

# Personal Communication Technologies for Smart Spaces

Guest Editors: Fawad Zaman, Sungchang Lee, Ali Kashif Bashir, and Abdul Razzaq



---



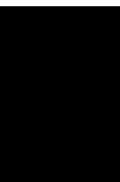
# **Personal Communication Technologies for Smart Spaces**

Mobile Information Systems

---

## **Personal Communication Technologies for Smart Spaces**

Guest Editors: Fawad Zaman, Sungchang Lee, Ali  
Kashif Bashir, and Abdul Razzaq




---

Copyright © 2021 Hindawi Limited. All rights reserved.






This is a special issue published in "Mobile Information Systems." 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

Alessandro Bazzi , Italy



## Academic Editors

Mahdi Abbasi , Iran  
Abdullah Alamoodi , Malaysia  
Markos Anastassopoulos, United Kingdom  
Marco Anisetti , Italy  
Claudio Agostino Ardagna , Italy  
Ashish Bagwari , India  
Dr. Robin Singh Bhadoria , India  
Nicola Biccocchi , Italy  
Peter Brida , Slovakia  
Puttamadappa C. , India  
Carlos Calafate , Spain  
Pengyun Chen, China  
Yuh-Shyan Chen , Taiwan  
Wenchi Cheng, China  
Gabriele Civitarese , Italy  
Massimo Condoluci , Sweden  
Rajesh Kumar Dhanaraj, India  
Rajesh Kumar Dhanaraj , India  
Almudena Díaz Zayas , Spain  
Filippo Gandino , Italy  
Jorge Garcia Duque , Spain  
Francesco Gringoli , Italy  
Wei Jia, China  
Adrian Kliks , Poland  
Adarsh Kumar , India  
Dongming Li, China  
Juraj Machaj , Slovakia  
Mirco Marchetti , Italy  
Elio Masciari , Italy  
Zahid Mehmood , Pakistan  
Eduardo Mena , Spain  
Massimo Merro , Italy  
Aniello Minutolo , Italy  
Jose F. Monserrat , Spain  
Raul Montoliu , Spain  
Mario Muñoz-Organero , Spain  
Francesco Palmieri , Italy  
Marco Picone , Italy  
Alessandro Sebastian Podda , Italy  
Maheswar Rajagopal, India  
Amon Rapp , Italy  
Filippo Sciarrone, Italy  
Floriano Scioscia , Italy

Mohammed Shuaib , Malaysia  
Michael Vassilakopoulos , Greece  
Ding Xu , China  
Laurence T. Yang , Canada  
Kuo-Hui Yeh , Taiwan



# Contents

## **Parkinson's Disease Diagnosis in Cepstral Domain Using MFCC and Dimensionality Reduction with SVM Classifier**

Atiqur Rahman, Sanam Shahla Rizvi , Aurangzeb Khan, Aaqif Afzaal Abbasi, Shafqat Ullah Khan, and Tae-Sun Chung 



Research Article (10 pages), Article ID 8822069, Volume 2021 (2021)

## **Modelling Reachability in Transport Networks: Using Alternative Visual Representations in Interactive Linked#Views to Gain Valuable Insights**

Rehmat Ullah , Laiq Hasan, Farman Ullah, Ajmal Khan, and You-Ze Cho 



Research Article (14 pages), Article ID 8813163, Volume 2021 (2021)

## **WiFi-Based Virtual Access Network Scheduling for Downlink Traffic Dominated Smart Spaces**

Pin Lv , Siyu Pan , and Jia Xu 



Research Article (9 pages), Article ID 8848558, Volume 2020 (2020)

## **Development of Hepatitis Disease Detection System by Exploiting Sparsity in Linear Support Vector Machine to Improve Strength of AdaBoost Ensemble Model**

Wasif Akbar , Wei-ping Wu, Sehrish Saleem, Muhammad Farhan, Muhammad Asim Saleem, Ashir Javeed, and Liaqat Ali 



Research Article (9 pages), Article ID 8870240, Volume 2020 (2020)

## **A Lightweight Location-Aware Fog Framework (LAFF) for QoS in Internet of Things Paradigm**

Qaisar Shaheen , Muhammad Shiraz, Muhammad Usman Hashmi, Danish Mahmood, Zhu zhiyu, and Rizwan Akhtar 


Research Article (15 pages), Article ID 8871976, Volume 2020 (2020)

## **Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification**

Ashir Javeed, Sanam Shahla Rizvi , Shijie Zhou, Rabia Riaz, Shafqat Ullah Khan, and Se Jin Kwon 



Research Article (11 pages), Article ID 8843115, Volume 2020 (2020)

## **Evaluation of the Challenges in the Internet of Medical Things with Multicriteria Decision Making (AHP and TOPSIS) to Overcome Its Obstruction under Fuzzy Environment**

Muhammad Imran Tariq , Natash Ali Mian, Abid Sohail, Tahir Alyas, and Rehan Ahmad



Research Article (19 pages), Article ID 8815651, Volume 2020 (2020)

## **Personal Communication Technologies for Smart Spaces Density-Based Clustering for Content and Color Adaptive Tone Mapping**

Maleeha Javed, Hassan Dawood , Muhammad Murtaza Khan, Ameen Banjar, Riad Alharbey, and Hussain Dawood 


Research Article (10 pages), Article ID 8846033, Volume 2020 (2020)

## **Impact of Node Density on the QoS Parameters of Routing Protocols in Opportunistic Networks for Smart Spaces**

Puneet Garg , Ashutosh Dixit , Preeti Sethi, and Plácido Rogerio Pinheiro


Research Article (18 pages), Article ID 8868842, Volume 2020 (2020)

**Your Knock Is My Command: Binary Hand Gesture Recognition on Smartphone with Accelerometer**

Huixiang Zhang , Wenteng Xu, Chunlei Chen, Liang Bai, and Yonghui Zhang


Research Article (16 pages), Article ID 8864627, Volume 2020 (2020)


**Analysis and Evaluation of Braille to Text Conversion Methods**

Sana Shokat, Rabia Riaz, Sanam Shahla Rizvi, Khalil Khan, Farina Riaz, and Se Jin Kwon 

Research Article (14 pages), Article ID 3461651, Volume 2020 (2020)

**Certificateless Proxy Reencryption Scheme (CPRES) Based on Hyperelliptic Curve for Access Control in Content-Centric Network (CCN)**

Zahid Ullah, Asim Zeb , Insaf Ullah, Khalid Mahmood Awan, Yousaf Saeed, M. Irfan Uddin, Mahmoud

Ahmad Al-Khasawneh , Marwan Mahmoud, and Mahdi Zareei


Research Article (13 pages), Article ID 4138516, Volume 2020 (2020)

**Robust Spectrum Sensing via Double-Sided Neighbor Distance Based on Genetic Algorithm in Cognitive Radio Networks**

Noor Gul, Muhammad Sajjad Khan, Junsu Kim, and Su Min Kim 


Research Article (10 pages), Article ID 8876824, Volume 2020 (2020)

**Provenance Transmission through a Two-Dimensional Covert Timing Channel in WSNs**

Qinbao Xu, Li Liu, Rizwan Akhtar, Muhammad Asif Zahoor Raja, and Changda Wang 


Research Article (9 pages), Article ID 8818374, Volume 2020 (2020)

**Towards the Design of Context-Aware Adaptive User Interfaces to Minimize Drivers' Distractions**

Inayat Khan and Shah Khusro 




Research Article (23 pages), Article ID 8858886, Volume 2020 (2020)



**Access and Use of Mobile Phone in Daily Life Activities by Rural Women of Gilgit-Baltistan, Pakistan**

Sabit Rahim , Sadruddin Bahadur Qutoshi, Syeda Abida, Faqeer Muhammad, and Imtiaz Hussain

Research Article (11 pages), Article ID 8835877, Volume 2020 (2020)


**A Novel Fuzzy Logic-Based Medical Expert System for Diagnosis of Chronic Kidney Disease**

Jimmy Singla , Balwinder Kaur, Deepak Prashar, Sudan Jha , Gyanendra Prasad Joshi , Kyungyun

Park , Usman Tariq, and Changho Seo 


Research Article (13 pages), Article ID 8887627, Volume 2020 (2020)

**New Method for Forest Resource Data Collection Based on Smartphone Fusion with Multiple Sensors**

Guangpeng Fan, Yanqi Dong, Danyu Chen, and Feixiang Chen 

Research Article (11 pages), Article ID 5736978, Volume 2020 (2020)

**A CRC-Based Classifier Micro-Engine for Efficient Flow Processing in SDN-Based Internet of Things**


Mahdi Abbasi , Navid Mousavi, Milad Rafiee, Mohammad R. Khosravi, and Varun G. Menon

Research Article (8 pages), Article ID 7641073, Volume 2020 (2020)

## Contents

---



### **A Review of Deep Learning Security and Privacy Defensive Techniques**

Muhammad Imran Tariq , Nisar Ahmed Memon, Shakeel Ahmed, Shahzadi Tayyaba, Muhammad Tahir Mushtaq, Natash Ali Mian, Muhammad Imran, and Muhammad W. Ashraf  
Review Article (18 pages), Article ID 6535834, Volume 2020 (2020)



## Research Article

# Parkinson's Disease Diagnosis in Cepstral Domain Using MFCC and Dimensionality Reduction with SVM Classifier

Atiqur Rahman,<sup>1</sup> Sanam Shahla Rizvi ,<sup>2</sup> Aurangzeb Khan,<sup>1</sup> Aaqif Afzaal Abbasi,<sup>3</sup> Shafqat Ullah Khan,<sup>4</sup> and Tae-Sun Chung <sup>5</sup>

<sup>1</sup>Department of Computer Science, University of Science and Technology Bannu, Bannu, Pakistan

<sup>2</sup>Raptor Interactive (Pty) Ltd., Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa

<sup>3</sup>Department of Software Engineering, Foundation University Islamabad, Islamabad 44000, Pakistan

<sup>4</sup>Department of Electronics, University of Buner, Buner, Pakistan

<sup>5</sup>Department of Software, Ajou University, Suwon, Republic of Korea

Correspondence should be addressed to Tae-Sun Chung; [tschung@ajou.ac.kr](mailto:tschung@ajou.ac.kr)

Received 15 July 2020; Revised 13 October 2020; Accepted 4 March 2021; Published 26 March 2021

Academic Editor: Carlos Tavares Calafate

Copyright © 2021 Atiqur Rahman 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.

Parkinson's disease (PD) is one of the most common and serious neurological diseases. Impairments in voice have been reported to be the early biomarkers of the disease. Hence, development of PD diagnostic tool will help early diagnosis of the disease. Additionally, intelligent system developed for binary classification of PD and healthy controls can also be exploited in future as an instrument for prodromal diagnosis. Notably, patients with rapid eye movement (REM) sleep behaviour disorder (RBD) represent a good model as they develop PD with a high probability. It has been shown that slight speech and voice impairment may be a sensitive marker of preclinical PD. In this study, we propose PD detection by extracting cepstral features from the voice signals collected from people with PD and healthy subjects. To classify the extracted features, we propose to use dimensionality reduction through linear discriminant analysis and classification through support vector machine. In order to validate the effectiveness of the proposed method, we also developed ten different machine learning models. It was observed that the proposed method yield area under the curve (AUC) of 88%, sensitivity of 73.33%, and specificity of 84%. Moreover, the proposed intelligent system was simulated using publicly available multiple types of voice database. Additionally, the data were collected from patients under on-state. The obtained results on the public database are promising compared to the previously published work.

## 1. Introduction

After Alzheimer's disease (AD), Parkinson's disease (PD) is the world's second most prevalent neurodegenerative disorder [1–3]. It has been reported that PD prevails at a rate of 0.3% in of the entire population in industrialized countries, while in elder population (60 or above age), the PD prevalence rate is 1% [1]. Impairments in voice have been reported to be the early biomarkers of the disease. Additionally, the proposed intelligent system has the capability to be used as an instrument for prodromal diagnosis. Notably, patients with REM sleep behaviour disorder (RBD) represent a good model as they develop PD with a high

probability. It has been shown that slight speech and voice impairment may be a sensitive marker of preclinical PD [4–7].

People with PD face numerous symptoms including movement impairments (gait and tremors), poor balance, bradykinesia which is slowness of movement, and rigidity [8–12]. As discussed above, the lack of reliable tests for diagnosis of PD has made the diagnosis of PD a challenging task [13–15]. However, recent research reported that PD patients manifest impairments in voice and speech. However, these voice defects cannot be detected in clinics by medical practitioners. Hence, automated signal processing tools are required to capture these impairments in voice and

to detect PD in its early stages. Recent research shows that machine learning and signal processing algorithms are successful in automated disease detection through automated risk factors extraction and classification [16–19]. Motivated by these studies, in this paper, we also attempt to develop a method based on machine learning and signal processing algorithms for PD detection.

The automated disease detection methods discussed above motivated us to develop automated model for PD detection using signal processing algorithms for feature extraction from voice signals and machine learning algorithms for classification. Hence, we collected a voice dataset, namely, Pak-Voice-PD that contains multiple types of vowel phonations for two types of subjects, i.e., healthy and PD patients. Numerical features are extracted using mel-frequency cepstral coefficients (MFCCs). In order to obtain better PD detection performance, we project the MFCC features to lower dimensional space using linear discriminant analysis (LDA) approach. Finally, numerous machine learning models are developed with the goal of obtaining an optimal learning model. Through performance analysis, we pointed out that support vector machine with linear and radial basis function (RBF) kernels provide optimal performance. Hence, in this study, we propose automated PD detection based on MFCC-LDA-SVM hybrid approach. The working of the proposed MFCC-LDA-SVM model is depicted in Figure 1.

The main contributions of this study are as follows:

- (1) Collection of a relatively larger dataset: the collected database has relatively larger number of multiple types of voice phonations or samples.
- (2) Construction of unbiased machine learning models for the automated detection of PD.
- (3) In this paper, we developed MFCC-LDA-SVM model for PD detection problem. To the best of our knowledge, no previous studies have explored development of MFCC-LDA-SVM model for PD detection based on voice data.
- (4) The proposed method, namely, MFCC-LDA-SVM has better performance than ten other machine learning models and many recently published studies.

The remaining of the manuscript presents related work in Section 2 and material and methods in Section 3. The evaluation and validation methods are briefly discussed in Section 4. Section 5 presents results of the proposed model and its discussion. Section 6 is about conclusion.

## 2. Related Work

During the last decade, various machine learning systems are proposed for the automated diagnosis of Parkinson's disease (PD) [20]. Resul [13] conducted a comparative study of different classification methods for effective diagnosis of the PD. Decision Tree, Regression, DMneural, and Neural Networks were evaluated for PD detection on the basis of performance scores. Neural network obtained the highest

classification score of 92.9% as compared to rest of classifiers. Tsanas et al. [21] presented speech signal processing algorithms for the prediction of PD symptom severity using random forests and support vector machines. The proposed algorithms were reported to have achieved classification accuracy of 99% using 10 dysphonia features. Kaya et al. [22] developed an entropy-based discretization method where support vector machines, C4.5, k-nearest neighbors, and naive boys were used as classifiers for the detection of PD. The proposed method was developed without using any preprocessing method. The discretization method improved the classification for diagnosis of PD by 4.1% to 12.8%.

Manda and Sairam [23] proposed a method for the early diagnosis of the PD based on the detection of dysphonia. A novel inference system measures the severity of disease through feature selection method based on support vector machines and ranker search method. Hariharan et al. [24] presented a hybrid intelligent system that consists of preprocessing through model-based clustering, feature selection using sequential forward selection, and linear discriminant analysis. For the classification purpose, least-square support vector machine (LS-SVM), probabilistic neural network (PNN), and general regression neural network (GRNN) are deployed. The maximum classification accuracy of 100% was achieved by the proposed method for Parkinson's dataset. Bhalchandra et al. [25] designed a system for early detection of Parkinson's disease (PD) using image processing to compute cheap-based features. Parkinson's progression markers initiative (PPMI) dataset was used along with a striatal binding ratio (SBR) to differentiate between the two types of subjects using discriminant analysis (DA) and support vector machine (SVM). The newly developed system observed the classification accuracy of 99.42%.

Saloni and Gupta [26] developed an algorithm for the detection of PD using clinical voice data. Voice features were used for the classification through support vector machines. The proposed algorithm achieved the accuracy of 100% for subset of features derived from the algorithm. Huang et al. [27] presented a framework for the prediction of Alzheimer's disease (AD) using nonlinear supervised sparse regression-based random forest (RF). The probabilistic paths are assigned using proposed soft-split technique to test sample in RF for more accurate prediction. The proposed soft-split sparse regression-based RF helped to estimate the missing scores. The proposed method demonstrated superior performance as compared to the traditional RF and regression models. Al-Fatlawi et al. [28] adopted deep belief network (DBN) for automated diagnosis of Parkinson's disease (PD). Voice data of Parkinson's disease patients are used for the experiments. The DBN classifier was composed of two stacked restricted Boltzmann machines (RBMs). The first stage is an unsupervised learning that used RBMs to eliminate the problems of the random value of initial weight. The second stage is a supervised learning based on the backpropagation algorithm for fine tuning. The accuracy reported by the proposed method was 94%.

