitive clinical data in a single database, which is undesirable due to data privacy concerns. Instead, federated learning emerges as a distributed framework that performs collaborative learning while keeping all the sensitive data locally, providing a privacy-preserving solution for connecting the fragmented healthcare data on the edge devices. Many works that used FL in smart healthcare have been proposed in recent years. Qayyum et al. [9] proposed a clustered FL-based method to

process clinical visual data at the edge so as to allow remote hospitals to benefit from multimodal data in a privacy-preserving manner. Brisimi et al. [10] predicted hospitalizations for patients with heart diseases by solving distributed sparse Support Vector Machine problems using FL. Xu et al. [11] gave a review for FL methods and pointed out the implications and potentials of FL in healthcare particularly. Zhang et al. [12] employed differential private generative adversarial network (DPGAN) to generate diverse patient data in a privacy-preservation way and leveraged FL to train COVID-19 models in collaboration with multiple hospitals. But these works [9–12] all need a central server to aggregate and distribute model parameters during the federated learning, which is vulnerable to a single point of failure attack.

To address this vulnerability, blockchain is introduced to enable full decentralized FL, which is also the idea of this paper. El Rifai et al. [13] integrated the FL and blockchain for the first time in a medical setting and proposed a smart contract to realize transparency and immutability while sharing knowledge. Passerat-Palmbach et al. [14] presented an advanced blockchain-orchestrated federated learning framework in medicine and outlined some challenges. Połap et al. [4] designed a multiagent medical system based on FL and blockchain that can separate complicated tasks into individual objects and process medical data in real time. While the integration of blockchain and FL can resist a single point of failure and enable life-cycle governance of the training process, due to the transparency of blockchain, it raises concerns with regard to the privacy of the model parameters. To this end, Liu et al. [15] proposed a blockchain-based secure FL framework for 5G networks, in which differential privacy noise was added on the updates to prevent inference attacks. Kumar et al. [16] proposed a blockchain-based FL framework that trained a collaborative deep learning model for COVID-19 detection using clinical data from multiple hospitals and added Laplace noise to the local gradients to ensure privacy. Rahman et al. [17] proposed a hybrid FL framework for the Internet of Health Things (IoHT) that supported lightweight DP to realize the privacy and anonymization of the IoHT data. However, differential privacy used by [15–17] will cause some loss of data utility, which will reduce the availability of smart healthcare. In this paper, we design an adaptive differential privacy algorithm to achieve a balance between data privacy and data utility.

On the other hand, poisoning attack [18] launched by malicious users is also another challenge faced by blockchain-based FL methods. Although Liu et al. [15] executed smart contracts to identify malicious participants who initiated poisoning attacks, they assumed that there was a public test dataset in advance, which was unrealistic for smart healthcare with private data of patients. In this paper, we present a simple gradient verification method that does not need a public test dataset to detect poisoning attacks.

3. Proposed Model

3.1. Threat Model. In this section, we give the threats faced by smart healthcare.

Threat 1. Potential data privacy leakage. AI models built on clinical data may be attacked by adversaries to infer patient privacy.

Threat 2. Single point of failure. Existing smart healthcare models rely on a central server to store the clinical data or exchange the model parameters. Once the central server is crashed, the model training will end with failure.

Threat 3. Poisoning attacks. Due to the vulnerability of the MIoT, adversaries may launch poisoning attacks on the MIoT device's data or local model parameters, which will compromise the correctness of FL.

For ease of understanding, the main symbols used in this paper are listed in Table 1.

3.2. System Architecture. Our smart healthcare system is mainly composed of the user layer and edge node layer. The user layer mainly includes wearable sensors, MIoT devices, and mobile terminals. They are used to monitor patients' physiological condition, collect clinical data, and train FL model locally. Edge nodes are mainly composed of base stations equipped with edge computing servers that have powerful computation and communication capabilities. They maintain the blockchain as miners, receive and store the model parameters, and authenticate the parameters by consensus protocol. The training process in one round is shown in Figure 2.

As shown in Figure 2, a complete training process of one round can be formulated as the following steps:

- (1) Hospitals determine and send the training task to the blockchain, and then the genesis block is created and distributed to all the MIoT devices and edge nodes to perform model initialization. The genesis block mainly contains the following information: ① initial model parameters w_0 and total training rounds T ; ② public keys of all parties that are used to create signatures; ③ initial reputation value of all edge nodes and MIoT devices; ④ reputation update function.
- (2) The MIoT device trains the model locally based on its collected clinical data and adds DP noise to the local gradient (see details in Section 3.3.1) so as to cope with Threat 1.
- (3) The MIoT device uploads the noised gradient along with the signature to its associated edge node in the form of transactions.
- (4) After receiving data from devices within their coverage, the edge nodes first verify the legality of the signature and then elect a verification committee to detect whether the local gradients are poisoned update (see details in Section 3.3.2) so as to cope with Threat 3.
- (5) A leader is randomly selected to generate a new block containing the necessary model parameters for this training round. The verification committee verifies the new block and broadcasts the valid one to

TABLE 1: Main symbols.

Symbols	Definitions
T	The total number of training rounds
K	The number of MIoT devices
M	The size of the verification committee
$g_{i,t}$	The local gradient of i -th device in the t -th training round
$\bar{g}_{i,t}$	The noised gradient of i -th device in the t -th training round
C_t	The gradient clipping threshold in the t -th training round
G	Prior threshold
ϵ	Privacy budget
δ	Violation probability of the “pure” differential privacy

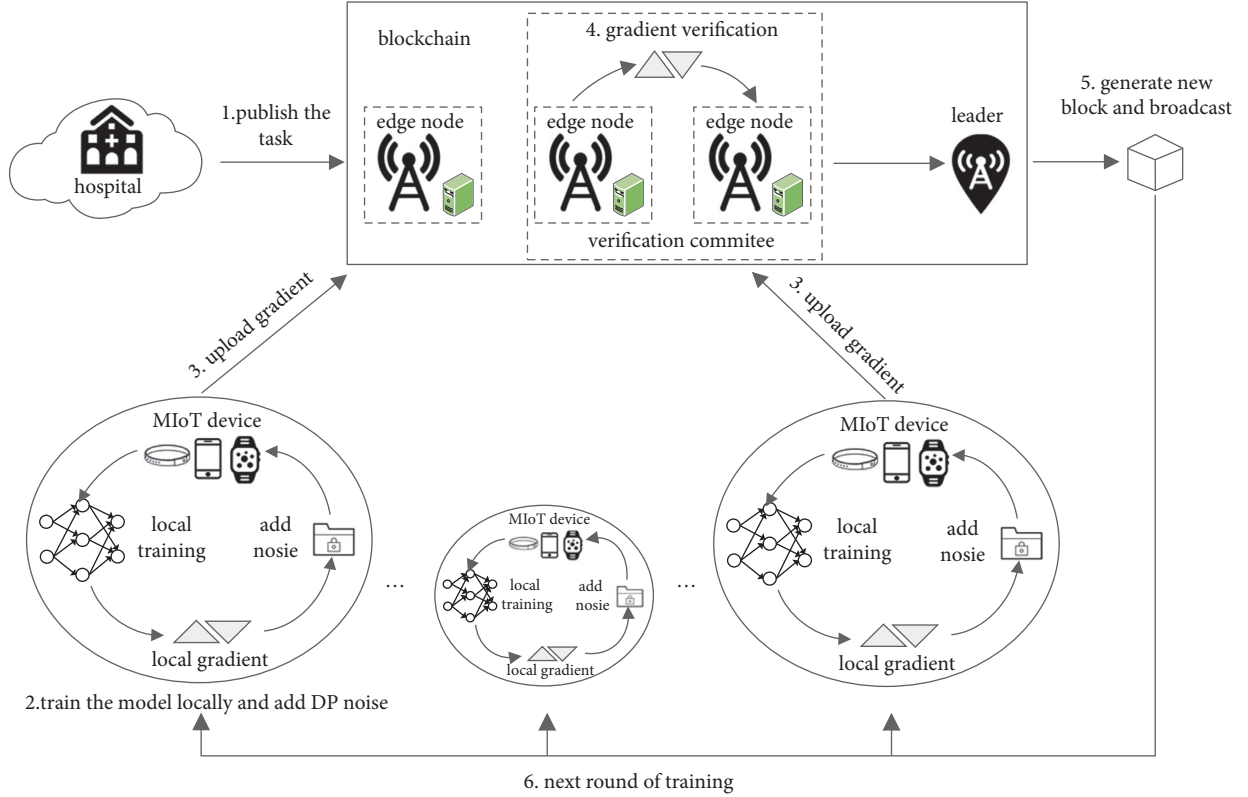


FIGURE 2: Training process in one round.

synchronize the ledgers of all edge nodes (see details in Section 3.3.2) so as to cope with Threat 2.

- (6) The MIoT device downloads the latest block from its associated edge node and updates its local model by the global gradient stored in the block. The next training round starts from step 2 until the model converges or the maximum rounds are reached.

Next, we will introduce the main construction of our method in detail.

3.3. Construction of Method

3.3.1. Adaptive Differential Privacy Algorithm. The advanced privacy attacks such as model inversion [19] and model extraction attack [20] have shown that the model parameters stored in the blockchain are not enough to

protect the privacy of raw clinical data. References [18, 21] used Shamir secret sharing and threshold Paillier encryption to protect local gradients, respectively, but both consume large computation overhead. In contrast, differential privacy technology needs less computation overhead, which is more suitable for MIoT devices with limited resources.

In principle, DP is a strictly provable mathematical framework whose basic idea is to add carefully designed noise to the input or output of a function so that the modification of any individual sample in the dataset will not have a significant impact on the output. The related definitions are as follows.

Definition 1 (differential privacy [22]). A randomized algorithm $A: D \rightarrow R$ is (ϵ, δ) -differentially private if for any two datasets D and D' differing in an individual sample and any output $O \in R$:

$$\Pr[A(D) = O] \leq e^\epsilon \times \Pr[A(D') = O] + \delta, \quad (1)$$

where ϵ is the privacy budget. A small ϵ means a higher level of privacy preservation but greater accuracy loss for algorithm A and vice versa. δ is the probability that measures the violation of the “pure” differential privacy, which is usually a small value.

Definition 2 (sensitivity [22]). For any real-valued function $f: D \rightarrow \mathbb{R}^d$ with D as the input dataset and \mathbb{R}^d as the d -dimensional vector output, the sensitivity of f is

$$\Delta f = \max_{D, D'} \|f(D) - f(D')\|_p, \quad (2)$$

where D and D' are two adjacent datasets differing in an individual sample and $\|\bullet\|_p$ denotes the L_p norm.

Definition 3 (Gaussian mechanism [22]). Assume that L_2 norm is used to compute the sensitivity of function f . (ϵ, δ) -differentially privacy can be realized via adding Gaussian noise to the output of function f :

$$A(D) = f(D) + N(0, (\Delta f \sigma)^2 I), \quad (3)$$

where $N(0, (\Delta f \sigma)^2 I)$ is the Gaussian distribution with mean 0 and standard deviation $\Delta f \sigma$ and I is the identity matrix.

From the above definitions, we can see that the private information in a dataset can be hidden by adding noise, but at the same time, the noise will lower the data utility. Reference [23] added noise on the raw data by local differential privacy, but it reduced the model accuracy severely. Reference [24] added Gaussian noise on the clipped gradient but did not explain how to select the clipping threshold. The value of the threshold is important to the FL model: too large a value will add excessive noise and too small a value will over clip the gradient, both of which will cause serious accuracy loss. Aimed at this issue, we draw on the idea of the RMSProp optimization algorithm and propose an adaptive differential privacy algorithm for MIoT devices, which can flexibly adjust the clipping threshold according to the training process to reduce the negative impact of noise on the model accuracy.

RMSProp is a variant of gradient descent algorithm for machine learning, which speeds up the convergence rate by adjusting the step size. The iteration formula is as follows:

$$\begin{aligned} \mathbb{E}[g^2]_t &\leftarrow (1 - \gamma)\mathbb{E}[g^2]_{t-1} + \gamma(g_t)^2, \\ \theta_t &\leftarrow \theta_{t-1} - \eta \frac{g_t}{\sqrt{\mathbb{E}[g^2]_t + \epsilon_0}}, \end{aligned} \quad (4)$$

where θ_t is the model parameter in the t -th iteration, g_t is the gradient, η is the learning rate, $\mathbb{E}[g^2]_{t-1}$ is the cumulative square of the historical gradient, γ is an exponent of gradient accumulation, and ϵ_0 is to ensure that the divisor is not zero, generally set to 10^{-8} . Due to the continuity and gradualness of the convergence process [25], the historical gradient can usually be used to estimate the current gradient. Therefore,

$\mathbb{E}[g^2]_{t-1}$ in the RMSProp algorithm can be regarded as the prior knowledge of the current gradient.

The existing method [26] lets $C \approx \|\tilde{g}_t\|_2$ be the approximate optimal value of the clipping threshold. But according to the training process in Figure 2, the MIoT device cannot obtain the global gradient of the current training round before uploading the local gradient. So based on the idea of RMSProp, this paper uses the prior knowledge $\mathbb{E}[\tilde{g}^2]_{t-1}$ to predict the global gradient \tilde{g}_t of the current round and then sets \tilde{g}_t as the clipping threshold; that is, $C_t = \beta\sqrt{\mathbb{E}[\tilde{g}^2]_{t-1}}$, where β denotes the local clipping factor, and the prior knowledge $\mathbb{E}[\tilde{g}^2]_{t-1}$ is computed as follows:

$$\begin{aligned} \mathbb{E}[\tilde{g}^2]_0 &= \vec{0}, \\ \mathbb{E}[\tilde{g}^2]_{t-1} &\leftarrow (1 - \gamma)\mathbb{E}[\tilde{g}^2]_{t-2} + \gamma(\tilde{g}_{t-1})^2. \end{aligned} \quad (5)$$

Note that the prior knowledge $\mathbb{E}[\tilde{g}^2]_0 = 0$ in the first training round will result in $C_1 = \beta\sqrt{\mathbb{E}[\tilde{g}^2]_0} = 0$, which cannot be used for gradient clipping. Therefore, we set another prior threshold G : when the prior knowledge of the gradient is insufficient in the initial training stage (i.e., $\mathbb{E}[\tilde{g}^2]_{t-1} < G$), set the gradient clipping threshold as a fixed value C ; when the training continues until the prior knowledge satisfies $\mathbb{E}[\tilde{g}^2]_{t-1} > G$, set the gradient clipping threshold as $C_t = \beta\sqrt{\mathbb{E}[\tilde{g}^2]_{t-1}}$. G usually takes an empirical value according to the training process of the model, which may vary in different datasets, but a simple way is to set G as the prior knowledge $\mathbb{E}[\tilde{g}^2]_{t-1}$ in a certain training round. So we have

$$C_t = \begin{cases} C, & \text{when } \mathbb{E}[\tilde{g}^2]_{t-1} < G, \\ \beta\sqrt{\mathbb{E}[\tilde{g}^2]_{t-1}}, & \text{when } \mathbb{E}[\tilde{g}^2]_{t-1} > G. \end{cases} \quad (6)$$

Then, in the t -th training round, the MIoT device i ($1 \leq i \leq K$) clips the local gradient $g_{i,t}$ and adds DP noise as follows:

$$\bar{g}_{i,t} = \frac{g_{i,t}}{\max\{1, \|g_{i,t}\|_2/C_t\}} + N(0, C_t^2 \sigma^2). \quad (7)$$

Since the value of $\sqrt{\mathbb{E}[\tilde{g}^2]_{t-1}}$ in equation (6) decreases as the model converges, the local clipping threshold C_t will also decrease, making the DP noise $\xi \sim N(0, (C_t \sigma)^2 I)$ in equation (7) less, which contributes to the convergence of the model in the later training stage.

3.3.2. Consensus Protocol Based on Gradient Verification. Since MIoT devices are widely distributed in the open network edge, the clinical data they collect may be of low quality and even be poisoned by adversaries, and then the local gradient trained on this kind of data will deviate from the global convergence trend. To remedy the adverse effects of these malicious gradients on the blockchain-based FL, we integrate gradient verification with consensus protocol to carry out a consensus process among the edge nodes. Each edge node identifies and removes malicious gradients uploaded by its associated MIoT devices so as to only aggregate qualified gradients to generate

the global model and achieve reliable FL. Unlike the proof-of-work (PoW) protocol which consumes a lot of computing resources, our protocol is improved on Algorand [27]. In each round of training, only some miners are selected to verify the new block by Byzantine agreement protocol, and the communication overhead among miners is further reduced, so the consensus efficiency is high and the forking probability is extremely low. The specific details are as follows:

(1) *Initialization.* A group of edge nodes with powerful computation and communication capabilities are chosen as miners. These miners not only generate or verify block but also execute gradient verification. To ensure the security of the blockchain, we assume that, at any point, no more than 1/3 of the miners are malicious. In addition, we assign an initial reputation value to each miner. If a miner is identified by other miners to return a falsified verification result or a fake block, then its reputation value will decrease by 1.

(2) *Gradient Verification.* After receiving the data from its associated MIoT devices, the miner first verifies the legality of the sender by checking the digital signature. If the signature is valid, then the miner puts the local gradient into the transaction pool. Subsequently, some miners are selected to form a verification committee, which is responsible for identifying and filtering malicious gradients. In this paper, we present a reputation-based consistent hashing protocol to designate the verifier role to some miners. Specifically, given a hash ring whose space is assigned to miners in proportion to their reputation value, we repeatedly rehash the initial SHA-256 hash of the last block and map the result to the hash ring. The miner corresponding to the space where the hash lies is chosen to be the member of the verification committee. This step is repeated until the size of the committee M is reached, which is shown in Figure 3. The principle of the above process is similar to that of Algorand [27]: the probability of a party being selected is proportional to its reputation. Since the adversary cannot obtain the state of the block until it is generated, they cannot predict the output of the consistent hashing and launch targeted attacks.

The verification committee executes the multi-KRUM algorithm [28] on gradients in the transaction pool and accepts the top majority of the gradients in each training round. The specific process is as follows:

Step 1. Assume that R is the total number of gradients in the transaction pool and f is the number of Byzantine gradients. The verifier adds up the Euclidean distance of each gradient to its closest $R-f-2$ gradients and uses the sum as the quality score of the gradient:

$$s(i, t) = \sum_{i \rightarrow j} \|\bar{g}_{i,t} - \bar{g}_{j,t}\|. \quad (8)$$

Step 2. The verifier selects the $R-f$ gradients with the lowest scores as qualified gradients and signs them using its public key. To prevent some malicious verifier from arbitrarily accepting the gradients from its colluding MIoT devices, we require that an MIoT device's gradient must be signed by most verifiers before it is finally accepted.

(3) *Candidate Block Verification.* A miner is randomly chosen from the verification committee as the leader of the current training round. The leader collects qualified gradients in the transaction pool and generates a new block shown in Figure 4, from which it can be seen that, except for the hash value used to link the previous block, the block also contains all the qualified gradients and corresponding signatures of the verification committee. Then the new block along with the signature of the leader is sent to the verification committee to verify the validity of the block, mainly by checking the signature of the leader and verifiers. Only when more than 2/3 of the verifiers agree on the block, the block is determined to be valid and broadcasted to arrive at a consensus in the blockchain through the popular gossip protocol [29]. Otherwise, an empty block is created.

(4) *Global Model Training.* All the MIoT devices download the latest block from the blockchain, compute the global gradient by averaging all the qualified gradients stored in the block, and then update their local models. The next training round will begin until the model converges or reaches the maximum number of rounds. Note that, in each round of training, the reputation value of the MIoT device whose local gradient is identified as qualified and the verifier who returns the correct verification result will both increase by 1; otherwise, their reputation value will decrease by 1. When the reputation value decreases to zero, the entity (e.g., MIoT device or edge node) is put into the blacklist and prohibited from participating in the consensus.

The security of the above consensus protocol can be guaranteed from the following aspects: (1) In each round of training, we use consistent hashing to select different miners to verify the new block. The output of consistent hashing cannot be predicted by attackers in advance, so attackers cannot launch targeted attacks on specific verifiers. In addition, as designed by the consensus protocol, the probability of a miner being selected as a verifier is proportional to its reputation value, so attackers cannot increase the probability of being selected through Sybil attack without increasing its own reputation value, which further strengthens the security. (2) We require that an MIoT device's gradient can be identified as qualified only when it owns the signatures of most verifiers so as to prevent some malicious verifiers from colluding with some MIoT devices. (3) The consensus protocol follows Algorand [27], requiring that the newly generated block can only be identified as valid and broadcasted after it is approved by more than 2/3 of the verifiers, so its security is equivalent to that of Algorand.

3.4. Security Analysis. Our scheme uses a differential privacy mechanism to protect data privacy, so how to track the accumulated privacy loss during training under a given privacy budget is very important. In this paper, we use the privacy accountant proposed by Abadi et al. [24] to compute the privacy loss, which is used by many related works [15, 17]. Related definitions are as follows.

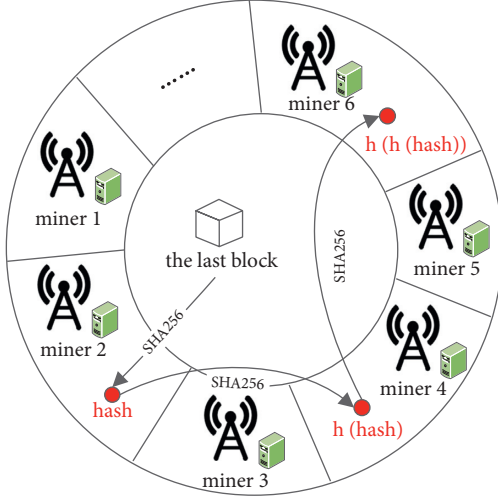


FIGURE 3: The selection of the verification committee.

previous block hash: 2fde7c3ac	time stamp
qualified gradient $\bar{g}_{1,t}$	qualified gradient $\bar{g}_{K',t}$
verifier sigs for $\bar{g}_{1,t}$	verifier sigs for $\bar{g}_{K',t}$
$(\bar{g}_{1,t})_{sign_1}$	$(\bar{g}_{K',t})_{sign_1}$
\vdots	\vdots
$(\bar{g}_{1,t})_{sign_j}$	$(\bar{g}_{K',t})_{sign_j}$

 FIGURE 4: Block contents at round t .

Definition 4 (privacy loss). Assume that $A: D \rightarrow R$ is a randomized algorithm, D and D' are adjacent datasets differing in an individual sample, and then the privacy loss of output $O \in R$ is

$$c(o, A, D, D') \triangleq \log \frac{\Pr[A(D) = o]}{\Pr[A(D') = o]}. \quad (9)$$

Definition 5 (moment accountant). The moment accountant of algorithm A at the λ th moment is defined as

$$\alpha(\lambda) \triangleq \max_{D, D'} \log E_{O \sim A(D)} [\exp(\lambda c(O, A, D, D'))]. \quad (10)$$

Theorem 1 (composability). Assume that algorithm A is composed of a sequence of subalgorithms A_1, A_2, \dots, A_k . For any moment λ , the moment accountant of A is bounded by the sum of moment accountant of A_1, A_2, \dots, A_k :

$$\alpha_A(\lambda) \leq \sum_{i=1}^k \alpha_{A_i}(\lambda). \quad (11)$$

Theorem 2 (tail bound). For any $\varepsilon > 0$, the algorithm A is (ε, δ) -differentially private for

$$\delta = \min_{\lambda} \exp(\alpha_A(\lambda) - \lambda\varepsilon). \quad (12)$$

According to Theorem 1, the privacy loss of our method is proportional to the number of MIoT devices and training rounds. Assume that the number of MIoT devices is K and training rounds is T . Let the overall moment accountant be $\alpha(\lambda)$ and the moment accountant of device i ($1 \leq i \leq K$) in the t -th round be $\alpha_{i,t}(\lambda)$. Based on Theorem 1, we have

$$\alpha(\lambda) \leq \sum_{t=1}^T \sum_{i=1}^K \alpha_{i,t}(\lambda), \quad (13)$$

where $\alpha_{i,t}(\lambda)$ mainly keeps track of the DP noise $\xi \sim N(0, (C_t\sigma)^2 I)$ added on the clipped gradient of devices, shown as equation (7). The computation of $\alpha_{i,t}(\lambda)$ is as follows.

Let μ_0 and μ_1 be the probability density function of Gaussian distribution $N(0, (C_t\sigma)^2)$ and $N(1, (C_t\sigma)^2)$, respectively. μ denotes the mixed Gaussian distribution $\mu = (1-q)\mu_0 + q\mu_1$ of μ_0 and μ_1 , where q is the sampling probability of local training. Then we need to compute $\alpha_{i,t}(\lambda) = \log \max(E_1, E_2)$, where

$$E_1 = \mathbb{E}_{x \sim \mu_0} \left[\left(\frac{\mu_0(x)}{\mu(x)} \right)^\lambda \right], \quad (14)$$

$$E_2 = \mathbb{E}_{x \sim \mu} \left[\left(\frac{\mu(x)}{\mu_0(x)} \right)^\lambda \right]. \quad (15)$$

Since the noise distribution $\xi \sim N(0, (C_t\sigma)^2 I)$ added on the local gradient is the same for all MIoT devices, the computation of $\alpha_{i,t}(\lambda)$, $1 \leq i \leq K$, $1 \leq t \leq T$ is the same for all devices. By equation 13, it suffices to compute or bound the overall moments $\alpha(\lambda)$ of our method. Then we can use the tail bound in Theorem 2 to convert the moment bounds to $(\varepsilon, \delta = \min_{\lambda} \exp(\alpha(\lambda) - \lambda\varepsilon))$ -differential privacy guarantee. Note that, in the execution of DP-based deep learning methods [24], the value range of integer λ is usually $0 \leq \lambda \leq 100$.

4. Experiments

We want to demonstrate the following points when designing the evaluation of our method: (1) Our method can make a tradeoff between the model accuracy and privacy preservation. (2) Given a reasonable privacy budget, the running time of our method is less than similar blockchain-based FL methods. (3) Our method is robust to poisoning attack.

- (1) Models and datasets: the experiments are conducted under Ubuntu 18.04 system with Intel i7-8700K CPU, GTX 1080T GPU, and 16 GB RAM. We implement a small blockchain prototype based on Ethereum in Go language and train the deep learning model in Python. *go-python v1.0* [30] library is used to interface between Python and Go. We use a convolutional neural network (CNN) composed of two 5×5 convolution layers, a full-connected layer and a softmax output layer (1,663,370 parameters), as the deep learning model, in which the model weights are initialized by normal distribution $N(0, 0.022)$ and the biases are initialized as 0. As for the experimental

dataset, a diabetes dataset from the American National Institute of Diabetes and Digestive and Kidney Diseases available online [31] is employed, which is composed of eight medical predictor variables and a target variable shown in Table 2, aiming to predict whether the patient has diabetes. We split the dataset into training and testing datasets in the ratio of 70:30. In order to simulate 20 distributed MIIoT devices in smart healthcare, we randomly shuffle and divide the dataset into 20 parts evenly, and each part is regarded as the local clinical data of an MIIoT device.

- (2) Hyperparameters and baselines: each MIIoT device trains the model locally with the batch size of 64 and local iterations of 20, and the gradient parameters are transformed into byte streams for transmission by pickle module. The hyperparameters in the adaptive DP algorithm are set as follows: $G = 10^{-6}$, $\beta = 1.2$, $\sigma = 4$, $\delta = 10^{-4}$, $\gamma = 0.1$, $C = 3$. Unless stated otherwise, we set the privacy budget $\epsilon = 3$ as default. In order to provide a comparison for our method, we choose two methods as the baseline: (1) BlockFL [32]: a blockchain-based FL method running on a device; (2) original FL [1]: the original federated learning method without any additional privacy-preserving strategies.

4.1. Model Accuracy. Given two different privacy budgets, we compare the model accuracy of our method with BlockFL and original FL, as shown in Figure 5.

We can find the following:

- (1) Our method exhausts the privacy budget $\epsilon = 2$ and $\epsilon = 3$ in the 36th and 53rd rounds and achieves model accuracy of 78.5% and 82.7%, respectively. It can be seen that the larger the privacy budget, the higher the model accuracy but the lower the level of privacy preservation simultaneously. In order to balance the model accuracy and data privacy, we set the privacy budget $\epsilon = 3$ in the rest of the experiments unless stated otherwise.
- (2) Original FL and BlockFL achieve higher model accuracy than our method; this is because our method adds DP noise on the gradient while the other two methods preserve the raw gradient. But given an appropriate privacy budget, our method protects the data privacy with only a slight accuracy loss. For example, when the privacy budget $\epsilon = 3$, our method achieves 82.7% accuracy, which is only slightly lower than 84.5% of original FL and 84% of BlockFL.

4.2. Running Time. In order to evaluate the introduction of blockchain on the training efficiency of federated learning, we compare the running time of the three methods, as shown in Figure 6, from which we can see that the running time of BlockFL and our method is greater than that of original FL. For example, when the training reaches 50 rounds, the running time of original FL, BlockFL, and our

TABLE 2: Experimental dataset.

November	Field name	Data type
1	Pregnanci	Integer
2	Glucose	Integer
3	BloodPressure	Integer
4	SkinThickness	Integer
5	Insulin	Integer
6	BMI	Integer
7	DiabetesPedigreeFunction	Integer
8	Age	Integer
9	Outcome	Integer

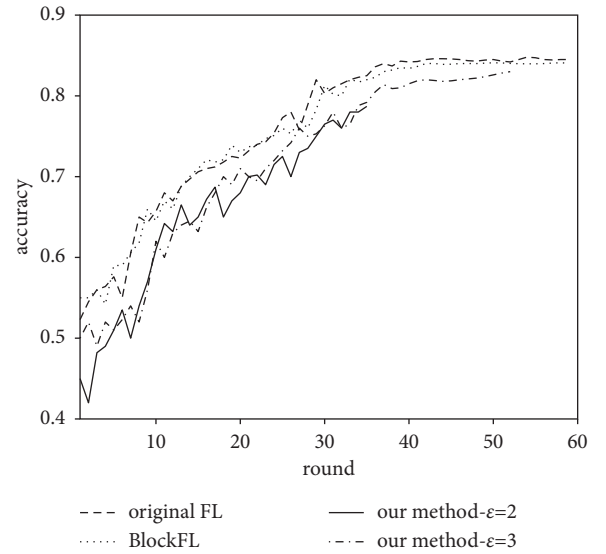


FIGURE 5: Accuracy of the three methods.

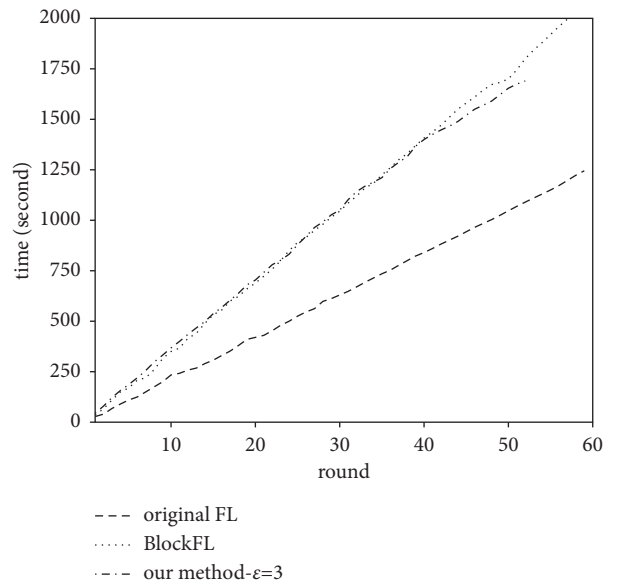


FIGURE 6: Running time of the three methods.

method is 1047 s, 1702 s, and 1624 s, respectively. This is because the consensus protocol in the blockchain involves time-consuming operations such as block generation,

verification, and broadcast. Therefore, blockchain-based FL methods (i.e., BlockFL and our method) achieve a series of security attributes such as auditability, reliability, and resistance to a single point of failure at the cost of some computation overhead, so they are more suitable for fields with high-security requirements, such as the medical field.