Benba et al. [29] studied the discrimination between the two groups of people (patients with PD and healthy subjects) based on multiple types of voice samples. Human factor

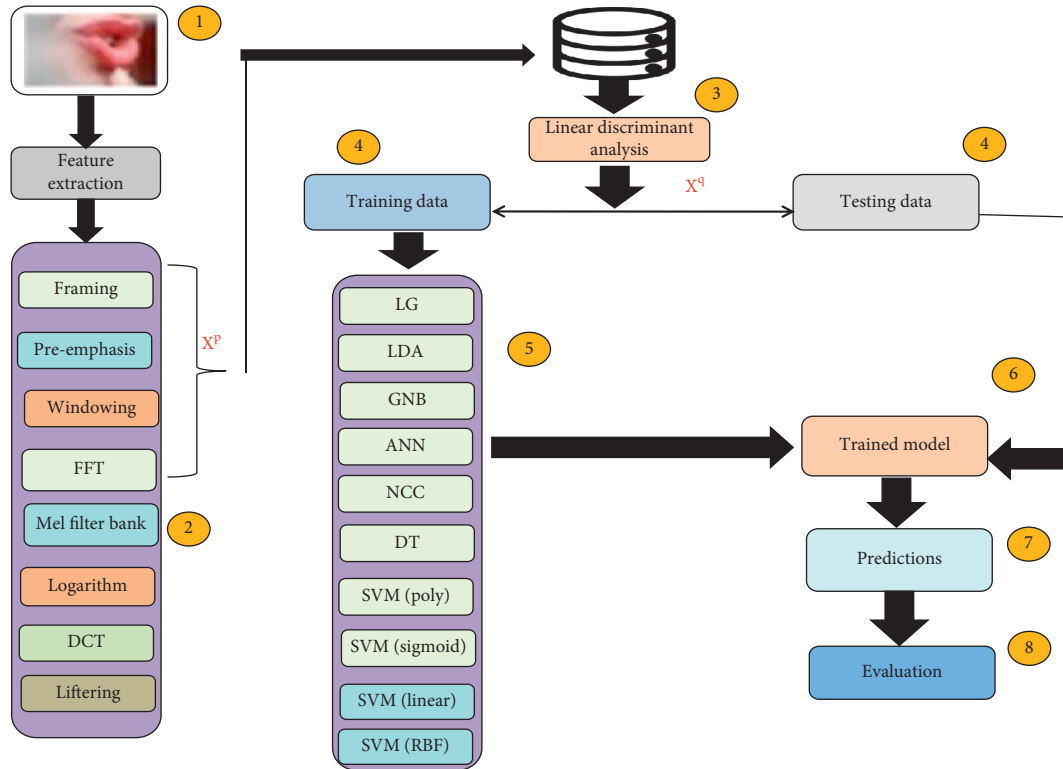


FIGURE 1: Block diagram of the proposed method with the numbering representing the flow of the whole process.

cepstral coefficients (HFCC) were used in the study. Voice print of the each voice recording was calculated for average value through the extracted HFCC. SVM with various kernels (RBF, Polynomial, Linear, and MLP) is deployed for the classification. The best accuracy of 87.5% was achieved through the linear kernel of SVM. Vaiciukynas et al. [30] adopted phonation corresponding to multiple types of vowel and speech tasks to pronounce short sentences in Lithuanian language. Random forest (RF) algorithm is utilized for the individual feature sets and decision-level fusion. It was pointed out that decision-level fusion provides better performance. Naranjo et al. [31] proposed a method for tracking Parkinson’s disease (PD) through Bayesian linear regression approach. The proposed method was suitable for the handling of replicated measurements. Li et al. [32] designed a hybrid feature learning algorithm for classification of PD. Hybrid features were developed through combining features and segments. Different methods were deployed for the selection of efficient hybrid features. The classification is made on the basis of selected hybrid features.

Zhang et al. [33] proposed a telediagnosis method through smart phone and machine learning-based Parkinson’s disease detection. Time frequency features, stack autoencoders (SAE), and k-nearest neighbor were used for the automated classification of the PD. The classification accuracy reported through proposed method was in the range from 94.00%–98.00%. In another study, Upadhyya et al. [34] adopted Single Taper Smooth (STS) window and Thomson Multitaper (TMT) windowing techniques for MFCC and PLP voice feature extraction. For classification,

neural network classifier was deployed for the classification of the subjects at the early stage of PD. Wu et al. [35] designed a feature learning technique for automatically learning about the extracted voice features. Spherical k-means model was deployed to train the two class sample space (PD patients and healthy subjects). The proposed method obtains the mean pooling accuracy of 95.35%. Ali et al. [20] studied the hand tremor abnormality detection associated with the risk of development of Parkinson’s disease using a Chi2-based feature selection and Adaboost-based classification. Khan et al. [36] proposed a method for the prediction of cancer and Parkinson’s disease. The proposed method utilized the wavelet-based neural networks for the prediction of cancer. The proposed evolutionary wavelet neural network was deployed on various biomedical benchmark datasets for breast cancer and Parkinson’s disease, while 10-fold cross-validation scheme was used for performance evaluation metric. The accuracy achieved by the proposed method was 90%.

Braga et al. [37] presented a methodology for early detection of Parkinson’s disease by using free-speech recording in uncontrolled background conditions. Machine learning (ML) algorithms along with signal and speech processing techniques were used for the early detection of the disease. For classification, support vector machine (SVM) and random forest (RF) were deployed. The accuracy reported by SVM (RBF) was 92.38% and 99.94% for RF. Recently, Ali [3] developed a hybrid intelligent system that carries out acoustic analysis of voice signals for automatically detecting Parkinson’s disease (PD). Linear discriminant

analysis (LDA) was adopted for the dimension reduction and genetic algorithm (GA) for fine tuning the parameter of neural network. Leave one subject out (LOSO) validation scheme was used to avoid the subject overlap. The proposed intelligent system achieved the classification accuracy of 80%. Mostafa et al. [38] presented a Multiple Feature Evaluation Approach (MFEA) and classification machine learning methods (Neural networks, Decision tree, SVM, and Random forest) based on the voice disorders analysis. The performance of the proposed method was evaluated through 10-fold cross-validation metric. The proposed system reported accuracy for SVM was 95.43%. Eskidere et al. [39] proposed a novel random subspace classifier ensemble and obtained 74.17% accuracy under 10-fold CV. Vadovský and Parali [40] utilized decision tree based methods, namely, C4.5, C5.0, Random Forest, CART, and obtained PD detection accuracy of 66.5% under 4-fold cross-validation. Kraipeerapun and Amornsamankul [41] proposed stacking of complementary neural networks (CMTNN) and obtained classification accuracy of 75% under 10-fold cross-validation.

The main problems in these studies were the inappropriate validation scheme that causes artificial subject overlap and biasedness in the developed models [2, 42]. Hence, the obtained results are biased due to the subject overlap between training and testing datasets. In order to develop unbiased machine learning models, Sarkar et al. proposed to use a more practical validation scheme, namely, Leave One Subject Out (LOSO) cross-validation [42]. Under their proposed LOSO approach, they trained and tested KNN and SVM classifiers on multiple types of speech data collected from two classes, i.e., healthy and PD patients and achieved 55% of PD detection accuracy, which are unbiased and more practical results. The same LOSO approach was adopted by Canturk and Karabiber in [43]. In order to improve the PD detection while developing unbiased machine learning methods, they explored integration of four different feature selection methods with six different machine learning models. They obtained best performance of 57.5 using LOSO approach. Recently, Ali et al. [44] proposed a multimodal approach under the LOSO approach and obtained unbiased performance of 70% classification accuracy using time frequency features.

### 3. Materials and Methods

**3.1. Data Acquisition.** In this study, we collected voice and handwritten-based database from two types of populations, i.e., PD patients and healthy subjects. The database was collected after the approval of ethical review board of Lady Reading Hospital (Medical Teaching Institution), Pakistan (Ref. No: 174/LRH, 2019). The database was collected from 160 subjects (60 PD patients and 100 age matched healthy subjects). The ages of the PD group of patients range from 43 to 88 with mean 68.3 and standard deviation of 10.4, while the ages of the healthy group of subjects range from 45 to 86 with mean 61.3 and standard deviation of 8.7. Moreover, the PD group contains data of 19 females and 41 males, while the data of healthy group contain 21 females and 79 males. The

data collection process was carried out using smart phones. The phone was kept at a distance of 10 cm from each subject during recording of the voice phonations. Each subject was asked to pronounce sustained phonations “a,” “o,” and “u.” Consequently, the database contains  $160 \times 3 = 480$  voice samples. Out of these 480 samples, 300 samples belong to healthy subjects and the remaining 180 samples belong to the patient group. The statistical information about the collected data have been reported in Table 1. Moreover, apart from using our own collected data, we also performed experiments on a bench mark dataset, namely, “multiple Types of Speech Dataset” [2].

**3.2. Proposed Method.** In this paper, we propose a three stage automated approach for PD detection. The first stage uses MFCC approach for feature extraction. The second stage is about dimensionality reduction through LDA, while the third stage is classification. In order to obtain better results, we explore the feasibility of various machine learning models at the third stage of the system. Hence, we developed ten different machine learning models. Based on the performance analysis, we pointed out that our proposed method, namely, MFCC-LDA-SVM approach, provides optimal PD detection. The proposed approach is depicted in Figure 1. The working of each stage of the proposed learning system is briefly discussed as follows.

**3.2.1. Feature Extraction through MFCC.** For extracting numerical features from the voice samples, we utilized the MFCC method. The MFCC algorithm establishes the relationship between perceived frequency and pitch of a pure tone as a function of its acoustic frequency. A subjective pitch is measured in the mel scale in units called mel. The mel for a given frequency  $f$  in Hz can be calculated using the following approximate formula [45]:

$$f_{\text{mel}} = 2595 \times \log_{10} \left( 1 + \frac{f_{\text{Hz}}}{700} \right). \quad (1)$$

**Framing:** according to [46], it takes a long period of time to examine the voice signals. This is because the voice signals are not stationary. Hence, it is necessary to move on with a short time analysis (generally, from 10 ms to 30 ms). The rate of movement of the voice articulators is limited by physiological limitations and can be considered stable within an interval from 10 to 30 ms. Therefore, the analysis of voice signal is carried out within uniform frames of this interval. In frame blocking, the voice signal is divided into frames of  $N$  samples. Neighboring frames should be separated by  $M$  ( $M < N$ ).

**Pre-emphasis:** in this step, we emphasize the higher frequencies by applying the first-order difference equation to the voice samples. This is to increase the energy in the voice signal.

The difference equation to voice signal ( $S_n, n = 1, \dots, N$ ) is given in equation (2) [47] as follows:

$$s'_n = s_n - k \times s_{n-1}, \quad (2)$$

TABLE 1: Summarized statistical information about the voice database.

Information	PD patients	Healthy subjects
No. of subjects	60	100
Age ( $\mu \pm \sigma$ )	68.3 $\pm$ 10.4	61.3 $\pm$ 8.7
Range of age	43–88	45–86
MDS-UPDRS-III ( $\mu \pm \sigma$ )	15.4 $\pm$ 4.7	—

where  $k$  is the pre-emphasis coefficient, and it should be within the range of  $0 \leq k < 1$ . Following the approach of [29], in this work, we used a pre-emphasis coefficient of  $k = 0.97$ .

Windowing: in order to minimize disrupts at the ends and make them continuous enough to correlate with the beginnings, windowing must be applied. Ideally, there exist several window functions (flat top window, hamming window, and rectangular window); however, the hamming window is used in our study for carrying out windowing. It is used to abate (decrease) signal to zero at the beginning and end of each frame and be represented as follows:

$$s'_n = \left\{ 0.54 - 0.46 \cdot \cos\left(\frac{2\pi(n-1)}{N-1}\right) \right\} \cdot s_n, \quad (3)$$

where  $s_n$  is the voice samples and  $n = 1, N$ .

Fast Fourier transform: the main purpose of FFT is to have a look at frequency domain information when the given signal information are in time domain. For this purpose, we will have to convert into frequency domain each frame having  $N$  samples. Compared to DFT, i.e., discrete Fourier transform, FFT is a faster algorithm on the given set of  $N$  samples [46, 47]:

$$s_n = \sum_{k=0}^{N-1} s_k e^{-2\pi jkn/N}, \quad (4)$$

where  $n = 0, 1, 2, \dots, N-1$ .

Mel scale/filter bank analysis: here, the approximation about the existing energy at each spot is determined. Thus, the spectrums calculated above are mapped on a mel scale using a triangular overlapping window, i.e., triangular filter bank (FB). The FB consists of a number of band pass filters with spacing along with bandwidth which is decided by steady mel frequency time. The mel frequency scale takes a linear spacing for frequency values below 1000 Hz and logarithmic spacing for values above 1000 Hz. To convert a given frequency ( $f$ ) to a mel frequency ( $m_f$ ), we used the approximate equation (1) [29].

Logarithm/DCT: with the intension of back conversion to spatial domain from the log mel spectrum, discrete cosine transform is brought into account for evaluating coefficients from the spectrum. Thus, we calculate the MFCC from the amplitudes of the log filter banks [15]:

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^N m_j \cdot \cos\left(\frac{\pi i}{N} (j - 0.5)\right). \quad (5)$$

Liftering: lack of correlation among the cepstral coefficients is the key advantage. However, the fact that the cepstral coefficients of higher order are fairly small is the

main problem. Hence, rescaling of the coefficients is necessary in order to have quite similar magnitudes [29, 45]. There is, therefore, the need to apply liftering to the cepstral coefficients using the following equation:

$$c'_n = \left(1 + \frac{L}{2} \times \sin\left(\frac{\pi \cdot n}{L}\right)\right) \times c_n, \quad (6)$$

where  $L$  is the cepstral sine lifter parameter.

3.2.2. *Linear Discriminant analysis (LDA)*. LDA is a supervised ML technique that is mostly used for classification and dimensionality reduction. The working of LDA is based on linear transformation of data (features) into small dimensional space, for maximum discrimination between classes [48]. LDA, in machine learning, is search for the vectors based on linear combination of features in vector space that separates two or more classes. Furthermore, original data values are plotted on the vectors for evaluation of the classes division. When classes are overlapped on the particular data values, then transformation mechanism is adopted by the LDA for better separation of the classes. To achieve the better separation between the classes, LDA deploys a rule known as the Fisher ratio. The maximum value of the Fisher ratio means maximum distance between the two classes. Equation (7) is the formulation of the Fisher ratio:

$$\frac{(v_1 - v_2)^2}{\rho_1^2 + \rho_2^2}, \quad (7)$$

where  $\rho_1$  and  $\rho_2$  denote the variance of 1<sup>st</sup> and 2<sup>nd</sup> class, while  $(v_1 - v_2)$  is the difference between the means of the two classes.  $\rho_1^2 + \rho_2^2$  is the sum of classes scatter. For example,  $\delta_m$  tries to compact two classes by reducing  $(v_1 - v_2)$  and  $\delta_s$  tries to minimize the class scatter. For detailed formulation and discussion about LDA, readers can refer to [3].

LDA has the following two benefits. Firstly, the performance of the predictive model is enhanced by LDA through transforming the original feature dimension into reduced dimensional space, where the class division is maximized. Secondly, time complexity of the predictive model reduced tremendously by LDA. Reduced dimensionality data by the LDA are supplied to the SVM for classification.

3.2.3. *Support Vector Machine*. Support vector machines (SVMs) are considered powerful learning methods and have been widely used in different biomedical- and health informatics-related problems [49]. During the training process, the output of an SVM model is an optimal hyperplane that could augment the distance of any class from the nearest training data points. The major reasons that motivate machine learning researchers to use SVM for their problems are as follows. (1) The first reason is that SVMs have powerful generalization capabilities to unseen data. (2) The second reason is the dependence of SVMs on a very small number of hyperparameters [50, 51].

Consider a dataset  $D_S$  with  $S$  instances,  $D = \{(x_i, y_i) | x_i \in R^Q, y_i \in \{-1, 1\}\}_{i=1}^S$ , where  $x_i$  stands for  $i^{\text{th}}$  instance,  $Q$  represents the dimension of the original features space of PD data, and  $y_i$  denotes the class labels, i.e., presence or absence of PD disease. The  $Q$  value is 20 for the PD dataset considered in this paper. The SVM model determines a hyperplane calculated by  $f(x) = \theta^T * x + \delta$ , where  $\delta$  represents the bias and  $\theta$  denotes the weight vector. Based on training data, the hyperplane  $f(x)$  of the SVM model augments the margin whereas curtails (reduces) the classification error. Sum of the distances to one of the closest negative and one of the closest positive instances is regarded as margin. The margin is defined as the sum of the distances between the closest negative and closest positive instances. That is, the hyperplane augments the margin distance  $2/\|\theta\|_2$ .

SVM uses a set of lax variables denoted by  $\xi_i, i = 1, \dots, S$ , and a penalty parameter, i.e.,  $C$ , and attempts to parity the minimization of  $\|\theta\|_2^2$  and minimization of the misclassification errors. This fact is formulated as follows:

$$\min_{\theta, \delta, \xi} \underbrace{\frac{1}{2}\|\theta\|_2^2}_{\text{Regularizer}} + C \underbrace{\sum_{i=1}^S \xi_i}_{\text{ErrororLoss}} \quad s.t. \begin{cases} y_i (\theta x_i + \delta) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, \dots, S \end{cases} \quad (8)$$

In equation (8),  $\xi$  is lax variable that calibrates the degree of misclassification and Euclidean norm or  $L_2$ -norm is the penalty term.

#### 4. Validation and Evaluation of the Proposed Approach

In order to validate the effectiveness of the proposed approach, we utilized leave-one-subject-out (LOSO) validation scheme in which the data of the one subject (all samples) are left out for testing and the proposed framework is trained on the remaining data. The process is repeated till the point where all the subjects have been tested. At the end, the final accuracy of the model is evaluated by calculating the mean accuracy for all the subjects.

To evaluate the performance of the proposed framework, we utilize some well-known statistical metrics, namely, Mathews Correlation Coefficient (MCC), sensitivity, specificity, and classification accuracy. Classification accuracy gives the precision with which the proposed method can classify all subjects (including patients and healthy). On the contrary, specificity tells us about how precise the model can classify healthy subjects and sensitivity tells us about how precise the developed model can classify patients. If  $A$  denotes the number of true positives,  $B$  denotes the number of true negatives,  $C$  denotes the number of false positives, and  $D$  denotes the number of false negatives, then the formulation of these evaluation metrics is given in equations (9)–(12):

$$\text{Acc} = \frac{A + B}{A + B + C + D}, \quad (9)$$

$$\text{Sn} = \frac{A}{A + D}, \quad (10)$$

$$\text{Sp} = \frac{B}{B + C}, \quad (11)$$

$$\text{MCC} = \frac{A \times B - C \times D}{\sqrt{(A + C)(A + D)(B + C)(B + D)}}, \quad (12)$$

where MCC is a value in the range  $-1$  to  $1$ , where  $-1$  denotes the worse case and  $1$  denotes the best case.

### 5. Experiment Results

In this section, we discuss the implementation details and the obtained performance of different developed machine learning models for the problem of PD detection based on the voice data. All the experiments were performed using Intel (R) Core (TM) m3-7Y30 CPU @ 1.00 GHz 1.61Ghz with memory of 8 GB and operating system of 64 bit Windows. All the experiments were performed using Python programming package and scikit-learn library.

The first experiment was performed by extracting the MFCC features from the voice phonations. The extracted MFCC was in the form of a matrix for each voice phonation. The matrix contained 20 columns which act as MFCC features. Following the approach of previous studies, we evaluated mean for each column or MFCC feature along the rows of the matrix. In this way, we obtained a feature vector of size equal to 20 for each voice phonation. Next, we used iterative feature selection before application of LDA for dimensionality reduction. After dimensionality reduction through the LDA model, we applied the resultant feature vectors at the input of machine learning models. The results for each of the developed machine learning models are given in Table 2.

After observing the results given in Table 2, it can be seen that the worst performance was produced by the GNB model and SVM with sigmoid kernel which are 48.12% accuracy and 46.87%, respectively, while the optimal performance is produced by the SVM model with RBF kernel which is 77.5% accuracy, 84% specificity, and 74.33% sensitivity. It means the proposed MFCC-LDA-SVM model can correctly classify 124 subjects out of the total 160 subjects. Similarly, the specificity value of 80% reveals that out of 100 healthy subjects, 80 are correctly classified, while the sensitivity rate of 73.33 reveals the fact that out of 60 PD patients, the proposed model can successfully detect 44 PD patients correctly. These statistical results are more clearly depicted in the confusion matrix given in Figure 2.

The performance of the MFCC-LDA-SVM model is further evaluated in terms of area under the curve (AUC) matrix which was calculated from the receiver operating characteristic curve (ROC curve). The ROC curve for the two models with worse performance and the ROC curve for the two models with optimal performance is given in Figures 3 and 4, respectively. It is important to note that a model with higher AUC is decided as a much better model than those models which are having lower values of AUC. Based on these evaluation criteria, we can see in the Figures 3 and 4 that the proposed MFCC-LDA-SVM is an optimal model when compared with other developed models. Additionally,

TABLE 2: Evaluation of different models in terms of PD detection based on voice data coeff, MFCC coefficients, selected: size of subset of features, Acc (%): percentage of correctly classified subjects, MCC: Mathews correlation coefficient, Sn: sensitivity and Sp: specificity, and HP: hyperparameters

Method	Coefficients	Acc (%)	Sp (%)	Sn (%)	Hyperparameters
LR	7	50.00	60.00	33.33	$C = 0.0001$
DT	12	51.25	60.00	36.66	$d = 10, l = 22$
GNB	1	48.12	60.00	28.33	—
LDA	15, 17, 19	62.25	60	66	—
NCC	11	50.00	50	50	—
ANN	11, 17, 18, 20	65.62	85	33	$H = 100$
SVM (Lin)	10, 15, 16	68.75	90	33.33	$C = 10$
SVM (Pol)	8	63.75	90	20	degree = 7
SVM (Sig)	14, 15	46.87	70	8.3	$C = 0.001$
SVM (RBF)	10, 12, 18	77.50	84	73.33	$C = 0.01, G = 0.0001$

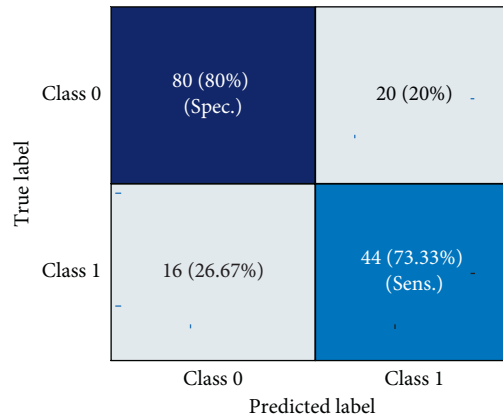


FIGURE 2: Graphical depiction of statistics of the obtained results for the proposed diagnostic system in terms of confusion matrix. Spec: Specificity and Sens: Sensitivity.

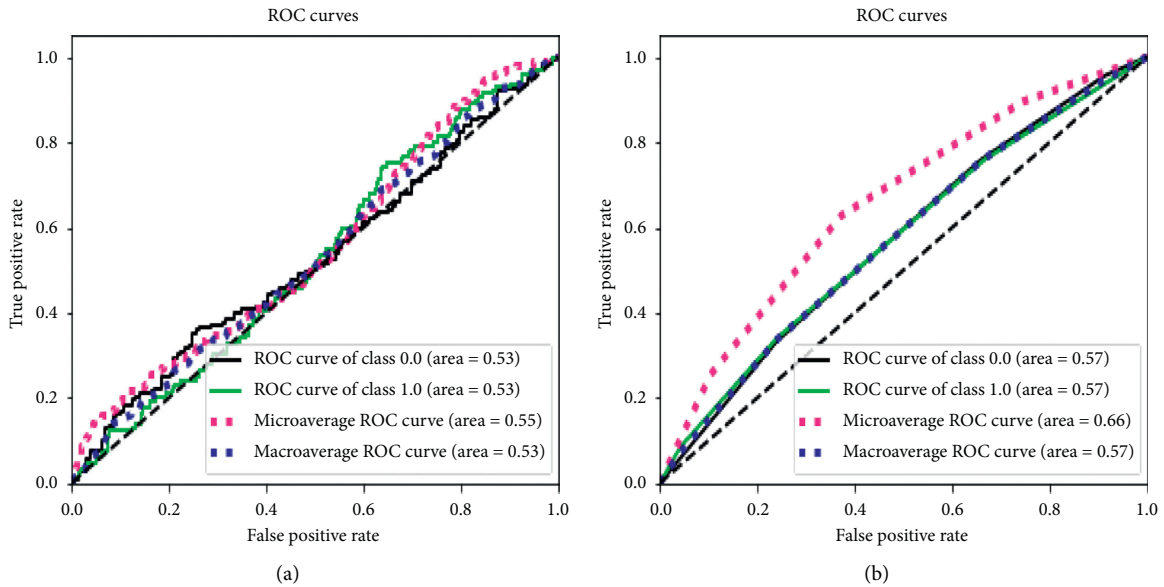


FIGURE 3: ROC charts of the two models with least performance. (a) ROC chart of the GNB model. (b) ROC chart of the LR model.

for further validation of the proposed approach, it is compared with recently published studies shown in Table 3.

The data were collected by different individuals who had different smart phones for recording the voice data. It is a

well-known fact that spectral characteristics of the microphone can highly influence the results, especially considering that MFCCs have been used in the study. These factors can degrade the performance of the proposed intelligent

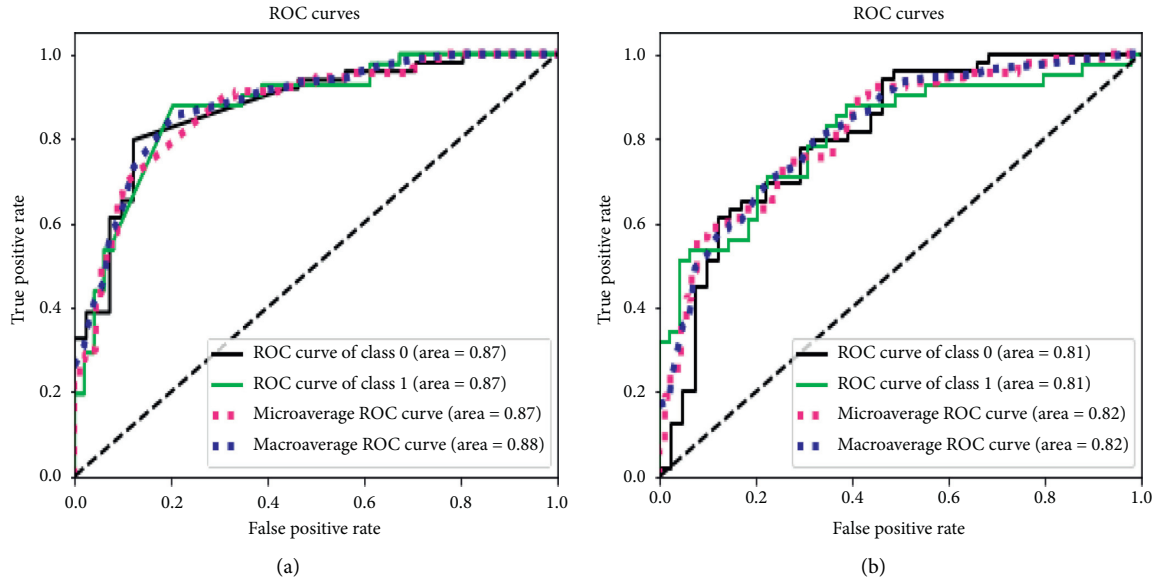


FIGURE 4: ROC charts of the two models with higher performance. (a) ROC chart of the SVM RBF model. (b) ROC chart of the SVM linear model.

TABLE 3: Performance comparison with recently published work.

Study	Method	Acc (%)	Sen. (%)	Spec. (%)
Sarkar et al. [42]	KNN + SVM	55.00 (LOSO on training database), 68.45 (LOSO on testing database)	60 (Training database)	50 (training database)
Canturk and Karabiber [43]	4 feature selection methods + 6 classifiers	57.5 (LOSO CV), 68.94 (10-fold)	54.28 (LOSO), 70.57 (10-fold)	80 (LOSO), 66.92 (10-fold)
Eskidere et al. [39]	Random subspace classifier ensemble	74.17 (10-fold CV)	Did not report	Did not report
Vadovský and Parali [40]	C4.5 + C5.0 + random forest + CART	66.5 (4-fold CV with pronouncing numbers), 65.86 (5-fold CV with pronouncing numbers)	Did not report	Did not report
Kraipeerapun and Amornsamankul [41]	Stacking + complementary neural Networks (CMTNN)	Average 75% (10-fold CV)	Did not report	Did not report
Ali et al. [44]	Multimodal approach	70	Not reported	Not reported
Benba et al. [29]	HFCC-SVM	87.5	90.00	85.00
Li et al. [32]	SVM + FS	82.50	85.00	80.00
Ali et al. [3]	LDA-NN-GA	95.00	95.00	95.00
Proposed method	MFCC-LDA-SVM	<b>97.5% (LOSO CV)</b>	<b>100.0%</b>	<b>97.5%</b>

system. Furthermore, the shorter length of phonations in PD could be another factor influencing cepstral analysis. To check the strength of our model, we simulated the same model on a publicly available dataset, namely, “Multiple Types of Speech Dataset” [42]. The proposed intelligent system, i.e., MFCC-LDA-SVM obtained outstanding results on the publicly available dataset. Using LOSO CV on the training dataset of the “Multiple Types of Speech Dataset,” we obtained 97.5% of accuracy, 100% sensitivity, and 95% specificity. Similarly, the proposed intelligent method produced accuracy of 89.28% on the testing dataset of the “Multiple Types of Speech Dataset.”

## 6. Conclusion

In this study, we considered the challenge of PD detection based on multiple types of voice signals. From each subject,

we recorded three different voice phonations. Signal processing algorithm (MFCC) was utilized to extract numerical features from the voice phonations. The extracted MFCC features were dimensionality reduced through the application of the linear discriminant analysis (LDA) model. At the final stage, numerous machine learning models were developed. It was pointed out that the MFCC-LDA-SVM method produces optimal performance in terms of PD detection. The performance comparison was carried out using different evaluation criteria including classification accuracy, area under the curve (AUC), and receiver operating characteristics curve. The proposed method produced AUC of 87%, PD detection accuracy of 78.5%, sensitivity of 73.33%, and specificity of 80%. Moreover, the proposed intelligent system was also simulated on the publicly available dataset. The obtained results were promising compared to the previous work.



## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

Atiqur Rahman would like to specially thank Dr. Amjad Iqbal from LRH Hospital for his support and guidance during data collection process and Dr. Akhtar Ali from the Northumbria University for his guidance and discussion. This work was supported by the Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1F1A1058548).

## References

- [1] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *The Lancet Neurology*, vol. 5, no. 6, pp. 525–535, 2006.
- [2] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.
- [3] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.
- [4] S. Arora, F. Baig, C. Lo et al., "Smartphone motor testing to distinguish idiopathic rem sleep behavior disorder, controls, and pd," *Neurology*, vol. 91, no. 16, pp. e1528–e1538, 2018.
- [5] J. Ruzs, J. Hlavnička, T. Tykalová et al., "Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder," *Sleep Medicine*, vol. 19, pp. 141–147, 2016.
- [6] J. Ruzs, M. Novotny, J. Hlavnivcka, T. Tykalova, and E. Ruuvzivcka, "High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1319–1321, 2016.
- [7] J. Ruzs, J. Hlavnicka, T. Tykalova et al., "Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 8, pp. 1495–1507, 2018.
- [8] L. Cunningham, S. Mason, C. Nugent, G. Moore, D. Finlay, and D. Craig, "Home-based monitoring and assessment of Parkinson's disease," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 47–53, 2011.
- [9] Z. A. Dastgheib, B. Lithgow, and Z. Moussavi, "Diagnosis of Parkinson's disease using electrovestibulography," *Medical & Biological Engineering & Computing*, vol. 50, no. 5, pp. 483–491, 2012.
- [10] G. Rigas, A. T. Tzallas, M. G. Tsipouras et al., "Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 478–487, 2012.
- [11] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig et al., "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.
- [12] S. K. Van Den Eeden, C. M. Tanner, A. L. Bernstein et al., "Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity," *American Journal of Epidemiology*, vol. 157, no. 11, pp. 1015–1022, 2003.
- [13] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568–1572, 2010.
- [14] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of Parkinson's disease," *Expert Systems with Applications*, vol. 110, pp. 182–190, 2018.
- [15] L. Naranjo, C. J. Pérez, J. Martín, and Y. Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications," *Computer Methods and Programs in Biomedicine*, vol. 142, pp. 147–156, 2017.
- [16] L. Ali, I. Wajahat, N. Amiri Golilarz, F. Keshtkar, and S. A. C. Bukhari, "LDA-GA-SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Computing and Applications*, 2020.
- [17] T. Meraj, A. Hassan, S. Zahoor et al., "Lungs nodule detection using semantic segmentation and classification with optimal features," *Neural Computing and Applications*, 2019.
- [18] L. Ali and S. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *IRBM*, 2020, In press.
- [19] L. Ali, S. U. Khan, N. A. Golilarz et al., "A feature-driven decision support system for heart failure prediction based on  $\chi^2$  statistical model and Gaussian naive bayes," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6314328, 8 pages, 2019.
- [20] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, "Reliable Parkinson's disease detection by analyzing handwritten drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model," *IEEE Access*, vol. 7, pp. 116480–116489, 2019.
- [21] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.
- [22] E. Kaya, O. Findik, I. Babaoglu, and A. Arslan, "Effect of discretization method on the diagnosis of Parkinson's disease," *International Journal of Innovative Computing, Information and Control*, vol. 7, pp. 4669–4678, 2011.
- [23] I. Mandal and N. Sairam, "Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system," *International Journal of Medical Informatics*, vol. 82, no. 5, pp. 359–377, 2013.
- [24] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 3, pp. 904–913, 2014.
- [25] N. A. Bhalchandra, R. Prashanth, S. D. Roy, and S. Noronha, "Early detection of Parkinson's disease through shape based