Since blockchain-based FL methods usually include local training phase and consensus phase, Figure 7 compares the average running time per round of each phase for BlockFL and our method under different number of local devices. It can be seen that, due to the designed adaptive DP algorithm, the local training phase of our method consumes slightly more time than that of BlockFL, which does not have any additional privacy-preserving mechanism, but their local training time does not increase with the number of devices. On the contrary, their consensus time is proportional to the number of devices, and the consensus time of our method is less than that of BlockFL. This is because the PoS consensus protocol used by BlockFL needs to continuously compute nonce until reaching the target condition, which is time-consuming, while our method uses a more efficient consistent hashing protocol and gradient verification method.

4.3. Privacy Budget Consumption. In the designed adaptive differential privacy algorithm, we adjust the clipping threshold C_t according to the training process. Figure 8 shows the change of C_t during the training. We can find that, in the first 10 rounds, C_t keeps unchanged, which is because we fix it as 3 in the initial training stage according to equation (6). As the training goes on, the value of C_t gradually decreases.

In order to further measure the effect of the designed adaptive differential privacy algorithm in reducing privacy budget consumption, we compare our method with the conventional DP-based method, which fixes the clipping threshold as $C=3$. We record the privacy budget consumed by the two methods when reaching the specified model accuracy, as shown in Table 3, where ϵ_D and ϵ_A denotes the privacy budget consumed by the conventional method and our method, respectively.

From Table 3, we can see that our method consumes much less privacy budget than the conventional method to reach the same model accuracy. For example, when the model accuracy is 80%, 82%, and 84%, our method reduces the privacy budget by 57%, 96%, and 81%, respectively, compared with the conventional method.

Figure 9 further shows how the privacy budget of the two methods consumes during the training process. It can be seen that the curves of the two methods almost overlap at the beginning, but the increase of the privacy budget of our method gradually decreases in the later stage of training, while the privacy budget of the conventional method still increases linearly. This proves that our method uses the same fixed clipping threshold as the conventional method due to insufficient prior knowledge at the beginning but adopts the adaptive DP algorithm in the latter to reduce the consumption of the privacy budget.

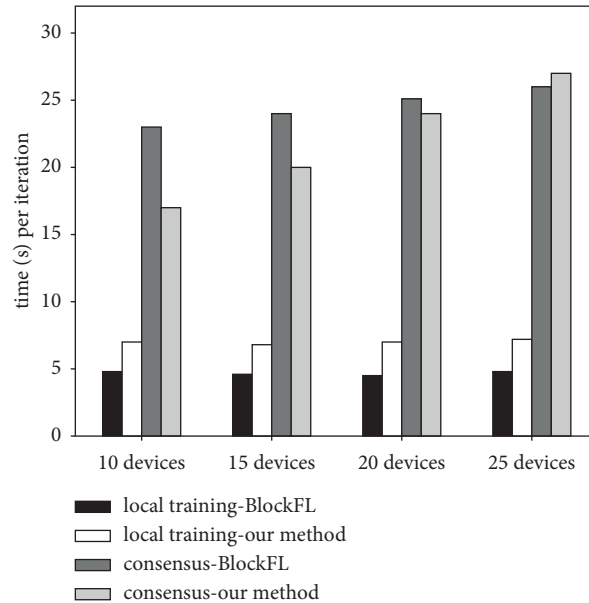


FIGURE 7: Time comparison of each training stage of BlockFL and our method.

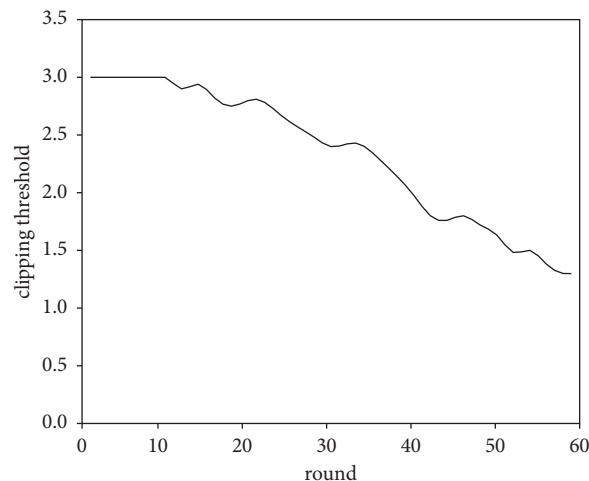


FIGURE 8: The change of clipping threshold during the training.

4.4. Resistance to Poisoning Attack. Since MIoT devices are usually located at the edge of an open network, they may face poisoning attacks from adversaries. In order to evaluate the ability of our method to resist poisoning attacks, we use a label flipping attack to generate poisoned samples by changing the labels of training samples and keeping the sample features unchanged. Then, we assign the poisoned samples to the designated MIoT devices and define the attack success rate as the proportion of incorrectly predicted samples on the test dataset. We set the proportion of poisoned MIoT devices as 30% and took an average of 20 experiments as the final results.

Figure 10 shows the training loss of the three methods, from which we can see that, due to the limit of privacy budget $\epsilon = 3$, our method converges in 53 rounds, but the other two methods cannot converge even within 70 rounds.

TABLE 3: Privacy budget consumed by our method and conventional method.

Accuracy	δ	ϵ_D	ϵ_A
0.75	10^{-4}	1.78	1.28
0.78		2.40	1.95
0.80		3.89	2.48
0.82		5.71	2.92
0.84		7.64	4.21

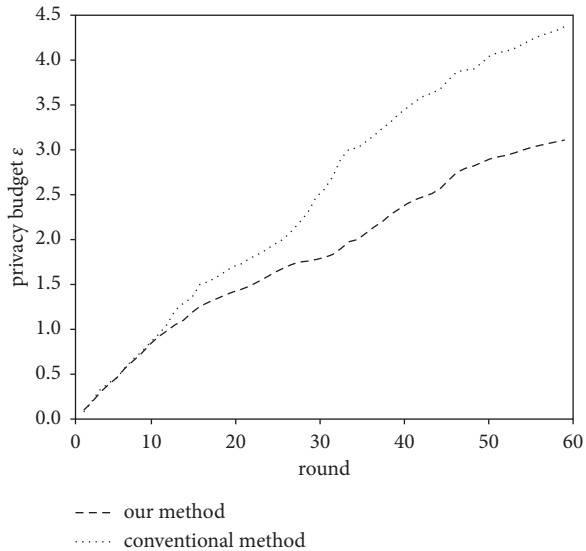


FIGURE 9: Privacy budget consumption of our method and conventional method.

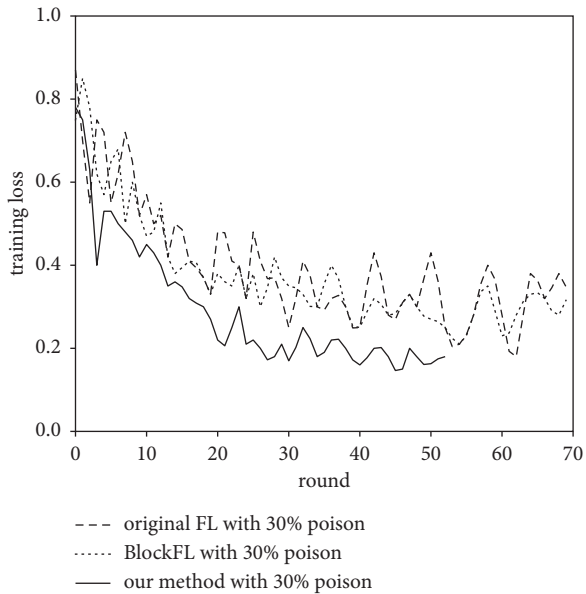


FIGURE 10: The training process of three methods under 30% poisoning attack.

Figure 11 further shows the attack success rate of the label flipping attack on the three methods. Since the original FL and BlockFL lack a mechanism to detect the poisoned

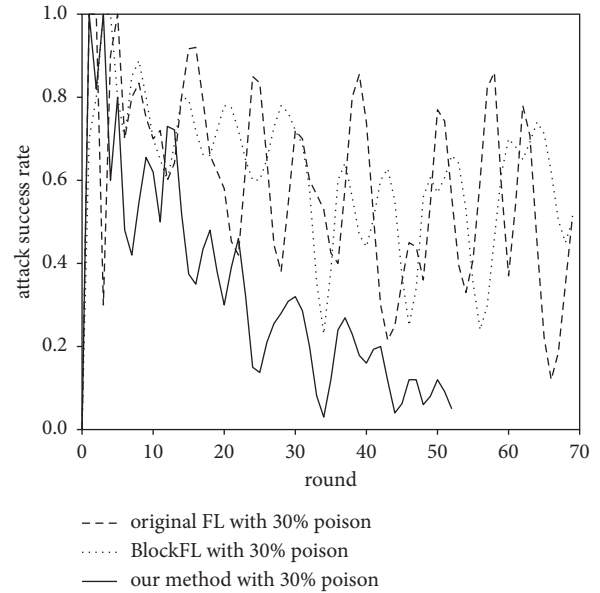


FIGURE 11: Attack success rate against three methods.

gradient, their attack success rate is almost always greater than 50%. However, our method limits the attack success rate to less than 20% in the later stage of training, and experimental data shows that four MIoT devices have been put into the blacklist at the end of the training, indicating that the consensus protocol based on gradient verification we design can effectively resist a certain proportion of poisoning attack.

5. Conclusions

In order to make full use of clinical data to improve the accuracy of disease diagnosis and medical service, smart healthcare based on MIoT has been widely exploited in recent years. However, it still faces challenges such as patient privacy leakage and various attacks from adversaries. To this end, we propose a blockchain-based federated learning method for smart healthcare. In particular, we design an adaptive differential privacy algorithm to carefully adjust the amount of noise added on the gradient to strike a balance between the privacy budget and accuracy degradation. The FL process is managed by a verification-based consensus protocol to prevent poisoning attacks and single point of failure. The experimental results on a real-world diabetes dataset show that our method can achieve similar accuracy to the original FL in acceptable running time. We also illustrate its ability to reduce the privacy budget consumption and withstand poisoning attacks. In the future, we will continue to explore and advance our method with public partners to make more improvements in smart healthcare.

Data Availability

The diabetes dataset is publicly available at <https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Other data in this paper come from the data statistics of the test process. All the data are real and can be used.

Conflicts of Interest

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

Acknowledgments

This work was supported by the Science and Technology Project of Henan Province (no. 162102310304).

References

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017*, pp. 1273–1282, Fort Lauderdale, FL, USA, April 2017.
- [2] Y. Zhao, J. Zhao, L. Jiang et al., "Privacy-preserving blockchain-based federated learning for IoT devices," *IEEE Internet of Things Journal*, vol. 8, no. 3, 2020.
- [3] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4177–4186, 2019.
- [4] D. Polap, G. Srivastava, and K. Yu, "Agent architecture of an intelligent medical system based on federated learning and blockchain technology," *Journal of Information Security and Applications*, vol. 58, Article ID 102748, 2021.
- [5] W. Dai, T. S. Brisimi, W. G. Adams, T. Mela, V. Saligrama, and I. C. Paschalidis, "Prediction of hospitalization due to heart diseases by supervised learning methods," *International Journal of Medical Informatics*, vol. 84, no. 3, pp. 189–197, 2015.
- [6] Y. J. Son, H. G. Kim, E. H. Kim, S. Choi, and S.-K. Lee, "Application of support vector machine for prediction of medication adherence in heart failure patients," *Healthcare informatics research*, vol. 16, no. 4, pp. 253–259, 2010.
- [7] A. Tariq, L. A. Celi, J. M. Newsome et al., "Patient-specific COVID-19 resource utilization prediction using fusion AI model," *NPJ Digital Medicine*, vol. 4, no. 1, pp. 1–9, 2021.
- [8] A. Sedik, A. M. Ilyyasu, A. El-Rahiem et al., "Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections," *Viruses*, vol. 12, no. 7, p. 769, 2020.
- [9] A. Qayyum, K. Ahmad, M. A. Ahsan, A. Al-Fuqaha, and J. Qadir, "Collaborative federated learning for healthcare: multi-modal COVID-19 diagnosis at the edge," 2021, <https://arxiv.org/abs/2101.07511>.
- [10] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," *International Journal of Medical Informatics*, vol. 112, pp. 59–67, 2018.
- [11] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021.
- [12] L. Zhang, B. Shen, A. Barnawi, S. Xi, N. Kumar, and Y. Wu, "FedDPGAN: federated differentially private generative adversarial networks framework for the detection of COVID-19 pneumonia," *Information Systems Frontiers*, pp. 1–13, 2021.
- [13] O. El Rifai, M. Biotteau, X. de Boissezon, I. Megdiche, F. Ravat, and O. Teste, "Blockchain-based federated learning in medicine," in *Proceedings of the International Conference on Artificial Intelligence in Medicine*, pp. 214–224, Springer, Minneapolis, MN, USA, August 2020.
- [14] J. Passerat-Palmbach, T. Farnan, M. McCoy et al., "Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data," in *Proceedings of the 2020 IEEE International Conference on Blockchain (Blockchain)*, pp. 550–555, IEEE, Rhodes Island, Greece, November 2020.
- [15] Y. Liu, J. Peng, J. Kang, A. M. Ilyyasu, D. Niyato, and A. A. A. El-Latif, "A secure federated learning framework for 5G networks," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 24–31, 2020.
- [16] R. Kumar, A. A. Khan, J. Kumar et al., "Blockchain-federated-learning and deep learning models for COVID-19 detection using ct imaging," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 16301–16314, 2021.
- [17] M. A. Rahman, M. S. Hossain, M. S. Islam, N. A. Alrajeh, and G. Muhammad, "Secure and provenance enhanced Internet of health things framework: a blockchain managed federated learning approach," *IEEE Access*, vol. 8, pp. 205071–205087, 2020.
- [18] M. Shayan, C. Fung, C. J. M. Yoon, and I. Beschastnikh, "Biscotti: a ledger for private and secure peer-to-peer machine learning," 2018, <https://arxiv.org/abs/1811.09904>.
- [19] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1322–1333, Denver, CO, USA, October 2015.
- [20] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction apis," in *Proceedings of the 25th {USENIX} Security Symposium ({USENIX} Security 16)*, pp. 601–618, Austin, TX, USA, August 2016.
- [21] J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "Deepchain: auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Transactions on Dependable and Secure Computing*, 2019.
- [22] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [23] Y. Qi, M. S. Hossain, J. Nie, and X. Li, "Privacy-preserving blockchain-based federated learning for traffic flow prediction," *Future Generation Computer Systems*, vol. 117, pp. 328–337, 2021.
- [24] M. Abadi, A. Chu, I. Goodfellow et al., "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, Vienna, Austria, October 2016.
- [25] C. Fang, Y. Guo, Y. Hu, B. Ma, L. Feng, and A. Yin, "Privacy-preserving and communication-efficient federated learning in Internet of Things," *Computers & Security*, vol. 103, Article ID 102199, 2021.
- [26] C. Xu, J. Ren, D. Zhang, Y. Zhang, Z. Qin, and K. Ren, "GANobfuscator: mitigating information leakage under GAN via differential privacy," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 9, pp. 2358–2371, 2019.
- [27] Y. Gilad, R. Hemo, S. Micali, G. Vlachos, and N. Zeldovich, "Algorand: scaling byzantine agreements for cryptocurrencies," in *Proceedings of the 26th Symposium on Operating Systems Principles*, pp. 51–68, Shanghai, China, October 2017.
- [28] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: byzantine tolerant gradient descent," in *Proceedings of the 31st International*

- Conference on Neural Information Processing Systems*, pp. 118–128, Long Beach, CA, USA, December 2017.
- [29] E. Buchman, “Tendermint: byzantine fault tolerance in the age of blockchains,” Ph.D. thesis, University of Guelph, Guelph, Canada, , 2016.
- [30] go-python, 2019, <https://github.com/sbinet/go-python>.
- [31] <https://www.kaggle.com/uciml/pima-indians-diabetes-database>.
- [32] H. Kim, J. Park, M. Bennis, and S.-L. Kim, “Blockchained on-device federated learning,” *IEEE Communications Letters*, vol. 24, no. 6, pp. 1279–1283, 2019.

Research Article

Research on PM_{2.5} Spatiotemporal Forecasting Model Based on LSTM Neural Network

Fang Zhao,¹ Ziyi Liang,² Qiyan Zhang,² Dewen Seng ² and Xiyuan Chen²

¹School of Computer Science and Technology, Zhejiang Shuren University, Hangzhou 310015, China

²School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

Correspondence should be addressed to Dewen Seng; sengdw@hdu.edu.cn

Received 17 May 2021; Accepted 27 August 2021; Published 19 October 2021

Academic Editor: Yassine Maleh

Copyright © 2021 Fang Zhao 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.

Accurate monitoring of air quality can no longer meet people's needs. People hope to predict air quality in advance and make timely warnings and defenses to minimize the threat to life. This paper proposed a new air quality spatiotemporal prediction model to predict future air quality and is based on a large number of environmental data and a long short-term memory (LSTM) neural network. In order to capture the spatial and temporal characteristics of the pollutant concentration data, the data of the five sites with the highest correlation of time-series concentration of PM_{2.5} (particles with aerodynamic diameter ≤ 2.5 μm) at the experimental site were first extracted, and the weather data and other pollutant data at the same time were merged in the next step, extracting advanced spatiotemporal features through long- and short-term memory neural networks. The model presented in this paper was compared with other baseline models on the hourly PM_{2.5} concentration data set collected at 35 air quality monitoring sites in Beijing from January 1, 2016, to December 31, 2017. The experimental results show that the performance of the proposed model is better than other baseline models.

1. Introduction

In recent years, with the rapid development of society, the pressure on the environment has become more and more serious, and some serious air pollution problems have seriously threatened people's health. In the case of cardiovascular disease, exposure to PM_{2.5} (fine particulate matter with a particle size of less than 2.5 μm) as short as several hours to several weeks can lead to an increase in mortality, which can lead to a reduction in lifespan for up to several years. Lowering the concentration of PM_{2.5} can effectively reduce the above risks [1]. In addition, there are more and more studies on the relationship between gas pollution and neurodegenerative diseases such as Alzheimer's disease. A 2015 study showed that prolonged exposure to PM_{2.5} can lead to an advance in the first outpatient time for neurodegenerative diseases, with more than 3 million premature deaths per year due to PM_{2.5} exposure [2–5]. Therefore, accurate prediction of its mass concentration plays a key role in atmospheric management decisions [6]. Predicting air

pollutant concentrations in advance is the basis for enhancing air pollution prevention and achieving comprehensive environmental management, which is important for public health and government decision-making [7].

However, predicting air quality concentrations is difficult, and it is not only susceptible to other factors, such as meteorological factors (temperature, relative humidity, wind speed, and precipitation), traffic pollution, industrial emissions, and so on. It is also affected by the concentration of other pollutants in the air. Specifically, the temperature has an effect on atmospheric and ventilating conditions. Humidity and precipitation have an effect on the deposition of particulate matter, while wind speed contributes to the diffusion of particulate matter [8]. Traffic pollution and industrial emissions will produce some harmful gases such as SO₂, NO₂, O₃, and CO. Some related analyses show that O₃ will inhibit the growth of PM_{2.5}, PM₁₀, NO₂, CO, and SO₂, and PM_{2.5}, PM₁₀, and CO are strongly correlated in each season (correlation coefficient is generally >0.5) [9]. So these effects also pose challenges for air quality predictions.

At the same time, air quality prediction is not only spatially dependent. In the time dimension, the PM_{2.5} concentration of the experimental site is also affected by the air mass concentration of the site in the past, so we need to capture both spatial dependence and time dependence.

In order to solve the above challenges, the author extracted the concentration data of the five sites with the closest correlation with the experimental site and the highest correlation of PM_{2.5} concentration sequence data to extract its spatial dependence; then the meteorological factors and other pollutant data at the same time were merged as the next input. Next, the previously integrated time-series data with hysteresis p were input into the LSTM neural network with n hidden layers and one fully connected layer for training. The LSTM neural network is used to extract its time dependence, which will be combined to extract spatiotemporal features and output future predictions of PM_{2.5}. LSTM is a time-cycle neural network that effectively solves long-term dependency problems and avoids gradient disappearance and explosion. It has been proposed to predict future output with past inputs. Compared with the traditional recurrent neural network (RNN) [10], it is unique in that it is designed with a loop body structure that has proven to be very suitable for prediction based on time-series data. And the disappearing gradient problem has better performance than RNN [11–13].

The contribution of this paper mainly includes three aspects: first, this paper proposes a new air quality spatiotemporal prediction model to predict future air quality. Second, integrating historical time air quality data, air quality data from nearest neighbors, meteorological data, and other pollutant data can improve prediction accuracy and help models better predict changes in air quality. Third, the paper evaluates the proposed model on the hourly concentration data set from January 1, 2016, to December 31, 2017, in Beijing. The experiment demonstrates the effectiveness of the method.

The rest of the paper is organized as follows: Section 2 describes the related work; the data and methods used in the experiments are detailed in Section 3; Section 4 reports on the results and discussion; and summary and outlook are drawn in Section 5.

2. Related Work

In recent years, more and more researchers use deep learning neural network technology to overcome the problems in the fields of big data and artificial intelligence, which is mainly due to its ability to realize effective learning of feature representation from massive input data and to deeply analyze the potential deep-seated features between data. Because the problem of environmental pollution has become more and more serious in recent years, people pay more attention to health problems, and the relevant environmental health departments have also strengthened management and monitoring, which has led more and more researchers to focus on the research of air quality. Researchers have carefully studied and put forward many prediction models of air quality, which can be roughly

divided into three types: single prediction model. It uses the existing methods to predict the air quality data. For the improved prediction model, there are various deficiencies in the prediction of a single model, so the research institute can improve the prediction performance of the existing methods by improving the corresponding weight parameters, adding optimization algorithms, or adding various auxiliary data on the basis of the existing method research: joint prediction model. Although the improved model can make up for the shortcomings of a single model to a certain extent, there are still some limitations. Therefore, researchers continue to make in-depth exploration, combine two or more single models together, give full play to their respective advantages, learn from each other, and combine to further improve the prediction accuracy of the model.

2.1. Spatiotemporal Prediction. In recent years, as air pollution has been paid more and more attention, researchers have also proposed many spatiotemporal prediction models to achieve future predictions of air quality. Qi et al. [14] proposed a novel combined prediction scheme based on CNN and LSTM for urban PM_{2.5} concentration; the model uses CNN to extract the spatial characteristics of inputs between monitoring stations and uses LSTM to predict future air pollution concentrations by learning the characteristics contained in past air pollution concentration time-series data. Zhou et al. [15] proposed a hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory; the model applies a graph convolutional network (GCN) to extract the spatial dependence between different sites and LSTM to capture the time dependence between observations at different times. Wen et al. [16] proposed a deep multi-output LSTM (DM-LSTM) neural network model that was incorporated with three deep learning algorithms (i.e., mini-batch gradient descent, dropout neuron, and L2 regularization) to configure the model for extracting the key factors of complex spatiotemporal relations. Wang and Song [17] proposed a novel spatiotemporal convolutional long short-term neural network for air pollution prediction; high-level spatiotemporal features are extracted by a combination of convolutional neural network (CNN) and long short-term memory neural network (LSTM-NN), and meteorological data and aerosol data are integrated to improve model prediction performance. Box and Jenkins [18] proposed a deep spatiotemporal ensemble model for air quality prediction; the model combines the collection method of the partitioning strategy based on the weather pattern and finds the spatial correlation by analyzing the causal relationship between the sites and generating the spatial data as relative sites and relative regions; finally, the depth LSTM-based time predictor is used to learn the long- and short-term dependence of air quality. These models achieve spatiotemporal prediction by analyzing spatiotemporal data, but the model proposed in this paper is different from the above mentioned. This paper proposes a new air quality spatiotemporal prediction model to integrate experimental air quality data of the site, air quality data of nearest neighbors,

meteorological data, and other pollutant data and combine LSTM deep neural network to extract spatiotemporal feature and ultimately achieve future predictions.

2.2. Classical Model for Time-Series Prediction. Forecasting flow in a spatiotemporal network can be viewed as a time-series prediction problem. Existing time-series models such as the autoregressive integrated moving average model (ARIMA [19]), seasonal ARIMA [20], and the vector autoregressive model [21] can capture the temporal dependencies very well, yet it fails to handle spatial correlations.

2.3. Neural Networks for Sequence Prediction. Neural networks and deep learning [22] have gained numerous successes in the fields such as compute vision [23], speech recognition [24], and natural language understanding [25]. Recurrent neural networks (RNNs) have been used successfully for sequence learning tasks [26]. The incorporation of long short-term memory (LSTM) [27] or gated recurrent unit (GRU) [28] enables RNNs to learn long-term temporal dependency. Some researchers have come up with some bold ideas that combine recursive neural networks with recurrent neural networks to process time-series data, which may better capture the spatiotemporal characteristics of the data. However, as the depth of the network increases, the training cost will also increase greatly, and training will become more and more difficult. Is there any way to improve the accuracy of model prediction without increasing the difficulty of training? It is the direction of future researchers and the problems to be solved.

3. Data and Method

3.1. Research Areas and Data. The research area is Beijing, and the data come from the hourly data of 35 air monitoring stations and meteorological monitoring stations in Beijing. The air quality data of Beijing from January 1, 2016, to December 31, 2017, came from the website of Beijing Environmental Protection Testing Center (<https://www.bjmec.com.cn/>). The location maps of Beijing and 35 monitoring stations are shown in Figure 1. This paper renumbered the samples and predicted the PM2.5 concentration in representative sites. The S1 station is the urban environmental assessment point; the S17 and S23 stations are the suburban environmental assessment points; the S29 station is the control point and the regional point; and S31 is the traffic pollution monitoring point. Air quality data is collected every hour; there are around 17,000 records for each site. Auxiliary data includes simultaneous meteorological data (temperature, dew point, pressure, wind direction, and wind speed) and other pollutant data (SO₂, NO₂, O₃, and CO), which have also been shown to be highly correlated with PM2.5 concentration [29–33]. The meteorological data come from the National Climate Data Center (NCDC), and other pollutant data are also from the Beijing Environmental Protection Testing Center website. The data sets used in the research are available directly from the

website <https://beijingair.sinaapp.com/>. The data set is first filled with outliers and missing values, normalized, and scaled to [0,1]. The data record of each site is different. This paper selects 67% of the total number of data records as the test set, and the remaining 33% records as the test set.

3.2. Extraction for Spatial Factors. According to Tobler's First Law of Geography, everything is related to other things, and similar things are more closely related, that is, the influence of the neighboring sites on the experimental site is greater than that of the distant sites. To illustrate the spatial characteristics of the PM2.5 concentration sequence, the authors calculated the distance between two sites and the Pearson correlation coefficient for the PM2.5 concentration sequence at each of the two sites.

The Haversine formula is used as recommended by Wikipedia to calculate the distance between two sites based on the latitude and longitude of each site. This formula uses a sine function to maintain enough valid numbers even if the distance is small. The formula is as follows:

$$\text{haver sin}\left(\frac{d}{R}\right) = \text{haver sin}(\varphi_2 - \varphi_1) + \cos(\varphi_2)\text{haver sin}(\Delta\lambda), \quad (1)$$

where

$$\text{haver sin}(\theta) = \sin^2\left(\frac{\theta}{2}\right) + \cos(\varphi_2) = \frac{(1 - \cos(\theta))}{2}, \quad (2)$$

where $\text{haver sin}(\cdot)$ indicates the distance between stations, R is the radius of the Earth and can take an average of 6371 km, φ_1 and φ_2 indicate the latitude between two points, and $\Delta\lambda$ represents the difference between two latitudes.

The Pearson correlation coefficient is used to measure the linear correlation strength between continuous variables. The formula is as follows:

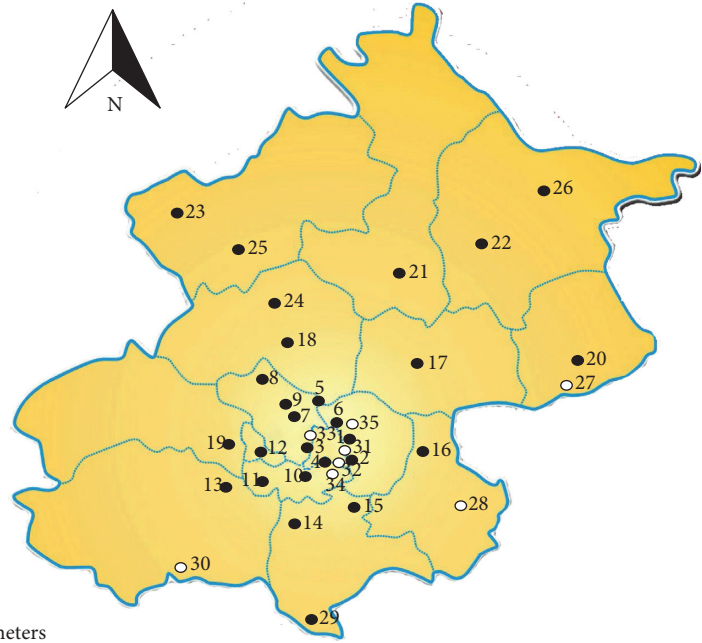
$$r(s_i, s_j) = \frac{\text{Cov}(s_i, s_j)}{\sigma(s_i)\sigma(s_j)}, \quad (3)$$

where r represents the correlation coefficient of the sequence of PM2.5 concentration between sites, Cov is the covariance, σ is the standard deviation.

The correlation coefficients of the 10 stations with the highest PM2.5 concentration sequence correlation between each of the 35 sites are shown in Figure 2. It can be observed from the figure that the value of the correlation coefficient of most stations is greater than 0.7, so adjacent stations can be used to improve the prediction accuracy of the station. The process of spatial factor extraction is shown in Figure 3. The initial data set used in this paper is the time-series data collected by 35 stations per hour. The data set includes five features, PM2.5 (time average concentration of particles with aerodynamic diameter ≤ 2.5 mm), PM2.5_24h (daily average concentration of particles with aerodynamic diameter ≤ 2.5 mm), PM10 (time average concentration of particles with aerodynamic diameter ≤ 10 mm), PM10_24h (daily average concentration of particles with aerodynamic

Schematic diagram of the distribution of air quality automatic detection system in Beijing

- Air quality monitoring point: 35
- Urban environmental assessment point (12):
 Dongsi: 1 Tiantan: 2 Guanyuan: 3 Wanshouxigong: 4
 Aoti-Center: 5 Nongzhanguan: 6 Wanliu: 7 Northern
 New District: 8 Botanical garden: 9 Fengtai Garden: 10
 Yungang: 11 Ancient city: 12
- Suburban environmental assessment point (11):
 Fangshan: 13 Daxing: 14 Yizhuang: 15 Tongzhou: 16
 Shunyi: 17 Changping: 18 Mentougou: 19 Pinggu: 20
 Huairou: 21 Miyun: 22 Yanqing: 23
- Control point and regional point (7):
 Dingling: 24 Badaling: 25 Miyun Reservoir: 26
 Donggao Village: 27 Yongle store: 28 Yufa: 29
 Liuli River : 30
- Traffic pollution control point (5):
 Qianmen: 31 Yongdingmennei: 32 Xizhimenbei: 33
 South Third Ring: 34 East Fourth Ring: 35



- Original site (27)
- Add site (8)

FIGURE 1: Research area and monitoring site distribution map.

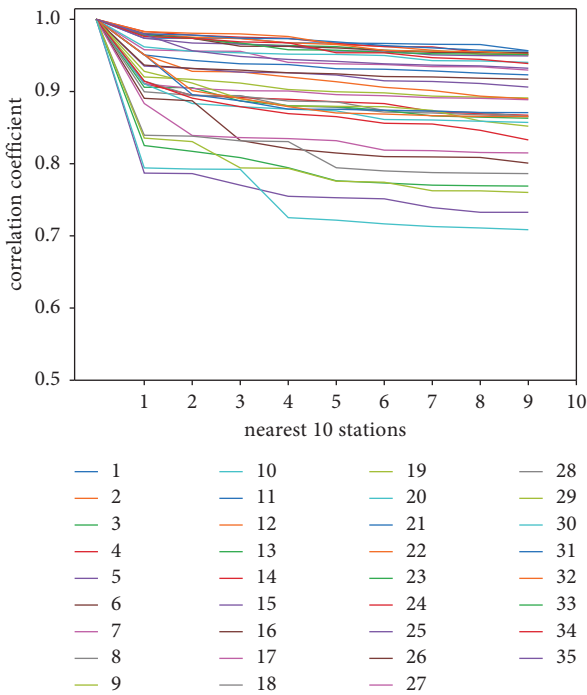


FIGURE 2: Correlation coefficient of 10 stations with the highest correlation of PM2.5 concentration sequences in 35 stations.

diameter ≤ 10 mm), and AQI (air quality index), then we will calculate the correlation coefficient between the stations by the formula mentioned above, and extract the concentration

of the top 5 stations with the highest correlation with the experimental site, and finally obtain the separate time-series data of 35 stations. Each data record includes six characteristics (self PM2.5 concentration, adjacent station 1_PM2.5 concentration, adjacent station 2_PM2.5 concentration, adjacent station 3_PM2.5 concentration, adjacent station 4_PM2.5 concentration, adjacent station 5_PM2.5 concentration), as the initial data for the next stage for use. As for the temporality of PM2.5 distribution, relevant research has pointed out that the current moment of the station has a good correlation with a certain moment in the past. In order to further reflect the spatiotemporal correlation between the sites, the site timing data obtained above is combined with the auxiliary data including meteorological data and other pollutant data, and the delayed timing values are input into the model. Then long- and short-term memory neural networks are applied to extract their spatiotemporal correlation [34].

3.3. ST_LSTM Model. The prediction framework of the model proposed in this paper is shown in Figure 4. The input of the model includes the fusion of three parts of data, including site autocorrelation concentration and adjacent site concentration data, meteorological data (temperature, dew point, pressure, wind direction, and wind speed), and other pollutant data (SO₂, NO₂, O₃, and CO). The output is the predicted value of the experimental site PM2.5 at $(t + 1, t + 2, \dots, t + N)$. The model is divided into three parts: extraction of site autocorrelation concentration and related

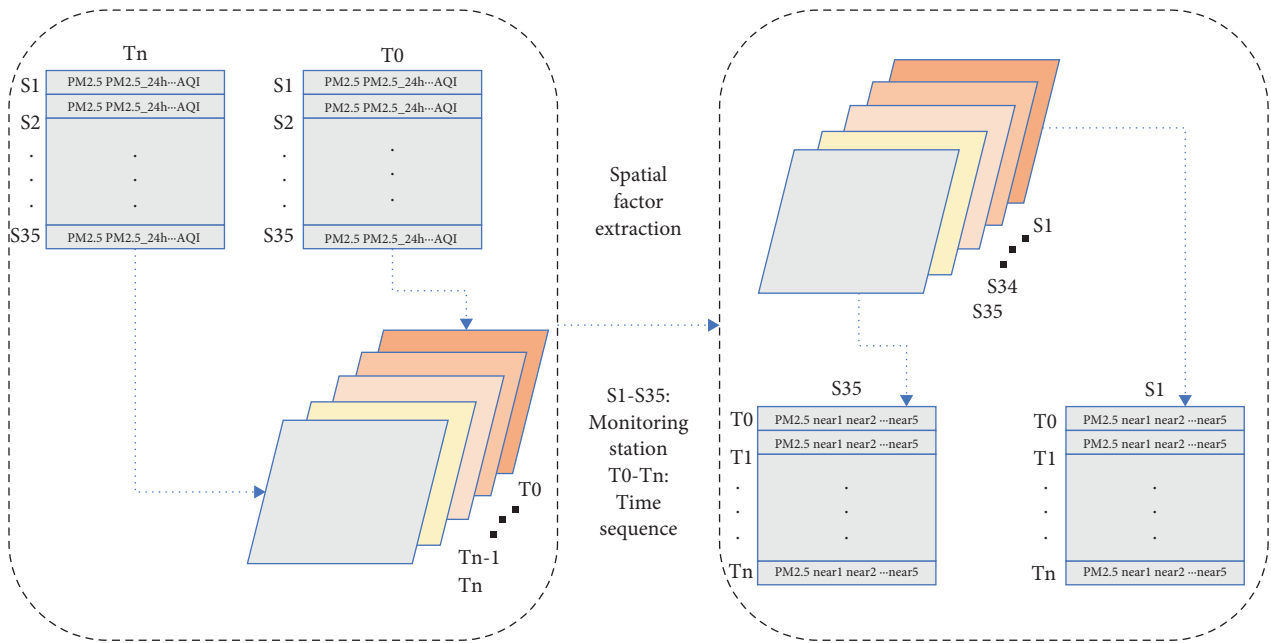


FIGURE 3: Spatial factor extraction.

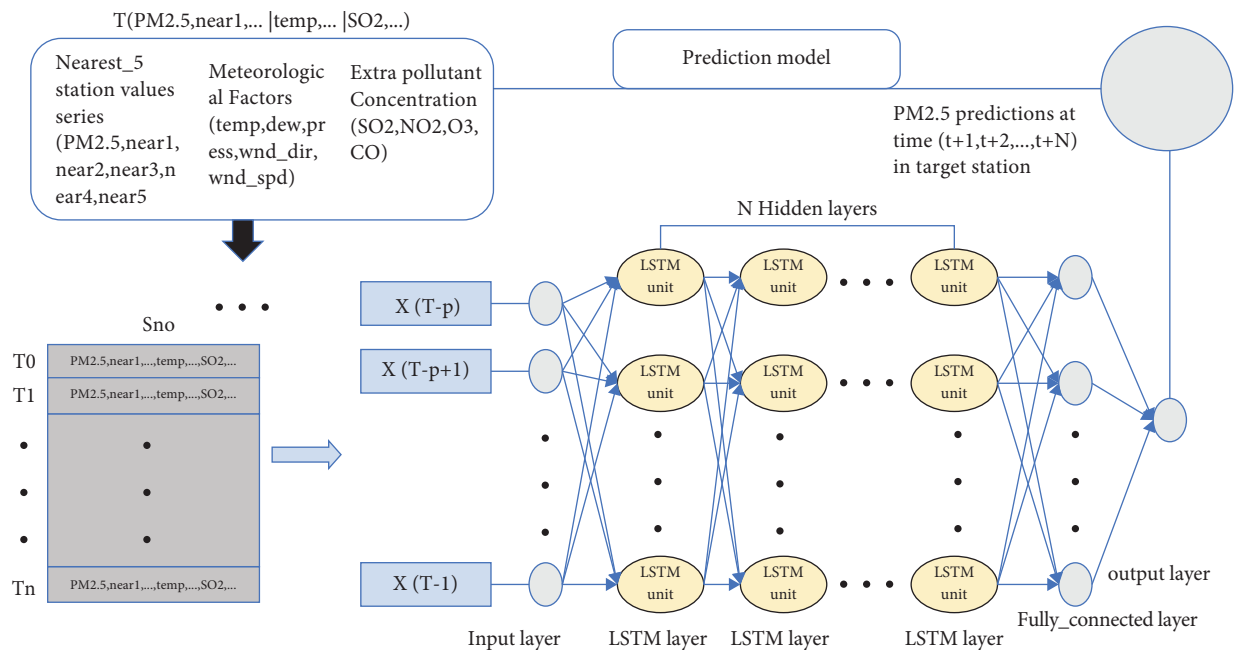


FIGURE 4: Overall model prediction architecture.

concentration data of adjacent sites, the fusion of auxiliary data and extraction of spatiotemporal features, and prediction of future PM2.5 concentration.

The first part is the extraction of spatial factors. That is to say, the site self-correlation concentration and the concentration data of the adjacent sites are extracted. The specific content is described in detail in Section 3.2.

The second part is the fusion of auxiliary data. Auxiliary data are added to extract more spatiotemporal features when the model is trained. All data are processed through cleaning and missing values before use, and records with outliers are

deleted. For the PM2.5 concentration values, meteorological data, and other pollutant data for each site, the authors used the method of mean filling, fixed value filling, and interpolation filling to process the missing values. The merged data is normalized as an input to the next stage.

The last part is the extraction of spatiotemporal features and the prediction of future PM2.5 concentration. The temporal and spatial features of the normalized time-series data are extracted using an LSTM model with multiple hidden layers. The predicted sequence value at the time of $(t + 1, t + 2, \dots, t + N)$ is predicted using data with a lag of past time t .

4. Experiments

4.1. Evaluation. In order to evaluate the performance of the model proposed in this paper, the author used three evaluation indicators, namely the mean absolute error (MAE), root-mean-square error (RMSE), and the decision coefficient (R squared, R^2). Because of the limitations of RMSE and MAE, that is, the same algorithm model, solving different problems cannot reflect the pros and cons of this model for different problems. Because the data is different in different practical applications, it is impossible to directly compare the predicted values, so it is impossible to judge which model is more suitable for predicting which problem. Therefore, the prediction results are converted into accuracy, and the results are all between [0,1]. For the prediction accuracy of different problems, it can be compared and judged which model is more suitable for predicting which problem. R^2 is the best indicator of linear regression. The calculation formula for the three indicators is as follows:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_{\text{test}}^{(i)} - \hat{y}_{\text{test}}^{(i)}|, \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_{\text{test}}^{(i)} - \hat{y}_{\text{test}}^{(i)})^2}, \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_{\text{test}}^{(i)} - \hat{y}_{\text{test}}^{(i)})^2}{\sum_{i=1}^m (y_{\text{test}}^{(i)} - \bar{y}_{\text{test}})^2} = 1 - \frac{\text{MSE}(\hat{y}_{\text{test}}, y_{\text{test}})}{\text{Var}(y_{\text{test}})}, \quad (6)$$

where $y_{\text{test}}^{(i)}$ indicates the predicted value and y_{test} represents the true value. The smaller the value of MAE and RMSE, the smaller the model error and the better the prediction performance. The larger the value of R_2 the better the model effect, the maximum value is 1; when R_2 is 1, the prediction model does not have any mistakes; when R_2 is 0, the model is equal to the reference model; when R_2 is less than 0, it means that the learned model is not as good as the benchmark model.

4.2. Settings. In the prediction architecture proposed in this study, several super parameters are preset, including the number of LSTM layers, the number of neurons in each LSTM layer, the number of fully connected layers, the number of neurons in each fully connected layer, and time step. While fixing other parameters, the impact of each parameter on the prediction performance of the model is checked to determine the best parameters.

Table 1 details the error size of the prediction results obtained using different hidden layers. The data show that when the number of hidden layers is 3, the error is the smallest. Therefore, this study uses an LSTM network with three hidden layers. The number of neurons in each layer is 100; the number of fully connected layers is 1; and the number of neurons in each layer is 1. In addition, other parameter settings in the research are as follows: Adam algorithm is used as the optimization algorithm; Mae function is used as the cost function; the batch size is 128; the

tanh function is used as the excitation function of this study; the maximum number of iterations is 100; the learning rate is 0.01; and the performance of the model is the best. In this study, MAE, RMSE, and R^2 are used as indicators to determine the impact of time step on prediction performance. Relevant studies have pointed out that a small time step cannot ensure sufficient long-term memory input of the model, but a large time step allows too many irrelevant inputs to be added [50]. Table 2 shows the impact of different time steps on prediction performance. It can be observed from the table that when the time step is 14, the performance of the model is the best.

4.3. Baselines. In order to test the overall performance of the current model, the author conducted a series of comparative experiments on two types of baseline models: nondeep learning model and deep learning model.

- (1) Nondeep learning baseline model. This includes linear regression (LR) models, support vector regression (SVR) models, random forest (RF), and autoregressive moving average (ARMA) models.
- (2) Deep learning baseline model. That is, it contains the different components in the model presented in this paper. It includes LSTM_N model (with only spatial factor data), LSTM_NW model (with spatial and meteorological data), LSTM_NE model (with spatial and other pollutant data), LSTM_WE model (with meteorological and other pollutant data), LSTM_S model (without any auxiliary data), and ST_LSTM model (model presented in this paper).

5. Results and Discussion

5.1. Prediction Performance. After determining the optimal network architecture for the current prediction task, the training set is used to train the current ST_LSTM model until convergence and then evaluated on the test set. This paper predicts the PM2.5 concentration value for the next hour for the monitoring station numbered 1 in Beijing and compares the predicted value of the model with the real value. The Beijing No. 1 site drawn using the ST-LSTM model proposed in this paper predicts the PM2.5 concentration value and the observed PM2.5 concentration value for the next hour as shown in Figure 5. It can be observed from the figure that the predicted value is substantially consistent with the observed value. The R^2 value between the observed and predicted data indicates that the model can capture 93% of the interpreted variance. The feasibility and accuracy of the proposed model are verified.

At the same time, the author also plotted the curve of the true value and the predicted value on the test set. The plot of the true value and the predicted value of the test site of Beijing No. 1 is shown in Figure 6. From the figure, the author observes that the trend of the two curves is roughly the same, and the degree of fitting is better. It is indicated that the model proposed in this paper can accurately capture the temporal and spatial variation of PM2.5 and achieve a

TABLE 1: Comparison of different hidden layers on model performance.

Hidden layers	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	R^2
1	7.57	12.261	0.944
2	7.54	13.250	0.944
3	7.34	11.963	0.947
4	7.99	12.260	0.944
5	7.69	12.407	0.943

TABLE 2: Comparison of model performance with different time lags.

Parameter	Parameter (h)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	R^2
Time lag	1	7.81	12.727	0.92
	2	8.26	13.020	0.92
	4	7.31	12.374	0.92
	8	7.26	12.219	0.93
	10	7.06	12.132	0.93
	12	7.03	12.074	0.93
	14	6.83	12.048	0.93
	16	6.99	12.082	0.93
	24	7.33	12.393	0.92

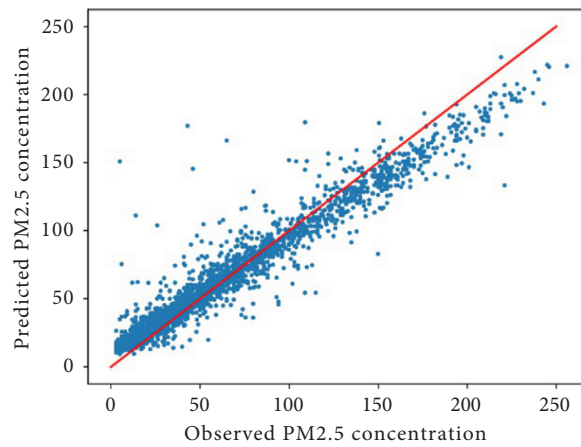


FIGURE 5: Scatter plot of PM2.5 observations and predicted values.

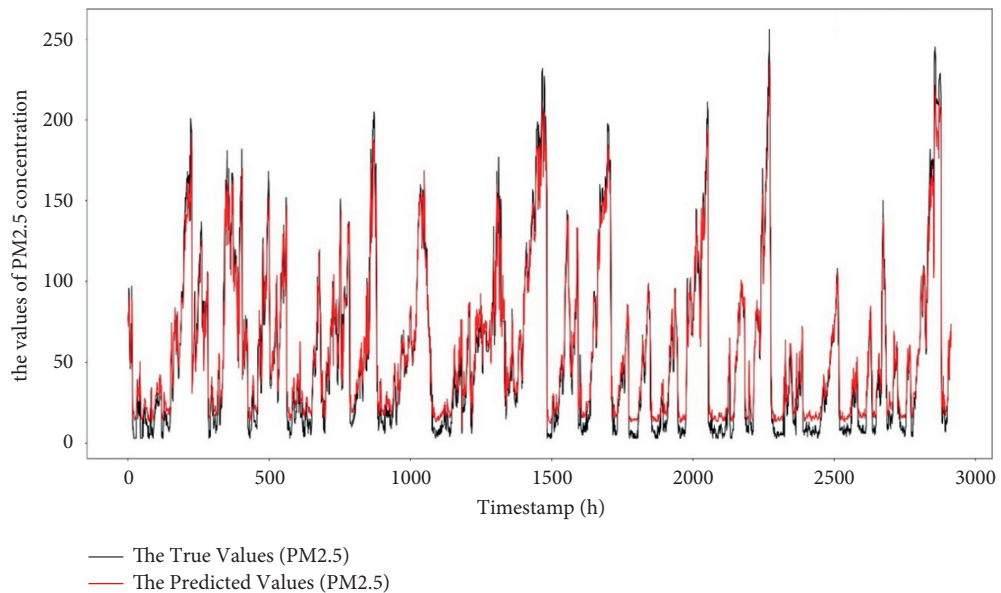


FIGURE 6: Fit curve of PM2.5 real value and predicted value.

TABLE 3: Comparison of various baseline models with the model’s indicators for the next hour.

Baseline model	MAE ($\mu\text{g}/\text{m}^3$)	RMES ($\mu\text{g}/\text{m}^3$)	R^2
LR	13.08	23.707	0.903
RF	8.02	15.131	0.918
SVM	34.69	68.106	0.255
ARMA	7.46	13.958	0.923
LSTM_N	7.11	13.943	0.922
LSTM_NW	7.33	13.567	0.926
LSTM_NE	7.79	12.757	0.919
LSTM_WE	7.21	12.353	0.926
LSTM_S	7.81	14.664	0.916
ST_LSTM	6.81	12.080	0.930

TABLE 4: Mean absolute error (MAE) values for different models predicted for the 1st to 6th hour of each baseline.

MAE ($\mu\text{g}/\text{m}^3$)	1st hour	2nd hour	3rd hour	4th hour	5th hour	6th hour
LSTM_N	7.02	10.78	14.09	17.27	20.25	22.40
LSTM_NW	7.33	11.02	14.03	16.71	19.15	21.19
LSTM_NE	7.79	10.33	13.24	15.54	17.88	20.26
LSTM_WE	6.99	10.37	13.18	16.98	18.09	19.85
LSTM_S	7.81	12.19	14.59	17.95	19.81	22.56
ST_LSTM	6.81	10.22	12.87	15.92	18.53	19.69

relatively accurate prediction of air quality in the future (predicted future concentration of PM2.5).

5.2. Comparison of Experiments. The comparison of the predicted performance of the model proposed in this paper with the other eight baselines at the next hour is shown in Table 3 on the three evaluation indicators of MAE, RMSE, and R^2 . It can be found that the deep learning baseline model performs better than the nondeep learning baseline model, where the SVM nondeep learning baseline model performs the worst. Comparing the three evaluation indicators of all deep learning baseline models, it is found that the proposed model performs optimally in predicting performance. Due to data limitations, the authors only conducted a comparison of the predicted results of the next 1 to 6 hours of the deep learning baseline model. The numerical values of the above six deep learning models on the MAE indicator are shown in Table 4 and are used to predict the air quality from the 1st hour to the 6th hour. From the numerical values, we can observe the predictions in the next 1 to 3 hours and the 6th hour. The error of the model proposed in this paper is the smallest. The error of the LSTM_NE model is the smallest at the fourth and fifth moments. The performance of the prediction model proposed in this paper ranks second; the possible reason is that other pollutant factors have a greater impact on PM2.5 concentration. The LSTM_NE model only considers other contaminant factors, and the model proposed in this paper not only considers other pollutant factors but also considers meteorological factors. Considering more factors, it weakens the influence of other pollutant factors. From the comparison of the data of the two models LSTM_NW and LSTM_NE, the author observed that the prediction performance of the second to sixth moments except for the 1st hour, the error of the LSTM_NE model is smaller than the error of the LSTM_NW model. It also

echoes the above conjecture that the influence of other pollutant factors on PM2.5 concentration is greater than that of meteorological factors on PM2.5 concentration.

6. Conclusions and Outlook

This paper presents a spatiotemporal prediction model for future prediction of air quality based on long short-term memory (LSTM) neural network. The model achieves more accurate and stable future predictions by integrating historical time air quality data, air quality data from nearest neighbors, meteorological data, and other pollutant data into the model. At the same time, the author used the real data sets of Beijing from January 1, 2016, to December 31, 2017, to evaluate the model proposed in this paper using MAE, RMSE, and R^2 evaluation indicators. The validity of the model is presented in the paper.

Generally, the proposed model is suitable for processing data from multiple monitoring sites in a single city as an input to a time series that can combine the interaction of multiple sites with the time dependence of air pollutants in the prediction system. However, there are still some limitations: (1) it can only predict the air pollutant concentration of a single site in a single city and cannot achieve the overall forecast of the city. In the future, it is hoped that all the site data and site prediction data in the city can be combined to achieve a comprehensive forecast for the entire city. (2) The model proposed in the article is only evaluated on the Beijing data set and has certain limitations. In the future, it is hoped that more monitoring data of other urban monitoring sites can be collected to further verify the performance of the model. (3) In the future work, I hope to consider more impact factors, such as traffic flow. This will allow you to better capture changes in air quality and obtain more accurate predictions.

Data Availability

The data used and analyzed in this paper are available at <https://www.bjmemc.com.cn/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Zhejiang Province under Grant no. LY20F020013, the Fundamental Public Welfare Research Program of Zhejiang Province under Grant nos. LGF19F020015 and LGG21F020006, and Hangzhou Dianzi University under Grant nos. JXGG2020YB009 and JXALK2020001.

References

- [1] R. D. Brook, S. Rajagopalan, C. A. Pope et al., "Particulate matter air pollution and cardiovascular disease," *Circulation*, vol. 121, no. 21, pp. 2331–2378, 2010.
- [2] B. Veronesi and M. Oortgiesen, "Neurogenic inflammation and particulate matter (pm) air pollutants," *Neurobehavioral Toxicology*, vol. 22, no. 6, pp. 795–810, 2002.
- [3] L. Calderón-Garcidueñas, M. Franco-Lira, R. Torres-Jardón et al., "Pediatric respiratory and systemic effects of chronic air pollution exposure: nose, lung, heart, and brain pathology," *Toxicologic Pathology*, vol. 35, no. 1, pp. 154–162, 2007.
- [4] M. A. Kioumourtzoglou, J. D. Schwartz, M. G. Weisskopf, S. J. Melly, Y. Wang, and F. Dominici, "Long-term pm2.5 exposure and neurological hospital admissions in the northeastern United States," *Environmental Health Perspectives*, vol. 124, no. 1, 124 pages, 2015.
- [5] G. Sermin, Z. Zeynep, S. H. Fuss, and G. Kursad, "The adverse effects of air pollution on the nervous system," *Journal of Toxicology*, vol. 2012, Article ID 782462, 2012.
- [6] Y. Zhang and Z. Li, "Remote sensing of atmospheric fine particulate matter (pm2.5) mass concentration near the ground from satellite observation," *Remote Sensing of Environment*, vol. 160, pp. 252–262, 2015.
- [7] Y. Zheng, X. W. Yi, M. Li, R. U. Li, Z. Q. Shan, and E. Chang, "Forecasting fine-grained air quality based on big data," in *Proceedings of the 21th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*, pp. 2267–2276, New York, NY, USA, August 2015.
- [8] S. Fangxia, Z. Tianle, and N. Mutong, "Advances in research on proinflammatory effects of biochemical components of atmospheric particulate matter," *Chinese Science Bulletin*, vol. 63, pp. 968–978, 2018.
- [9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
- [10] A. Al-Qaisi, M. S. AlTarawneh, A. ElSaid, and Z. Alqadi, "A hybrid method of face feature extraction, classification based on MLBP and layered-recurrent network," *Traitement du Signal*, vol. 37, no. 4, pp. 555–561, 2020.
- [11] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2017.
- [12] X. Yan and X. Song, "An image recognition algorithm for defect detection of underground pipelines based on convolutional neural network," *Traitement du Signal*, vol. 37, no. 1, pp. 45–50, 2020.
- [13] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, and B. Zhang, "A novel combined prediction scheme based on CNN and LSTM for urban pm2.5 concentration," *IEEE Access*, vol. 7, pp. 20050–20059, 2019.
- [14] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of pm2.5 based on graph convolutional neural network and long short-term memory," *The Science of the Total Environment*, vol. 664, pp. 1–10, 2019.
- [15] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, and Y.-S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts," *Journal of Cleaner Production*, vol. 209, pp. 134–145, 2019.
- [16] C. Wen, S. Liu, X. Yao et al., "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *The Science of the Total Environment*, vol. 654, pp. 1091–1099, 2019.
- [17] J. Wang and G. Song, "A deep spatial-temporal ensemble model for air quality prediction," *Neurocomputing*, vol. 314, pp. 198–206, 2018.
- [18] G. E. P. Box and G. M. Jenkins, "Time series analysis, forecasting and control," *Journal of the American Statistical Association*, vol. 134, no. 3, 1971.
- [19] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.
- [20] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 53–72, 2009.
- [21] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *News in Physiological Sciences*, vol. 25, pp. 1097–1105, 2012.
- [22] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649, Vancouver, BC, Canada, May 2013.
- [23] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the 31th International Conference on Machine Learning*, vol. 32, no. 2, pp. 1188–1196, Beijing China, June 2014.
- [24] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," 2014, <https://arxiv.org/abs/1409.3215v3>.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, and H. Schwenk, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, <https://arxiv.org/abs/1406.1078v3>.
- [27] D. A. Chu, Y. J. Kaufman, G. Zibordi, J. D. Chern, and B. N. Holben, "Global monitoring of air pollution over land from the earth observing system-terra moderate resolution imaging spectroradiometer (MODIS)," *Journal of Geophysical Research Atmospheres*, vol. 108, no. 21, 2003.
- [28] J. A. Engel-Cox, C. H. Holloman, B. W. Coutant, and R. M. Hoff, "Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air

- quality,” *Atmospheric Environment*, vol. 38, no. 16, pp. 2495–2509, 2004.
- [29] R. B. A. Koelemeijer, C. D. Homan, and J. Matthijsen, “Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe,” *Atmospheric Environment*, vol. 40, no. 27, pp. 5304–5315, 2006.
- [30] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu et al., “A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile,” *Atmospheric Environment*, vol. 42, no. 35, pp. 8331–8340, 2008.
- [31] P. E. Saide, G. R. Carmichael, S. N. Spak et al., “Forecasting urban pm10 and pm2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF-chem co tracer model,” *Atmospheric Environment*, vol. 45, no. 16, pp. 2769–2780, 2011.
- [32] K. Pearson, “Note on regression and inheritance in the case of two parents,” *Proceedings of the Royal Society of London*, vol. 58, no. 1, pp. 240–242, 2006.
- [33] X. Li, L. Peng, X. Yao et al., “Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation,” *Environmental Pollution*, vol. 231, pp. 997–1004, 2017.
- [34] D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, “Spatiotemporal prediction of air quality based on LSTM neural network,” *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2021–2032, 2021.

Research Article

Medical Text Classification Using Hybrid Deep Learning Models with Multihead Attention

Sunil Kumar Prabhakar ¹ and Dong-Ok Won ²

¹Department of Artificial Intelligence, Korea University, Seongbuk-gu, Seoul 02841, Republic of Korea

²Department of Artificial Intelligence Convergence, Hallym University, Chuncheon, Gangwon 24252, Republic of Korea

Correspondence should be addressed to Dong-Ok Won; dongok.won@hallym.ac.kr

Received 11 August 2021; Accepted 31 August 2021; Published 24 September 2021

Academic Editor: Yassine Maleh

Copyright © 2021 Sunil Kumar Prabhakar and Dong-Ok Won. 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.

To unlock information present in clinical description, automatic medical text classification is highly useful in the arena of natural language processing (NLP). For medical text classification tasks, machine learning techniques seem to be quite effective; however, it requires extensive effort from human side, so that the labeled training data can be created. For clinical and translational research, a huge quantity of detailed patient information, such as disease status, lab tests, medication history, side effects, and treatment outcomes, has been collected in an electronic format, and it serves as a valuable data source for further analysis. Therefore, a huge quantity of detailed patient information is present in the medical text, and it is quite a huge challenge to process it efficiently. In this work, a medical text classification paradigm, using two novel deep learning architectures, is proposed to mitigate the human efforts. The first approach is that a quad channel hybrid long short-term memory (QC-LSTM) deep learning model is implemented utilizing four channels, and the second approach is that a hybrid bidirectional gated recurrent unit (BiGRU) deep learning model with multihead attention is developed and implemented successfully. The proposed methodology is validated on two medical text datasets, and a comprehensive analysis is conducted. The best results in terms of classification accuracy of 96.72% is obtained with the proposed QC-LSTM deep learning model, and a classification accuracy of 95.76% is obtained with the proposed hybrid BiGRU deep learning model.

1. Introduction

There is a huge increase in the total number of electronic documents available online due to the development of information and Internet technology. This huge and unstructured form of text enables the automated text classification to a great extent [1]. In the field of NLP, text classification is one of the most important fields, and it helps in the assignment of the text documents to proper classes depending on their content. Many challenges and solutions are exhibited by the publicly available documents, and its classification is mainly intended for web classification, unstructured text classification, sentiment classification, spam e-mail filtering, and author identification [2]. Supervised classification techniques, like support vector machine (SVM) or naïve Bayesian classifier (NBC), are employed for

extraction of features when done by the most common bag of words approach [3]. As some words can be neglected easily along with the small training data here, it may suffer from sparsity problem. Therefore, recent studies concentrate on focusing of more complex features. In the text classification field, a special emphasis is always given to the medical text classification as a lot of medical records along with medical literature are contained in the medical text [4]. The medical records include the doctor's examination, diagnosis procedures, treatment protocols, and notification of improvement of the disease in the patient. The entire medical history along with the prescription effect of the medicine on the patient is also stored in the medical record. The medical literature includes the oldest and recent documents of the medical techniques used for diagnosis and treatment of a particular disease [5]. Both these two information resources

are very important in the field of clinical medicine. Due to the advent of information technology, tremendous quantity of electronic medical records and literature have been found online, which provides good resources of data mining in the medical field. Text classification in medical field is quite challenging because of two main issues: first, it has a few grammatical mistakes, and second, a lot of medical techniques are presented in the text [6]. With the advent of deep learning, such as convolutional neural networks (CNN) and recurrent neural network (RNN) being used widely in image, signals, and other applications, it has been equally successful in medical text classification [7].

Medical data can be classified on word, sentence, and even document levels in some works [8]. A good amount of medical data is available online, and these data provide useful information about the disease, symptoms, treatment, patient history, medication, and so on. To imbibe the most useful information, they need to be classified into their respective classes. An important step towards further implementation, such as classification and design of an automated medical diagnosis tool, is enabled by this task. Only very few works with respect to medical text classification has been proposed in literature, and only a handful amount of works have addressed multiclass medical text classification and some works have concentrated on binary medical text classification [9]. The majority of the medical text classification models are either on the word level or the sentence level classification rather than the document level classification. A recent comprehensive survey on text classification from shallow to deep learning was discussed in [8], and a survey on text classification algorithms was thoroughly analyzed recently in [9]. These two survey papers are very much useful as all the past techniques, associated working methodologies, datasets analysis, and comparison of all the results along with the possible future works are discussed well, thereby making it a nonnecessity for other authors to repeat the past works over and over again. However, a few essential works, which deal with medical text classification, are discussed in this work as follows. A famous work for medical text classification, which is being cited by almost every researcher in the medical text classification, was done by Hughes et al., where they have used more complex schemes to specify the classification features using CNN [10]. A systematic literature review along with the open issues present in it, exclusively in the field of clinical text classification research trends, was analyzed comprehensively by Mujtaba et al. [11]. A novel neural network-based technique using a BiGRU model [12], a paradigm using weak supervision and deep representation [13], a rule-based feature representation with knowledge guided CNN [14], a deep learning-based model using hybrid BiLSTM [15], and an improved distributed document representation with medical concept description for traditional Chinese medicine clinical text classification [16] are some of the famous works in the medical text classification. A cancer hallmark text classification using CNN was proposed by Baker et al., where the medical datasets were thoroughly investigated [17]. Other works in medical text classification, providing

some interesting results, include the integration technique of attentive rule construction with neural networks [18], genetic programming with the data driven regular expressions evolution methodology [19], improving multilabel medical text classification by means of efficient feature selection analysis [20], multilabel learning from medical plain text with convolutional residual models [21], and ontology based two-stage approach with particle swarm optimization (PSO) [22]. A medical social media text classification integrating consumer health technology [23], NLP-based instrument for medical text classification [24], efficient text augmentation techniques for clinical case classification [25], and hybridizing the idea of deep learning with token selection for the sake of patient phenotyping [26] are some of the applications related to medical text classification in general health technology aspects. The application of medical text classification in clinical assessments deals with the works, such as time series modelling using deep learning in the intensive care unit (ICU) data [27], phenotype prediction for multivariate time series clinical assessment using LSTM [28], hybridizing of RNN, LSTM, GRU, and BiLSTM for extraction of the clinical concept from texts [29], and identifying of the depression status in youth using unstructured text notes along with deep learning in [30]. An automatic text classification scheme, known as FasTag, which deals with unstructured medical semantics, was proposed recently in [31]. Similarly, many NLP tools are available in literature with specific observations, source codes, frameworks, and licenses, such as CLAMP, MPLUS, KMCI, SPIN, and NOBLE [32]. Every work proposed in literature has its own merits and demerits. No method is consistently successful at all times and no method is a consistent failure too. On analyzing this important point, after several trial-and-error attempts, in this work, two efficient deep learning models for medical text classification are proposed as a boost to the existing methods reporting some good results. Therefore, the main contributions in the paper are as follows:

- (i) A quad channel hybrid LSTM deep learning model has been implemented, and to the best of our knowledge, no one has ever developed, such a type of model for medical text classification. The main intention to develop a quad channel hybrid LSTM model is because, with four input channels, the characteristic diversity of the input can be greatly improved, thereby enhancing the classification accuracy of the model.
- (ii) A hybrid BiGRU model with a multihead attention is also successfully developed, and the primary intention to develop such a model is that the effective features in multiple subspaces can be well explored and the concatenation of the convolutional layers with the BiGRU layer can definitely provide good classification accuracy.