- features from <sup>123</sup>I-Ioflupane SPECT imaging,” in *Proceedings of the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 963–966, IEEE, Brooklyn, NY, USA, April 2015.
- [26] R. Saloni and A. Gupta, “Detection of Parkinson disease using clinical voice data mining,” *International Journal of Circuits, Systems and Signal Processing*, vol. 9, 2015.
- [27] L. Huang, Y. Jin, Y. Gao et al., “Longitudinal clinical score prediction in Alzheimer’s disease with soft-split sparse regression based random forest,” *Neurobiology of Aging*, vol. 46, pp. 180–191, 2016.
- [28] A. H. Al-Fatlawi, M. H. Jabardi, and S. H. Ling, “Efficient diagnosis system for Parkinson’s disease using deep belief network,” in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1324–1330, IEEE, Vancouver, Canada, July 2016.
- [29] A. Benba, A. Jilbab, and A. Hammouch, “Using human factor cepstral coefficient on multiple types of voice recordings for detecting patients with Parkinson’s disease,” *IRBM*, vol. 38, no. 6, pp. 346–351, 2017.
- [30] E. Vaiciukynas, A. Verikas, A. Gelzinis, and M. Bacauskiene, “Detecting Parkinson’s disease from sustained phonation and speech signals,” *PLoS One*, vol. 12, no. 10, Article ID e0185613, 2017.
- [31] L. Naranjo, C. J. Pérez, and J. Martín, “Addressing voice recording replications for tracking Parkinson’s disease progression,” *Medical & Biological Engineering & Computing*, vol. 55, no. 3, pp. 365–373, 2017.
- [32] Y. Li, C. Zhang, Y. Jia, P. Wang, X. Zhang, and T. Xie, “Simultaneous learning of speech feature and segment for classification of Parkinson disease,” in *Proceedings of the 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, IEEE, Dalian, China, October 2017.
- [33] Y. Zhang, “Can a smartphone diagnose parkinson disease? a deep neural network method and telediagnosis system implementation,” *Parkinson’s Disease*, vol. 2017, Article ID 6209703, 11 pages, 2017.
- [34] S. S. Upadhyaya, A. N. Cheeran, and J. H. Nirmal, “Thomson multitaper MFCC and PLP voice features for early detection of Parkinson disease,” *Biomedical Signal Processing and Control*, vol. 46, pp. 293–301, 2018.
- [35] K. Wu, D. Zhang, G. Lu, and Z. Guo, “Learning acoustic features to detect Parkinson’s disease,” *Neurocomputing*, vol. 318, pp. 102–108, 2018.
- [36] M. M. Khan, A. Mendes, and S. K. Chalup, “Evolutionary wavelet neural network ensembles for breast cancer and Parkinson’s disease prediction,” *PLoS One*, vol. 13, no. 2, Article ID e0192192, 2018.
- [37] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, “Automatic detection of Parkinson’s disease based on acoustic analysis of speech,” *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 148–158, 2019.
- [38] S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., “Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson’s disease,” *Cognitive Systems Research*, vol. 54, pp. 90–99, 2019.
- [39] Ö. Eskidere, A. Karatutlu, and C. Ünal, “Detection of Parkinson’s disease from vocal features using random subspace classifier ensemble,” in *Proceedings of the 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)*, pp. 1–4, IEEE, Almaty, Kazakhstan, September 2015.
- [40] M. Vadovský and J. Paralič, “Parkinson’s disease patients classification based on the speech signals,” in *Proceedings of the 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMII)*, p. 000321, January 2017, Article ID 000326.
- [41] P. Kraipeerapun and S. Amornsamankul, “Using stacked generalization and complementary neural networks to predict Parkinson’s disease,” in *Proceedings of the 2015 11th International Conference on, Natural Computation (ICNC)*, pp. 1290–1294, IEEE, Zhangjiajie, China, August 2015.
- [42] B. E. Sakar, M. E. Isenkul, C. O. Sakar et al., “Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.
- [43] İ. Cantürk and F. Karabiber, “A machine learning system for the diagnosis of Parkinson’s disease from speech signals and its application to multiple speech signal types,” *Arabian Journal for Science and Engineering*, vol. 41, no. 12, pp. 5049–5059, 2016.
- [44] L. Ali, S. U. Khan, M. Arshad, S. Ali, and M. Anwar, “A multi-model framework for evaluating type of speech samples having complementary information about Parkinson’s disease,” in *Proceedings of the 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, IEEE, Swat, Pakistan, July 2019.
- [45] A. Benba, A. Jilbab, A. Hammouch, and S. Sandabad, “Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson’s disease,” in *Proceedings of the 2015 International Conference on Electrical and Information Technologies (ICEIT)*, pp. 300–304, IEEE, Marrakech, Morocco, March 2015.
- [46] C. S. Kumar and P. M. Rao, “Design of an automatic speaker recognition system using MFCC, vector quantization and LBG algorithm,” *International Journal on Computer Science and Engineering*, vol. 3, no. 8, p. 2942, 2011.
- [47] S. Gupta, J. Jaafar, W. F. wan Ahmad, and A. Bansal, “Feature extraction using mfcc,” *Signal & Image Processing: An International Journal*, vol. 4, no. 4, pp. 101–108, 2013.
- [48] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*, Springer-Verlag, New York, NY, USA, 2001.
- [49] F. S. Ahmad, L. Ali, Raza-Ul-Mustafa et al., “A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs),” *Journal of Ambient Intelligence and Humanized Computing*, 2020.
- [50] S. Maldonado, J. Pérez, R. Weber, and M. Labbé, “Feature selection for support vector machines via mixed integer linear programming,” *Information Sciences*, vol. 279, pp. 163–175, 2014.
- [51] L. Ali, A. Niamat, J. A. Khan et al., “An optimized stacked support vector machines based expert system for the effective prediction of heart failure,” *IEEE Access*, vol. 7, pp. 54007–54014, 2019.

## Research Article

# Modelling Reachability in Transport Networks: Using Alternative Visual Representations in Interactive Linked-Views to Gain Valuable Insights

Rehmat Ullah <sup>1</sup>, Laiq Hasan,<sup>1</sup> Farman Ullah,<sup>2</sup> Ajmal Khan,<sup>2</sup> and You-Ze Cho <sup>3</sup>

<sup>1</sup>Department of Computer Systems Engineering, University of Engineering and Technology Peshawar, Peshawar, Pakistan

<sup>2</sup>Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Attock, Pakistan

<sup>3</sup>School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, Republic of Korea

Correspondence should be addressed to You-Ze Cho; [yzcho@ee.knu.ac.kr](mailto:yzcho@ee.knu.ac.kr)

Received 29 July 2020; Revised 22 October 2020; Accepted 22 January 2021; Published 8 February 2021

Academic Editor: Juan Carlos Cano

Copyright © 2021 Rehmat Ullah 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.

Most people use maps for navigation. Geographic maps visually represent physical distance between locations. These maps sometimes provide a false impression of travel times. Two cities geographically close to each other might be “far apart” in terms of travel time because of slower connections, whereas two cities geographically distant might be “nearby” in terms of travel time because of faster connections. Under such circumstances, visualizing a transport network using time as a distance measure can make the transport network more understandable. This study integrates several (carto)graphic representations—a time line, a distance line, a time prism, a time cartogram, and a geographic map—in an interactive linked-views environment to model reachability in transport networks. A prototype is implemented in a web environment using D3.js. The implementation can be applied to any transport network. In this research, the approach is illustrated with railroad network data for the Dutch province of Overijssel. The solution provides an alternative and insightful perspective for analyzing the data. In addition to complementing a wide variety of methods to visualizing travel times, the approach could be applied in areas such as spatial analysis and transport planning.

## 1. Introduction and Background

Sources of geographical data are abundant, and its temporal component is often associated with change or movement. Movement can be continuous like wind and currents in the oceans or discrete having a clear origin and/or destination. This last type of movement can follow a fixed network, such as rail or road network, or be free like movement of animals. Due to the complexity of the data, there is a need for the development, application, and evaluation of interactive analytical cartographic representations in order to produce meaningful insights about the movement data and effectively support spatiotemporal inference and decision-making by people [1–3]. Of the many different available representations, most work very well with the locational and attribute component of data, whereas representations that also work with the data’s

temporal component have not been sufficiently developed [4, 5].

Most people use maps for navigation. When making a trip, travel time between two locations can in some cases be more important than physical distance between them. In geographic maps, however, travel time and distance do not correlate equally across the map. For instance, city A can be geographically close to city B, while city C is rather distant from B, but in terms of travel time, A might be farther apart than C because of slower or longer connections. To make these insightful, different methods can be used to visualize travel times.

A table can be provided to show travel times. However, in such a case, a user loses the geographical context. On a map, labels can be used to display travel times along segments of roads, in which case a user has to read the

individual labels to calculate travel times (see Figure 1(a)). Besides, labels take up space in the map and can make the map crowded because of a large number of labels. Another approach is to use labels at destinations only. In Figure 1(b), all stations reachable from Enschede have a label with the travel time needed to reach the particular station. A disadvantage of this solution is that it is only valid from a single location, here Enschede. In Figure 1(c), ten-minute isochrones have been drawn and time zones between the lines have been filled with color, applying the visual variable “value.” The darker the zone is, the more time it takes to reach the zone. This has the same disadvantage as with labels at destinations (Figure 1(b)). An alternative approach is the use of time cartograms [6–9]. In a time cartogram, the geographic distance between locations is replaced by travel time. This potentially distorts the geography accordingly. This distortion can give a clearer picture of what is nearby or distant in terms of travel time (see Figure 1(d)). The distortion of the geography, however, can make the area of interest difficult to recognize [8, 10, 11]. Similar to the examples in Figures 1(b) and 1(c), this solution is also valid from a single location only. However, this may not be a problem in an interactive environment where the user can change the starting location. Alternative to distorting the base map, only network segments can be distorted, keeping the locations fixed (see [12, 13]). However, the resulting maps from the transformation approach proposed by Buchin et al. [12] do not suit networks in which segments are dense and plenty because the curves might intersect, making the cartograms harder to read (see Figure 2(a)). Similarly, the resulting cartograms from the approach proposed by Wu and Hung [13] do not suit networks with long segments far away from each other or with irregular patterns (see Figure 2(b)). These observations made us look for other supplemental graphic representations to model reachability in transport networks.

For this purpose, we integrated several (carto)graphic representations—a time line, a distance line, a time prism, a time cartogram, and a geographic map—in an interactive linked-views environment (as shown in Figure 3). The departure time picker (labelled 1) indicates the time during the day. The picker can be moved back and forth to a particular time of the day. This will show the time of the first train leaving (e.g., 05:16 AM as indicated by 2). The geographic map (labelled 3) visualizes the railroad network and the spatial relationships of stations of the Dutch province of Overijssel. The travel origin or starting station (i.e., Enschede) is shown by a black square on the map and the other stations are represented by dots. The stations that are reachable at a selected time instant (here 05:16 AM) are colored green, and those that are not reachable are colored red. The map also shows both reachable and non-reachable subnetworks at a selected moment in time. The time cartogram (labelled 4) shows the reachable subnetwork from Enschede at a selected time instant. The concentric circles depict the travel times in steps of 10 minutes from Enschede to other destinations in the reachable network. The time line (labelled 5) shows the travel times in minutes from Enschede to only those stations that are reachable at a selected moment

in time. The distance line (labelled 6) shows the physical distances, based on crow flies, in kilometers from Enschede to all stations. On both lines, the stations that are reachable at a selected time instant are colored green. The distance line also shows the stations that are not reachable in red.

The check box (indicated by 7) can be used to select one station to see how the reachability of the selected station changes over the day. A station can be selected by checking the check box and then clicking on the desired station in the geographic map. The time budget picker (labelled 8) can be used to choose a time budget (the time that one could spare to travel, excluding sightseeing). Label 9 displays the chosen time budget. An individual can travel from the starting station to all reachable stations (indicated on the time line) within the time budget. The stations dynamically appear or disappear on the time line as soon as the time budget is increased or decreased using the time budget picker. To better use and understand the network and explore patterns in space and time, we linked all the views in the visualization environment. For example, whenever one moves the mouse over a station in any of the views, the corresponding station is also highlighted in the other views (e.g., Ommen in Figure 4).

The visualization can also be animated. The animation runs from 12:00 AM to 11:59 PM to show how the reachability of stations changes over time. An animate/pause button is provided to run/pause the animation at any moment in time.

Reachability in time-geography theory is indicated by what Hägerstrand [14] referred to as the space-time prism. The space-time prism determines the ability of individuals to travel to different locations in a network in a limited interval of time. Later, several researchers have attempted to operationalize and apply the space-time prism in spatial analysis and planning (e.g., [15–19]). However, the full potential of the space-time prism has not been realized in locational analysis and transportation planning [20–22].

This research introduces a linear version of the space-time prism by mapping the transport network onto time and distance lines. The modified version is termed *time prism*. An example of the time prism can be seen in Figures 5 and 6. Figure 5 shows the input network: the travel origin (i.e., Enschede) and all other stations are indicated. The stations that are reachable at the selected time instant are indicated by green dots, while the stations that are not reachable are indicated by red dots. The time line shows the travel times in minutes from Enschede to only those stations that are reachable. The distance line shows the physical distances in kilometers from Enschede to all stations. Figure 6 shows the network after the application of the prism with time budgets of 40 and 80 minutes, respectively. The time lines in both cases indicate the subnetworks that are reachable. One can travel from the origin to any station in the reachable subnetwork within the selected time budget. As evident, more stations can be reached when the time budget is increased from 40 minutes to 80 minutes (compare Figure 6(a) versus Figure 6(b)).

A prototype of the visualization system was implemented in a web environment using D3.js. The implementation can be applied to any transport network. In this

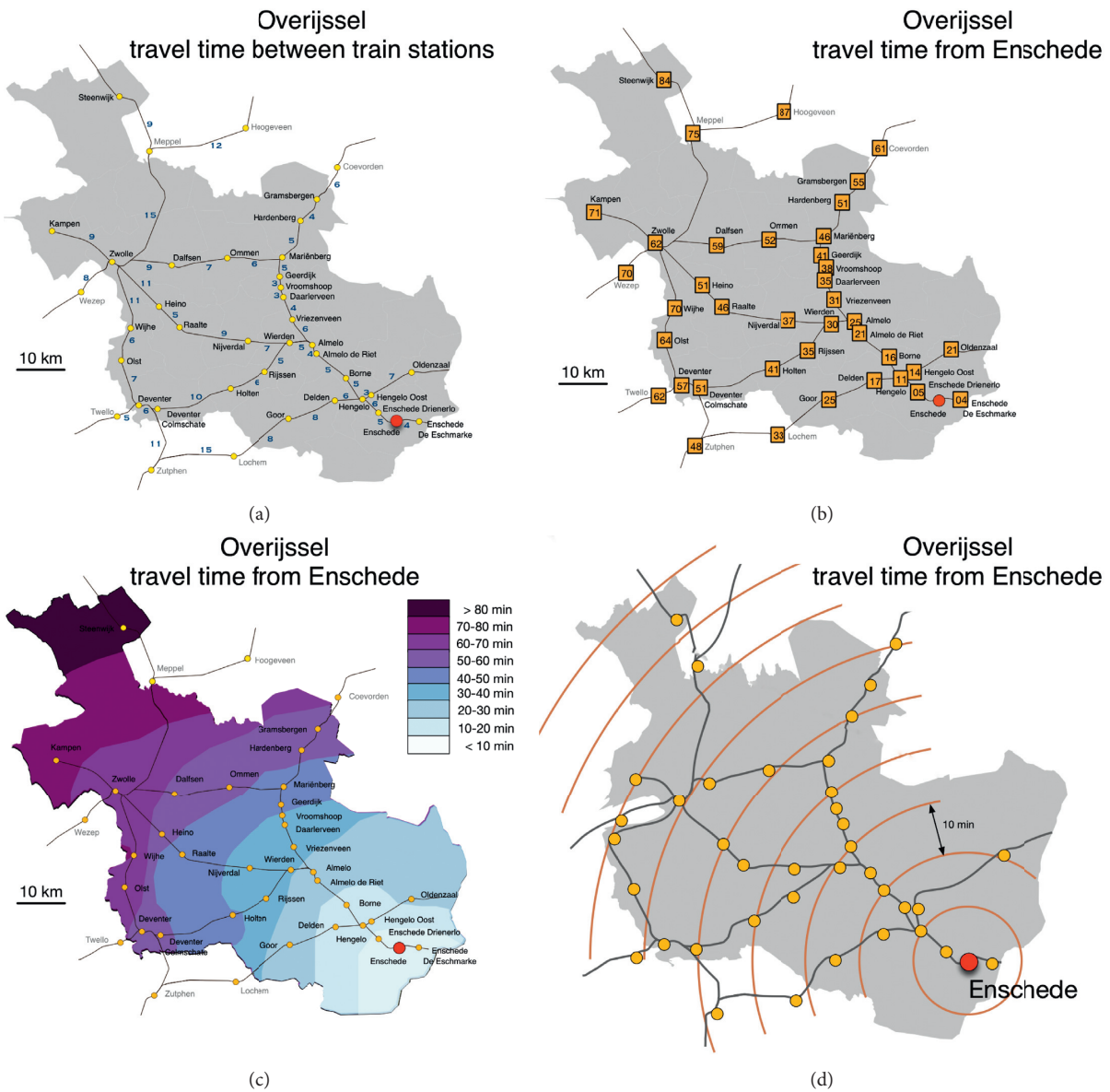


FIGURE 1: Mapping travel time by train from the city of Enschede to other towns in the Dutch province of Overijssel. (a) Labels along network segments. (b) Labels at destinations. (c) Isochrones. (d) A time cartogram [6].

study, the approach is illustrated with railroad network data for the Dutch province of Overijssel. Our solution provides an alternative and insightful perspective for analyzing the network. The proposed method solves the clutter and overplotting problems found in other representations, like the labelling. In addition to complementing a wide variety of methods to visualizing travel times, our approach could be applied in areas such as spatial analysis and transport planning. It can help transport planners to know how the reachability of stations changes over time and also to investigate which segments of the transport network are heavily-served and which are under-served. Using the visualization environment, the transport planners could figure out which stations take more time to reach despite the fact that they are geographically closer and could improve the connections if needed.

The remainder of this paper is organized as follows: Section 2 discusses the data used to illustrate the visualization and Section 3 explains the design and implementation of the visualization environment. In Section 4, we show how our visualization environment can be used to answer the user questions (listed in Table 1). Finally, in Section 5, we provide conclusions and look at possibilities for future work.