The organization of the paper is as follows. In Section 2, the two deep learning models for medical text classification are proposed, and the results and discussion are present in Section 3, followed by the conclusion in Section 4.

2. Design of Deep Learning Models for Classification of Medical Text

The hybrid deep learning models developed for medical text classification include two methods such as a quad channel hybrid LSTM model as a first method and hybrid BiGRU model as the second method.

2.1. Proposed Method 1: Quad Channel Hybrid LSTM Model.

Generally, there is a limitation of the semantic features as only the word level embedding is used often by the traditional CNN and RNN networks. There is a very limited capability when utilized by these models especially when the semantics has to be determined by these words. Therefore, expansion of channels is quite necessary, and the usage of multilevel embedding is required so that the characteristic diversity of the input is improved. For every word in the text, the relative importance is quite contrasting from the modality, which is conveyed. Few words can give a lot of contribution to modality, while some words have less contribution to modality. Therefore, to learn the characteristics of every word in a detailed manner, hybrid attention is added after LSTM, so that a tradeoff is achieved for various words with contrasting emotions. Thereby, the learning potential of the LSTM representation is improved. Overall, the unique characteristics in the learning of neural network representation are also improved. This specific aspect helps to refine the generalization, so that the overfitting can be easily prevented [33]. The division of the proposed model is done in the following parts, such as word embedding, hybridizing of CNN with LSTM, and hybrid attention scheme followed by the design of quad channel LSTM.

2.1.1. Word Embedding. For the representation of the word, an unsupervised learning algorithm called GloVe is utilized in order to obtain vector representations for words [34]. It is a count-based word representation natural language processing tool, and it utilizes the overall statistics. In between the words, the main semantic properties are captured by a vector of real numbers. By analyzing the Euclidean distance or the cosine similarity, the semantic similarity between the two words is computed easily. Word and character levels are two kinds of word segmentation considered in this model. Word2vec model proposed in [34] uses the related attributes between words, so that the semantic accuracy is increased. To deal with the dimensionality problem, a low dimensional space representation is utilized. CBOw and skip-gram are the two architectures used for word embedding in Word2vec. The surrounding words are utilized to predict the center word by CBOw method, and the central words are utilized to predict the surrounding words by skip-gram method. CBOw is fast in terms of swiftness for training the word embedding when compared to skip-gram. Skip-gram seems to be better with regard to accuracy when the semantic detail is expressed. Therefore, to train the word embedding, Word2vec model dependent on skip-gram is utilized in this paper. Figure 1 expresses the structure of a word embedding module.

2.1.2. CNN with LSTM Module. One of the primary algorithms of deep learning techniques is CNN. It is a famous feed forward neural network with a deep structure, which has convolution calculation, and it has been successfully implemented in computer vision and NLP. In this work, the hybrid combination of CNN along with LSTM is utilized. For processing the sequential data, RNN is used widely. The past output and the current input are concatenated together by this RNN model. The activation function tanh is used to control it, so that the sequence states can be considered. At a time t , the RNN derivative will spread and communicate to time $t - 1, t - 2, \dots, l$, thereby leading to the existence of a multiplication coefficient. Gradient explosion and disappearance occur when there is continuous multiplication occurring. During the forward process, the input of the start sequence has a very small or negligible effect on the late occurring sequences, and therefore, it is considered as a main problem of loss distance dependence. By means of introducing several gates, LSTM problem can be easily solved. The memorization of the input is done in a selective manner by the LSTM gate structures [35]. The memorization of the most vital information is done, and the less important information is forgotten completely. Thereby, the assessment of the next new information that could be saved in the current state is generated successfully. To a sigmoid function, the preceding state output h_{t-1} and the contemporary input in a function X_t are fed as an input so that a value between 0 and 1 is generated, thereby determining the current new information that could be retained easily. The complete state C_t of the next moment is obtained with the help of forget gate and the input gate, and it is utilized for the inception of the hidden layer h_t of the succeeding state, thereby forming the output of the present unit. The determination of the output is done by the output gate with respect to the information obtained from the cell state. A sigmoid function homogeneous to input gate, which generates a value o_t between 0 and 1, shows the amount of cell state information determined to project it as output. When the multiplication of the cell state information happens with o_t , it is activated by means of utilizing tanh layer, and so, the output details of the LSTM representation h_t are modeled. Figure 2 shows the illustration of a typical LSTM unit with suitable inputs and outputs. For the LSTM, the corresponding alliances between the various gates are mathematically expressed as follows:

$$\begin{aligned}
 z_t &= \tanh(W_z[h_{t-1}, X_t] + b_z), \\
 i_t &= \text{sigmoid}(W_i[h_{t-1}, X_t] + b_i), \\
 f_t &= \text{sigmoid}(W_f[h_{t-1}, X_t] + b_f), \\
 o_t &= \text{sigmoid}(W_o[h_{t-1}, X_t] + b_o), \\
 c_t &= f_t \cdot C_{t-1} + i_t \cdot z_t, \\
 h_t &= o_t \cdot \tanh(c_t).
 \end{aligned} \tag{1}$$

Figure 3 expresses the illustration of a LSTM unit utilized in this work. The gradient problem explosion will surely occur if the length of the input sequences is longer, thereby making it hectic to learn the information from a long-time

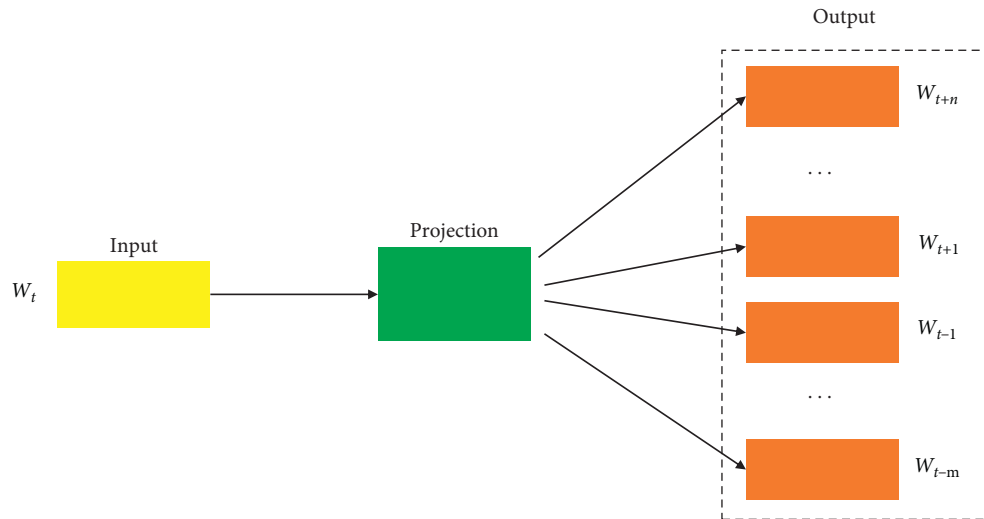


FIGURE 1: Structure of word embedding module.

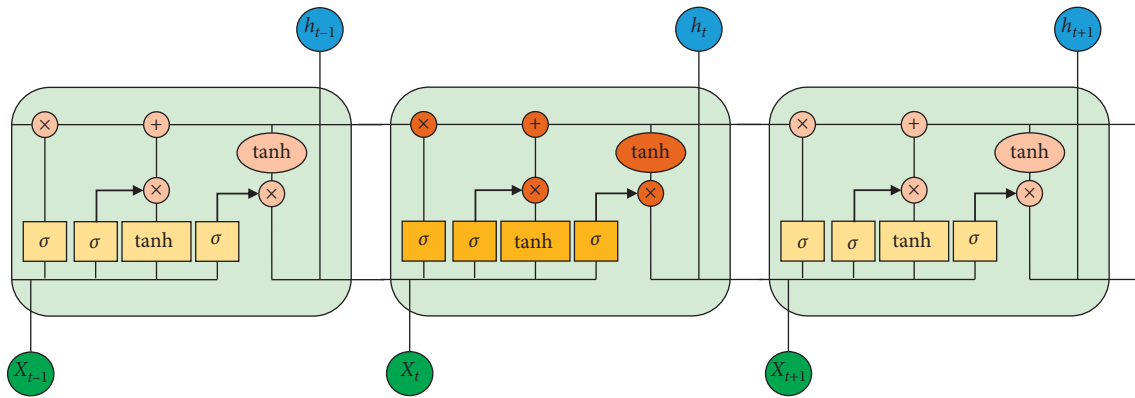


FIGURE 2: Illustration of a typical LSTM unit with suitable inputs and outputs.

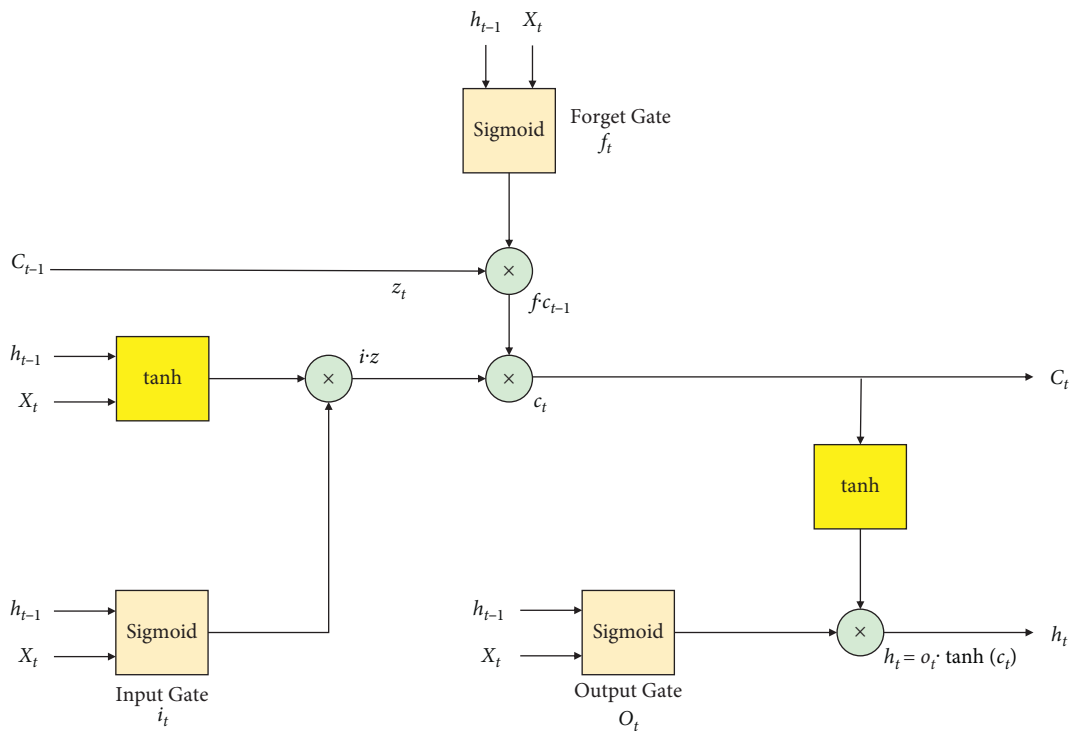


FIGURE 3: Illustration of a LSTM unit utilized in this work.

context. To solve this issue, the most popular variation of RNN that can be utilized is LSTM, and by means of launching a gate structure in every LSTM unit, this problem can be easily solved. The discarding information from a cell state is decided by the forget gate, and the assessment of new inputs is determined by the input gate. Depending on the present state of the cell, the determination of output value is done based on the information added to the cell state. A four-channel mechanism is introduced in the CNN-LSTM model by means of giving multiple labels of embeddings as input simultaneously at a given instant of time, so that multiple aspects of features are acquired. Therefore, the extraction of both word level and character level features can be done easily and at the same time. Based on the embedding granularity, the structure is split into the character and word levels. In each channel, the structure of model is sequential, and it is divided into two unique but different parts, such as CNN and LSTM neural network. For the input sequence X , the convolution result c is computed along with the convolution kernel K and is mathematically represented as

$$c = \text{conv}(X, K) + b. \quad (2)$$

For simplification of the representation, the LSTM procedure is unified as $LSTM(x)$. Series and parallel structures can be utilized for CNN and LSTM neural networks. Generally, series structures are commonly used in spite of the information loss due to the nature of the convolution process. Many time series characteristics are lost with the series structure, and so, compressed information is received with LSTM neural network. Therefore, series structure is replaced by parallel structure, and the results obtained are pretty good. In every channel, the recording of the structure is done, and it is expressed as

$$\text{channel}(x) = [\text{conv}(x) \oplus LSTM(x)]. \quad (3)$$

The basic explanation of the character and word levels is obtained from (3). With x representing the input and the output, expressed as C_{out} and W_{out} , it can be expressed as follows:

$$C_{out} = \text{channel}_{\text{embedding}} = v_w(x), \quad (4)$$

$$W_{out} = \text{channel}_{\text{embedding}} = v_c(x).$$

The word level embedding vectors trained is expressed as v_w , and the character level embedding vector trained is expressed as v_c , respectively. The interpretation and outcome of the output of the four channels are merged as a hidden layer output and is represented as

$$h = [C_{out} \oplus W_{out}]. \quad (5)$$

To the fully connected (FC) layer, this hidden layer result is sent, and finally for the classification output, the Softmax layer is used and is represented as

$$\hat{y} = \text{softmax}(\text{dense}(h)). \quad (6)$$

The four-channel representation is explained in the following sections, respectively.

2.1.3. Hybrid Attention Model. A vital constituent of the dynamic pliable weight structure is represented by the weight score w and its computation is expressed as

$$\begin{aligned} e_i &= v_a^T \tanh(W_r h_i + b), \\ h_i &= \langle h'_i; c_i \rangle, \end{aligned} \quad (7)$$

where h'_i indicates the LSTM output at a specific time t , h_i expresses the hidden layer output, c_i indicates the states in LSTM, v_a indicates the random initialization vector, b represents the bias, which is randomly initialized, and W_r indicates the random initialization weight matrix. The computation of score w is done as follows:

$$w = \frac{\exp(e_i)}{\varepsilon} \left[\sum_{k=1}^{T_x} \exp(e_i k) \right], \quad (8)$$

where the sequence length is expressed as x .

The dynamic adaptive weight is weighted to an output vector c_i and is represented as

$$c_i = \sum_{j=1}^{T_x} w \cdot h_j. \quad (9)$$

2.1.4. Design of the Quad Channel Hybrid LSTM Model.

The input text is first embedded, and then, the vector representation of these sequences is obtained to get a better semantic depiction and extricate the best text features. After the vector portrayal of these sequences are obtained, then these sequences are convolved by utilizing the convolution layer. The word-level semantic features can be well extracted by this model, and so, the input data along with the output size can be reduced by means of mitigating the overfitting aspect. The convolutional layer processes the data efficiently, and it sends it to the LSTM layer, so that the timing characteristics of the data can be well analyzed. Therefore, to increase the classification accuracy and avoid the secondary information of context semantics, this architecture is achieved. Figure 4 illustrates the quad channel hybrid attention model.

2.2. Proposed Method 2: Hybrid BiGRU Model with a Multihead Attention.

To the word embedding layer, the data processing results are fed as input, and the corresponding word vectors are obtained as output, which has rich semantics and a very low dimensionality. To extract the local features, CNN has a very strong ability, and parallel computing is enabled by it, so that a high training speed is achieved. To get the feature maps, multiple filters with suitable filter sizes are adopted. The features obtained from convolution are dealt with much efficiency here by means of applying maximum pooling and average pooling approaches, so that a good feature information is captured, and then, it is concatenated thereby the sentences are represented well. In order to get exact and more accurate semantic information, BiGRU is applied, so that the context information is extracted. The main reason for implementing the

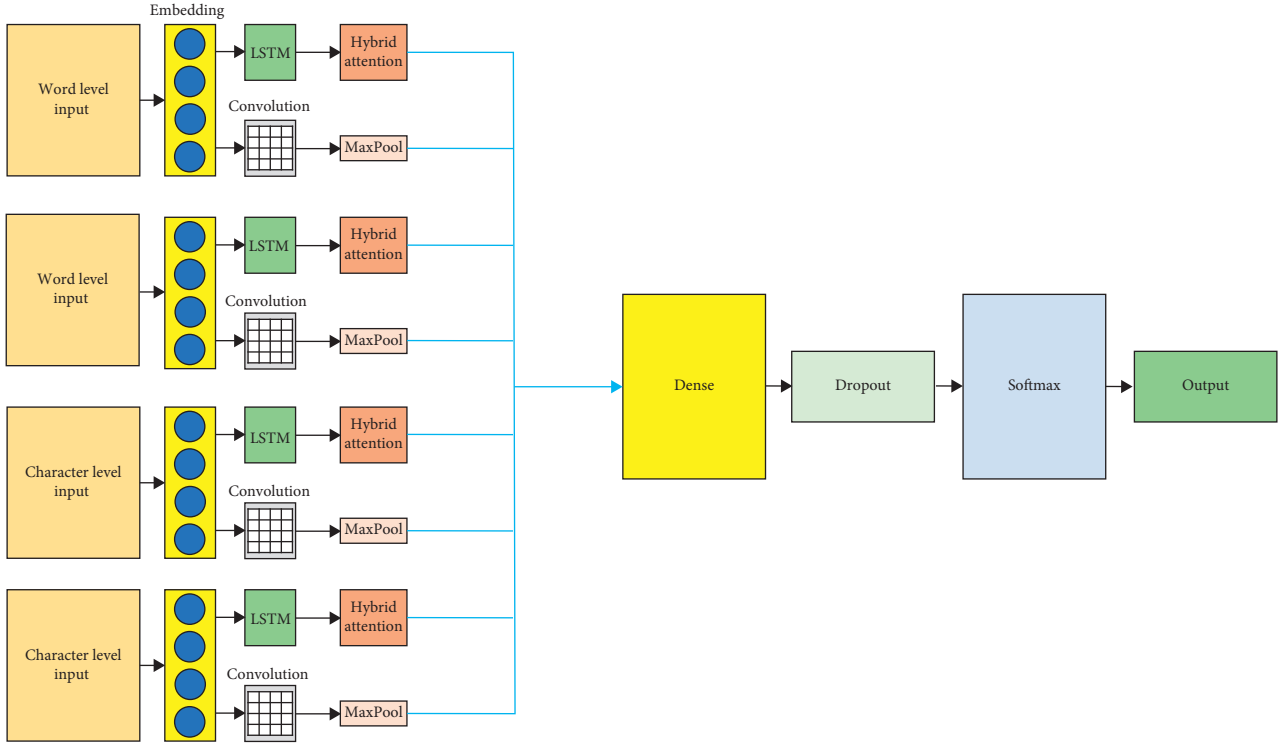


FIGURE 4: Proposed quad channel hybrid attention model.

BiGRU is due to the inability of CNN to capture context information and the gradient explosion problem caused by the simple RNN. In multiple subspaces, more effective and potential features can be obtained by the multihead attention rather than using single-head attention. The multihead attention layer outputs are nothing but the weighted word vector representation. Many global features are obtained by means of implementing the maximum and average pooling techniques, so that the word vector can be represented more accurately. Depending on the distinct attributes of CNN, BiGRU along with the multihead attention, the features are merged or concatenated as final features, and then, it is fed to the FC layer. Finally, the Softmax classifier is utilized to perform the classification process.

2.2.1. CNN and Text CNN. By means of imitating the biological visual perception mechanism, CNN was constructed, and so, both supervised learning and unsupervised learning are done easily [36]. With a very small amount of calculation, the lattice point features are obtained by CNN as the sparsity of the connections between the layers is enabled along with the sharing of parameters of convolutional kernel in the hidden layer. Figure 5 explains the structure of the CNN and it comprises of input layer, convolutional layer, and pooling layer along with a FC layer.

A text classification model known as text CNN is developed in [37] by making some preliminary adjustments or modifications in the input layer of the traditional CNN, and this work has been partly inspired by it and has been used in our work too. After the padding, the length of the sentence is considered to be n , the filter size is denoted by h , and the

word embedding dimension is denoted by d . The successful merging of words such as $x_i, x_{i+1}, \dots, x_{i+h-1}$ is expressed in every sentence as $x_{i:i+h-1}$. By means of using a nonlinear function, the resulting of a feature t_i is obtained from a collection of words $x_{i:i+h-1}$ and it is represented as follows:

$$t_i = f(wgx_{i:i+h-1} + b_i). \quad (10)$$

The bias term is represented as b_i and $w \in \mathfrak{R}^{hd}$ is a filter kernel. In the sentence representation $[x_1, x_2, \dots, x_{n-h+1}]^T$, this filter is used to each window of words, so that a feature map $[t_1, t_2, \dots, t_{n-h+1}]^T$ is obtained, and thus, the feature extraction from a filter is expressed by the previously mentioned process. The extraction of local features of various sizes is done by means of utilizing the diverse characteristics of the different filter kernel size. In this work, maximum and average pooling techniques are implanted to the features, which are obtained from the convolution layer, so that more features are extracted. Figure 6 expresses the proposed hybrid BiGRU deep learning model architecture.

2.2.2. Description of BiGRU Utilized in the Work. A famous kind of RNN is GRU [38], and to fathom issues like long-term memory along with gradients in the backpropagation process, this technique was utilized to solve the problem and it is more or less similar to LSTM. With sequential data as input, recursion is performed in the evolutionary direction of sequences by this class of RNN and the connection of all the neurons are in a chain. The information can be well received by the neurons from their own historical moments because of the cyclic factors addition in the hidden layer. The

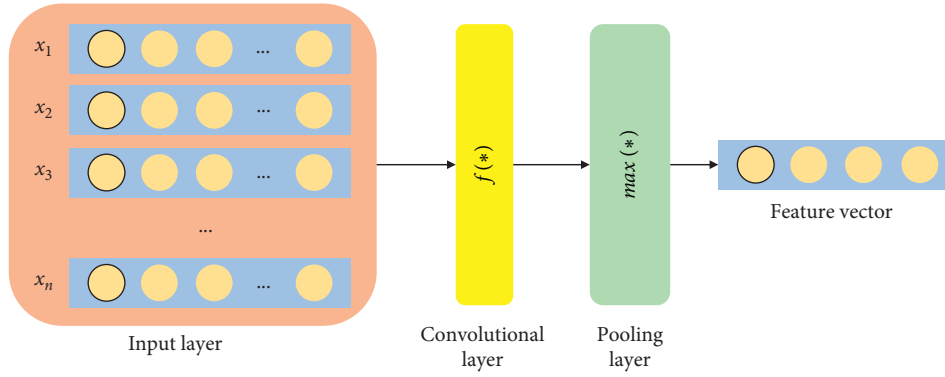


FIGURE 5: Structure of a CNN.

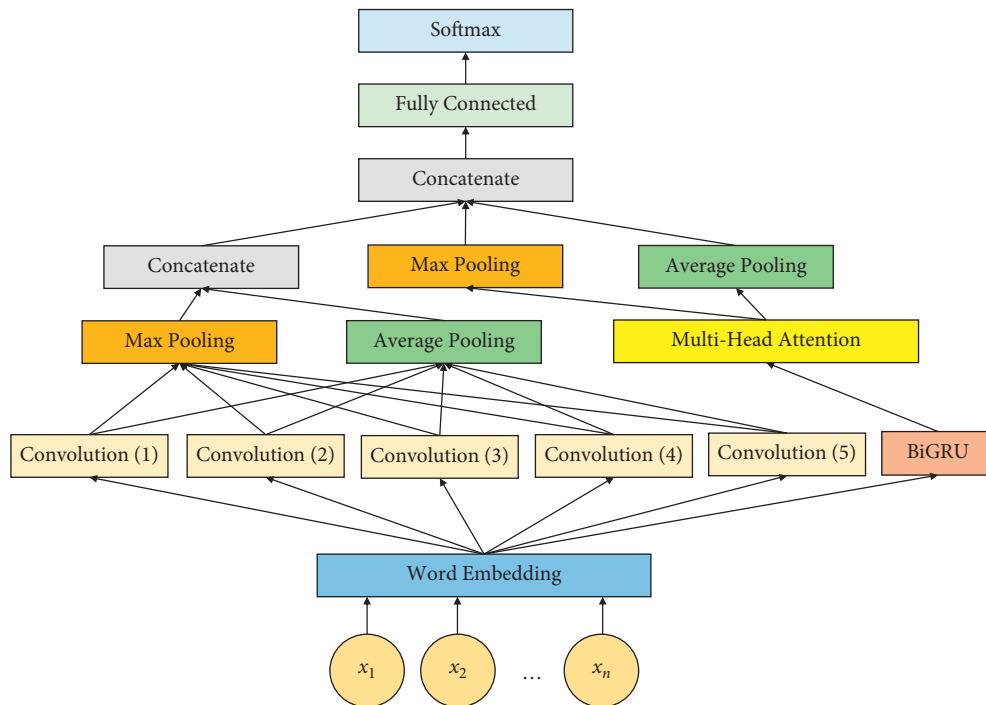


FIGURE 6: Proposed hybrid BiGRU deep learning model.

traits of sharing both memory and parameters are present in the RNN. In order to deal with the nonlinear feature learning of several data, RNN seems to be quite superior. The RNN gradient disappearance is a huge problem, and so, long-term historical load features cannot be learnt and LSTM is proposed by researchers, as in between the long short-term sequence data, the correlation information can be easily learnt. To deal with LSTM and its huge parameters along with a very slow or moderate convergence rate, GRU has been procured. Thus, a famous alternative of LSTM is GRU as it has very less parameters and can achieve a high convergence rate along with a good learning performance too [38]. Internally, the GRU model comprises of update gate and reset gate. The input gate and forget gate of LSTM are replaced by the update gate of GRU. The effect of output information of the hidden layer neuron is represented by the update gate at the preceding moment in the hidden layer

neurons of the present moment. The influence degree is pretty high when the value of updating gate is large. At the preceding moment, the hidden layer outputs are indicated by the reset gate, and less information is generally ignored when the reset gate value is large. A typical illustration of a GRU is depicted in Figure 7.

Using the following formulae, the hidden layer unit can be computed:

$$\begin{aligned}
 z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]), \\
 r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]), \\
 \tilde{h}_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]), \\
 \hat{h}_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t.
 \end{aligned}
 \tag{11}$$

where z_t represents the update gate and r_t represents the reset gate.

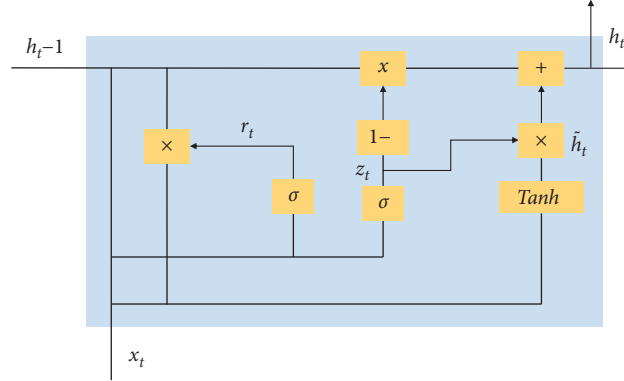


FIGURE 7: An illustration of a GRU.

The sigmoid function is represented by σ . The hyperbolic tangent is expressed by \tanh . The training parameter metrics considered here are W_r, W_z along with U_r, U_z , and U . The training parameter metrics W and U , resetting gate r_t , input x_t at the current moment, and output h_{t-1} at the previous moment of the hidden layer neuron are used to assess the candidate activation state \tilde{h}_t at the present moment. To grasp the association between current load along with the past and future load effecting components, a good capacity is present in the BiGRU network as the deep features of the load data can be conductively extracted. The structural representation of BiGRU is shown in Figure 8.

Its computations are as follows:

$$Y_2 = g(VA_2 + V'A_2'). \quad (12)$$

The computation of A_2' is as follows:

$$\begin{aligned} A_2 &= f(WA_1 + Ux_2), \\ A_2' &= f(W'A_3' + Ux_2). \end{aligned} \quad (13)$$

The hidden layer value S_t is highly affiliated to S_{t-1} in the forward calculation. The hidden layer value S_t is also highly concomitant to S_{t-1} in the reverse calculation. Depending on the success of both the forward and reverse calculations, the computation on final output is obtained. For the bidirectional RNN, the computation is as follows:

$$\begin{aligned} o_t &= g(VS_t + V'S_t'), \\ S_t &= f(Ux_t + WS_{t-1}), \\ S_t' &= f(U'x_t + W'S_{t-1}'). \end{aligned} \quad (14)$$

2.2.3. Implementation of Cross-Entropy Loss Function.

For classification issues, the implementation of the cross-entropy loss function is usually done [39]. The probability of each category is computed by the cross-entropy, and it materializes with sigmoid or softmax function mostly. Sigmoid function is usually expressed as follows:

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \quad (15)$$

The following function is obtained once the sigmoid function $\sigma(z)$ is derived and represented as

$$\sigma'(z) = \frac{e^{-z}}{(1 + e^{-z})^2} = \delta(z)(1 - \delta(z)). \quad (16)$$

The sigmoid function curve is smoother if the value of x is large or small, which specifies that the derivative $\sigma'(x)$ is inclined closely to zero. The model needs to predict two cases in the dichotomy situations. For each of these categories, the prediction probabilities are p and $1 - p$. The expression of cross-entropy loss function at this time is given as

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)], \quad (17)$$

where the label of sample i is indicated by y_i , negative and positive classes are indicated by 0 and 1, and p_i represents the likelihood that the sample i is anticipated to be positive.

2.2.4. Incorporation of the Multihead Attention Mechanism Scheme.

A famous brain signal processing procedure similar to vision of humans is the visual attention mechanism. In order to procure the specific area that needs to be carefully pivoted on, the global image is scanned quickly by the human vision and is termed as focus of attention. To get more detailed information, attention resources are fully set to this area so that the necessary attention is paid, and the useless information is avoided completely. Therefore, from a huge amount of information, the information with high values can be easily screened out with very limited attention resources. The efficiency and accuracy of visual information processing are improved to a great extent by means of using human visual attention mechanism. To different fields of deep learning, attention mechanism has been applied, such as image processing tasks, NLP tasks, and speech recognition tasks. Therefore, to understand the development of deep learning methodology, the working of attention mechanism is quite important. When similar sentences appear, then the model will be prompted by the attention mechanism to focus more on the words, so that the learning capability along with its generalization ability of the model is enhanced. A very

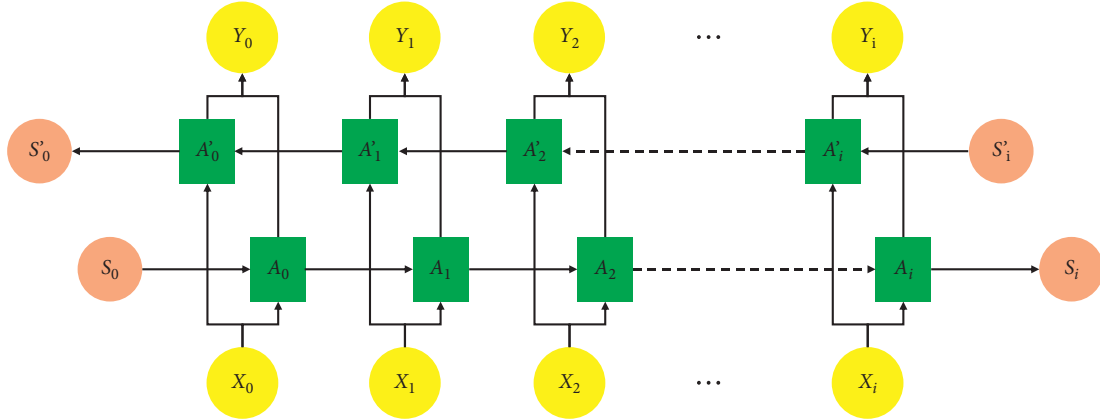


FIGURE 8: Illustration of a BiGRU.

special case of the general attention mechanism is the self-attention mechanism. The attention-related query matrix is represented by Q , the key matrix is represented by K , and the value matrix is represented by V . The condition of $Q = K = V$ is satisfied in the self-attention mechanism. The distance between the words is completely ignored, and the dependency relationship is calculated directly. The internal structure of a sentence can be learnt well, and a good attention can be paid to the interdependence between the internal words. To enhance the learning model ability and increase the neural network interpretability, the RNN is combined with the CNN model. Figure 9 explains the basic structure of multihead attention. A variation of the general attention is nothing but the scaled dot product attention at the central position. The computation of the scaled dot product attention for given matrices $Q \in \mathfrak{R}^{n \times d}$, $K \in \mathfrak{R}^{n \times d}$, and $V \in \mathfrak{R}^{n \times d}$ is given as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V, \quad (18)$$

where the total number of hidden units in the neural network model is expressed as d .

The self-attention mechanism is adopted by the multi-head attention implying that $Q = K = V$ as projected in Figure 9. Therefore, to apprehend the dependencies within a full series pattern, the calculation of the current position information along with the other position's information is done because of this mechanism. For instance, if the input is considered as a sentence, then every word in it should be managed with attention calculation. On the inputs Q, K , and V , a linear transformation is performed by the multihead attention. The scaled dot product attention computation is performed multiple times as it is a multihead attention mechanism [40]. For every head calculation, the linear projections of Q, K and V are quite divergent from each other. The number of calculations is actually meant by the number of heads. If the i^{th} head is considered as an example, then it is represented as follows:

$$\begin{aligned} Q' &= Q * W_i^Q, \\ K' &= K * W_i^K, \\ V' &= V * W_i^V. \end{aligned} \quad (19)$$

The output of the BiGRU layer is received by this layer, and so it is represented as

$$Q = K = V = y_t. \quad (20)$$

The ultimate result of this head is represented as

$$M_i = \text{softmax}\left(\frac{Q'K'^T}{\sqrt{d}}\right)V'. \quad (21)$$

3. Results and Discussion

In this section, the evaluation indices and datasets utilized along with the respective analysis of the two proposed deep learning models is analyzed comprehensively.

3.1. Evaluation Index. The evaluation indices considered in this work are accuracy, precision, recall, and F score. Their respective formulae are as follows:

$$\begin{aligned} \text{accuracy} &= \frac{T_P + T_N}{T_P + F_N + T_N + F_P}, \\ \text{recall} &= \frac{T_P}{T_P + F_N}, \\ \text{precision} &= \frac{T_P}{T_P + F_P}, \\ F_1 &= \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}, \end{aligned} \quad (22)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

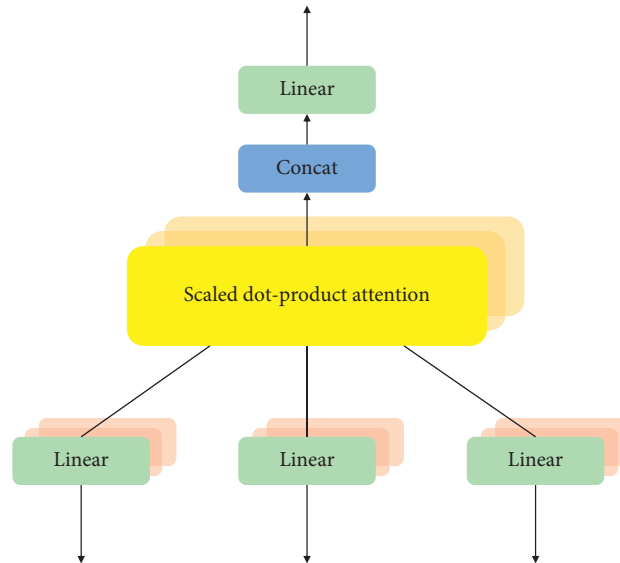


FIGURE 9: Multihead attention scheme.

3.2. Datasets Utilized. In order to conduct the performance evaluation of the proposed approach for medical text classification, the experiments were tested on two important benchmarking medical literature datasets, such as the Hallmarks dataset and AIM dataset. These datasets are available in [17] and is nothing but a group of biomedical publication abstracts, which are annotated for the hallmark of cancer. In about 1852 biomedical publication abstracts, three hallmarks of cancer are contained in this dataset, such as activating invasion and metastasis, deregulating cellular energetics, and tumor promoting inflammation. AIM dataset, known as activating invasion, and metastasis database has two sets of categories, such as positive and negative. All the in-depth details of the dataset can be obtained in [17]. The details of the datasets are tabulated in Table 1.

3.3. Analysis with Proposed Model 1: Quad Channel Hybrid LSTM. For the proposed quad channel hybrid LSTM, the experiments were analyzed with various parameters in order to get the consolidated best results. The batch size was set as 512, and the filter size was chosen in the ranges between [1, 3, 5, 7, 9]. The feature map number was assigned as 200, and the activation function was experimented with various combinations, such as ReLU, sigmoid, SoftPlus, and hard sigmoid. The values of the LSTM output were set as 128, respectively. The learning rate analyzed in this experiment is tested with various values, such as 0.1, 0.01, 0.001, and 0.0001, in order to assess the performance, and the dropout rate was analyzed with 0.2, 0.3, 0.4, and 0.5, to check out for the best results. The loss considered is binary cross entropy and the optimizer utilized is again experimented with Adam, SGD, Nadam, and AdaGrad, to provide a comprehensive analysis.

The experiment was tried for various convolution kernels, and the results are reported in Table 2. The best results are obtained when the kernel filter size is set as [1, 3, 5] as an accuracy of 72.18, precision of 70.52, recall value of 69.69, and $F1$ -score of 70.10 are obtained for the Hallmarks dataset, and an accuracy of 94.12, precision of 85.71, recall value of 83.93, and $F1$ -score of 84.81 are obtained for the AIM dataset. If the kernel filter size is further increased, then it leads to degradation in the performance computation measures.

The convergence of the objective function to the local minimum is determined by the learning rate, and it serves as a significant hyper parameter in almost all the deep learning applications. In order to make sure that the convergence of the objective function is successfully implemented to the local minimum in a specific interval of time, choosing the learning rate should be done wisely. The convergence would be very slow or moderate if the learning rate is quite small. A cost function oscillation will occur if the learning rate is too high. The experiment was tried for different learning rates, and the results are reported in Table 3. The best results are obtained when the learning rate is set as 0.01 as an accuracy of 73.18, precision of 71.95, recall value of 72.67, and $F1$ -score of 72.30 are obtained for the Hallmarks dataset, and an accuracy of 95.72, precision of 86.77, recall value of 83.98, and $F1$ -score of 85.35 are obtained for the AIM dataset. The experiment was started with the learning rate of 0.1, but it did not provide satisfactory results. However, when learning rate was set as 0.01, the best result was obtained and if the learning rate is further decreased, then there is a degradation in the performance metrics measures.

The experiment was tried for different dropout rates, and the results are reported in Table 4. The best results are obtained when the dropout rate was gradually increased from 0.2 to 0.5. When the dropout rate was set as 0.5, the best

TABLE 1: Dataset details.

Datasets	Classes	Sentence length	Vocabulary size	Dataset size	Training set	Validation set	Test set
Hallmarks	3	833	29141	8472	5931	1694	847
AIM	2	833	29141	2646	1853	529	264

TABLE 2: Analysis with different convolution kernel conditions for the proposed quad channel hybrid model.

Kernel filter sizes	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
[1, 3]	Hallmarks	62.58	55.76	54.82	55.28
	AIM	81.44	75.36	79.21	77.23
[1, 3, 5]	Hallmarks	72.18	70.52	69.69	70.10
	AIM	94.12	85.71	83.93	84.81
[1, 3, 5, 7]	Hallmarks	71.49	62.98	64.98	63.96
	AIM	79.90	73.98	72.81	73.39
[1, 3, 5, 7, 9]	Hallmarks	64.71	59.37	55.91	57.58
	AIM	57.36	52.90	54.27	53.57

TABLE 3: Analysis of results with different learning rates for the proposed quad channel hybrid model.

Learning rate	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
0.1	Hallmarks	59.85	61.76	60.78	61.26
	AIM	75.47	78.31	79.16	78.73
0.01	Hallmarks	73.18	71.95	72.67	72.30
	AIM	95.72	86.77	83.98	85.35
0.001	Hallmarks	72.19	70.99	72.98	71.97
	AIM	84.65	79.99	79.81	79.89
0.0001	Hallmarks	71.73	74.31	73.81	74.05
	AIM	82.36	80.91	79.73	80.31

TABLE 4: Analysis of results with different dropout rates for the proposed quad channel hybrid model.

Dropout rate	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
0.2	Hallmarks	69.75	61.65	62.89	62.26
	AIM	90.17	87.38	85.12	86.23
0.3	Hallmarks	71.76	63.63	61.79	62.69
	AIM	56.23	51.80	52.67	52.23
0.4	Hallmarks	71.19	65.99	63.18	64.55
	AIM	79.71	73.99	72.91	73.44
0.5	Hallmarks	72.93	70.95	69.67	70.30
	AIM	94.73	87.81	88.98	88.39

results are obtained, and an accuracy of 72.93, precision of 70.95, recall value of 69.67, and *F1*-score of 70.30 are obtained for the Hallmarks dataset, and an accuracy of 94.73, precision of 87.81, recall value of 88.98, and *F1*-score of 88.39 are obtained for the AIM dataset.

Table 5 shows the analysis of results with different optimizers for the proposed quad channel hybrid model. The best results are obtained when Adam optimizer is used instead of SGD, Nadam, and AdaGrad as a high accuracy of 72.98, precision of 69.65, recall value of 71.61, and *F1*-score of 70.61 are obtained for the Hallmarks dataset, and an accuracy of 95.12, precision of 87.17, recall value of 85.99, and *F1*-score of 86.57 are obtained for the AIM dataset.

Table 6 shows the analysis of results with different activation functions for the proposed quad channel hybrid

model. The best results are obtained when ReLU activation function is used instead of Sigmoid, SoftPlus, and hard sigmoid as a high accuracy of 71.92, precision of 70.92, recall value of 68.62, and *F1*-score of 69.75 are obtained for the Hallmarks dataset, and an accuracy of 92.17, precision of 88.83, recall value of 86.91, and *F1*-score of 87.85 are obtained for the AIM dataset.

Table 7 shows the consolidated analysis of the proposed quad channel hybrid model with the best combinations of values. The best results are obtained when the convolution filter size is [1, 3, 5], learning rate is 0.01, dropout rate is 0.5, optimization function used in Adam along with ReLU activation function is used instead of Sigmoid, SoftPlus, and hard sigmoid, and the results are interpreted. The proposed quad channel LSTM model produces an accuracy of 75.98,

TABLE 5: Analysis of results with different optimizers for the proposed quad channel hybrid model.

Optimizer	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Adam	Hallmarks	72.98	69.65	71.61	70.61
	AIM	95.12	87.17	85.99	86.57
SGD	Hallmarks	71.72	63.58	61.84	62.69
	AIM	82.35	78.98	77.21	78.08
Nadam	Hallmarks	71.43	65.98	63.28	64.60
	AIM	86.76	72.98	72.31	72.64
AdaGrad	Hallmarks	70.61	61.56	60.81	61.18
	AIM	94.41	85.13	85.26	85.19

TABLE 6: Analysis of results with different activation functions for the proposed quad channel hybrid model.

Optimizer	Dataset (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ReLU	Hallmarks	71.92	70.92	68.62	69.75
	AIM	92.17	88.83	86.91	87.85
Sigmoid	Hallmarks	69.71	64.72	60.81	62.70
	AIM	55.31	50.90	52.21	51.54
SoftPlus	Hallmarks	70.49	64.97	66.91	65.92
	AIM	77.73	74.91	72.82	73.85
Hard sigmoid	Hallmarks	64.35	61.75	60.81	61.27
	AIM	89.48	86.22	85.22	85.71

TABLE 7: Consolidated results for the best combination values of the proposed quad channel hybrid model.

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Quad channel LSTM	Hallmarks	75.98	71.95	70.67	71.30
	AIM	96.72	88.87	86.98	87.91

precision of 71.95, recall value of 70.67, and $F1$ -score of 71.30, which are obtained for the Hallmarks dataset, and an accuracy of 96.72, precision of 88.87, recall value of 86.98, and $F1$ -score of 87.91 are obtained for the AIM dataset.

3.4. Analysis with Proposed Model 2: Hybrid BiGRU Model with Multihead Attention. The analysis with the proposed second model deals with analysis of various filter sizes of CNN, analysis with different layers of BiGRU, analysis with different learning rates, analysis with different dropout rates, analysis with different optimizers, and analysis with different activation functions. To test the model, various convolution kernel filter sizes were effectively utilized, and the results are tabulated in Table 8. The efficiency of the model training will be greatly affected by the different kernel sizes, and so, the accuracy of the experimental results can vary. The best results are obtained when the kernel filter size is set as [1, 3, 5] as an accuracy of 73.88, precision of 69.54, recall value of 68.67, and $F1$ -score of 69.10 are obtained for the Hallmarks dataset, and an accuracy of 95.29, precision of 89.88, recall value of 84.13, and $F1$ -score of 86.90 are obtained for the AIM dataset. If the kernel filter size is further increased, then it leads to degradation in the performance computation measures.

For the developed hybrid BiGRU model, the analysis is done with a single layer and multiple layers too, and the analysis of these results is tabulated in Table 9. The best

results are obtained when the number of layers is set as two, as an accuracy of 72.11, precision of 71.87, recall value of 72.75, and $F1$ -score of 72.30 are obtained for the Hallmarks dataset, and an accuracy of 94.01, precision of 88.95, recall value of 84.91, and $F1$ score of 86.88 are obtained for the AIM dataset. If the layer size is slightly increased, then it leads to a degradation in the performance computation measures.

Multiple learning rates are assessed in this experiment for this architecture also, so that an optimal learning rate is found out, and the results are tabulated in Table 10. The best results are obtained when the learning rate is set as 0.01, as an accuracy of 74.18, precision of 73.58, recall value of 72.93, and $F1$ -score of 73.25 are obtained for the Hallmarks dataset, and an accuracy of 95.12, precision of 87.87, recall value of 87.27, and $F1$ -score of 87.56 are obtained for the AIM dataset. The experiment was started with the learning rate of 0.1 but it did not provide satisfactory results. However, when learning rate was set as 0.01, the best result was obtained. If the learning rate is further decreased, then there is degradation in the performance metrics measures, similar to the proposed first model.

The experiment was tried for different dropout rates, and the results are reported in Table 11. The best results are obtained when the dropout rate was gradually increased from 0.2 to 0.5. When the dropout rate was set as 0.4, the best results are obtained, and an accuracy of 72.82, precision of 71.56, recall value of 70.92, and $F1$ -score of 71.23 are

TABLE 8: Analysis with different convolution kernel conditions for the proposed hybrid BiGRU model.

Kernel filter sizes	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
[1, 3]	Hallmarks	71.85	61.79	60.70	61.24
	AIM	93.27	89.36	85.29	87.27
[1, 3, 5]	Hallmarks	73.88	69.54	68.67	69.10
	AIM	95.29	89.88	84.13	86.90
[1, 3, 5, 7]	Hallmarks	72.52	69.72	65.99	67.80
	AIM	82.79	76.91	73.78	75.31
[1, 3, 5, 7, 9]	Hallmarks	70.48	67.72	63.85	65.72
	AIM	58.39	51.84	57.18	54.37

TABLE 9: Analysis of results with different BiGRU layers.

Layers	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	Hallmarks	67.81	59.16	57.19	58.15
	AIM	91.11	83.35	83.26	83.30
2	Hallmarks	72.11	71.87	72.75	72.30
	AIM	94.01	88.95	84.91	86.88
3	Hallmarks	71.91	69.27	65.91	67.54
	AIM	81.16	76.18	75.37	75.77
4	Hallmarks	72.73	66.17	64.98	65.56
	AIM	69.69	54.99	57.21	56.07

TABLE 10: Analysis of results with different learning rates.

Learning rate	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
0.1	Hallmarks	63.85	56.16	55.81	55.98
	AIM	88.17	82.36	81.28	81.81
0.01	Hallmarks	74.18	73.58	72.93	73.25
	AIM	95.12	87.87	87.27	87.56
0.001	Hallmarks	74.41	69.91	69.98	69.94
	AIM	82.64	77.09	76.91	76.99
0.0001	Hallmarks	70.71	64.71	65.91	65.30
	AIM	61.36	55.16	57.73	56.41

TABLE 11: Analysis of results with different dropout rates for the proposed hybrid BiGRU model.

Dropout rate	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
0.2	Hallmarks	61.88	57.19	58.84	58.00
	AIM	89.72	83.58	82.97	83.27
0.3	Hallmarks	71.51	66.73	64.70	65.69
	AIM	62.39	57.18	55.23	56.18
0.4	Hallmarks	72.82	71.56	70.92	71.23
	AIM	94.22	89.87	88.21	86.12
0.5	Hallmarks	72.45	68.97	67.88	68.42
	AIM	81.61	75.98	74.92	75.44

obtained for the Hallmarks dataset, and an accuracy of 94.22, precision of 89.87, recall value of 88.21, and F1-score of 86.12 are obtained for the AIM dataset.

Table 12 shows the analysis of results with different optimizers for the proposed hybrid BiGRU model. The best results are obtained when Adam optimizer is used instead of SGD, Nadam, and AdaGrad, as a high accuracy of 70.52, precision of 74.78, recall value of 73.16, and F1-score of 73.96 are obtained for the Hallmarks dataset, and an accuracy of

93.98, precision of 88.63, recall value of 89.08, and F1-score of 88.85 are obtained for the AIM dataset.

Table 13 shows the analysis of results with different activation functions for the proposed quad channel hybrid model. The best results are obtained when sigmoid activation function is used instead of ReLU, SoftPlus, and hard sigmoid, as a high accuracy of 74.69, precision of 72.29, recall value of 71.91, and F1-score of 72.09 are obtained for the Hallmarks dataset, and an accuracy of 94.29, precision of

TABLE 12: Analysis of results with different optimizers for the proposed hybrid BiGRU model.

Optimizer	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Adam	Hallmarks	70.52	74.78	73.16	73.96
	AIM	93.98	88.63	89.08	88.85
SGD	Hallmarks	69.72	65.23	61.83	63.48
	AIM	69.31	66.57	68.27	67.40
Nadam	Hallmarks	70.13	68.84	68.64	68.73
	AIM	80.62	75.96	72.56	74.22
AdaGrad	Hallmarks	65.15	62.46	55.81	58.94
	AIM	86.15	81.59	81.28	81.43

TABLE 13: Analysis of results with different activation functions for the proposed hybrid BiGRU model.

Optimizer	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ReLU	Hallmarks	71.64	67.39	66.69	67.03
	AIM	88.39	82.20	83.61	82.89
Sigmoid	Hallmarks	74.69	72.29	71.91	72.09
	AIM	94.29	89.74	88.56	89.14
SoftPlus	Hallmarks	73.12	71.03	71.82	71.42
	AIM	83.39	81.91	82.84	82.37
Hard sigmoid	Hallmarks	61.22	58.92	57.41	58.15
	AIM	85.87	83.19	81.49	82.33

TABLE 14: Comparison of results for the best combination values of the proposed BiGRU hybrid model.

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Hybrid BiGRU model	Hallmarks	74.71	70.82	68.99	69.89
	AIM	95.76	88.38	84.15	86.21

89.74, recall value of 88.56, and *F1*-score of 89.14 are obtained for the AIM dataset.

Table 14 shows the consolidated analysis of the proposed hybrid BiGRU model with the best combinations of values. The best results are obtained when the convolution filter size is [1, 3, 5], learning rate is 0.01, dropout rate is 0.4, optimization function used is Adam along with sigmoid activation function instead of ReLU, SoftPlus, and hard sigmoid, and the results are interpreted. The proposed hybrid BiGRU model produces an accuracy of 74.71, precision of 70.82, recall value of 68.99, and *F1*-score of 69.89, which are obtained for the Hallmarks dataset, and an accuracy of 95.76, precision of 88.38, recall value of 84.15, and *F1*-score of 86.21 are obtained for the AIM dataset.

3.5. Baseline Methods and Overall Comparison with Other Methods. For text classification, the following baseline methods are used for comparison, such as CNN, LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM, logistic regression, naïve Bayesian classifier (NBC), SVM, and BiGRU. Table 15 reports the compared classification accuracy of the two proposed architectures against other machine and deep learning models on the two datasets. Actually, just one or two works published in high quality peer-reviewed journals are available online for comparison of the proposed deep learning models with the results of the other deep learning models on the same dataset. Therefore, in

this work, the results have been computed and then compared by analyzing the proposed deep learning model results with the standard and conventional deep learning techniques.

The developed two models have obtained very good results and crossed the performance of the state of art literature compared with some deep learning models. In machine learning and deep learning, it has to be observed that the final classification accuracies may range from a plus or minus two to three percent, but the working methodology and interpretation of the result are more important than trying to prove or obtain slightly higher classification accuracy than the other methods. Therefore, with this understanding the proposed quad channel hybrid LSTM model produced a high classification accuracy of 75.98% for the Hallmark dataset, and the same model produced a classification accuracy of 96.72% for the AIM dataset. The high performance is due to the development of four channels, so that the inherent features can be learnt and observed well through those channels, thereby enhancing the characteristic diversity of the input. Similarly, the hybrid BiGRU with multihead attention model produced a high classification accuracy of 74.71% for the Hallmark dataset, and the same model produced a classification accuracy of 95.76% for the AIM dataset. This is due to the effective capturing of the features by the hybrid model along with the careful selection of appropriate hyperparameters.

TABLE 15: Performance comparison of accuracy (%) of the proposed deep learning models with other deep learning models on the same dataset.

Methods	Hall mark dataset	AIM dataset
CNN	68.55	82.17
LSTM	70.76	83.16
BiLSTM	72.58	87.77
CNN-LSTM	71.81	91.98
CNN-BiLSTM	73.99	93.06
Logistic regression	61.91	72.92
NBC	65.35	73.84
SVM	66.99	84.55
BiGRU	69.34	89.98
Proposed method 1: quad channel hybrid LSTM model	75.98	96.72
Proposed method 2: hybrid BiGRU with multihead attention model	74.71	95.76

4. Conclusion and Future Work

By means of extracting the structured information, such as specification of the diseases and the pathological conditions associated with it, the information embedded in the clinical text is unlocked by using automated clinical text classification. By means of using symbolic techniques/statistical techniques, the tackling of the medical text classification is done. Handcrafted expert rules are usually needed every time with symbolic techniques, and they are quite expensive and cumbersome to develop. Statistical techniques, like machine learning, seem to be quite effective for the medical text classification tasks. However, it still requires extensive human efforts in order to label a large set of training data. In this paper, two deep learning models have been developed, and it has been successfully validated on two datasets too. When the proposed quad channel hybrid LSTM is implemented to Hallmarks dataset, a classification accuracy of 75.98% is obtained, and when it is implemented to AIM dataset, a classification accuracy of 96.72% is obtained. When the proposed hybrid BiGRU model is implemented to Hallmarks dataset, a classification accuracy of 74.71% is obtained, and when it is implemented to AIM dataset, a classification accuracy of 95.76% is obtained. Future works aim to develop more effective hybrid deep learning models for the efficient classification of medical texts. Future works also aim to explore content-based features and a variety of other domain specific features and plans to amalgamate it with very efficient hybrid deep learning techniques to get a good classification accuracy.

Data Availability

All the programming codes will be made available to the researchers upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by Hallym University Research Fund, 2020 (HRF-202011-006).

References

- [1] C. Zhou, C. Sun, and Z. Liu, "A C-LSTM neural network for text classification," 2015, <https://arxiv.org/abs/1511.08630>.
- [2] P. Zhou, Z. Qi, S. Zheng, and J. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," in *Proceedings of the COLING*, pp. 3485–3495, Osaka, Japan, December 2016.
- [3] T. Joachims, "Text categorization with support vector machines: learning with many relevant features," in *Proceedings of the European Conference on Machine Learning*, pp. 137–142, Chemnitz, Germany, April 1998.
- [4] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Met thesaurus: the MetaMap program," in *Proceedings of the Annual Symposium of the American Medical Informatics Association (AMIA '01)*, pp. 17–21, Washington, DC, USA, November 2001.
- [5] G. Luo, J. Liu, C. C. Yang, L. Deléger, and C. Grouin, "Detecting negation of medical problem in French clinical notes," in *Proceedings of the 2nd ACM SIGHT International Health Informatics Symposium*, New York, NY, USA, January 2012.
- [6] J. T. Polletini, S. R. G. Panico, J. C. Daneluzzi, R. Tinós, J. A. Baranauskas, and A. A. Macedo, "Using machine learning classifiers to assist healthcare-related decisions: classification of electronic patient records," *Journal of Medical Systems*, vol. 36, no. 6, pp. 3861–3874, 2012.
- [7] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 544–551, 2011.
- [8] Q. Li, H. Peng, J. Li, S. Lichao, S. Y. Philip, and H. Lifang, "A survey on text classification: from shallow to deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 1–21, 2020.
- [9] K. Kowsari, K. J. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: a survey," *Information*, vol. 10, no. 4, p. 150, 2019.
- [10] M. Hughes, I. Li, S. Kotoulas, and T. Suzumura, "Medical text classification using convolutional neural networks," *Studies in Health Technology and Informatics*, vol. 235, pp. 246–250, 2017.
- [11] G. Mujtaba, "Clinical text classification research trends: systematic literature review and open issues," *Expert Systems with Applications*, vol. 116, pp. 494–520, 2019.
- [12] L. Qing, W. Linhong, and D. Xuehai, "A novel neural network-based method for medical text classification," *Future Internet*, vol. 11, no. 12, p. 255, 2019.
- [13] Y. Wang, S. Sohn, S. Liu et al., "A clinical text classification paradigm using weak supervision and deep representation,"

- BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 1, 2019.
- [14] L. Yao, C. Mao, and Y. Luo, "Clinical text classification with rule-based features and knowledge-guided convolutional neural networks," *BMC Medical Informatics and Decision Making*, vol. 19, no. S3, p. 71, 2019.
- [15] Z. Shen and S. Zhang, "A novel deep learning-based model for medical text classification," in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, New York; NY, USA, October 2020.
- [16] L. Yao, Y. Zhang, B. Wei, Z. Li, and X. Huang, "Traditional Chinese medicine clinical records classification using knowledge-powered document embedding," in *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1926–1928, IEEE, Piscataway, NJ, USA, December 2016.
- [17] S. Baker, A. Korhonen, and S. Pyysalo, "Cancer hallmark text classification using convolutional neural networks," in *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*, pp. 1–9, Osaka, Japan, December 2016.
- [18] X. Li and M. Cui, "A Hybrid Medical Text Classification Framework: Integrating Attentive Construction and Neural Network," *Neurocomputing*, vol. 443, 2021.
- [19] J. Liu, R. Bai, Z. Lu, P. Ge, U. Aickelin, and D. Liu, "Data-driven regular expressions evolution for medical text classification using genetic programming," in *Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8, Glasgow, UK, July 2020.
- [20] K. Glinka, R. Wozniak, and D. Zakrzewska, "Improving multi-label medical text classification by feature selection," in *Proceedings of the 2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pp. 176–181, Poznan, Poland, June 2017.
- [21] X. Zhang, R. Henao, Z. Gan, Y. Li, and L. Carin, "Multi-label learning from medical plain text with Convolutional residual models," *Computer Science*, vol. 9, 2018.
- [22] M. Abdollahi, X. Gao, Y. Mei, S. Ghosh, and J. Li, "An ontology-based two-stage approach to medical text classification with feature selection by Particle Swarm optimisation," in *Proceedings of the 2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 119–126, Wellington, New Zealand, June 2019.
- [23] K. Liu and L. Chen, "Medical social media text classification integrating consumer health terminology," *IEEE Access*, vol. 7, pp. 78185–78193, 2019.
- [24] M. Khachidze, M. Tsintsadze, and M. Archuadze, "natural language processing based instrument for classification of free text medical records," *BioMed Research International*, vol. 2016, Article ID 8313454, 10 pages, 2016.
- [25] A. Ollagnier and H. Williams, "Text augmentation techniques for clinical case classification," *CLEF*, vol. 22–25, 2020.
- [26] Z. Yang, M. Dehmer, O. Yli-Harja, and F. Emmert-Streib, "Combining deep learning with token selection for patient phenotyping from electronic health records," *Scientific Reports*, vol. 10, no. 1, p. 1432, 2020.
- [27] A. N. Jagannatha and H. Yu, "Bidirectional RNN for medical event detection in electronic health records," in *Proceedings of the Conference. Association for Computational Linguistics*, p. 473, Stroudsburg, PA, USA, July 2016.
- [28] Z. Che, D. Kale, W. Li, M. T. Bahadori, and Y. Liu, "Deep computational phenotyping," in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 507–516, New York, NY, USA, August 2015.
- [29] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," in *Proceedings of the International Conference on Learning Representations (ICLR)*, Vancouver, BC, Canada, May 2016.
- [30] J. Geraci, P. Wilansky, V. de Luca, A. Roy, J. L. Kennedy, and J. Strauss, "Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression," *Evidence-Based Mental Health*, vol. 20, no. 3, pp. 83–87, 2017.
- [31] G. R. Venkataraman, A. L. Pineda, O. J. Bear Don't Walk IV et al., "FasTag: automatic text classification of unstructured medical narratives," *PLoS One*, vol. 15, no. 6, Article ID e0234647, 2020.
- [32] H. J. Lee, H. Xu, J. Wang, Y. Zhang, S. Moon, and J. Xu, "UTHealth at SemEval-2016 task 12: an end-to-end system for temporal information extraction from clinical notes," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 1292–1297, San Diego, CA, USA, June 2016.
- [33] C. W. Chen, S. P. Tseng, T. W. Tang, and J. F. Wang, "Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital," *Information*, vol. 11, no. 2, 2020.
- [34] J. Pennington, R. Socher, and C. D. Manning, "GloVe: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, October 2014.
- [35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [36] S. Lai, L. Xu, and K. Liu, "Recurrent convolutional neural networks for text classification," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2015)*, pp. 2267–2273, Austin, TX, USA, January 2015.
- [37] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, Doha, Qatar, October 2014.
- [38] C. Zhang, D. Wang, L. Wang et al., "Temporal data-driven failure prognostics using BiGRU for optical networks," *Journal of Optical Communications and Networking*, vol. 12, no. 8, pp. 277–287, 2020.
- [39] K. Hu, Z. Zhang, X. Niu et al., "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, 2018.
- [40] E. Wu, D. Talbot, and F. Moiseev, "Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5797–5808, Florence, Italy, July 2019.