## 2. Data

A dataset of the railroad network of the Dutch province of Overijssel was used to illustrate the method. The dataset consists of the intercity and regular train stations and the trains' time tables plus an administrative map of the province. This network has 33 train stations and is shown in

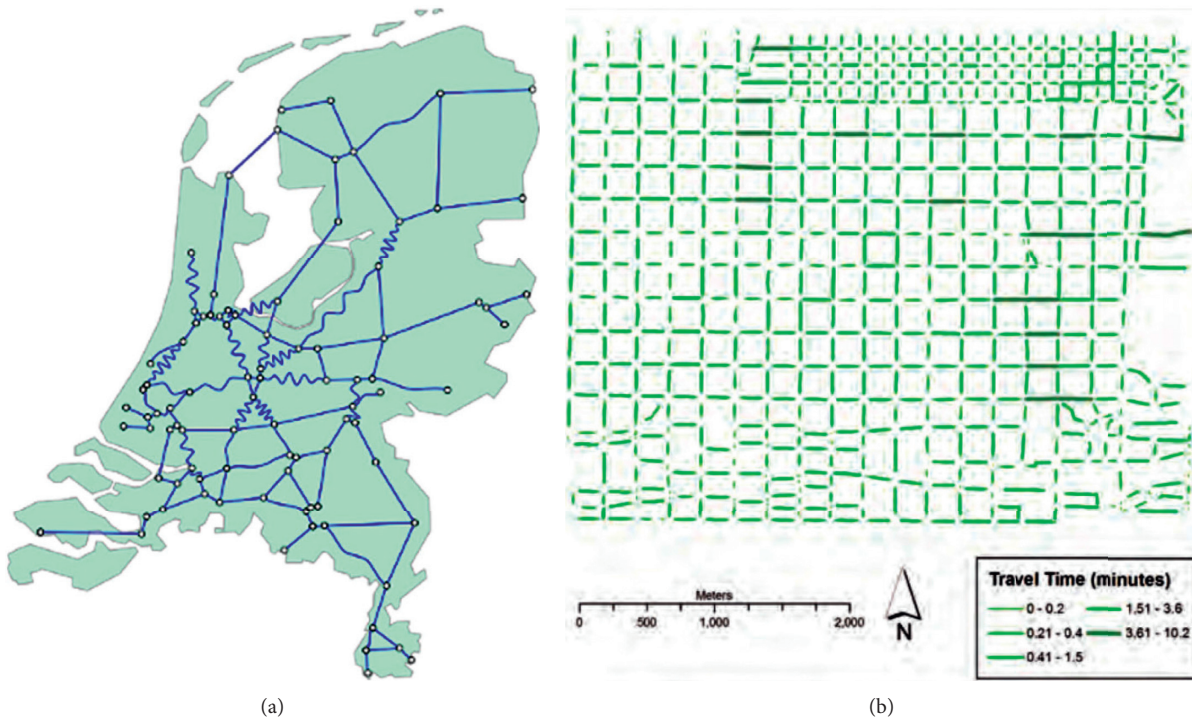


FIGURE 2: (a) A linear cartogram with fixed vertex locations showing travel times for the Dutch railroad network. The map and the locations are fixed, but edges are drawn as sinusoid curves to indicate travel times [12]. (b) A nonconnective linear cartogram showing travel times in Salt Lake City, Utah. Such cartograms do not show the connectivity between line segments. Lengths and widths of road segments are modified according to travel times [13].

Figure 1(a). We chose Enschede as the starting station. Both intercity and stop trains were considered to calculate the travel times from Enschede to all other destinations.

### 3. Design and Implementation

Unlike the present design approaches on visualizing travel times (e.g., [6, 7, 9, 12, 13, 23–31]), the design and implementation of the proposed visualization environment were approached by using a systematic, user-centric, and task-oriented visualization design framework (see, e.g., [32–35]). In the first step, a set of user questions (listed in Table 1) were formulated in coordination with both frequent and casual train travelers. The travelers, from 25 different nationalities, were postgraduate students and staff members of the Faculty of Geo-Information Science and Earth Observation at the University of Twente in the Netherlands. For details on this study, see our article “Usability evaluation of centered time cartograms” in the June 2016 issue of the *Journal of Open Geosciences*. Questions 1 to 5 are typical travel time questions, while others are more analytical questions. The questions were chosen to be real-world, with the intention to assess the practical value of the visualization approach. In the second step, the design and implementation of the visualization environment (largely influenced by the user

questions and users’ ease in interacting with it) were carried out to answer the user questions.

No one single visualization can completely address all the above questions without suffering from clutter and overplotting [35, 36]. These problems can partly be solved by combining different visual representations in an interactive linked-views environment [37]. In addition, the use of different interactive linked visual representations can be helpful in revealing patterns otherwise missed because they each provide their own unique perspective on the data [4, 38–40].

To implement the visualization, we used HTML5 for encoding web pages, CSS3 for styling and layout, SVG for vector graphics, and JavaScript for interaction and animation. Several JavaScript frameworks and libraries support Open Web Platform. We chose the D3 library [41]. D3.js is a JavaScript library for manipulating documents programmatically based on data. It allows binding arbitrary data to a Document Object Model (DOM) and then applying data-driven transformations to it. Its code structure, based on JavaScript framework jQuery, helps to produce dynamic and interactive data visualizations using the full capabilities of modern web standards such as HTML5, SVG, and CSS3 in modern web browsers [36]. Furthermore, D3 is fast and efficient, even for large datasets [42].

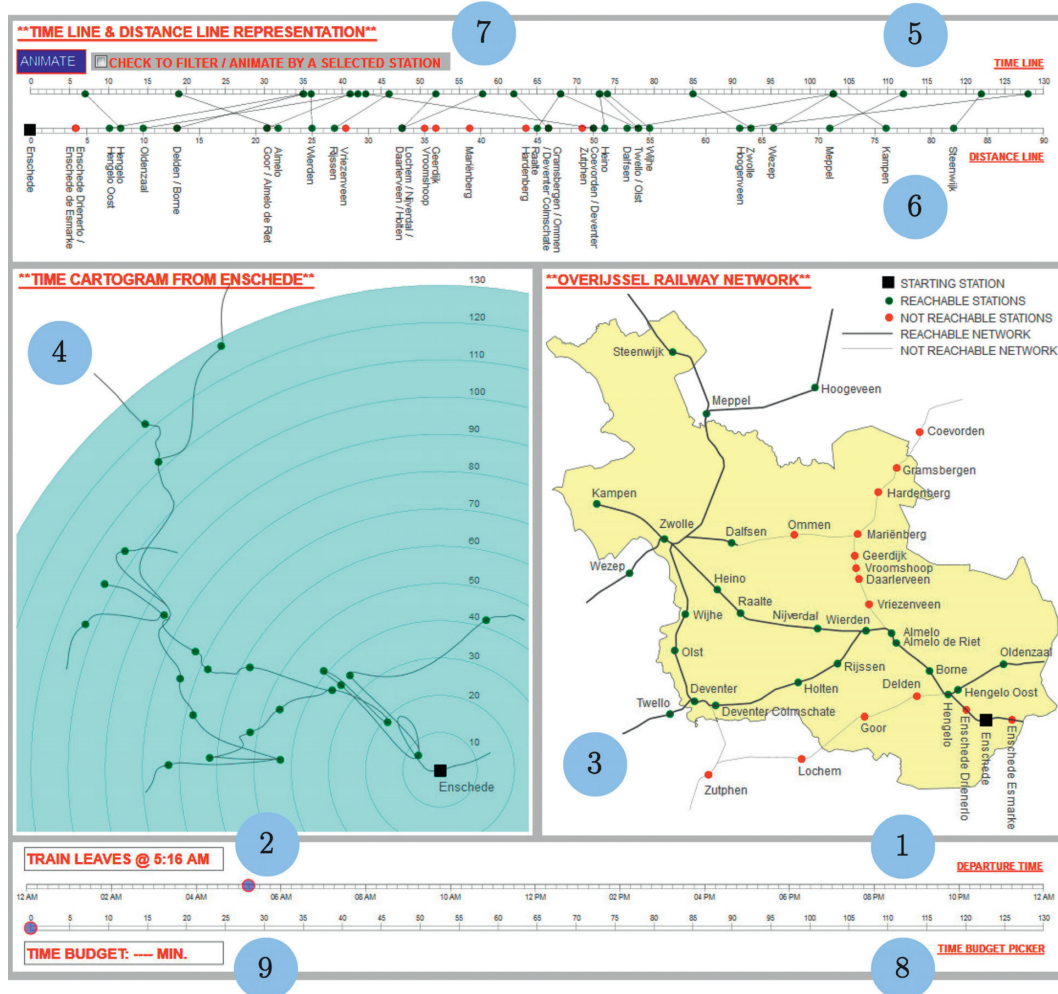


FIGURE 3: The visualization environment integrating the time line (labelled 5), the distance line (labelled 6), the time prism (labelled 1, 2, 5, 6, 8, and 9), the geographic map (labelled 3), and the time cartogram (labelled 4). All views are interactive and linked: moving the mouse over a station in any of the views highlights the corresponding station in the other views.

### 4. Working with the Visualization

Figure 3 shows the different elements of the visualization environment and their functionalities. Below, we discuss how our visualization environment can be used to answer the posed user questions (see Table 1).

4.1. UQ1: How Long Does It Take to Reach Station X with the First Train Departing after Time T? Figure 4 shows how one can find the travel time to station X with the first train departing after time T. Select the time instant on the time slider (e.g., 06:00 AM in Figure 4). The first train after the selected moment leaves at 06:04 AM. The time line and the time cartogram show the travel times of all the stations that are reachable from Enschede with the train leaving at 06:04 AM. The geographic map and the distance line show all the stations with the reachable stations in green and those not reachable in red. By selecting any reachable station in the map or time cartogram or on time line or distance line, one can see how much time is needed to reach the station. In

Figure 4, Ommen is selected. The time line and the time cartogram show that it takes 93 minutes to reach Ommen from Enschede with the train leaving at 06:04 AM. The distance line shows the physical distance from Enschede to Ommen (46 kilometers).

4.2. UQ2: Which Stations Are Reachable with the Train Departing at Time T? Clicking on the animate/pause button plays/stops the animation. The animation runs from 12:00 AM to 11:59 PM and shows how the reachability of stations changes over time. The time slider can also be moved forward and backward to interact with the animation. Figures 4 and 7 explain how an individual can know which stations are reachable with the train leaving at time T. Figure 4 shows all stations that are reachable with the train departing at 06:04 AM, while Figure 7 shows all stations that are reachable at 06:16 AM. The reachable stations are colored green, whereas those not reachable are colored red. One can see that fewer stations are reachable with the train departing at 06:16 AM than at 06:04 AM.

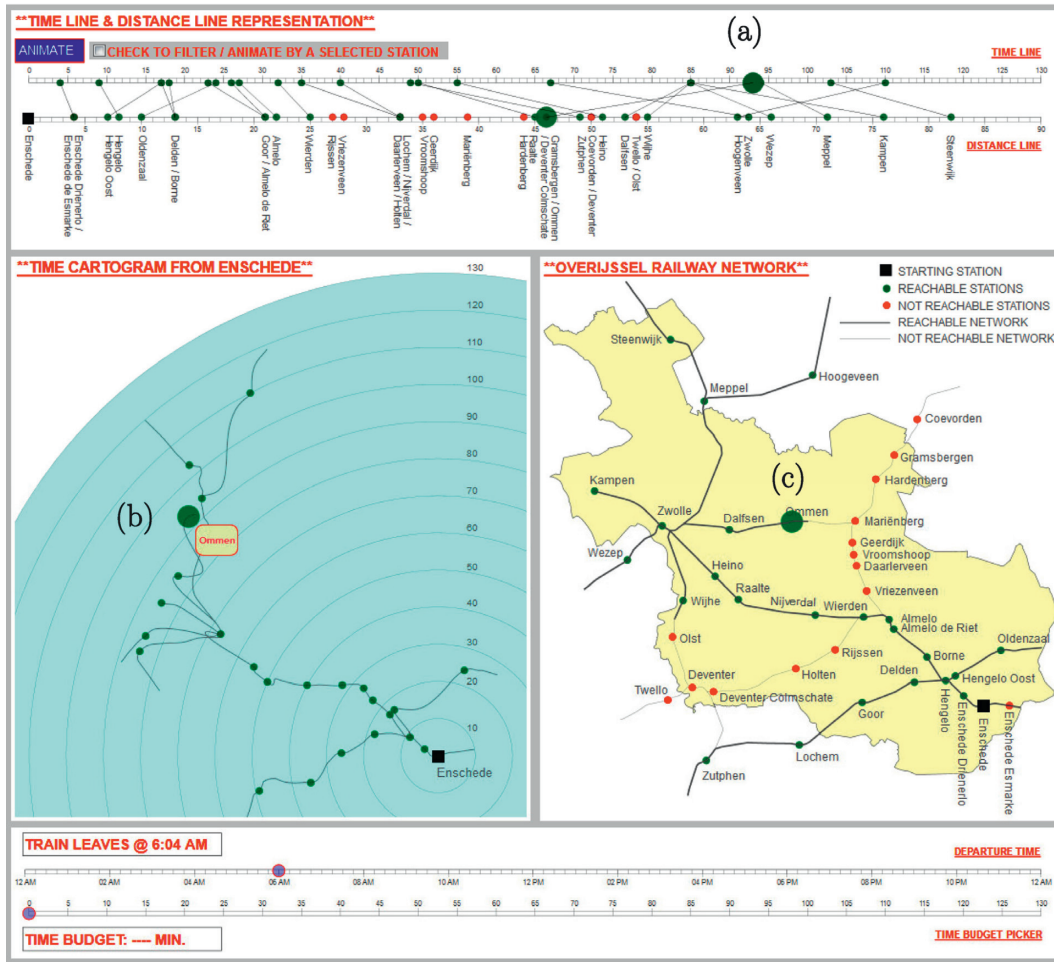


FIGURE 4: Travelling from Enschede with the train departing at 06:04 AM. The reachable stations are indicated by green dots, whereas those not reachable are indicated by red dots. The time line (a) and the time cartogram (b) show that it takes 93 minutes to reach the highlighted station (Ommen) from Enschede when leaving at 06:04 AM.

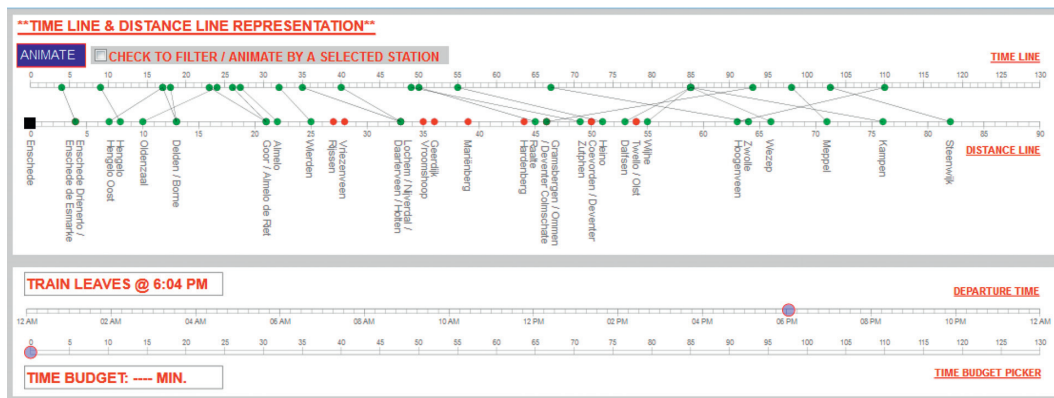
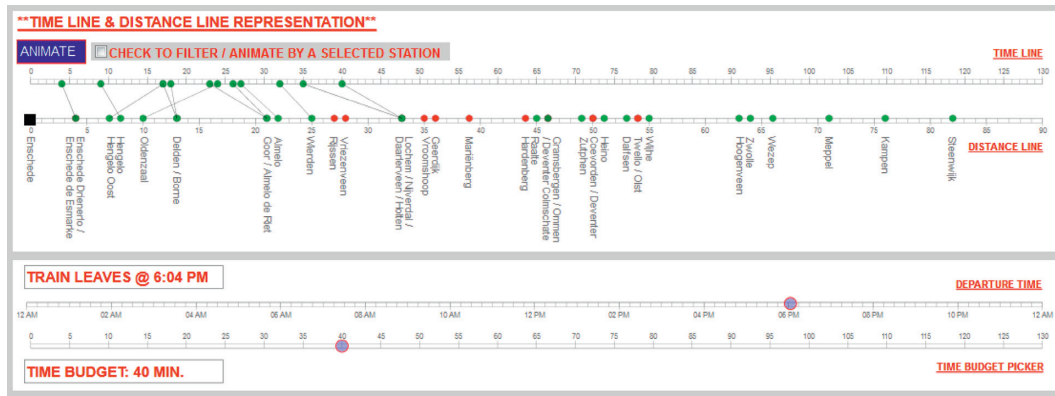
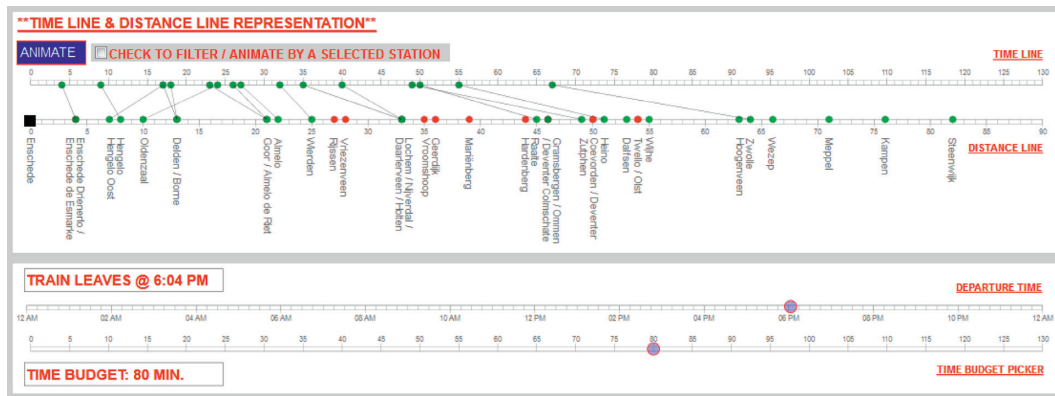


FIGURE 5: Travelling from Enschede with the train departing at 06:04 PM with no time budget constraints. The reachable stations at the selected moment in time are indicated by green dots, while the non-reachable stations are indicated by red dots. The time line shows the travel times in minutes from Enschede to only those stations that are reachable. The distance line shows the physical distances in kilometers from Enschede to all stations.