Research Article

A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning

Baha Ihnaini ¹, **M. A. Khan** ², **Tahir Abbas Khan**,³ **Sagheer Abbas** ³,
Mohammad Sh. Daoud,⁴ **Munir Ahmad** ³, and **Muhammad Adnan Khan** ⁵

¹Department of Computer Science, College of Science and Technology, Wenzhou-Kean University, Wenzhou 325060, China

²Riphah School of Computing and Innovation, Faculty of Computing, Riphah International University, Lahore Campus, Lahore 54000, Pakistan

³School of Computer Science, National College of Business Administration and Economics, Lahore 54660, Pakistan

⁴College of Engineering, Al Ain University, Abu Dhabi 112612, UAE

⁵Pattern Recognition and Machine Learning Lab, Department of Software, Gachon University, Seongnam 13557, Republic of Korea

Correspondence should be addressed to Munir Ahmad; munir@ncbae.edu.pk and Muhammad Adnan Khan; adnan@gachon.ac.kr

Received 12 June 2021; Revised 9 July 2021; Accepted 6 September 2021; Published 17 September 2021

Academic Editor: Ahmed A. Abd El-Latif

Copyright © 2021 Baha Ihnaini 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 prediction of human diseases precisely is still an uphill battle task for better and timely treatment. A multidisciplinary diabetic disease is a life-threatening disease all over the world. It attacks different vital parts of the human body, like Neuropathy, Retinopathy, Nephropathy, and ultimately Heart. A smart healthcare recommendation system predicts and recommends the diabetic disease accurately using optimal machine learning models with the data fusion technique on healthcare datasets. Various machine learning models and methods have been proposed in the recent past to predict diabetes disease. Still, these systems cannot handle the massive number of multifeatures datasets on diabetes disease properly. A smart healthcare recommendation system is proposed for diabetes disease based on deep machine learning and data fusion perspectives. Using data fusion, we can eliminate the irrelevant burden of system computational capabilities and increase the proposed system's performance to predict and recommend this life-threatening disease more accurately. Finally, the ensemble machine learning model is trained for diabetes prediction. This intelligent recommendation system is evaluated based on a well-known diabetes dataset, and its performance is compared with the most recent developments from the literature. The proposed system achieved 99.6% accuracy, which is higher compared to the existing deep machine learning methods. Therefore, our proposed system is better for multidisciplinary diabetes disease prediction and recommendation. Our proposed system's improved disease diagnosis performance advocates for its employment in the automated diagnostic and recommendation systems for diabetic patients.

1. Introduction

A recent development in biotechnologies and high throughout computing progressively contribute to quick and affordable e-healthcare data collections and disease diagnosis. The efficiency and reliability are dependent on accurate model building from e-healthcare big data. One of many life-threatening diseases is diabetes disease (DD) [1–3].

Diabetes disease arrests 422 million adults all over the world [4]. The death rate due to diabetes disease is 1.5 million, and 3.7 million deaths are due to diabetes and high blood pressure [4]. Diabetes disease is a multidisciplinary disease that arrests the human body's significant parts like the kidney, eyes, lungs, and heart. The diagnosis of this disease was done either manually by a medical practitioner or by any automatic device.

All of these types of measurements for diabetes disease have some benefits and some drawbacks. Any experienced medical expert cannot manually find the diabetes disease early due to some hidden side effects on the human body. With the intelligent recommendation system's help and application of deep machine learning (DML) and artificial intelligent methods, this disease can be predicted at the earlier stage [5–9] with a minimal error rate.

There are some healthcare automated systems for detection and recommendation of human diseases in recent researches. Myocardial infarction [10] is an acute disease for blood circulation in the heart. In this paper, deep CNN is applied for detection and to prevent humans from a heart attack. A computer-aided diagnosis (CAD) system [11] is used by applying the transfer learning technique for accurate and timely response to reduce the extensive calculation. In this paper, a CAD system works efficiently and accurately to detect and prevent heart attacks. Internet of Healthcare Things (IoHT) and Decentralization Interoperable Trust (DIT) [12] framework are a better healthcare system. In this paper, blockchain is used for data privacy and security. In this research, data is collected via IoHT at each point and transformed through blockchain for smooth and accurate healthcare data for better system accuracy.

Many ensemble learning models have been used in recent healthcare researches for better accuracy. For example, hepatocellular carcinoma [13] is a hazardous cancer disease in the human body. In this paper, an automated prediction system is developed using a stack learning approach for deep learning and examining healthcare data about this deadly disease. Stack learning is an ensemble learning technique. In this paper, evolutionary computational techniques are also used to examine the healthcare data about hepatocellular carcinoma disease. In cervical cancer [14] diagnosis, an ensemble machine learning approach is used. In this paper, two approaches are used for predicting disease on an images dataset.

In mobile edge computing [15], an automated recommendation system is proposed for the joint computation of multiuser offloading and task caching. In this paper, Q-learning and Deep-Q-Network-based algorithms are proposed for this system. Multilevel vehicular edge cloud computing [16], secure federated learning for 5G [17], and augmented Coronavirus disease detection [18] used an advanced ensemble deep learning approach for better results.

For detection of COVID-19 disease, a deep learning approach is adopted with an augmented approach [18] and it achieved 100% accuracy. In industrial mobile edge computing [19], the deep ensemble learning approach is used for resource allocation and data security.

The importance of a smart healthcare recommendation system is increasing day by day for better and timely prediction. To minimize the risk of life-threatening human diseases, we need an efficient system for diagnosing and effectively recommending life-threatening diseases such as diabetes. Electronic health records (EHRs) play an essential role in smart healthcare recommendation systems for predicting life-threatening diseases, especially for multidisciplinary and life-threatening diabetes diseases. However, the data collected from sensors and EHRs are unstructured. To

manage adequately such kinds of multisourced data for further examining is a challenging task.

Further, extracting the critical features and fusing them in a structured form is also a hectic and skills-demanding task. Therefore, this section is further divided into two parts: a wearable sensors-based diabetes prediction system and extracting information from EHRs textual data. Then, data fusion is essential for better results and accurate prediction of diabetes disease with DML.

Many recommendation systems for healthcare are already proposed in recent researches. The significant contribution of this research is to enrich the healthcare dataset for the best prediction of multidisciplinary diabetes disease. We have collected the patients' data through wearable sensors and EHRs in the textual record form of each patient. After collecting the records of each patient, essential data from both ends are fused to enrich the healthcare dataset. The Ensemble deep learning approach works accurately and produces better results in larger healthcare datasets. Finally, we have developed a better recommendation system by collecting patients' records and applying an ensemble machine learning approach for accurate and timely prediction and recommendation of multidisciplinary diabetes disease patients.

The organization of the paper is as follows: Section 2 describes the most recent developments of diabetes disease detection and recommendation from the literature; Section 3 provides research methodology, data fusion, and proposed DML model; Section 4 presents the dataset selection, pre-processing, data fusion, and results and discussion; Section 5 describes the conclusion and future work; and the last section is devoted to references.

2. Related Work

Many researchers contributed to diagnosing diabetes disease. They have used machine learning (ML) classifier and artificial intelligence (AI) assistance for the prediction of diabetes disease. With the help of artificial intelligence, we can easily collect healthcare data. After collecting the big data from the healthcare center, we can easily predict human diseases, including multidisciplinary diabetes diseases.

In early detection of diabetes disease [20], the k-nearest neighbor (KNN) classifier model was used and the result was compared with that of the support vector machine (SVM) model achieving 85.6% accuracy. In this paper, the KNN machine learning classifier was used for predicting diabetes disease. The results comparison was made with another machine learning classifier called SVM to authenticate the work. For the detection of diabetes type II disease [21], the authors used a convolution neural network (CNN) and compared their work with the linear regression (LR) model and multilayer perceptron (MLP). In this paper, the neural network was used to diagnose diabetes type II disease. For results comparison, two machine learning classifiers were used for the authenticity of the work. An accuracy of 77.5% was achieved in the area under the curve (AUC). Analysis of early detection of diabetes disease with feature selection technique [22] was carried out using SVM classifier and their results were compared with random forest (RF), naïve Bayes (NB), decision tree (DT), and KNN classification models. The highest

accuracy achieved with SVM was 77.73%. In this paper, the feature selection technique was adopted. With the help of the feature selection technique, we can reduce the system's computational capability and improve accuracy. Multiple machine learning models were applied for comparison and authenticity of results. Bloodless techniques for the prediction of diabetes disease are used with computational tools [23]. The accuracy achieved through this technique was 91.67%. In diabetic retinopathy detection [24], the deep (DNN) technique was adopted. The accuracy achieved via CNN was 74.4%. Detection of multiclass retinal disease [25] was done with with AI. The CNN classifier was used and it achieved 92% accuracy.

A data-driven approach is used for predicting diabetes and cardiovascular diseases [26] with ML. This paper adopted an extreme gradient boost and compared it with the LR, SVM, RF, and weighted ensemble model. An accuracy of 95.7% was achieved in the area under the ROC curve. In type II diabetes disease prediction [27], an ensemble classification model was adopted. The accuracy achieved via this model was 82.2% in the AUC. A new methodology, smartphone-based diabetes detection [28], was presented. In this paper, image data was considered for diagnoses and further directions. A microcontroller-based agent [29] was used to measure the blood glucose level of patients. A sensor integrated therapy [30] for diabetes disease was used to monitor glucose levels in a diabetic patient. A self-recommendation smart app [31, 32] was used and trained on recorded health data like patients' daily physical activities and other important parameters related to diabetes.

The valuable information extraction from wireless sensor data and the patient's electronic medical record is also challenging for predicting multidisciplinary disease. To handle this challenge, different models have been presented for extracting the most valuable information from the healthcare textual data [33–35] for making a dataset for the prediction of diabetic disease. The textual dataset collected through EHR's was pre-processed and converted into a meaningful format as per smart healthcare recommendation system demands. The wireless sensor data of healthcare was also collected through wireless devices. After collecting data through a wireless communication device, data was preprocessed for removing noisy wireless data to make a rich and accurate healthcare dataset. In this way, we can easily apply machine learning algorithms for predicting multidisciplinary diabetes disease.

After collecting and converting meaningful data from textual data and fusing preprocessed wireless sensor data [36–38] for making a rich healthcare dataset of diabetes, Table 1 shows the comprehensive limitation of previously published approaches.

3. Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients

This section explains the overall structure of a smart healthcare recommendation system for multidisciplinary diabetes disease patients (SHRS-M3DP) in detail. The proposed approach is divided into distinct layers for an

accurate description of each layer working. In the end, the ensemble DML structure is presented, which is further deployed in the whole SHRS-M3DP to predict and recommend diabetes disease in the patients. It should deliver a concise and accurate representation of the experimental results, explanation, and the conclusion of experiments that can be drawn.

3.1. Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients. This part explains the overall structure of a smart healthcare recommendation system for multidisciplinary diabetes disease patients (SHRS-M3DP). Initially, the general structure of the proposed SHRS-M3DP is described. Then, a proposed system's structure is divided into distinct layers for an accurate description of each layer working. Finally, the ensemble deep learning model structure is presented, and it is further deployed in the whole SHRS-M3DP to predict and recommend diabetes disease in the patients.

The proposed structure of the SHRS-M3DP system is exhibited in Figure 1. It is divided into two main segments: (1) training phase and (2) validation phase. These phases are essential for the accurate prediction of multidisciplinary diabetes disease. These phases communicate via a cloud. The training phase comprises seven levels: (i) sensory layer, (ii) EHRs layer, (iii) raw feature layer-1, (iv) raw feature layer-2, (v) fused raw feature layer, (vi) preprocessing layer, and (vii) application layer. The sensory layer comprises input parameters, including age, family history, glucose, skin thickness, blood pressure (BP), pregnancies, insulin, and body mass index (BMI). The sensory layer's input values are collected and transferred to the database, raw feature layer-1, through the Internet of medical things (IoMT). Because wireless communication is applied, data collected from multiple feature nodes and stored in the database may be inaccurate. For that reason, we considered such kind of data as feature raw data. EHRs layer consists of lab reports, questions, observation, and the patient's medical history. All the data collected from the EHRs layer are reports and need some methodology to convert them into a structured format for further processing.

The Framingham risk factors (FRFs) methodology used in the smart healthcare monitoring system for heart disease prediction [39] is adopted to extract data from the EHRs, as shown in Figure 2, and stored in raw feature layer-2.

The data fusion approach is then applied for fusing the common features of both sensory data and EHRs to generate enhanced healthcare data on multidisciplinary diabetes disease.

These fused feature data are then stored in the fused feature layer for further processing to predict diabetes disease.

The following preprocessing layer plays a crucial role in the model. All deficiencies received through the sensory layer previously via wireless communication and EHRs layer are preprocessed in this phase. These missing values are managed by applying moving averaging and normalization methodology to mitigate noisy results. Subsequently, after

TABLE 1: Summary of existing literature reviews.

Paper	Classification methodology adopted	Limitations	Advantages
[12]	ML	<ol style="list-style-type: none"> 1. Single dataset 2. No data fusion 3. Only structured data 	Only the optimal feature selection technique was adopted
[13]	ML and AI	<ol style="list-style-type: none"> 1. Single dataset 2. No data fusion 3. Only DR image data 	The bloodless technique was adopted
[18]	DML and generalized linear model	<ol style="list-style-type: none"> 1. Single EHRs dataset 2. No data fusion 	<ol style="list-style-type: none"> 1. Electronic health record 2. Data fusion 3. Feature selection
[19]	AI	<ol style="list-style-type: none"> 1. Single dataset 2. No data fusion 3. Only DR image data 	<ol style="list-style-type: none"> 1. Automated software 2. Smartphone-based DR and sight-threatening detection
[20]	AI and ML	<ol style="list-style-type: none"> 1. Single dataset 2. No data fusion 	Incorporating wearable devices and IoT to collect and manage big data
[29]	Supervised ML	<ol style="list-style-type: none"> 1. Single dataset 2. No data fusion 	<ol style="list-style-type: none"> 1. Combine structured and unstructured data 2. Feature selection

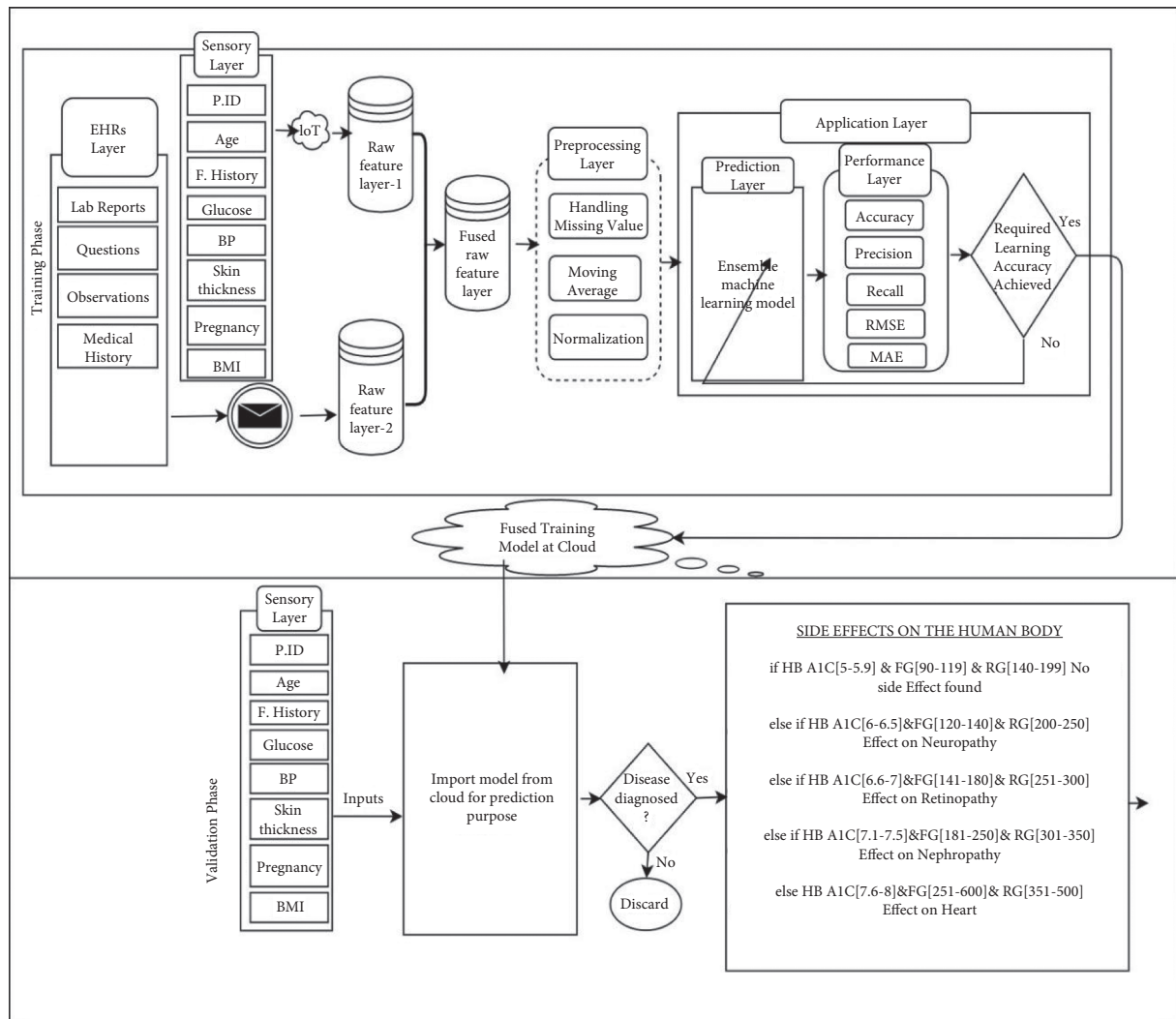


FIGURE 1: Proposed SHRS-M3DP model.

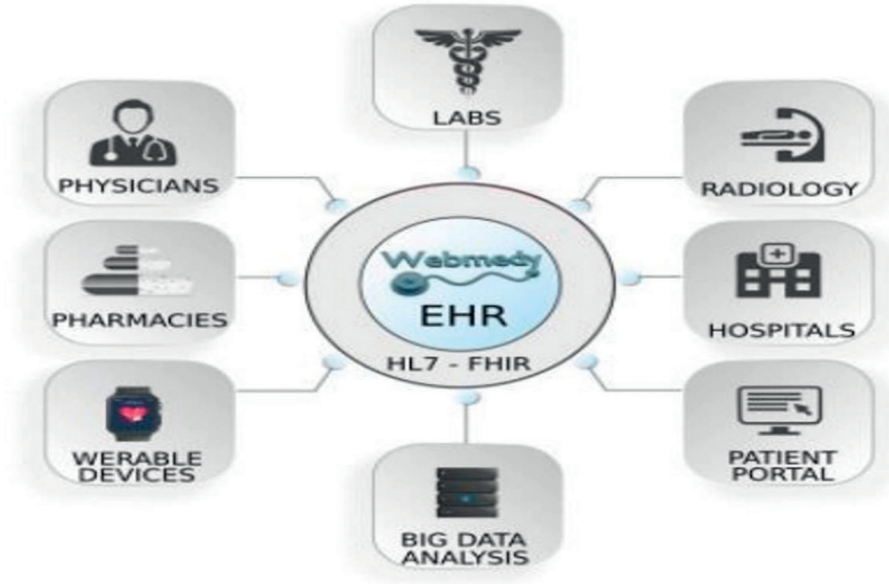


FIGURE 2: EHRs record conversion flow chart [40].

preprocessing, the fused feature dataset is forwarded to the application layer. This layer is further divided into two sublayers: (i) prediction layer and (ii) performance layer. In the prediction layer, an ensemble DML model is applied to predict multidisciplinary diabetes disease.

The ensemble deep learning combines several individual models to obtain better generalization performance of any predictive classification problem. The convergence process of the ensemble ML model is implemented in three ways: (1) max voting, (2) averaging, and (3) weighted average for classification.

The ensemble ML model used an advanced boosting technique for regularizing, limiting the overfitting issue, and producing better accuracy compared to other ML models. As a result, the response rate of the ensemble ML model is ten times faster compared to other ML models. In the ensemble ML model's boosting technique, the trained dataset is divided into multiple weak learners. The average error rate of one weak learner is updated in the next weak learner. Resultantly, the final strong learner was found to have a minimal error rate for prediction.

Ensemble deep machine learning classifier can be expressed as follows:

$$F(x) = \sum_m \alpha_m h_m(x), \quad (1)$$

where $F(x)$ is a strong learner of ensemble classifiers, α_m is weight calculated by considering the last iteration's error, and $h_m(x)$ is a weak learner

In this way, we can achieve maximum prediction accuracy. The operational flow of the advanced boosting ensemble DML model is shown in Figure 3 (Algorithm 1).

The results are then sent to the performance layer. In this layer, data received from the previous layer is calculated. The performance layer results are evaluated based on accuracy, precision, recall, F1, root mean square error (RMSE), and

mean average error (MAE) achieved by the SHRS-M3DP model. After comparing results, "YES" indicates that our proposed SHRS-M3DP model successfully predicted diabetes disease, and "NO" means the prediction layer of the proposed SHRS-M3DP model will be modified till the learning criteria objectives are attained. After successfully training the proposed SHRS-M3DP model, the trained fused model moves to a cloud to further import and predict diabetes disease.

The validation phase comes in the last where trained fused SHRS-M3DP model is imported for prediction to authenticate whether the patient is affected with multidisciplinary diabetes disease based on the results.

The results are then sent to the next layer, called the performance layer. In this layer, the data received from the prediction layer is evaluated.

The performance layer results are then evaluated based on accuracy, precision, recall, F1, root mean square error (RMSE), and mean average error (MAE) achieved by the SHRS-M3DP model. After comparing the results, "YES" indicates that our proposed SHRS-M3DP model successfully predicted diabetes disease, and "NO" means the prediction layer of the proposed SHRS-M3DP model will be updated until the learning criteria are achieved. After successfully training the proposed SHRS-M3DP model, the trained fused model moves to a cloud to further import and predict diabetes disease.

The last phase is the validation phase. In this phase, the trained fused SHRS-M3DP model is imported for recommendation to validate whether the patient is affected with multidisciplinary diabetes disease based on the results.

4. Experiments

The data collected from EHRs and sensors were discussed previously in the proposed SHRS-M3DP model. In addition, the fused feature database was also discussed in the last

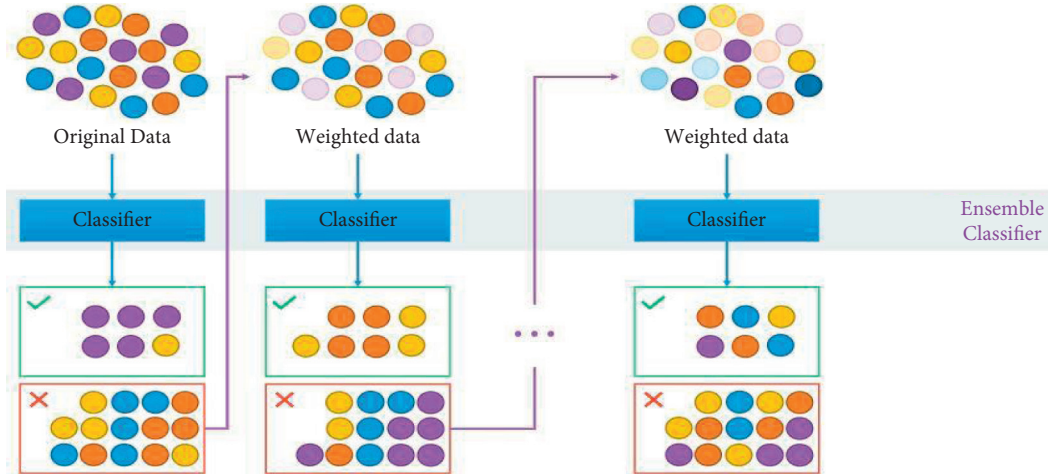


FIGURE 3: Flow chart of ensemble ML algorithm [41].

Input: training set $\{(x_i - y_i)\}_{i=1}^n$ and a differentiable **loss Function** $L(y_i, F(x))$, number of iterations M .

Output: A targeted person affected by diabetes disease or safe from the diabetes disease

Step 1:

- (1) Initialize model with a constant value: $F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma)$
- (2) For $m = 1$ to M :
 - (1) Compute so-called pseudo-residuals: $r_{im} = -[\sigma L(y_i, F(x_i))/\sigma F(x_i)]_{F(x)=F(x_{m-1}(x))}$ for $i = 1, \dots, n$
 - (2) Fit a base learner (or weak learner, i.e., tree) $h_m(x)$ to (pseudo residuals), train it using the training set $\{(x_i - y_i)\}_{i=1}^n$
 - (3) Compute multiplier r_m by solving the following one dimensional optimization problem : $r_m = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$
 - (4) Update the model: $F_m(x) = F_{m-1}(x) + r_m h_m(x)$
- (3) Output $F_m(x)$

ALGORITHM 1: Ensemble ML-based diabetes prediction.

section. In this section, the proposed SHRS-M3DP model's performance is evaluated, and the results are discussed.

4.1. Dataset. The proposed SHRS-M3DP model is simulated with two different diabetes disease datasets: Hospital Frankfurt Germany diabetes dataset [42] and Pima Indians diabetes dataset. The Hospital Frankfurt Germany diabetes dataset consists of 2000 cases with eight features. The Pima Indians diabetes dataset consists of 768 patients with eight features. The fused features dataset for an experiment was made with the combination of features of both datasets. Both datasets' cases with some missing values are managed with proposed filtering, and normalization techniques were discussed earlier. The combined, fused dataset features are shown in Table 2. The fused features attributes, measuring units, and their ranges are also mentioned.

The fused features dataset consists of 2768 cases with eight features. A deep machine learning model cannot be utilized for the small dataset having nominal values. Therefore, all the nominal data is converted into numeric values for utilizing the ensemble deep learning model. Detailed features description of fused features is shown in Table 2.

TABLE 2: Feature information about the diabetes disease.

Sr. no.	Attribute	Unit	Ranges
1	Age	Year	01–120
2	Family history	Yes (1), no (0)	0, 1
3	Glucose	mg/Dl	37–380
4	Skin thickness	Mm	0–210
5	Blood pressure (BP)	mm Hg	90–190
6	Pregnancies	Number (0–9)	0–8
7	Insulin	uU/ml	0–764
8	BMI	Kg/m ²	14–80.6
9	Diagnosis result	Positive (1), negative (0)	1, 1

4.2. Performance Evaluation. The experiment was carried out to indicate the proposed SHRS-M3DP model's performance for diagnosing diabetic disease. Initially, the data was collected from sensors, which were transferred through IoMT to the feature database. Similarly, the patients' data collected through lab reports, questions, observations, and medical history were converted from unstructured format to structure format for further preprocessing. After collecting the features from sensors and EHRs, both datasets' features were combined to make a rich health dataset for better prediction and recommendation of diabetes disease. Finally,

the processing module analyzed the final combined, fused feature dataset for further processing.

Furthermore, the Hospital Frankfurt Germany diabetes dataset and Pima Indians diabetes dataset were then utilized for training the diabetes disease prediction model. For evaluation purposes, the proposed ensemble deep learning model was compared with some other classifiers: SVM, LR, KNN, NB, RF, and DT. The proposed SHRS-M3DP model was used before and after the feature selection and performance was compared. The datasets were divided randomly into 80% and 20%, respectively, to train and test the models mentioned above in the proposed model.

4.3. Evaluation Metrics. Dissimilar evaluation metrics were used to conclude the model's overall efficiency, as shown in Table 3. With the accuracy metric's help, we can present the proposed deep learning model's overall predictive ability. In the confusion matrix, true positive (TP) and true negative (TN) determine the proposed classifier's capability to predict the absence and presence of diabetes disease. false negative (FN) and false positive (FP) identify the proposed model's total false prediction. Recall metric and precision metric calculate the sensitivity and success of the diabetes disease presented model individually.

The function measure (FM) metric is used for prediction accuracy. Root mean square error (RMSE) and mean absolute error (MAE) calculate the difference and absolute variations among the predicted and the actual values. The values of y_i , \hat{y}_i denote the total numbers of observations of the predicted values and the actual values, respectively.

4.4. Results. This section presents the results of the above-mentioned proposed model and a comparison with other classifiers, respectively. The complete details of all classifiers for diabetes prediction are divided into three parts: prediction of diabetes disease, Pima Indians diabetes dataset consisting of 768 cases with eight features, Hospital Frankfurt Germany diabetes dataset consisting of 2000 patients with eight features, and finally with fused features dataset having 2768 cases with eight features as shown in Table 2, respectively.

The proposed SHRS-M3DP ensemble deep learning model prediction accuracy with other baseline classifiers is shown in Figure 4. The comprehensive explanation of each classifier before data fusion and after data fusion is as follows:

- (i) Learner regression classifier (LR): LR accuracy in dataset 1 is 74.6% for predicting diabetes disease with 786 cases. In dataset 2, LR performed better, with 77.7% accuracy with 2000 cases. Still, in the final data fusion dataset, the accuracy for LR's prediction of diabetes disease is decreased to 75.2% with 2786 cases.
- (ii) Naïve Bayes (NB): NB classifier's accuracy in dataset 1 is 72% for predicting diabetes disease with 786 cases. In dataset 2, NB performed better, with 76.5% accuracy with 2000 cases. Still, in the final data

fusion dataset, the accuracy for LR's prediction of diabetes disease is decreased to 74% with 2786 cases.

- (iii) Random forest (RF): RF classifier's accuracy in dataset 1 is 74.8% for predicting diabetes disease with 786 cases. In dataset 2, RF also performed better, with 81.2% accuracy with 2000 cases. Still, in the final data fusion dataset, the accuracy of random forest's prediction of diabetes disease is decreased to 80.5% with 2786 cases.
- (iv) K-nearest neighbor (KNN): KNN classifier's accuracy in dataset 1 is 73.3% for predicting diabetes disease with 786 cases. In dataset 2, KNN also performed better, with 77.7% accuracy with 2000 cases. In the final data fusion dataset, the accuracy of predicting diabetes disease with KNN is also increased up to 80.8% with 2786 cases.
- (v) Decision tree (DT): DT classifier's accuracy in dataset 1 is 74% for predicting diabetes disease with 786 cases. In dataset 2, DT performed better, with 83.7% accuracy with 2000 cases. In the final data fusion dataset, the accuracy of prediction of diabetes disease with DT is 84.3% with 2786 cases.
- (vi) Support vector machine (SVM): SVM classifier's accuracy in dataset 1 is 74.6% for predicting diabetes disease with 786 cases. In dataset 2, SVM performed better, with 84% accuracy with 2000 patients. In the final data fusion dataset, the accuracy for predicting diabetes disease with SVM is 84.3% with 2786 cases.

The proposed ensemble deep machine learning model performed outstandingly as compared to all the baseline classifiers. The proposed ensemble DML classifier's accuracy in dataset 1 is 72.7% for predicting diabetes disease with 786 cases. However, it is low due to the small dataset. In dataset 2, ensemble ML performed better and achieved 91% accuracy with 2000 cases, higher than all other classifiers. In the final data fusion dataset, the accuracy of prediction of diabetes disease with the proposed ensemble ML model is 99.6%, having a minimal error rate.

The summary of performance metrics, accuracy, precision, recall, F1, root mean square error, and mean absolute error of selected datasets individually and data fusion datasets, is presented in Tables 4–6.

Different DML classifiers are compared with the proposed model by various evaluation metrics, as shown in Table 4. In this experiment, only Pima Indians diabetes dataset is considered, without feature selection technique. The performance of each metric on a given dataset is precisely shown in Table 4. The proposed model's overall performance is less compared to the other classifiers due to the small dataset and the absence of feature selection technique.

In Table 5, only the Hospital Frankfurt Germany diabetes dataset is considered, without feature selection technique. Different DML classifiers are compared with the proposed model by various evaluation metrics, as shown in Table 5.