(a)



(b)

FIGURE 6: An example of the time prism showing the reachable subnetworks from Enschede with the train departing at 06:04 PM after the application of prism with time budgets of (a) 40 minutes and (b) 80 minutes.

TABLE 1: List of user questions.

User Questions	
UQ1	How long does it take to reach station X with the first train departing after time T? <i>Example: How long does it take to reach Ommen with the first train departing after 06:00 AM?</i>
UQ2	Which stations are reachable with the train departing at time T? <i>Example: Which stations are reachable with the train departing at 06:04 AM?</i>
UQ3	Which stations can be reached within X minutes with the train departing at time T? <i>Example: List all stations reachable within 60 minutes with the train departing at 06:04 AM.</i>
UQ4	When does it take the least or the most time to reach station X during time interval T1-T2? <i>Example: When does it take the least or the most time to reach Ommen between 06:00 AM and 07:00 AM?</i>
UQ5	Given a time budget of X minutes, which stations can be reached with the first train departing after time T? <i>Example: Robert is in Enschede to attend a 3-day workshop at the University of Twente in the Netherlands. The workshop runs from 09:00 AM to 05:00 PM daily. As he has never been in the Netherlands before, on one of the workshop days he wants to visit another place that is close to Enschede. He can only spare 40 minutes (single trip time only, excluding sightseeing) to travel and wants to know which cities are reachable by train from Enschede within his time budget.</i>
UQ6	How does the reachability of station X change during the day? <i>Example: How does the travel time to Almelo de Riet change over time?</i>
UQ7	Which station(s) or network segment(s) is/are heavily-served (receive more trains) or under-served (receive less trains)? <i>Example: Which segment of the network is heavily-served: segment A or segment B?</i>
UQ8	Why does it take less time to reach station X than station Y, despite X being farther away than Y in geography? <i>Example: Why does travelling by train from Enschede to Deventer Colmschate take longer than travelling from Enschede to Deventer (62 minutes versus 43 minutes), even though Deventer Colmschate is closer to Enschede than Deventer (46 kilometers versus 50 kilometers)?</i>

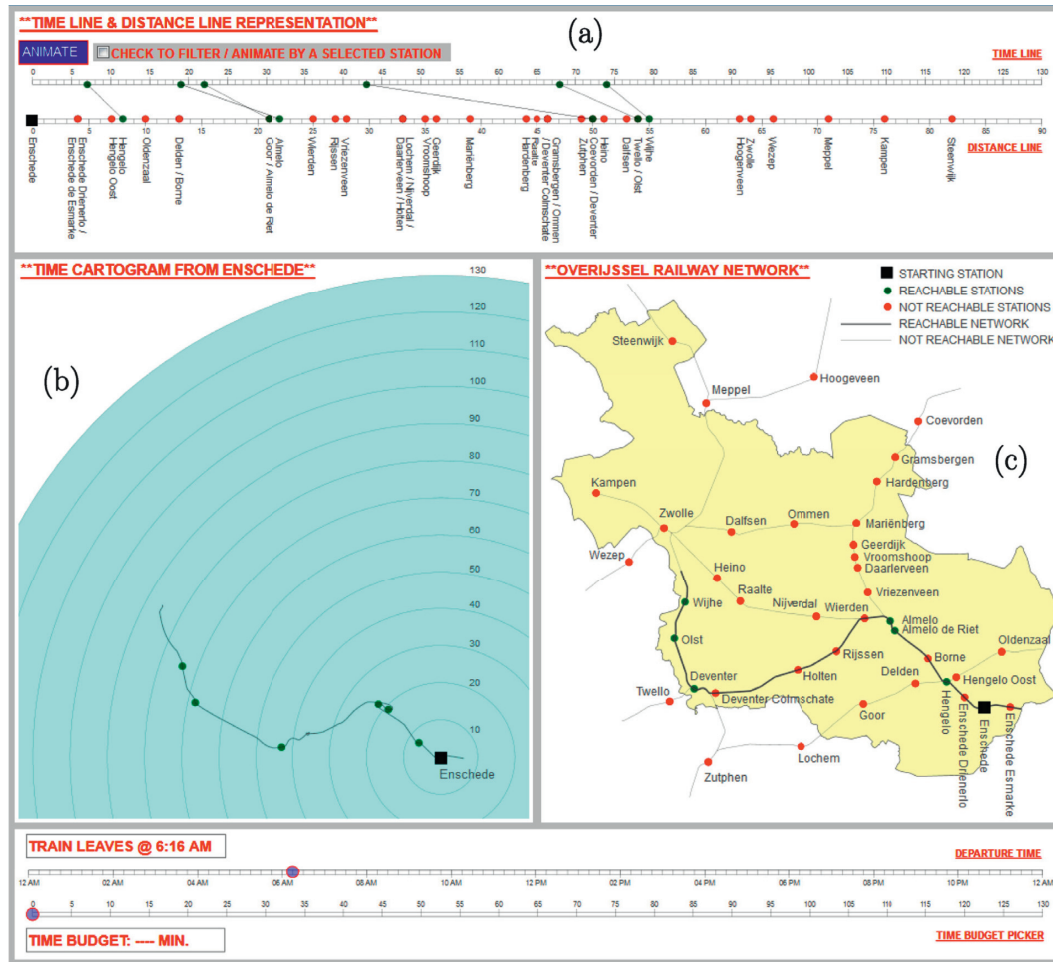


FIGURE 7: Travelling from Enschede with the train departing at 06:16 AM. The reachable stations are indicated by green dots, whereas those not reachable are indicated by red dots.

4.3. UQ3: Which Stations Can Be Reached within  $X$  Minutes with the Train Departing at Time  $T$ ? A traveler can in some cases be interested to find out which stations are reachable within  $X$  minutes with the train departing at time  $T$ . The numbers along the time line and the concentric circles in the time cartogram map show travel times in minutes of the reachable stations at a selected time instant. For instance, one can see that fifteen stations (namely, Enschede Drienerlo, Hengelo, Hengelo Oost, Oldenzaal, Borne, Delden, Almelo de Riet, Goor, Almelo, Wierden, Nijverdal, Lochem, Zutphen, Raalte, and Heino) are reachable within 60 minutes from Enschede with the train leaving at 06:04 AM (see Figures 4(a) and 4(b)). At 06:16 AM, only four stations (namely, Hengelo, Almelo de Riet, Almelo, and Deventer) are reachable within 60 minutes (see Figures 7(a) and 7(b)).

4.4. UQ4: When Does It Take the Least or the Most Time to Reach Station  $X$  during Time Interval  $T_1$ - $T_2$ ? The visualization environment can help an individual to figure out when it takes the most or the least time to reach station  $X$  during time interval  $T_1$ - $T_2$ . The time slider can be moved forward and backward to interact with the animation.

Clicking on the animate/pause button plays/stops the animation. By default, the animation runs from 12:00 AM to 11:59 PM and shows how the reachability of stations changes over time. The time slider and the animate/pause button can be used to run the animation for a particular time interval. In the graphics in Figures 4 and 8, we played the animation from 06:00 AM to 07:00 AM and the following patterns were observed. At 06:04 AM, it takes 93 minutes to reach Ommen (see Figure 4). At 06:19 AM, it takes 63 minutes to reach Ommen (see Figure 8). The geographic map and the time cartogram explain why it takes more time to reach Ommen at 06:04 AM than at 06:19 AM. At 06:04 AM, Ommen is reachable via Zwolle, while at 06:19 AM, Ommen is reachable via Mariënberg. This also indicates that the shortest path to Ommen is travelling via Mariënberg, but it takes the longest when travelling via Zwolle.

4.5. UQ5: Given a Time Budget of  $X$  Minutes, Which Stations Can Be Reached with the First Train Departing after Time  $T$ ? Sometimes travelers want to know which stations they can reach with the first train leaving after time instant  $T$

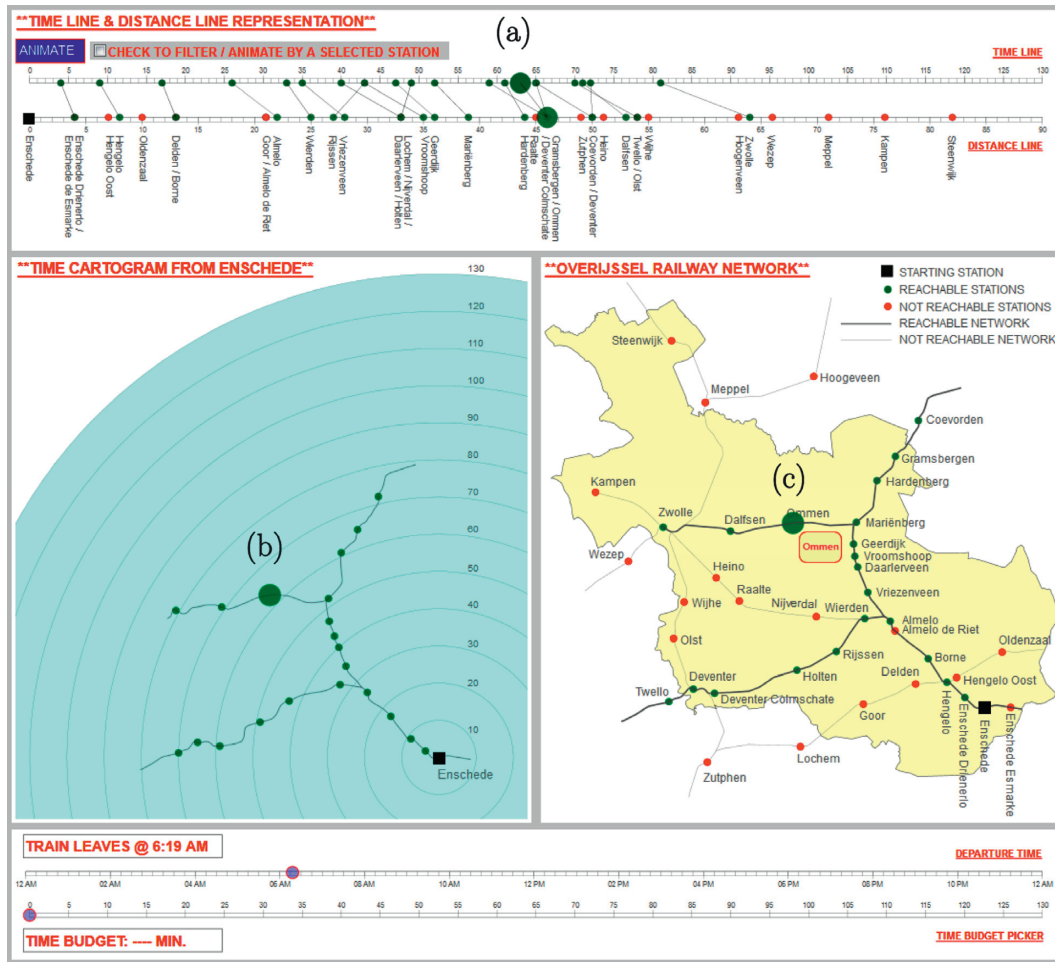
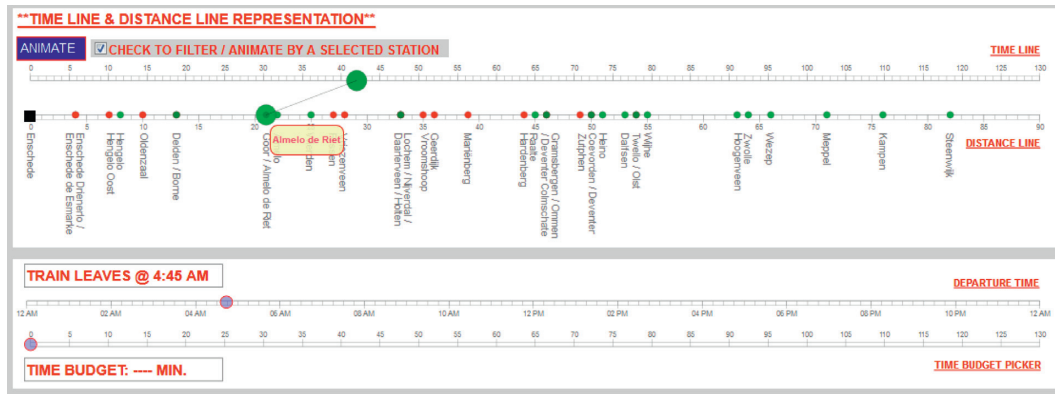


FIGURE 8: Travelling from Enschede at 06:19 AM. The reachable stations are indicated by green dots, whereas those not reachable are indicated by red dots. The time line and the time cartogram show that it takes 63 minutes to reach the highlighted station (Ommen) from Enschede when leaving at 06:19 AM.

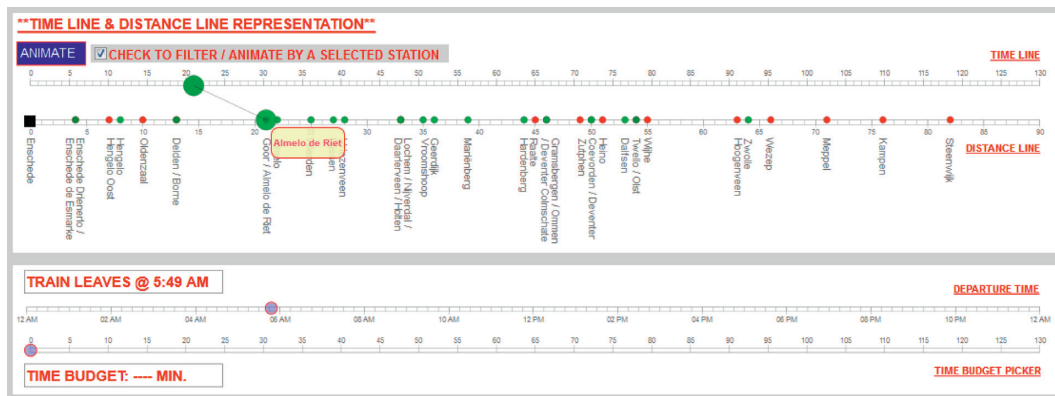
provided a time budget of  $X$  minutes. An example could be the following: Robert is in Enschede to attend a 3-day workshop at the University of Twente in the Netherlands. The workshop runs from 09:00 AM to 05:00 PM daily. As he has never been in the Netherlands before, on one of the workshop days he wants to visit another place that is close to Enschede. He can only spare 40 minutes to travel (single trip time only, excluding sightseeing) and wants to know the cities that are reachable by train from Enschede within his time budget. The graphics in Figures 5 and 6 explain how the visualization environment can be used to address such queries. If Robert is at the Enschede train station at 06:04 PM with no travel budget constraints, he can reach twenty-four stations (as shown on the time line in Figure 5). At 06:04 PM, but with a time budget of 40 minutes, he can reach twelve stations (Figure 6(a)). If the time budget is increased to 80 minutes, he can travel to sixteen stations (Figure 6(b)).

4.6. UQ6: How Does the Reachability of Station  $X$  Change during the Day? An individual or the transport planners might be interested in knowing how the reachability of a station  $X$  changes over time. This can be investigated by selecting the filter and the desired station on the geographic map. Running the animation then shows how the reachability of the selected station changes over time. In Figure 9, the selected station is Almelo de Riet. The animation shows some interesting patterns. For instance, it takes 42 minutes to reach Almelo de Riet from Enschede at 04:45 AM and 21 minutes at 05:49 AM.

4.7. UQ7: Which Station(s) or Network Segment(s) Is/Are Heavily-Served or Under-Served? The transport planners might also be interested in knowing which stations or segments of the network are heavily-served and which are under-served. Running the animation from 12:00 AM to



(a)



(b)

FIGURE 9: Travelling to Almelo de Riet from Enschede at 04:45 AM (a) and 05:49 AM (b).

11:59 PM reveals interesting findings. One can find out that the stations on the railroad segments labelled as A and B in Figure 10 are under-served; i.e., a fewer number of trains travel in these directions. In addition, the travel times to these stations are longer because of the slower connections. On the other hand, all stations on the railroad segments labelled as C and D in Figure 10 are heavily-served; i.e., more trains travel in these directions. The stations on these railroad segments have shorter travel times due to faster connections.

*4.8. UQ8: Why Does It Take Less Time to Reach Station X Than Station Y, despite X Being Farther away Than Y in Geography?* Sometimes a station (e.g., X) takes less travel time to reach than another station (e.g., Y), despite X being farther away than Y in geography. This is because of the different train types: those that stop at every station (stop trains) and those that stop only at the main stations (intercity trains). Consider Figure 11. Travelling by train from Enschede to

Deventer Colmschate takes longer time than travelling from Enschede to Deventer (62 minutes versus 43 minutes) because of slower connections, even though Deventer Colmschate is closer to Enschede than Deventer (46 kilometers versus 50 kilometers). Deventer is an intercity or main station and is served by intercity trains. On the other hand, Deventer Colmschate is a regular station and is served by stop trains. Similarly, it takes more time to reach Almelo de Riet than Almelo (41 minutes versus 19 minutes), despite Almelo de Riet being closer than Almelo in geography (21 kilometers versus 22 kilometers). This can be seen in Figure 12. Our visualization environment can be helpful for transport planners to figure out which stations take longer time to reach despite the fact that they are geographically closer and improve the connections if needed. These patterns can be discovered from the crossing lines on time line/distance line (as seen in Figures 11(a) and 12(a)) and also self-crossing railroad lines in the time cartogram (as seen in Figures 11(b) and 12(b)).