TABLE 3: Performance metrics.

Name of metric	Description	Equation #
Accuracy (Acc)	$(TP + TN) / (TP + TN + FP + FN)$	(a)
Precision (Pre)	$TP / (TP + FP)$	(b)
Recall (Rec)	$TP / (TP + FN)$	(c)
F1-measure (FM)	$((2 * Pre * Rec) / (Pre + Rec))$	(d)
RMSE	$\sqrt{1/N \sum_{i=1}^N (y_i - \hat{y}_i)^2}$	(e)
MAE	$\sqrt{1/N \sum_{i=1}^n y_i - \hat{y}_i }$	(f)

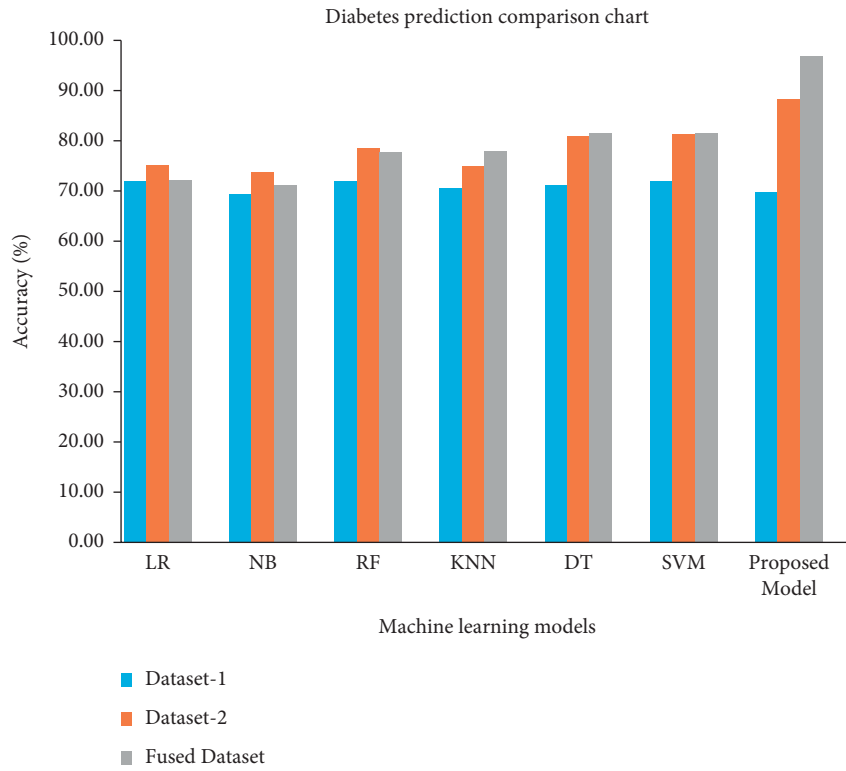


FIGURE 4: Accuracy of models before and after data fusion.

TABLE 4: Comparison results of the proposed model with other classifiers before data fusion on the Pima Indians diabetes dataset.

Classifier model	Acc (%)	Pre (%)	Rec (%)	FM (%)	RMSE	MAE
Logistic regression	74.68	0.68	0.52	0.59	0.25	0.50
Naïve Bayes	72.08	0.62	0.52	0.57	0.28	0.53
Random forest	74.68	0.69	0.50	0.58	0.25	0.50
K-nearest neighbors	73.38	0.67	0.48	0.56	0.27	0.52
Decision tree	74.03	0.63	0.63	0.63	0.26	0.51
Support vector machine	74.68	0.70	0.48	0.57	0.25	0.50
*Proposed model	72.73	0.63	0.56	0.59	0.27	0.52

TABLE 5: Comparison results of the proposed model with other classifiers before data fusion on the Hospital Frankfurt Germany diabetes dataset.

Classifier model	Acc (%)	Pre (%)	Rec (%)	FM (%)	RMSE	MAE
Logistic regression	77.75	0.71	0.58	0.64	0.22	0.47
Naïve Bayes	76.50	0.67	0.61	0.64	0.24	0.48
Random forest	81.25	0.75	0.69	0.71	0.19	0.43
K-nearest neighbors	77.75	0.71	0.59	0.65	0.22	0.47
Decision tree	83.75	0.73	0.83	0.78	0.16	0.40
Support vector machine	84.00	0.79	0.73	0.76	0.16	0.40
*Proposed model	91.00	0.89	0.84	0.86	0.09	0.30

TABLE 6: Comparison results of the proposed model with other classifiers after data fusion on a fused dataset.

Classifier model	Acc (%)	Pre (%)	Rec (%)	FM (%)	RMSE	MAE
Logistic regression	75.27	0.67	0.55	0.61	0.25	0.50
Naïve Bayes	74.01	0.64	0.55	0.59	0.26	0.51
Random forest	80.51	0.75	0.65	0.70	0.19	0.44
K-nearest neighbors	80.87	0.77	0.63	0.70	0.19	0.44
Decision tree	84.30	0.77	0.78	0.77	0.16	0.40
Support vector machine	84.30	0.81	0.71	0.76	0.16	0.40
*Proposed model	99.64	1.00	0.99	0.99	0.00	0.06

TABLE 7: Summary of comparison using diabetes datasets with existing methods.

Sr. no.	Author/year	Fusion method	Classifier/compared with classifier	Overall accuracy
1	Mohebbi et al. [10]/2017	—	CNN/LR, MLP	77.5%
2	Sneha and Gangil [11]/2019	Yes/—	SVM/RF, NB, KNN, DT	77.7%
3	Nguyen et al. [16]/2019	—	Ensemble ML/—	82.2%
4	Aminah and saputro [9]/2019	—	KNN/SVM	85.6%
5	Sah and sarma [12]/2018	—	AI	91.6%
6	Dinh et al. [15]/2019	—	XGB/LR, SVM, RF, weight ensemble model	95.7%
7	Proposed method	Yes/yes	Ensemble deep machine learning classifier/LR, NB, DT, KNN, RF, SVM	99.6%

The proposed model’s overall performance is more outstanding compared to the other classifiers due to the large dataset, as shown in Figure 4. Therefore, our proposed deep machine learning model can achieve more accurate results concerning a higher dataset ratio. In this experiment, the proposed model achieved 91% accuracy, much higher compared to other DML classifiers. On the other hand, in this experiment, RMSE and MAE are also very low compared with different DML classifiers’s RMSE and MAE.

In Table 6, both diabetes datasets are being considered with the data fusion technique. Different DML classifiers are compared with the proposed model by various evaluation metrics, as shown in Table 3. The performance of each metric on a given dataset is precisely shown in Table 6. The proposed model’s overall performance is much higher compared to the other classifiers due to the fused technique by making a rich healthcare dataset. Our proposed model performed outstandingly due to the data fusion technique and produced extraordinary results. In this experiment, the proposed model achieved 99.64% accuracy, much higher compared to other DML classifiers. In this experiment, RMSE is 0%, and MAE is only 0.06%. As shown in Table 6, all other classifiers produce results less than 85%.

Our proposed model accurately performed well with the help of data fusion technique. Our proposed ensemble DML model has achieved higher accuracy compared to other studies done in the recent past. The details of recent studies on diabetes with their authors are also summarized and shown in Table 7. The significant contribution for higher accuracy in our model is due to data fusion. In this way, we have made a rich healthcare dataset for the prediction of multidisciplinary diabetes disease. In this way, we have achieved higher accuracy compared to other studies, which is 99.6%.

5. Conclusion

The prediction of human diseases, particularly multidisciplinary diabetes, is challenging for better and timely treatment. A multidisciplinary diabetes illness is a life-threatening disease worldwide which attacks major essential human body parts. A proposed SHRS-M3DP model is presented to predict and recommend multidisciplinary diabetes disease in the patients quickly and efficiently. The ensemble deep ML model and data fusion technique are used for fast response and better accuracy rate. The proposed model efficiently predicted and recommended whether the patient is a victim of multidisciplinary diabetes disease or not. The proposed SHRS-M3DP model can also identify the effect of human body parts: Neuropathy, Retinopathy, Nephropathy, or Heart. The proposed SHRS-M3DP model simulation is made by using Python language. Finally, the study of this research concluded that the proposed SHRS-M3DP model’s overall performance is 99.6%, which is outstanding compared to previously published approaches.

5.1. Contribution. Many recommendation systems for healthcare have already been proposed in recent researches. The significant contribution of this research is to enrich the healthcare dataset for the best prediction of multidisciplinary diabetes disease. We have collected the patients’ data through wearable sensors and EHRs in the textual record form of each patient. After collecting the records of each patient, essential data from both ends are fused to enrich the healthcare dataset. The ensemble deep learning approach works accurately and produces better results in larger healthcare datasets. Finally, we have developed a better recommendation system by collecting patients’ records and applying an ensemble machine learning approach for

accurate and timely prediction and recommendation for multidisciplinary diabetes disease patients. The overall performance of our recommendation system is 99.6%. In this way, future academic research and practices will be helpful for new researchers in this medical field, especially for automated prediction and recommendation systems for human diseases.

5.2. Future Work. The proposed SHRS-M3DP recommendation system achieved overall good performance. However, there is still a need to work for a better generalized efficient prediction and recommendation system for all human diseases. The complexity of the deep ensemble algorithm will also be considered in the near future for accurate and quick results of this algorithm.

Data Availability

The data used in this paper can be requested from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this work.

Authors' Contributions

B. I., M. A. K., and T. A. K. collected data from different resources. M. A. K. and T. A. K. performed formal analysis and Simulation, M. A. K., B. I., M. A., and M. S. D. contributed to writing—original draft preparation; M. A. K. and M. A. contributed to writing—review and editing. M. A. K. and S. A. supervised the paper. B. I., M. S. D., and T. A. K. drafted pictures and tables. M. A. K. and M. A. performed revision and improved the quality of the draft. All authors have read and agreed to the published version of the manuscript.

References

- [1] N. S. Artzi, S. Shilo, E. Hadar et al., "Prediction of gestational diabetes based on nationwide electronic health records," *Nature Medicine*, vol. 26, no. 1, pp. 71–76, 2020.
- [2] V. Raman, P. Then, and P. Sumari, "Proposed retinal abnormality detection and classification approach: computer-aided detection for diabetic retinopathy by machine learning approaches," in *Proceedings of the 2016 8th IEEE International Conference on Communication Software and Networks*, Beijing, China, June 2016.
- [3] J. Ramsingh and V. Bhuvaneshwari, "An efficient map reduce-based hybrid NBC-TFIDF algorithm to mine the public sentiment on diabetes mellitus—a big data approach," *Journal of King Saud University Computer and Information Sciences*, 2018, In press.
- [4] WHO, *Global Health Risks Report 2016*, <https://www.who.int/diabetes/global-report>, World Health Organisation, Geneva, Switzerland, 2020, <https://www.who.int/diabetes/global-report>.
- [5] N. Sharma and A. Singh, "Diabetes detection and prediction using machine learning/IoT: a survey," *Communications in Computer and Information Science*, vol. 18, pp. 471–479, 2018.
- [6] S. Afzali and O. Yildiz, "An effective sample preparation method for diabetes prediction," *The International Arab Journal of Information Technology*, vol. 15, no. 6, 2018.
- [7] N. Theera-Umpon, I. Poonkasem, S. Auephanwiriyaikul, and D. Patikulasila, "Hard exudate detection in retinal fundus images using supervised learning," *Neural Computing & Applications*, vol. 32, 2020.
- [8] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, 2018.
- [9] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford exercise testing (FIT) project," *PLoS One*, vol. 12, 2017.
- [10] M. Hammad, M. H. Alkinani, B. B. Gupta, and A. A. A. El-Latif, "Myocardial infarction detection based on the deep neural network on imbalanced data," *Multimedia Systems*, vol. 27, 2021.
- [11] A. Alghamdi, M. Hammad, H. Ugail et al., "Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities," 2020, <https://arxiv.org/abs/1906.09358>.
- [12] E. M. Abou-Nassar, A. M. Iliyasu, P. M. El-Kafrawy, O.-Y. Song, A. K. Bashir, and A. A. A. El-Latif, "DITrust chain: towards blockchain-based trust models for sustainable healthcare IoT systems," *IEEE Access*, vol. 8, pp. 111223–111238, 2020.
- [13] W. Książek, M. Hammad, P. Pławiak, U. Rajendra Acharya, and R. Tadeusiewicz, "Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection," *Biocybernetics and Biomedical Engineering*, vol. 40, pp. 1512–1524, 2020.
- [14] V. Chandran, M. G. Sumithra, A. Karthick et al., "Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images," *BioMed Research International*, vol. 2021, Article ID 5584004, 2021.
- [15] I. A. Elgendy, W.-Z. Zhang, H. He, B. B. Gupta, and A. A. Abd El-Latif, "Joint computation offloading and task caching for multi-user and multi-task MEC systems: reinforcement learning-based algorithms," *Wireless Networks*, vol. 27, no. 3, pp. 2023–2038, 2021.
- [16] M. Khayyat, I. A. Elgendy, A. Muthanna, A. S. Alshahrani, S. Alharbi, and A. Koucheryavy, "Advanced deep learning-based computational offloading for multilevel vehicular edge-cloud computing networks," *IEEE Access*, vol. 8, pp. 137052–137062, 2020.
- [17] Y. Liu, J. Peng, J. Kang, A. M. Iliyasu, D. Niyato, and A. A. A. El-Latif, "A secure federated learning framework for 5G networks," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 24–31, 2020.
- [18] A. Sedik, M. Hammad, F. E. Abd El-Samie, B. B. Gupta, and A. A. A. El-Latif, "Efficient deep learning approach for augmented detection of Coronavir disease," *Neural Computing & Applications*, vol. 8, 2021.
- [19] I. A. Elgendy, A. Muthanna, M. Hammoudeh, H. Shaiba, D. Unal, and M. Khayyat, "Advanced deep learning for resource allocation and security aware data offloading in industrial mobile edge computing," *Big Data*, vol. 9, pp. 1–14, 2021.

- [20] R. Aminah and A. H. Saputro, "Diabetes prediction system based on iridology using machine learning," in *Proceedings of the 2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, Semarang, Indonesia, September 2019.
- [21] A. Mohebbi, T. B. Aradottir, A. R. Johansen, H. Bengtsson, M. Fraccaro, and M. Mørup, "A deep learning approach to adherence detection for type 2 diabetics," in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, Korea, July 2017.
- [22] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *Journal of Big Data*, vol. 6, 2019.
- [23] P. Sah and K. K. Sarma, "Bloodless technique to detect diabetes using the soft computational tool," in *Ophthalmology IGI Global*, Hershey, PA, USA, 2018.
- [24] M. Arora and M. Pandey, "Deep neural network for diabetic retinopathy detection," in *Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 189–193, Faridabad, India, February 2019.
- [25] S. Karthikeyan, P. Sanjay Kumar, R. J. Madhusudan, S. Sundaramoorthy, and P.-K.-K. Namboori, "Detection of multi-class retinal diseases using artificial intelligence: an expeditious learning using deep CNN with minimal data," *Biomedical and Pharmacology Journal*, vol. 12, 2019.
- [26] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Medical Informatics and Decision Making*, vol. 19, 2019.
- [27] B. P. Nguyen, H. N. Pham, H. Tran et al., "Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records," *Computer Methods and Programs in Biomedicine*, vol. 182, 2019.
- [28] R. Rajalakshmi, R. Subashini, R. M. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," *Eye*, vol. 32, 2018.
- [29] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, and C.-H. Youn, "5G-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 16–23, 2018.
- [30] P. Choudhary, S. De Portu, A. Arrieta, J. Castañeda, and F. M. Campbell, "Use of sensor-integrated pump therapy to reduce hypoglycemia in people with type 1 diabetes: a real-world study in the UK," *Diabetic Medicine*, vol. 36, 2019.
- [31] A. Steinert, E. Steinhagen-Thiessen, and M. Haesner, "App-basiertes selbstmonitoring bei Typ-2-diabetes," *Zeitschrift für Gerontologie und Geriatrie*, vol. 50, no. 6, pp. 516–523, 2017.
- [32] R. Davoodi and M. H. Moradi, "Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier," *Journal of Biomedical Informatics*, vol. 79, 2018.
- [33] L. B. Fazlic, A. Hallawa, A. Schmeink, A. Peine, L. Martin, and G. Dartmann, "A novel NLP-fuzzy system prototype for information extraction from medical guidelines," in *Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, May 2019.
- [34] C. W. Song, H. Jung, and K. Chung, "Development of a medical big-data mining process using topic modeling," *Cluster Computing*, vol. 22, 2019.
- [35] E. A. Bernal, X. Yang, Q. Li et al., "Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors," *IEEE Transactions on Multimedia*, vol. 20, 2018.
- [36] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Information Fusion*, vol. 53, pp. 155–164, 2020.
- [37] H. F. Nweke, Y. W. Teh, U. R. Alo, and G. Mujtaba, "Analysis of multi-sensor fusion for mobile and wearable sensor-based human activity recognition," in *Proceedings of the International Conference on Data Processing and Applications 2018*, Guangzhou, China, May 2018.
- [38] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Applied Computing and Informatics*, vol. 17, 2019.
- [39] F. Ali, S. El-Sappagh, S. M. R. Islam et al., "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, 2020.
- [40] "Electronic Health Record," 2021, <https://webmedy.com/ehr/>.
- [41] G. S. Engine, "Boosting (machine learning)," 2020, [https://en.wikipedia.org/wiki/Boosting\(machine_learning\)#/media/File:Ensemble_Boosting.svg](https://en.wikipedia.org/wiki/Boosting(machine_learning)#/media/File:Ensemble_Boosting.svg).
- [42] H. F. Germany, "Diabetes data set," 2020, <https://www.kaggle.com/johndasilva/diabetes>.

Research Article

Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning

Rohit Bharti ¹, Aditya Khamparia ², Mohammad Shabaz ³, Gaurav Dhiman ⁴,
Sagar Pande ¹ and Parneet Singh ⁵

¹School of Computer Science and Engineering, Lovely Professional University, Phagwara, India

²Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India

³Arba Minch University, Arba Minch, Ethiopia

⁴Department of Computer Science, Government Bikram College of Commerce, Patiala, India

⁵All India Institute of Medical Science, Rishikesh, India

Correspondence should be addressed to Mohammad Shabaz; mohammad.shabaz@amu.edu.et

Received 16 May 2021; Revised 15 June 2021; Accepted 21 June 2021; Published 1 July 2021

Academic Editor: Ahmed A. Abd El-Latif

Copyright © 2021 Rohit Bharti 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 correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time. In this paper different machine learning algorithms and deep learning are applied to compare the results and analysis of the UCI Machine Learning Heart Disease dataset. The dataset consists of 14 main attributes used for performing the analysis. Various promising results are achieved and are validated using accuracy and confusion matrix. The dataset consists of some irrelevant features which are handled using Isolation Forest, and data are also normalized for getting better results. And how this study can be combined with some multimedia technology like mobile devices is also discussed. Using deep learning approach, 94.2% accuracy was obtained.

1. Introduction

Heart disease describes a range of conditions that affect your heart. Today, cardiovascular diseases are the leading cause of death worldwide with 17.9 million deaths annually, as per the World Health Organization reports [1]. Various unhealthy activities are the reason for the increase in the risk of heart disease like high cholesterol, obesity, increase in triglycerides levels, hypertension, etc. [1]. There are certain signs which the American Heart Association [2] lists like the persons having sleep issues, a certain increase and decrease in heart rate (irregular heartbeat), swollen legs, and in some cases weight gain occurring quite fast; it can be 1-2kg daily [3]. All these symptoms resemble different diseases also like it occurs in the aging persons, so it becomes a difficult task to get a correct diagnosis, which results in fatality in near future.

But as time is passing, a lot of research data and patients records of hospitals are available. There are many open sources for accessing the patient's records and researches can

be conducted so that various computer technologies could be used for doing the correct diagnosis of the patients and detect this disease to stop it from becoming fatal. Nowadays it is well known that machine learning and artificial intelligence are playing a huge role in the medical industry. We can use different machine learning and deep learning models to diagnose the disease and classify or predict the results. A complete genomic data analysis can easily be done using machine learning models. Models can be trained for knowledge pandemic predictions and also medical records can be transformed and analyzed more deeply for better predictions [4–6].

Many studies have been performed and various machine learning models are used for doing the classification and prediction for the diagnosis of heart disease. An automatic classifier for detecting congestive heart failure shows the patients at high risk and the patients at low risk by Melillo et al. [7]; they used machine learning algorithm as CART which stands for Classification and Regression in which

sensitivity is achieved as 93.3 percent and specificity is achieved as 63.5 percent. Then for improving the performance electrocardiogram (ECG) approach is suggested by Rahhal et al. [8] in which deep neural networks are used for choosing the best features and then using them. Then, for detecting heart failures, a clinical decision support system is contributed by Guidi et al. [9] for preventing it at an early stage. They tried to compare different machine learning models and deep learning models especially neural networks, as support vector machine, random forest, and CART algorithms. An 87.6 percent accuracy was achieved by random forest and CART, which outperformed everyone used in the classification. Combining the natural language processing with the rule-based approach, Zhang et al. [10] achieved 93.37 percent accuracy when the NYHA HF class was found from the unstructured clinical notes. SVM techniques used for detecting patients who already have diabetes and then predicting heart disease by Parthiban and Srivatsa [11] achieved a 94.60 percent accuracy rate, and the features taken were common like blood sugar level, age of the patient, and their blood pressure data.

In machine learning, a common problem is the high dimensionality of the data; the datasets which we use contain huge data and sometimes we cannot view that data even in 3D, which is also called the curse of dimensionality [12]. So, when we perform operations on this data, we require a huge amount of memory, and sometimes the data can also grow exponentially and overfitting can happen. The weighting features can be used, so the redundancy in the dataset can be decreased which in turn also helps in decreasing the processing time of the execution [13–17]. For decreasing the dimensionality of the dataset, there are various feature engineering and feature selection techniques which can be used to remove that data not having that much importance in the dataset [18].

In literature, when feature engineering and feature selection are applied, the results improve, both for classification as well as predictions. Dun et al. [19] tried various machine learning and deep learning techniques for detecting the heart disease and also performed hyperparameters tuning for increasing the results accuracy. Neural networks achieved high accuracy of 78.3 percent, and the other models were logistic regression, SVM, and ensemble techniques like Random Forest, etc. For reducing the cardiovascular features, Singh et al. [20] used generalized discriminant analysis for extracting nonlinear features; a binary classifier like an extreme learning machine for less overfitting and increasing the training speed and the ranking method used for all these was Fisher. The accuracy achieved was 100 percent for detecting coronary heart disease. Arrhythmias classification was done by Yaghouby et al. [21] for heart rate variability. A multilayer perceptron neural network was used for doing the classification and 100 percent accuracy is achieved by reducing the features or Gaussian Discriminant Analysis. Asl et al. [22] used Gaussian discriminant analysis for reducing the HRV signal features to 15 and 100 percent precision is achieved using the SVM classifier.

For dealing with data that are of high variance or high dimensional data, by using appropriate dimensionality

reduction techniques like PCA, we can store valuable information in new components [23]. PCA is used by many researchers as the first preference while dealing with high dimensionality data. Rajagopal and Ranganathan [24] used five different dimensionality reduction techniques which are unsupervised (linear and nonlinear), and neural network is used as a classifier for classifying cardiac arrhythmia. FastICA (used for independent component analysis) with a minimum of 10 components was able to achieve an F1 score of 99.83 percent. Zhang et al. [25] used the AdaBoost algorithm which is based on PCA for detecting breast cancer. Negi et al. [26] combined uncorrelated discriminant analysis with PCA so that the best features that are used for controlling the upper limb motions can be selected and the results were great. Avendaño-Valencia et al. [27] tried to reduce heart sounds to increase performance by applying PCA techniques on time-frequency representations. Kamencay et al. [28] tried a new method for different medical images reaching an accuracy of 83.6 percent when trained on 200 images by using PCA-KNN which is a scale-invariant feature used in medical images for the scaling purpose. Ratnasari et al. [29] used a gray-level threshold of 150 based on PCA and ROI, all of these used for reducing features of the X-ray images.

The studies of the past are mainly based on a 13-feature dataset. The classification is common in every study to predict if a patient has heart disease or not, and also one most common pattern which can be seen is that the dataset commonly used is of Cleveland [30]. The results obtained achieved great accuracies like random forest with 89.2 percent accuracy [31]; decision tree with 89.1 percent accuracy [32]; ANN with 92.7 percent accuracy [33], 89 percent [33], and 89.7 percent accuracy [34]; and SVM accuracy with 88 percent [34]. A hybrid model is created which achieved an accuracy of 94.2 percent by GA β NN [35]. PCA models achieved an accuracy of 92 and 95.2 percent as PCA β regression and PCA1 β NN [36]. The dimensionality reduction was the main focus here for learning three things: (i) selection of the best features, (ii) validation of performance, and (iii) use of six different classifiers for calculating the 74 features which are selected.

Heart disease is very fatal and it should not be taken lightly. Heart disease happens more in males than females, which can be read further from Harvard Health Publishing [37]. Researchers found that, throughout life, men were about twice as likely as women to have a heart attack. That higher risk persisted even after they accounted for traditional risk factors of heart disease, including high cholesterol, high blood pressure, diabetes, body mass index, and physical activity. The researchers are working on this dataset as it contains certain important parameters like dates from 1998, and it is considered as one of the benchmark datasets when someone is working on heart disease prediction. This dataset dates from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V, and the results achieved are quite promising.

The rest of the paper is divided into four sections. Section 1 consists of the introduction, Section 2 consists of the literature review, Section 3 consists of the methodology used,

Section 4 consists of the discussion, Section 4 consists of the results analysis, and Section 5 consists of conclusion and future scope.

2. Literature Review

The summary of the literature review can be seen in Table 1. Several approaches have been performed on this popular dataset, but the accuracy obtained by all the approaches is more with time computations.

3. Methodology

3.1. Description of the Dataset. The dataset used for this research purpose was the Public Health Dataset and it is dating from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 attributes, including the predicted attribute, but all published experiments refer to using a subset of 14 of them. The “target” field refers to the presence of heart disease in the patient. It is integer-valued 0 = no disease and 1 = disease. The first four rows and all the dataset features are shown in Table 1 without any preprocessing. Now the attributes which are used in this research purpose are described as follows and for what they are used or resemble:

- (i) Age—age of patient in years, sex—(1 = male; 0 = female).
- (ii) Cp—chest pain type.
- (iii) Trestbps—resting blood pressure (in mm Hg on admission to the hospital). The normal range is 120/80 (if you have a normal blood pressure reading, it is fine, but if it is a little higher than it should be, you should try to lower it. Make healthy changes to your lifestyle).
- (iv) Chol—serum cholesterol shows the amount of triglycerides present. Triglycerides are another lipid that can be measured in the blood. It should be less than 170 mg/dL (may differ in different Labs).
- (v) Fbs—fasting blood sugar larger than 120 mg/dl (1 true). Less than 100 mg/dL (5.6 mmol/L) is normal, and 100 to 125 mg/dL (5.6 to 6.9 mmol/L) is considered prediabetes.
- (vi) Restecg—resting electrocardiographic results.
- (vii) Thalach—maximum heart rate achieved. The maximum heart rate is 220 minus your age.
- (viii) Exang—exercise-induced angina (1 yes). Angina is a type of chest pain caused by reduced blood flow to the heart. Angina is a symptom of coronary artery disease.
- (ix) Oldpeak—ST depression induced by exercise relative to rest.
- (x) Slope—the slope of the peak exercise ST segment.
- (xi) Ca—number of major vessels (0–3) colored by fluoroscopy.

- (xii) Thal—no explanation provided, but probably thalassemia (3 normal; 6 fixed defects; 7 reversible defects).
- (xiii) Target (T)—no disease = 0 and disease = 1, (angiographic disease status).

3.2. Preprocessing of the Dataset. The dataset does not have any null values. But many outliers needed to be handled properly, and also the dataset is not properly distributed. Two approaches were used. One without outliers and feature selection process and directly applying the data to the machine learning algorithms, and the results which were achieved were not promising. But after using the normal distribution of dataset for overcoming the overfitting problem and then applying Isolation Forest for the outlier’s detection, the results achieved are quite promising. Various plotting techniques were used for checking the skewness of the data, outlier detection, and the distribution of the data. All these preprocessing techniques play an important role when passing the data for classification or prediction purposes.

3.2.1. Checking the Distribution of the Data. The distribution of the data plays an important role when the prediction or classification of a problem is to be done. We see that the heart disease occurred 54.46% of the time in the dataset, whilst 45.54% was the no heart disease. So, we need to balance the dataset or otherwise it might get overfit. This will help the model to find a pattern in the dataset that contributes to heart disease and which does not as shown in Figure 1.

3.2.2. Checking the Skewness of the Data. For checking the attribute values and determining the skewness of the data (the asymmetry of a distribution), many distribution plots are plotted so that some interpretation of the data can be seen. Different plots are shown, so an overview of the data could be analyzed. The distribution of age and sex, the distribution of chest pain and trestbps, the distribution of cholesterol and fasting blood, the distribution of ecg resting electrode and thalach, the distribution of exang and oldpeak, the distribution of slope and ca, and the distribution of thal and target all are analyzed and the conclusion is drawn as shown in Figures 2 and 3.

By analyzing the distribution plots, it is visible that thal and fasting blood sugar is not uniformly distributed and they needed to be handled; otherwise, it will result in overfitting or underfitting of the data.

3.2.3. Checking Stats of the Normal Distribution of Data. Checking the features which are important for heart disease and not important for heart disease is shown in Figures 4 and 5, respectively. Here the important factors show a different variation which means it is important.

The conclusion which can be drawn from these statistical figures is that we can see a Gaussian distribution which is

TABLE 1: Summary of the literature review.

Sr.no.	Author	Year	Findings
1	Gárate-Escamila et al. [38]	2020	DNN and ANN were used with the X^2 statistical model. The clinical data parameters were used for conforming the predictions.
2	Harvard Medical School [37]	2020	Hungarian-Cleveland datasets were used for predicting heart disease using different machine learning classifiers and PCA was used for dimensionality reduction and feature selection
3	Zhang et al. [25]	2018	AdaBoost classifier with PCA combination was used for the feature extraction and the accuracy of the prediction was increased
4	Singh et al. [20]	2018	Heart rate variability was for the detection of coronary artery disease. Fisher method and generalised discriminant analysis with binary classifiers were used for the detection of important features.
5	Chen et al. [16]	2018	A subspace feature clustering was used as a subset of stratified feature clustering and for doing a feature reduction of the clusters formed
6	Yang and Nataliani [15]	2018	A fuzzy clustering method especially fuzzy c -means was used for various feature weighted methods and features were reduced
7	Kumar [32]	2017	Different machine learning algorithms were applied for getting the results and then compared with each other
8	Rajagopal and Ranganathan [24]	2017	Combination of probabilistic neural network classifier, PCA, kernel PCA, and unsupervised dimensionality reduction was used so that feature reduction can be used and a domain expert was used for the correct analysis of the result
9	Zhang et al. [10]	2017	Support vector machine is used for the classification purpose of the clinical data which is matched with the codes of New York heart association; further findings are left for other researchers
10	Khan and Quadri [31]	2016	The main aim of this research was to summarize the best model and angiographic disease status by analyzing different unstructured data and using data mining techniques
11	Negi et al. [26]	2016	Uncorrelated linear discriminant analysis with PCA was used for studying the electrocardiogram and Wilson methods were also used for the distinction of upper limb motions
12	Dun et al. [19]	2016	They applied a variety of deep learning techniques and ensemble techniques and also performed hyperparameter tuning techniques for increasing the accuracy.
13	Rahhal et al. [8]	2016	ECG approach is used by consulting various domain experts and then MIT-BIH arrhythmia database as well as two other databases called INCART and SVDB, respectively
14	Imani and Ghassemian [17]	2015	There are several times when the data is not enough, so Imani approached a weighted training sample method including feature extraction for the spatial dimension of the images and the accuracy was increased
15	Guidi et al. [9]	2014	Neural networks, SVM, and fuzzy system approach are used and Random Forest is used as a classifier, for the prediction of heart failure by using a clinical decision support system
16	Santhanam and Ephzibah [36]	2013	A regression technique with PCA with its different versions like PCA1, PCA2, PCA3, and PCA4 was used and the features were extracted and the results were promising
17	Ratnasari et al. [29]	2013	The datasets used were Cleveland-Hungarian dataset and the UCI machine learning datasets were analyzed with feature selection techniques
18	Kamencay et al. [28]	2013	Object recognition was performed with scale-invariant feature transformation. Caltech 101 database was used for the evaluation purpose.
19	Melillo et al. [7]	2013	Two public Holster databases were used for finding high-risk and low-risk patients. Cart algorithm is applied for the classification purpose.
20	Amma [35]	2012	The dataset used was from University of California, Irvine. The genetic algorithm was used for the training purpose and neural network for the classification purpose.
21	Keogh and Mueen [12]	2012	How to break the curse of dimensionality using PCA, SVM, and other classifiers and reduce features.
22	Parthiban and Srivatsa [11]	2012	Diabetes is one of the main causes of heart disease. The classifiers used are Naïve Bayes and SVM for extracting important features and classification purpose.
23	Srinivas et al. [34]	2010	Prediction of heart diseases in the coal mines was the prime consideration, and decision tree, naïve Bayes, and neural networks were used for the classification
24	Das et al. [33]	2009	On Cleveland dataset, using a SAS-based software, a great accuracy was achieved with different ensemble techniques
25	Yaghouby et al. [21]	2009	Cardiac arrhythmias was considered using the MIT-BIH database. HRV similar to [20] was used.
26	Asl et al. [22]	2008	Generalised discriminant analysis and SVM were used for feature reduction and classification
27	Avendaño-Valencia et al. [27]	2009	Feature extraction was based upon the heart murmur frequency with time representation frequency and PCA was used for the analysis of the features
28	Guyon et al. [23]	2008	Book for doing feature extraction efficiently.

TABLE 1: Continued.

Sr.no.	Author	Year	Findings
29	UCI Machine Learning Repository [30]	1998	This dataset is used for many ML and deep learning benchmark results
30	Liu and Motoda [18]	1998	Feature importance and how to select them appropriately was discussed in this book
31	Wettschereck et al. [14]	1997	K-NN algorithm was used for the classification as they are mostly the derivatives for the lazy learning algorithms for the feature selection using weighted methods
32	Wettschereck and Dietterich [13]	1995	Different classification problems decision boundaries were analyzed, and the problem was tackled using nested generalized example

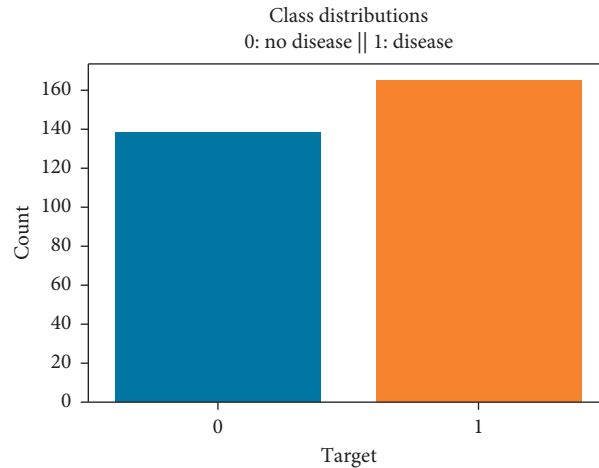


FIGURE 1: Class distribution of disease and no disease.

important for heart disease and no Gaussian distribution which is playing that much important role in heart disease.

3.2.4. Feature Selection. For selecting the features and only choosing the important feature, the Lasso algorithm is used which is a part of embedded methods while performing feature selection. It shows better predictive accuracy than filter methods. It renders good feature subsets for the used algorithm. And then for selecting the selected features, select from the model which is a part of feature selection in the scikit-learn library.

3.2.5. Checking Duplicate Values in the Data. The duplicates should be tackled down safely or otherwise would affect the generalization of the model. There might be a chance if duplicates are not dealt with properly; they might show up in the test dataset which is also in the training dataset. The duplicate values can be seen in Table 2.

3.3. Machine Learning Classifiers Proposed. The proposed approach was applied to the dataset in which firstly the dataset was properly analyzed and then different machine learning algorithms consisting of linear model selection in which Logistic Regression was used. For focusing on neighbor selection technique KNeighborsClassifier was used, then tree-based technique like DecisionTreeClassifier was used, and then a very popular and most popular technique of ensemble methods RandomForestClassifier

was used. Also for checking the high dimensionality of the data and handling it, Support Vector Machine was used. Another approach which also works on ensemble method and Decision Tree method combination is XGBoost classifier as shown in Figures 6 and 7.

3.4. Deep Learning Pseudocode.

- (i) Dataset of training
- (ii) Dataset of testing
- (iii) Checking the shape/features of the input
- (iv) The procedure of initiating the sequential layer
- (v) Adding dense layers with dropout layers and ReLU activation functions
- (vi) Adding a last dense layer with one output and binary activation function
- (vii) End repeat
- (viii) L (output)
- (ix) End procedure

3.5. Deep Learning Proposed. There are two ways a deep learning approach can be applied. One is using a sequential model and another is a functional deep learning approach. In this particular research, the first one is used. A sequential model with a fully connected dense layer is used, with the flatten and dropout layers to prevent the overfitting and the results are compared of the machine learning and deep

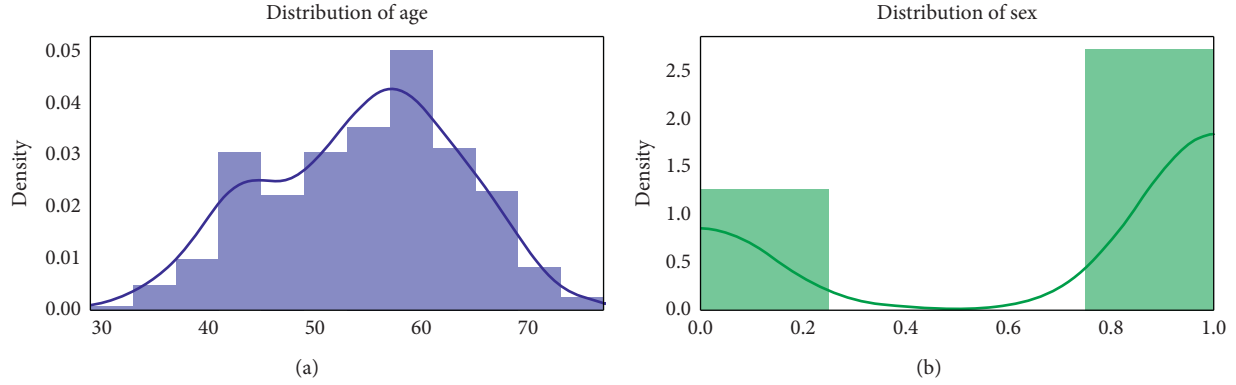


FIGURE 2: Distribution of age and sex.

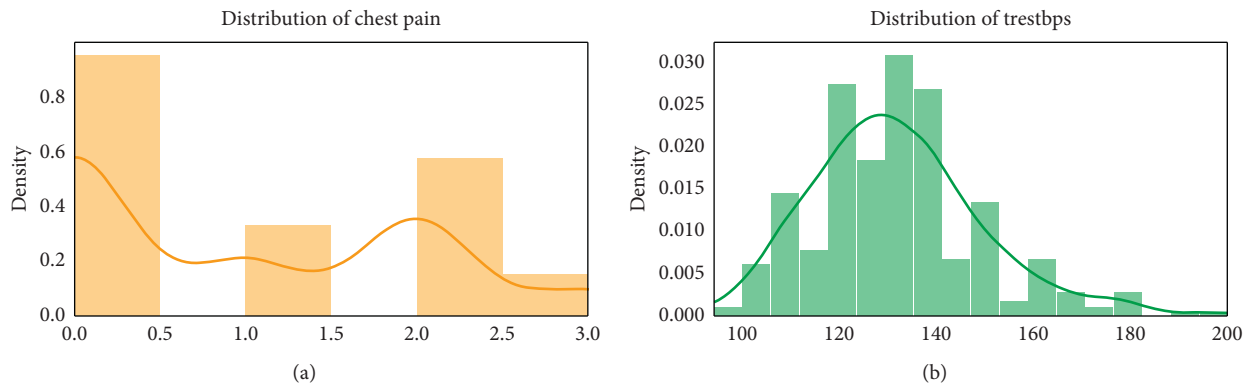


FIGURE 3: Distribution of chest pain and trestbps.

learning and variations in the learning including computational time and accuracy can be analyzed and can be seen in the figures further discussed in the Results section.

3.6. Evaluation Process Used. For the evaluation process, confusion matrix, accuracy score, precision, recall, sensitivity, and F1 score are used. A confusion matrix is a table-like structure in which there are true values and predicted values, called true positive and true negative. It is defined in four parts: the first one is true positive (TP) in which the values are identified as true and, in reality, it was true also. The second one is false positive (FP) in which the values identified are false but are identified as true. The third one is false negative (FN) in which the value was true but was identified as negative. The fourth one is true negative (TN) in which the value was negative and was truly identified as negative. The table is shown in Figure 8.

In Figure 8, P = positive, N = negative, TP = true positive, FN = false negative, FP = false positive, TN = true negative.

Then for checking how well a model is performing, an accuracy score is used. It is defined as the true positive values plus true negative values divided by true positive plus true negative plus false positive plus false negative. The formula is

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (1)$$

After accuracy there is specificity which is the proportion of true negative cases that were classified as negative; thus, it is a measure of how well a classifier identifies negative cases. It is also known as the true negative rate. The formula is

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (2)$$

Then there is sensitivity in which the proportion of actual positive cases got predicted as positive (or true positive). Sensitivity is also termed as recall. In other words, an unhealthy person got predicted as unhealthy. The formula is

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

3.7. Use of Multimedia. The whole knowledge which will be obtained could be transferred to the mobile devices means, when the person will input these symptoms in the mobile device in which the trained model will already be present and then can analyze the symptoms and could give the prescription accordingly. Different doctors could be taken under consideration and a complete autonomous system could be generated. We can also integrate the doctors' numbers if the model is showing high risk and they can consult the doctor. And if they are showing less symptoms, then medicines already prescribed by the doctors for a

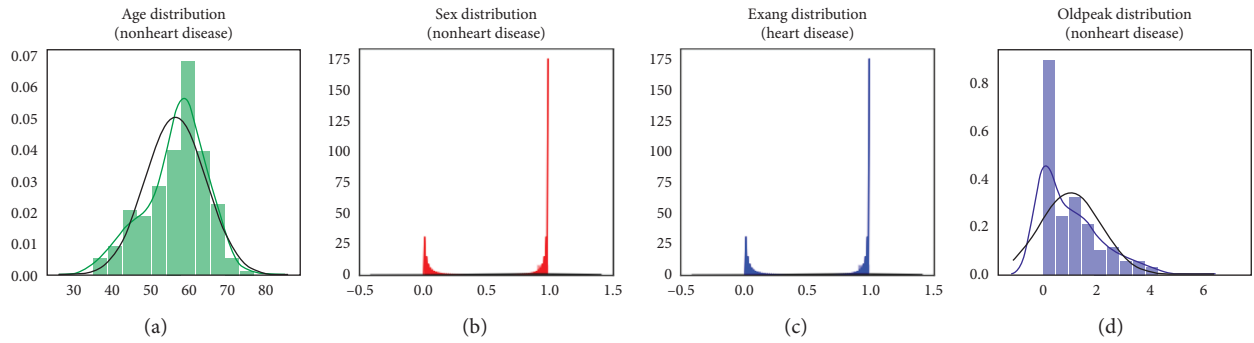


FIGURE 4: Features important for heart disease.

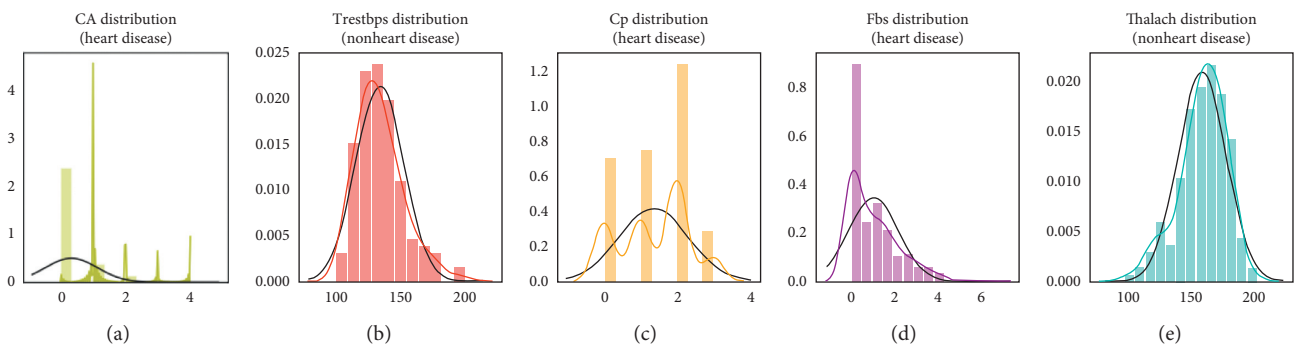


FIGURE 5: Features not important for heart disease.

TABLE 2: Duplicate values.

Age	Sex	Cp	Trest bps	Chol	Rest ecg	Thalach	Exang	Old peak	Slope	Ca	Thal	T
38	1	2	138	175	1	173	0	0.0	2	4	2	1

Using the pandas' function for dropping these values is the simplest. It is also an important part while performing data preprocessing.

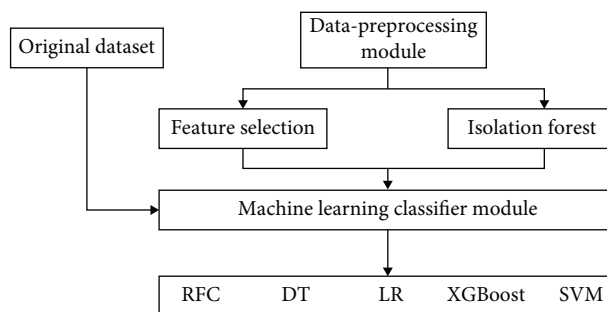


FIGURE 6: 1st schematic diagram of the proposed model.

certain range will be shown. This system will prove beneficial and the workload on the doctors would also be less. Also in these current times of coronavirus, we need more autonomous systems which would also help in keeping the virtuality between persons more. Thus we could create some applications with the help of doctors and make it work.

4. Analysis of Results

By applying different machine learning algorithms and then using deep learning to see what difference comes when it is applied to the data, three approaches were used. In the first approach, normal dataset which is acquired is directly used

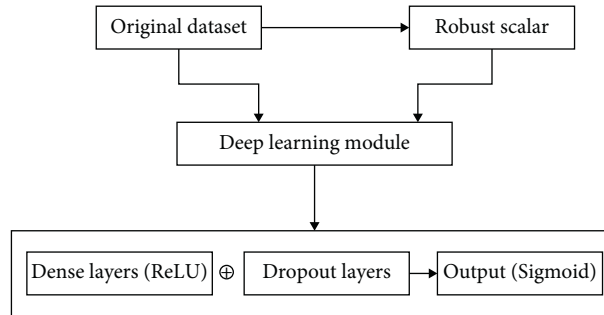


FIGURE 7: 2nd schematic diagram of the proposed model.

		Predicted value	
		P	N
True value	P	TP	FN
	N	FP	TN

FIGURE 8: Confusion matrix.

for classification, and in the second approach, the data with feature selection are taken care of and there is no outliers detection. The results which are achieved are quite promising and then in the third approach the dataset was normalized taking care of the outliers and feature selection; the results achieved are much better than the previous techniques, and when compared with other research accuracies, our results are quite promising.

4.1. Using the First Approach (without Doing Feature Selection and Outliers Detection). As can be seen in Figure 1, the dataset is not normalized, there is no equal distribution of the target class, it can further be seen when a correlation heatmap is plotted, and there are so many negative values; it can be visualized in Figure 9.

So, even if the feature selection is done, still, we have outliers which can be seen in Figure 10.

By applying the first approach, the accuracy achieved by the Random Forest is 76.7%, Logistic Regression is 83.64%, KNeighbors is 82.27%, Support Vector Machine is 84.09%, Decision Tree is 75.0%, and XGBoost is 70.0%. SVM is having the highest accuracy here which is achieved by using the cross-validation and grid search for finding the best parameters or in other words doing the hyperparameter tuning. Then after machine learning, deep learning is applied by using the sequential model approach. In the model, 128 neurons are used and the activation function used is ReLU, and in the output layer which is a single class prediction problem, the sigmoid activation function is used, with loss as binary cross-entropy and gradient descent optimizer as Adam. The accuracy achieved is 76.7%.

4.2. Using the Second Approach (Doing Feature Selection and No Outliers Detection). After selecting the features (feature

selection) and scaling the data as there are outliers, the robust standard scalar is used; it is used when the dataset is having certain outliers. In the second approach, the accuracy achieved by Random Forest is 88%, the Logistic Regression is 85.9%, KNeighbors is 79.69%, Support Vector Machine is 84.26%, the Decision Tree is 76.35%, and XGBoost is 71.1%. Here the Random Forest is the clear winner with a precision of 88.4% and an F1 score of 86.5%.

Then deep learning is applied with the same parameters before and the accuracy achieved is 86.8%, and the evaluation accuracy is 81.9%, which is better than the first approach.

4.3. Using the Third Approach (by Doing Feature Selection and Also Outliers Detection). In this approach, the dataset is normalized and the feature selection is done and also the outliers are handled using the Isolation Forest. The correlation comparison can be seen in Figure 10. The accuracy of the Random Forest is 80.3%, Logistic Regression is 83.31%, KNeighbors is 84.86%, Support Vector Machine is 83.29%, Decision Tree is 82.33%, and XGBoost is 71.4%. Here the winner is KNeighbors with a precision of 77.7% and a specificity of 80%. A lot of tips and tricks for selecting different algorithms are shown by Garate-Escamila et al. [38]. Using deep learning in the third approach, the accuracy achieved is 94.2%. So, the maximum accuracy achieved by the machine learning model is KNeighbors (83.29%) in the third approach, and, for deep learning, the maximum accuracy achieved is 81.9%. Thus, the conclusion can be drawn here that, for this dataset, the deep learning algorithm achieved 94.2 percent accuracy which is greater than the machine learning models. We also made a comparison with another research of the deep learning by Ramprakash et al. [39] in which they achieved 84% accuracy and Das et al. [33]

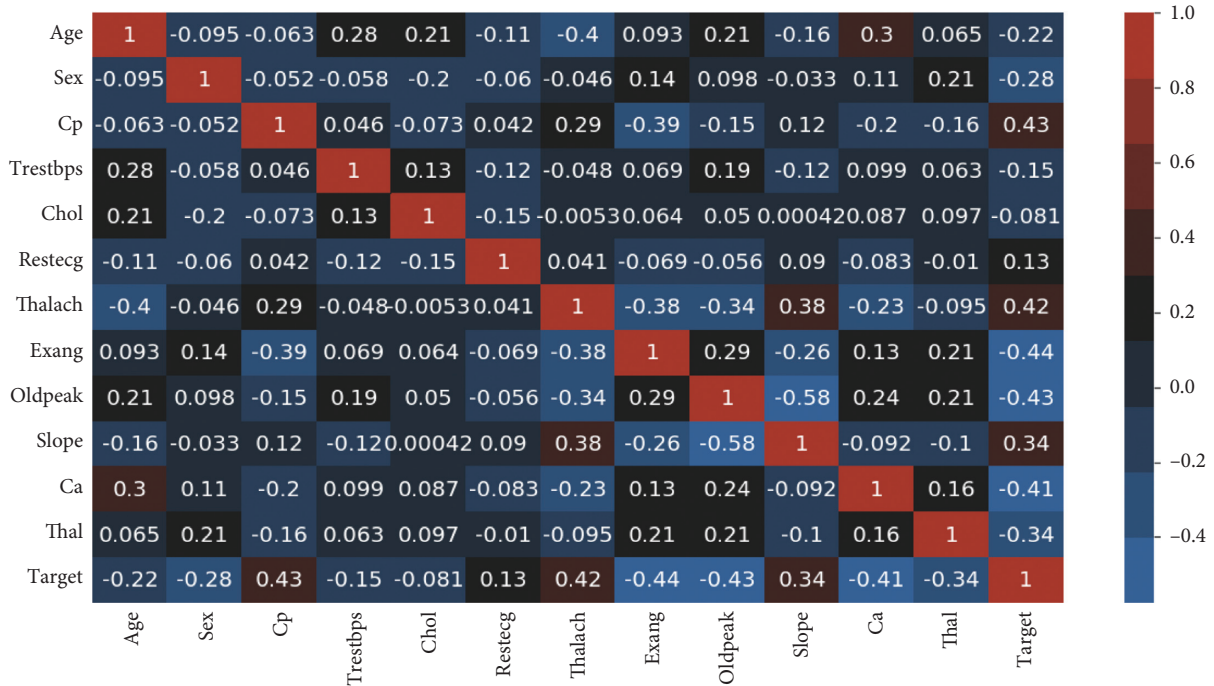


FIGURE 9: Correlation heatmap.

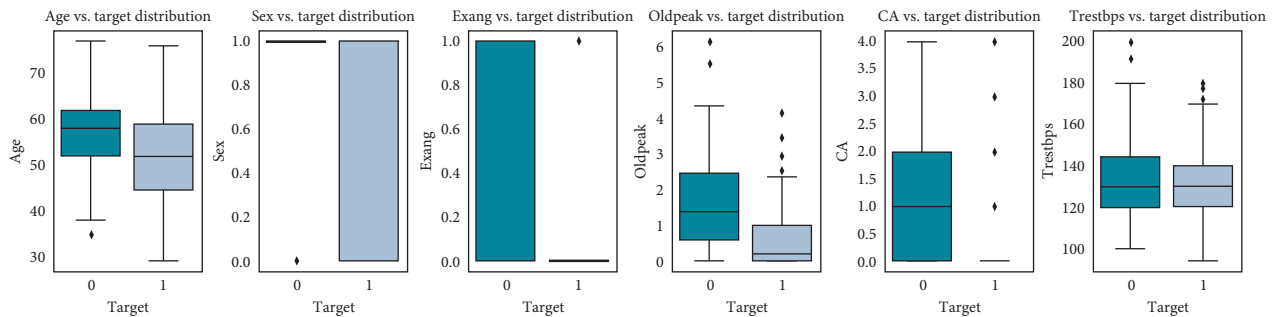


FIGURE 10: Feature selection on correlation heatmap.

TABLE 3: Comparative analysis.

Classifiers	Accuracy (%)	Specificity	Sensitivity
Logistic regression	83.3	82.3	86.3
K neighbors	84.8	77.7	85.0
SVM	83.2	78.7	78.2
Random forest	80.3	78.7	78.2
Decision tree	82.3	78.9	78.5
DL	94.2	83.1	82.3

achieved 92.7 percent accuracy. So our algorithm produced greater accuracy and more promising than other approaches [40, 41]. The comparison of different classifiers of ML and DL can be seen in Table 3.

4.4. *Architecture for Using Deep Learning Approach.* Here in this architecture, we used three dense layers: the first dense layer consists of 128 units, the second dense layer consists of 64 units, and the third dense layer consists of 32 units. For

the first layer, the Dropout Layer (HyperParameter) is 0.2 and for the second is 0.1.

5. Conclusion and Future Scope

In this paper, we proposed three methods in which comparative analysis was done and promising results were achieved. The conclusion which we found is that machine learning algorithms performed better in this analysis. Many researchers have previously suggested that we should use ML

where the dataset is not that large, which is proved in this paper. The methods which are used for comparison are confusion matrix, precision, specificity, sensitivity, and F1 score. For the 13 features which were in the dataset, KNeighbors classifier performed better in the ML approach when data preprocessing is applied.

The computational time was also reduced which is helpful when deploying a model. It was also found out that the dataset should be normalized; otherwise, the training model gets overfitted sometimes and the accuracy achieved is not sufficient when a model is evaluated for real-world data problems which can vary drastically to the dataset on which the model was trained. It was also found out that the statistical analysis is also important when a dataset is analyzed and it should have a Gaussian distribution, and then the outlier's detection is also important and a technique known as Isolation Forest is used for handling this. The difficulty which came here is that the sample size of the dataset is not large. If a large dataset is present, the results can increase very much in deep learning and ML as well. The algorithm applied by us in ANN architecture increased the accuracy which we compared with the different researchers. The dataset size can be increased and then deep learning with various other optimizations can be used and more promising results can be achieved. Machine learning and various other optimization techniques can also be used so that the evaluation results can again be increased. More different ways of normalizing the data can be used and the results can be compared. And more ways could be found where we could integrate heart-disease-trained ML and DL models with certain multimedia for the ease of patients and doctors.

Data Availability

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

Conflicts of Interest

All the authors declare that there are no conflicts of interest regarding the publication.

References

- [1] World Health Organization, *Cardiovascular Diseases*, WHO, Geneva, Switzerland, 2020, https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1.
- [2] American Heart Association, *Classes of Heart Failure*, American Heart Association, Chicago, IL, USA, 2020, <https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure>.
- [3] American Heart Association, *Heart Failure*, American Heart Association, Chicago, IL, USA, 2020, <https://www.heart.org/en/health-topics/heart-failure>.
- [4] S. Shalev-Shwartz and S. Ben-David, "Understanding machine learning," *From Theory to Algorithms*, Cambridge University Press, Cambridge, UK, 2020.
- [5] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," *Data Mining, Inference, and Prediction*, Springer, Cham, Switzerland, 2020.
- [6] S. Marsland, "Machine learning," *An Algorithmic Perspective*, CRC Press, Boca Raton, FL, USA, 2020.
- [7] P. Melillo, N. De Luca, M. Bracale, and L. Pecchia, "Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 727–733, 2013.
- [8] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.
- [9] G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A machine learning system to improve heart failure patient assistance," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1750–1756, 2014.
- [10] R. Zhang, S. Ma, L. Shanahan, J. Munroe, S. Horn, and S. Speedie, "Automatic methods to extract New York heart association classification from clinical notes," in *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1296–1299, IEEE, Kansas City, MO, USA, November 2017.
- [11] G. Parthiban and S. K. Srivatsa, "Applying machine learning methods in diagnosing heart disease for diabetic patients," *International Journal of Applied Information Systems*, vol. 3, no. 7, pp. 25–30, 2012.
- [12] E. Keogh and A. Mueen, "Curse of dimensionality," in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds., Springer, Cham, Switzerland, 2017.
- [13] D. Wettschereck and T. G. Dietterich, "An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms," *Machine Learning*, vol. 19, no. 1, pp. 5–27, 1995.
- [14] D. Wettschereck, D. W. Aha, and T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms," *Lazy Learning*, vol. 11, no. 1/5, pp. 273–314, 1997.
- [15] M.-S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 817–835, 2018.
- [16] R. Chen, N. Sun, X. Chen, M. Yang, and Q. Wu, "Supervised feature selection with a stratified feature weighting method," *IEEE Access*, vol. 6, pp. 15087–15098, 2018.
- [17] M. Imani and H. Ghassemian, "Feature extraction using weighted training samples," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1387–1391, 2015.
- [18] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection*, Springer, Cham, Switzerland, 1998.
- [19] B. Dun, E. Wang, and S. Majumder, "Heart disease diagnosis on medical data using ensemble learning," 2016.
- [20] R. S. Singh, B. S. Saini, and R. K. Sunkaria, "Detection of coronary artery disease by reduced features and extreme learning machine," *Medicine and Pharmacy Reports*, vol. 91, no. 2, pp. 166–175, 2018.
- [21] F. Yaghoubi, F. Yaghoubi, A. Ayatollahi, and R. Soleimani, "Classification of cardiac abnormalities using reduced features of heart rate variability signal," *World Applied Sciences Journal*, vol. 6, no. 11, pp. 1547–1554, 2009.
- [22] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial Intelligence in Medicine*, vol. 44, no. 1, pp. 51–64, 2008.

- [23] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, *Feature Extraction: Foundations and Applications*, Springer, Cham, Switzerland, 2008.
- [24] R. Rajagopal and V. Ranganathan, "Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification," *Biomedical Signal Processing and Control*, vol. 34, pp. 1–8, 2017.
- [25] D. Zhang, L. Zou, X. Zhou, and F. He, "Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer," *IEEE Access*, vol. 6, pp. 28936–28944, 2018.
- [26] S. Negi, Y. Kumar, and V. M. Mishra, "Feature extraction and classification for EMG signals using linear discriminant analysis," in *Proceedings of the 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall)*, September 2016.
- [27] D. Avendaño-Valencia, F. Martínez-Tabares, D. Acosta-Medina, I. Godino-Llorente, and G. Castellanos-Dominguez, "TFR-based feature extraction using PCA approaches for discrimination of heart murmurs," in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5665–5668, IEEE, Minneapolis, MN, USA, September 2009.
- [28] P. Kamencay, R. Hudec, M. Benčo, and M. Zachariasova, "Feature extraction for object recognition using PCA-KNN with application to medical image analysis," in *Proceedings of the 2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 830–834, IEEE, Rome, Italy, July 2013.
- [29] N. R. Ratnasari, A. Susanto, I. Soesanti, and Maesadji, "Thoracic X-ray features extraction using thresholding-based ROI template and PCA-based features selection for lung TB classification purposes," in *Proceedings of the 2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*, pp. 65–69, IEEE, Bandung, Indonesia, November 2013.
- [30] UCI Machine Learning Repository, "Heart disease data set," 2020, <http://archive.ics.uci.edu/ml/datasets/heart+disease>.
- [31] S. S. Khan and S. M. K. Quadri, "Prediction of angiographic disease status using rule based data mining techniques," *Biological Forum—An International Journal*, vol. 8, no. 2, pp. 103–107, 2016.
- [32] S. Kumar, "Predicting and diagnosing of heart disease using machine learning algorithms," *International Journal of Engineering and Computer Science*, vol. 6, no. 6, pp. 2319–7242, 2017.
- [33] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.
- [34] K. Srinivas, G. Raghavendra Rao, and A. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques," in *Proceedings of 2010 5th International Conference on Computer Science & Education*, pp. 1344–1349, IEEE, Hefei, China, August 2010.
- [35] N. G. B. Amma, "Cardiovascular disease prediction system using genetic algorithm and neural network," in *Proceedings of the 2012 International Conference on Computing, Communication and Applications*, February 2012.
- [36] T. Santhanam and E. P. Ephzibah, "Heart disease classification using PCA and feed forward neural networks," *Mining Intelligence and Knowledge Exploration*, Springer, Cham, Switzerland, 2013.
- [37] Harvard Medical School, "Throughout life, heart attacks are twice as common in men than women," 2020, <https://www.health.harvard.edu/heart-health/throughout-life-heart-attacks-are-twice-as-common-in-men-than-women>.
- [38] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrés, "Classification models for heart disease prediction using feature selection and PCA," *Informatics in Medicine Unlocked*, vol. 19, Article ID 100330, 2020.
- [39] P. Ramprakash, R. Sarumathi, R. Mowriya, and S. Nithyavishnupriya, "Heart disease prediction using deep neural network," in *Proceedings of the 2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 666–670, IEEE, Coimbatore, India, February 2020.
- [40] Kanksha, B. Aman, P. Sagar, M. Rahul, and K. Aditya, "An intelligent unsupervised technique for fraud detection in health care systems," *Intelligent Decision Technologies*, vol. 15, no. 1, pp. 127–139, 2021.
- [41] K. Divya, A. Sirohi, S. Pande, and R. Malik, "An IoMT assisted heart disease diagnostic system using machine learning techniques," in *Cognitive Internet of Medical Things for Smart Healthcare*, A. E. Hassanien, A. Khamparia, D. Gupta, K. Shankar, and A. Slowik, Eds., vol. 311, pp. 145–161, Springer, Cham, Switzerland, 2021.