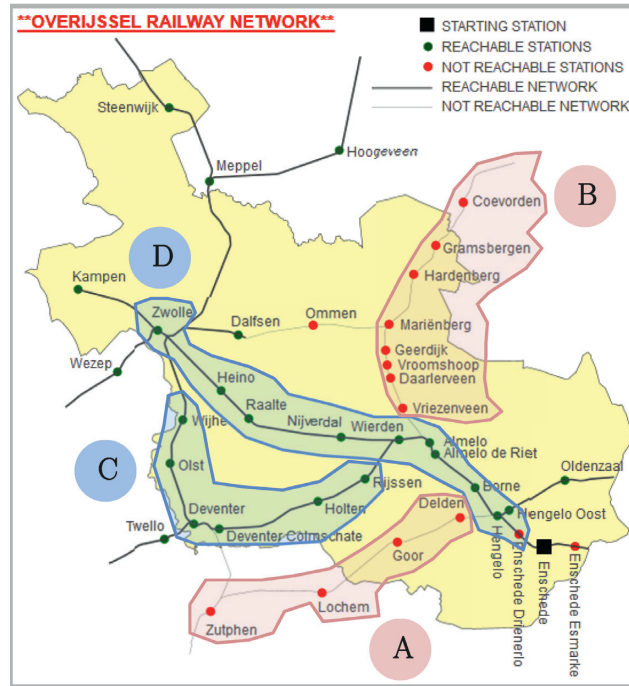


FIGURE 10: Under-served versus heavily-served stations and network segments. The stations on the railroad segments labelled as A and B are under-served, whereas the stations on the railroad segments labelled as C and D are heavily-served.

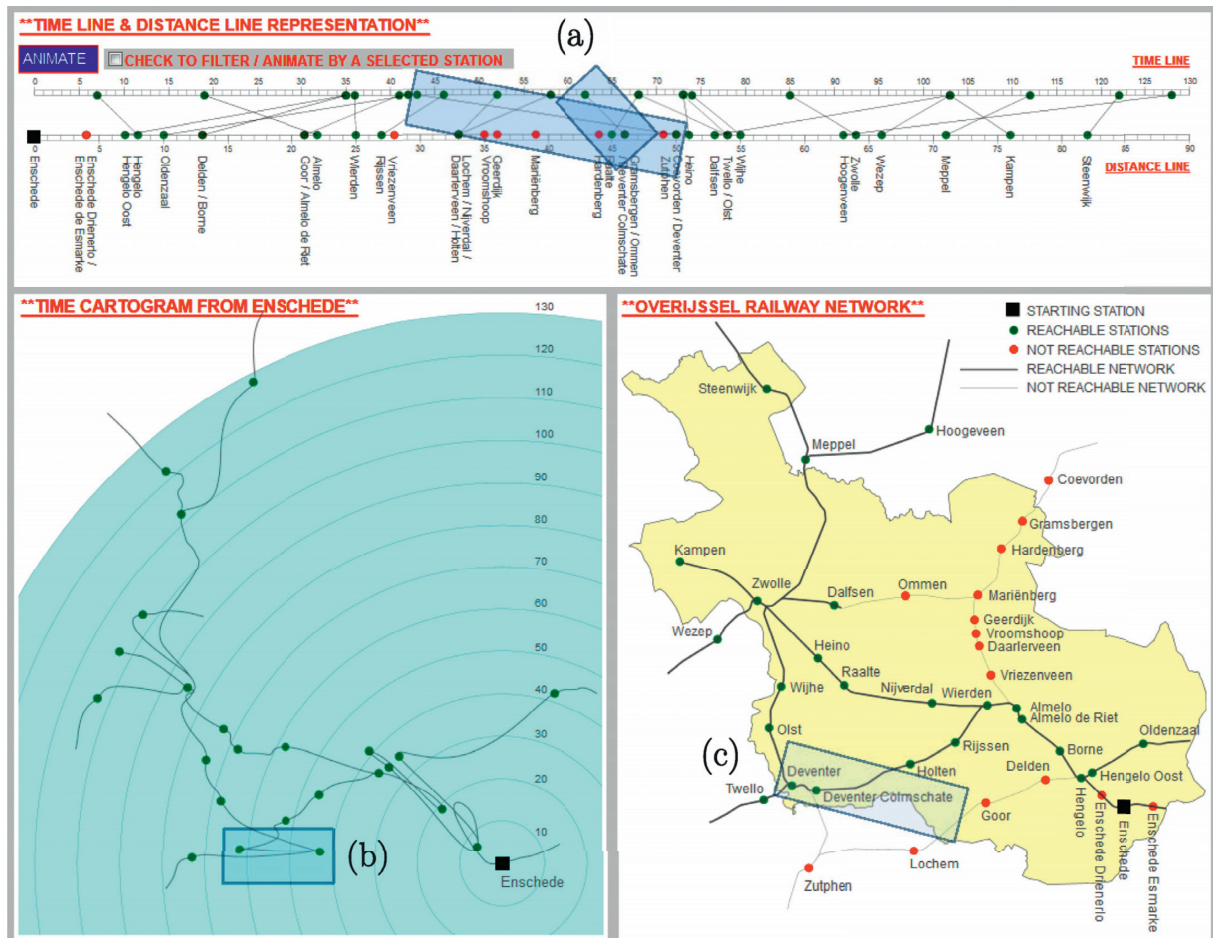


FIGURE 11: Travelling to Deventer and Deventer Colmschate from Enschede at 05:16 AM.

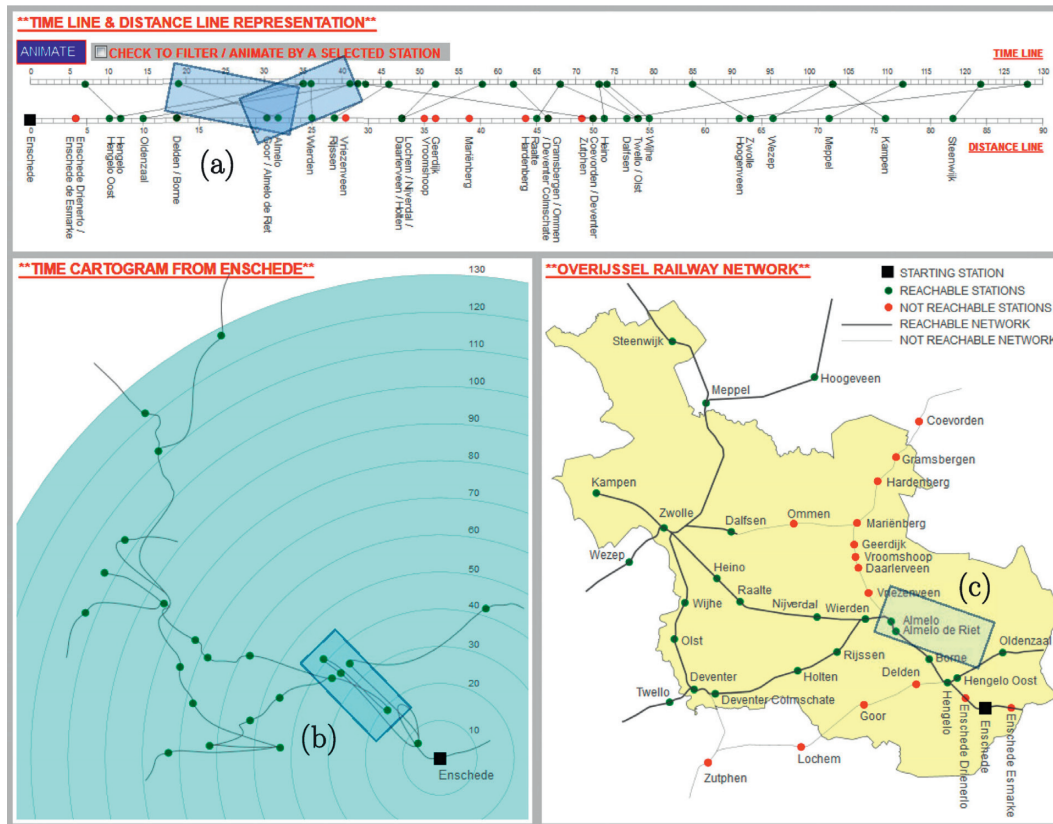


FIGURE 12: Travelling to Almelo and Almelo de Riet from Enschede at 05:16 AM.

## 5. Conclusions and Future Work

This research has presented the concept of a systematic, user-centric, and task-oriented visualization design framework for modelling reachability in transport networks. In the first step, a set of user questions were formulated in coordination with both frequent and casual train travelers. In the second step, the design and implementation of the visualization environment (largely influenced by the user questions and users' ease in interacting with it) were carried out to answer the questions posed by the users.

We integrated several (carto)graphic representations, the time line, the distance line, the time prism, the time cartogram, and the geographic map, in an interactive linked-views environment. A prototype was implemented in a web environment using D3.js. The implementation can be applied to any transport network. In this research, this has been achieved in the context of a large Dutch province of Overijssel with a fairly complex railroad network. This is a testament to the power and flexibility of our approach.

The approach as presented and used here provides an easily understood method for examining accessibility in transport networks. As is clear from the foregoing, the solution provides an alternative perspective for analyzing the network and gaining valuable insights. As a supplement to existing methods for visualizing travel times, our approach could be applied in areas such as

spatial analysis and transport planning. The approach can also be the basis of more sophisticated analysis techniques.

In the future, we intend to perform a comprehensive usability evaluation and further develop the visualization environment so that it can be used by train travelers to plan their trips and transport planners as a visual analytical tool to analyze the transport network in a more focused manner.

## Data Availability

The data that support the findings of this study are available from the first author, Rehmat Ullah (rehmatullah@uetpeshawar.edu.pk), upon reasonable request.

## Conflicts of Interest

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

## Acknowledgments

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (no. NRF-2018R1A6A1A03025109) and by the NRF grant funded by the Korea Government (MSIT) (no. NRF-2019R1A2C1006249).

## References

- [1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual Analytics of Movement*, Springer Science & Business Media, Berlin, Germany, 2013.
- [2] G. Andrienko, N. Andrienko, J. Dykes, M.-J. Kraak, and H. Schumann, "GeoVA (t)—geospatial visual analytics: focus on time," *Journal of Location Based Services*, vol. 4, no. 3-4, pp. 141–146, 2010.
- [3] G. Diansheng, C. Jin, A. M. MacEachren, and K. Liao, "A visualization system for space-time and multivariate patterns (VIS-STAMP)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1461–1474, 2006.
- [4] X. Li, "The time wave in time space: a visual exploration environment for spatio-temporal data," Ph. D thesis, University of Twente, Enschede, Netherlands, 2010.
- [5] X. Li and M.-J. Kraak, "The time wave. A new method of visual exploration of geo-data in time-space," *The Cartographic Journal*, vol. 45, no. 3, pp. 193–200, 2008.
- [6] R. Ullah and M.-J. Kraak, "An alternative method to constructing time cartograms for the visual representation of scheduled movement data," *Journal of Maps*, vol. 11, no. 4, pp. 674–687, 2014.
- [7] S. Bies and M. van Kreveld, "Time-space maps from triangulations," in *Graph Drawing*, W. Didimo and M. Patrignani, Eds., pp. 511–516, Springer, Berlin, Germany, 2013.
- [8] S. Hong, R. Kocielnik, Y. Min-Joon, S. Battersby, K. Juho, and C. Aragon, "Designing interactive distance cartograms to support urban travelers," in *Proceedings of the 2017 IEEE Pacific Visualization Symposium (PacificVis)*, Seoul, Republic of Korea, April 2017.
- [9] E. Shimizu and R. Inoue, "A new algorithm for distance cartogram construction," *International Journal of Geographical Information Science*, vol. 23, no. 11, pp. 1453–1470, 2009.
- [10] R. Han, Z. Li, P. Ti, and Z. Xu, "Experimental evaluation of the usability of cartogram for representation of GlobeLand30 data," *ISPRS International Journal of Geo-Information*, vol. 6, no. 6, p. 180, 2017.
- [11] R. Ullah, Z. Mengistu Eskedar, C. P. J. M. van Elzakker, and M.-J. Kraak, "Usability evaluation of centered time cartograms," *Open Geosciences*, vol. 8, p. 337, 2016.
- [12] K. Buchin, A. van Goethem, M. Hoffmann, M. van Kreveld, and B. Speckmann, "Travel-time maps: linear cartograms with fixed vertex locations," in *Geographic Information Science*, M. Duckham, E. Pebesma, K. Stewart, and A. Frank, Eds., pp. 18–33, Springer International Publishing, Berlin, Germany, 2014.
- [13] Y.-H. Wu and M.-C. Hung, "Non-connective linear cartograms for mapping traffic conditions," *Cartographic Perspectives*, vol. 65, pp. 33–50, 2012.
- [14] T. Hägerstrand, "What about people in regional science," *Papers in Regional Science*, vol. 24, pp. 7–24, 1970.
- [15] T. Gärling and E. Gärling, "Distance minimization in downtown pedestrian shopping," *Environment and Planning A: Economy and Space*, vol. 20, no. 4, pp. 547–554, 1988.
- [16] J. L. Horowitz, "Travel and location behavior: state of the art and research opportunities," *Transportation Research Part A: General*, vol. 19, no. 5-6, pp. 441–453, 1985.
- [17] U. Landau, J. N. Prashker, and B. Alpern, "Evaluation of activity constrained choice sets to shopping destination choice modelling," *Transportation Research Part A: General*, vol. 16, no. 3, pp. 199–207, 1982.
- [18] A. Pred, "The choreography of existence: comments on Hagerstrand's time-geography and its usefulness," *Economic Geography*, vol. 53, no. 2, pp. 207–221, 1977.
- [19] C. E. Sigal, A. A. B. Pritsker, and J. J. Solberg, "The stochastic shortest route problem," *Operations Research*, vol. 28, no. 5, pp. 1122–1129, 1980.
- [20] H. J. Miller, "Modelling accessibility using space-time prism concepts within geographical information systems," *International Journal of Geographical Information Systems*, vol. 5, no. 3, pp. 287–301, 1991.
- [21] H. J. Miller, *Time Geography and Space-Time Prism International Encyclopedia of Geography: People, the Earth, Environment and Technology*, John Wiley & Sons, Hoboken, NJ, USA, 2016.
- [22] Y. Song and H. J. Miller, "Simulating visit probability distributions within planar space-time prisms," *International Journal of Geographical Information Science*, vol. 28, no. 1, pp. 104–125, 2014.
- [23] N. Ahmed and H. J. Miller, "Time-space transformations of geographic space for exploring, analyzing and visualizing transportation systems," *Journal of Transport Geography*, vol. 15, no. 1, pp. 2–17, 2007.
- [24] C. Cauvin, "A systemic approach to transport accessibility. a methodology developed in strasbourg: 1982–2002," *Cybergeoe: European Journal of Geography*, 2005.
- [25] X. Chen, "Seeing differently: cartography for subjective maps based on dynamic urban data," M.Sc. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2011.
- [26] J.-C. Denain and P. Langlois, "Cartographie en anamorphose," *Mappemonde*, vol. 49, no. 1, pp. 16–19, 1998.
- [27] A. Goedvolk, "De nieuwe relatieve afstand voor het openbaar vervoer," *Nieuwe Geografenkrant*, vol. 10, pp. 6-7, 1988.
- [28] C. Kaiser, F. Walsh, C. Q. Farmer, and A. Pozdnoukhov, "User-centric time-distance representation of road networks," in *Geographic Information Science*, S. Fabrikant, T. Reichenbacher, M. Kreveld, and C. Schlieder, Eds., pp. 85–99, Springer, Berlin, Germany, 2010.
- [29] L. Ramaer, "De vervaardiging van temporele kartogrammen. 100 jaar veranderingen in de reistijd per trein in beeld," *Geo-Info*, vol. 11, pp. 11–13, 2011.
- [30] F. Tang, *A Comparative Study of Various Travel Time Representation Approaches for a Road Network*, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Yau Ma Tei, Hong Kong, 2012.
- [31] M. van Campenhout, "Travel time maps," M.Sc. thesis, Technical University Eindhoven, Enschede, Netherlands, 2010.
- [32] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visualizing time-oriented data—a systematic view," *Computers & Graphics*, vol. 31, no. 3, pp. 401–409, 2007.
- [33] A. Kölzsch, A. Slingsby, J. Wood, B. Nolet, and J. Dykes, "Visualisation design for representing bird migration tracks in time and space," in *Proceedings of the 2013 Workshop on Visualisation in Environmental Sciences (EnvirVis)*, Leipzig, Germany, July 2013.
- [34] J. I. Maletic, A. Marcus, and M. L. Collard, "A task oriented view of software visualization," in *Proceedings of the Visualizing Software for Understanding and Analysis First International Workshop*, Paris, France, June 2002.
- [35] M. Schots and C. Werner, "Using a task-oriented framework to characterize visualization approaches," in *Proceedings of the 2014 Second IEEE Working Conference on Software Visualization (VISSOFT)*, Victoria, Canada, February 2014.

- [36] M.-J. Kraak, B. Köbben, and Y. Tong, “Integrated time and distance line cartogram: a schematic approach to understand the narrative of movements,” *Cartographic Perspectives*, vol. 77, pp. 7–16, 2014.
- [37] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, *Mastering the Information Age—Solving Problems with Visual Analytics*, Eurographics Association, Norrköping, Sweden, 2010.
- [38] J. Dykes, A. M. MacEachren, and M. J. Kraak, *Exploring Geovisualization*, Elsevier, Amsterdam, Netherlands, 2005.
- [39] M.-J. Kraak, “Timelines, temporal resolution, temporal zoom and time geography,” in *Proceedings of the 22nd International Cartographic Conference*, A Coruña, Spain, December 2005.
- [40] J. C. Roberts, “Exploratory visualization with multiple linked views,” in *Exploring Geovisualization*, J. Dykes, A. M. MacEachren, and M.-J. Kraak, Eds., pp. 159–180, Elsevier, Amsterdam, Netherlands, 2005.
- [41] M. Bostock, V. Ogievetsky, and J. Heer, “D<sup>3</sup> data-driven documents,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.
- [42] B. Köbben, “Towards a national atlas of The Netherlands as part of the national spatial data infrastructure,” *The Cartographic Journal*, vol. 50, no. 3, pp. 225–231, 2013.



## Research Article

# WiFi-Based Virtual Access Network Scheduling for Downlink Traffic Dominated Smart Spaces

Pin Lv <sup>1,2,3</sup>, Siyu Pan <sup>1</sup> and Jia Xu <sup>1,2,3</sup>

<sup>1</sup>School of Computer Electronics and Information, Guangxi University, Nanning 530004, China

<sup>2</sup>Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China

<sup>3</sup>Guangxi Colleges and Universities Key Laboratory of Parallel and Distributed Computing, Nanning 530004, China

Correspondence should be addressed to Jia Xu; [xujia@gxu.edu.cn](mailto:xujia@gxu.edu.cn)

Received 25 March 2020; Revised 27 July 2020; Accepted 24 October 2020; Published 6 November 2020

Academic Editor: Sungchang Lee

Copyright © 2020 Pin Lv 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.

WiFi networks are widely and densely deployed as infrastructure in smart spaces. However, differentiated services with guaranteed access bandwidths are not supported in traditional WiFi networks. In this paper, wireless virtual access networks are established to provide guaranteed downlink bandwidths for primary users. For each primary user with a demanded access bandwidth, a group of APs are coordinated to serve it. In order to maximize network utilization, two wireless virtual access network scheduling algorithms are designed. One scheduling algorithm is designed based on the maximum independent set in the conflict graph, which has an exponential computation complexity. The other scheduling solution is based on a greedy strategy with linear complexity. Simulation results prove that both scheduling algorithms improve network utilization effectively, and the greedy algorithm is more suitable for practical use.

## 1. Introduction

As important infrastructure, WiFi networks are widely and densely deployed in many smart space scenarios [1, 2], such as smart home [3], smart building [4], and smart campus [5]. WiFi-enabled smart devices access the Internet as stations when they are associated with WiFi access points (APs). Since WiFi networks are based on IEEE 802.11 standards, distributed coordinate function (DCF) is adopted as the media access control (MAC) mechanism in most cases. With such MAC mechanism, all wireless nodes including both APs and stations have almost equal opportunity to randomly access the channel, and the quality of service (QoS) cannot be guaranteed due to contention and backoff. In many smart space cases, downlink traffic is far more than uplink traffic, and some users demand various but guaranteed downlink bandwidths. For example, a 4-channel HD (high-definition) monitor may demand a guaranteed downlink bandwidth of 16 Mbps, and a remotely controlled device may only require a downlink bandwidth of 100 kbps which is also guaranteed. Traditional WiFi networks cannot support such

differentiated and guaranteed service. Hence, wireless network virtualization is introduced into WiFi networks to meet the personalized bandwidth requirements of the users.

Wireless network virtualization is an emerging technology which has attracted much attention in recent years [6, 7], and WiFi network virtualization is studied for service isolation [8] or seamless roaming [9]. With wireless network virtualization, network infrastructure can be decoupled from the services that it provides, where differentiated services can coexist on the same infrastructure, maximizing its utilization [10]. In order to provide guaranteed QoS for the users, virtual networks are established and operated in isolation from each other. In order to manage and maintain virtual networks automatically, software-defined networking (SDN) technology is utilized to control network resources for data transmission [11]. At least one controller is deployed in SDN to take charge of resource allocation.

In this paper, two types of WiFi network users in downlink traffic dominated smart space are considered, which are the primary users and the secondary users. The primary users have specific access bandwidth requirements,

and the secondary users do not have such requirements. Wireless virtual access networks (VANs) are established in software-defined WiFi networks to provide primary users with differentiated and guaranteed access bandwidths. The residual bandwidth is allocated to the secondary users to provide them the best-effort service. In order to improve network utilization, the VANs should be scheduled subtly to maximize the residual bandwidth. Hence, two scheduling algorithms are designed. In the first scheduling algorithm which is called MISS, most VANs without conflict are found out based on the maximum independence set to transmit concurrently, which has an exponential computation complexity. The second scheduling algorithm is a greedy algorithm which has a linear computation complexity; thus, it is named LINS. Extensive simulations are conducted to evaluate the performances of MISS and LINS. Based on the simulation result analysis, MISS outperforms LINS slightly in network utilization maximization, but LINS has higher efficiency than MISS.

The contributions of this paper are summarized as follows:

- (1) A framework of the wireless virtual access network is proposed for downlink traffic dominated WiFi networks to provide differentiated services with guaranteed downlink bandwidths
- (2) Two wireless virtual access network scheduling algorithms, i.e., MISS and LINS, are designed to maximize the WiFi network utilization
- (3) A simulator is developed to evaluate the performances of the two scheduling algorithms, and a conclusion is drawn that LINS is more suitable for practical use

The remainder of the paper is organized as follows. Related work is summarized in Section 2. The system model is described in Section 3. In Section 4, two VAN scheduling algorithms are designed. Experimental results are analyzed in Section 5, and the paper is concluded in Section 6.

## 2. Related Work

Wireless network virtualization is considered a promising technology to improve resource utilization, QoS, security, and so forth, which has become a research hotspot in recent years [6, 12, 13]. In wireless network virtualization, a logical function is decoupled from the physical resource. According to the mapping relation between the physical resource and logical function, the related work can be divided into three types, that is, isolation, aggregation, and hybrid.

In the isolation type of wireless network virtualization, multiple virtual networks coexist on the same physical wireless network without mutual interference. It is a one-to-many relationship from a physical network to virtual networks. The virtual networks are isolated from each other in at least one dimension, such as time, frequency, space, and coding [14]. In [8], the virtual access network in the wireless mesh network is proposed to provide guaranteed access bandwidth, and the virtual access networks are assigned

different OFDMA subcarriers for isolation. In this paper, virtual access networks are constructed in commodity WiFi networks to provide primary users with guaranteed access bandwidths. The distinct feature of this paper is that virtual access networks are isolated in the time domain, and two scheduling algorithms are designed correspondingly.

In the aggregation type of wireless network virtualization, multiple physical devices serve as a virtual entity together. It is a many-to-one relationship from physical devices to virtual entity. For example, in [15], a mechanism that multiple APs are virtualized into a virtual AP is described. With the help of virtual AP, seamless roaming within the WiFi network is achieved, and AP diversity is employed to enhance the transmission rate. In this paper, we leverage the similar AP virtualization mechanism with [15] but extend it to support multiple virtual access networks with guaranteed bandwidths.

The hybrid type of wireless network virtualization is a combination of the above two types, in which the physical networks and virtual networks have a many-to-many relationship. In [16], for instance, the isolated virtual networks are supported by heterogeneous physical networks. The wireless network virtualization scheme in this paper also belongs to the hybrid type. In our proposed scheme, multiple APs are virtualized into a single AP, while the physical WiFi network accommodates multiple virtual access networks.

To the best of our knowledge, this is the first attempt to combine the virtual access network with AP aggregation virtualization.

## 3. System Model

In this paper, downlink traffic dominated smart space scenario is considered. In the scenario, a software-defined WiFi network is deployed to provide Internet access to devices. In the software-defined WiFi network, all the APs are managed by a controller. The wireless interfaces of all the APs are configured into identical parameters, including IP address, MAC address, channel, and ESSID. Hence, multiple APs act as a unique virtual AP [15]. If the APs are set to different nonoverlapping channels to increase the network capacity, the APs on the same channel are organized into a virtual AP. In this case, multiple virtual APs exist in the WiFi network, and each virtual AP operates in the way described in this paper. When a station is associated with the virtual AP, it accesses the Internet through the WiFi network. If the station transmits a packet to the virtual AP, all the APs within its transmission range receive it. These APs are organized into a group to serve the station. The AP group forms a wireless virtual access network (VAN) for the station. Each AP in the group periodically reports the signal strength of the station to the controller. When a downlink packet is needed to send to the station, it is firstly multicast to each member in the AP group, and one AP from the group is chosen by the controller to transmit the packet to the station according to the signal strength. Meanwhile, the controller commands other APs which may cause collisions to keep silent. Without the controller, multiple APs may contend for the channel to transmit packets, and collisions may cause

serious performance degradation. In order to avoid a collision caused by uplink transmissions, the APs always set the contention window to 0. Therefore, the downlink transmission has a higher priority than the uplink transmission.

The users in the WiFi network are classified into two types, i.e., the primary users (PUs) and the secondary users (SUs). The primary user  $PU_i$  has an access bandwidth requirement  $b_i$ , which should be guaranteed. In this paper, the normalized bandwidth is used to facilitate the analysis. For example, if a primary user demands an access bandwidth of 0.2, it means 20% of the access time should be allocated to the user. It is assumed that the sum of all primary users' demanded normalized bandwidth does not exceed 1. Hence, for each primary user, its VAN is assigned a certain active time  $A_i$ . In the active time, the AP group transmits packets to the primary user, and the VANs of other stations should not interfere with the transmission. It is feasible to guarantee access bandwidth by means of active time allocation [17, 18]. The advantage of our solution is that it is compatible with IEEE 802.11, and no modification is needed on the user-side devices.

An example is demonstrated in Figure 1. Four APs and two primary users exist in a WiFi network. The VAN of PU1 contains AP1 and AP2, and the AP group which serves PU2 includes AP2, AP3, and AP4. Since AP2 exists in the two AP groups at the same time, the VANs of PU1 and PU2 cannot be activated simultaneously.

The active time allocated to a primary user may contain several separated time intervals as follows:

$$A_i = a_{i1} \cup a_{i2} \cup \dots \cup a_{in}. \quad (1)$$

The total length of these time intervals should be equal to the access bandwidth requirement of the primary user, i.e.,

$$b_i = \sum_{j=1}^n L(a_{ij}). \quad (2)$$

After catering to the bandwidth requirements of all the primary users, the residual bandwidth is used by the secondary users, as shown in Figure 1. The residual bandwidth is denoted as  $B_s$  (bandwidth for secondary users), and the total bandwidth for primary users is  $B_p$ . They are computed as follows:

$$\begin{aligned} B_s &= 1 - B_p, \\ B_p &= S(\{b_i | i = 1, 2, \dots, m\}), \end{aligned} \quad (3)$$

where  $S(\{b_i | i = 1, 2, \dots, m\})$  means the total bandwidth allocated to  $m$  primary users under a certain scheduling scheme.

In order to improve the network utilization, the VANs should be scheduled skillfully to maximize the residual bandwidth. Therefore, the optimization objective is to maximize the residual bandwidth as follows:

$$\max B_s. \quad (4)$$

From another perspective, the concurrent transmissions of different VANs should be maximized, and the active time

for all primary users is minimized using appropriate scheduling:

$$\min B_p. \quad (5)$$

To determine whether VANs can transmit concurrently, a conflict graph should be constructed and maintained by the controller. In the conflict graph, a node represents a VAN. If two VANs conflict, an edge exists between the corresponding nodes.

Denote the interference radius of the AP as  $R_i$ . Two VANs do not interfere with each other if their distance is larger than  $R_i$ . The distance between two VANs,  $V_1$  and  $V_2$ , is defined as

$$D(V_1, V_2) = \min_{n_i \in V_1, n_j \in V_2} d(n_i, n_j), \quad (6)$$

where  $d(n_i, n_j)$  is the distance between wireless nodes  $n_i$  and  $n_j$  and a wireless node in a VAN is either an AP or a station.

#### 4. Wireless Virtual Access Network Scheduling Algorithms

Finding optimal scheduling for such an environment is an NP-hard problem [19]; two VAN scheduling algorithms are designed to find approximate optimal solutions. The first scheduling algorithm is based on the maximum independent set, and it is referred to as MISS. The second one is a greedy algorithm which has linear complexity, thus it is called LINS. The scheduling algorithm is executed by the controller.

**4.1. MISS.** The MISS algorithm explores the confliction relationships between the VANs of the primary users based on the conflict graph of the network. The maximum independent set of the conflict graph indicates the largest set of VANs in which any VAN does not interfere with each other. Thus, these VANs can be activated at the same time.

As shown in Algorithm 1, in the scheduling process, the maximum independent set of the conflict graph is found. Each VAN in the maximum independent set is allocated a time interval with the length equal to the shortest bandwidth requirement in the set. If the bandwidth requirement of  $v_i$  (the VAN for  $PU_i$ ) is satisfied, it is removed from the conflict graph. Repeat the process until all the VANs are provided their demanded bandwidths.

From Lines 1 to 4, the allocated active time and bandwidth requirement of each VAN are initialized. Line 5 indicates the termination condition of the loop is that each VAN is assigned enough service time. In Line 6, the maximum independent set is computed, which means the most VANs without mutual interference are found. Among these VANs, the one with the least unsatisfied bandwidth requirement determines the active interval length of the VANs in the maximum independent set. Hence, the length of the least unsatisfied bandwidth requirement is obtained in Line 7. From Lines 8 to 15, every VAN in the maximum independent set is allocated an active interval. If the bandwidth requirement of a VAN is satisfied, it is removed from the candidate set.

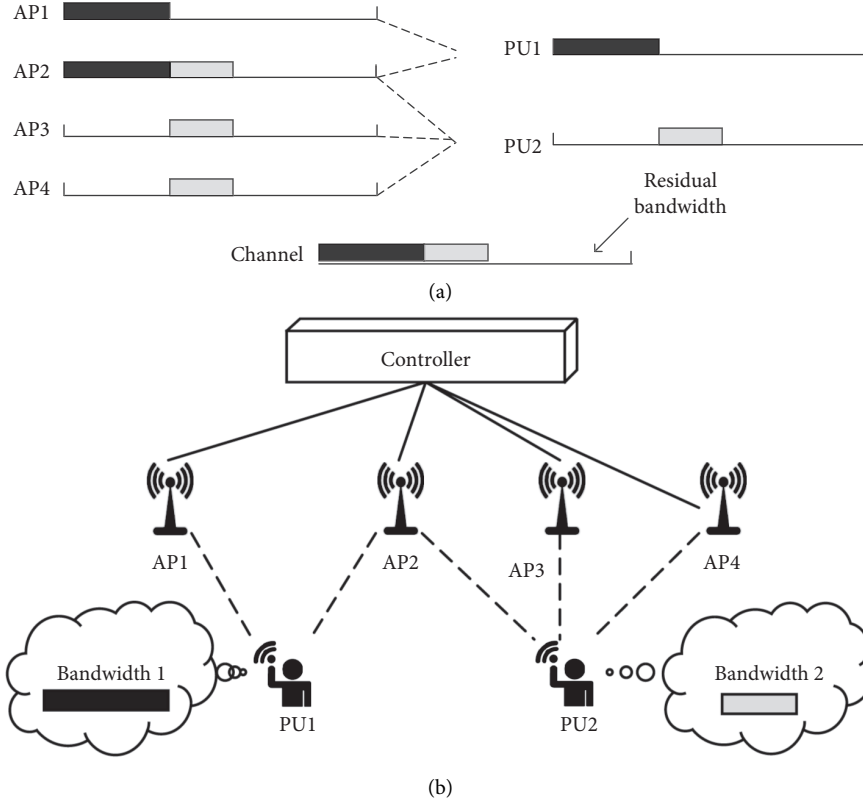
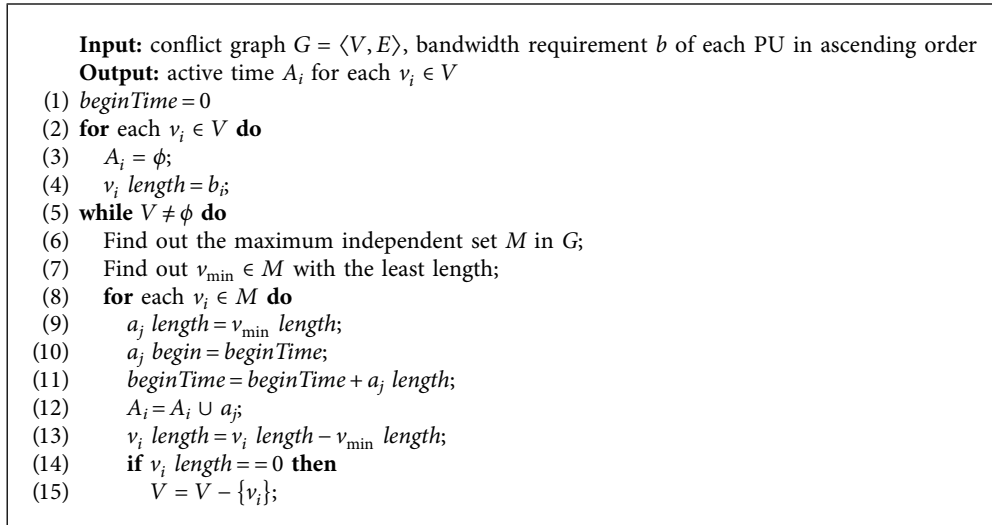


FIGURE 1: An example of wireless virtual access networks.



ALGORITHM 1: MISS.

Since the process of finding the maximum independent set has a complexity of  $O(2^n)$ , the computation complexity of MISS is  $O(2^n)$ , which is exponential.

**4.2. LINS.** To achieve low computation complexity, LINS is designed based on a greedy strategy. LINS schedules the VANs in the descending order of demanded bandwidths.

The VAN with the largest demanded bandwidth is allocated active time first. For the following VAN, it is assigned as early as possible active time in which it does not interfere with other scheduled VANs, as shown in Algorithm 2. Different from MISS, each VAN is allocated a continuous service time interval under the scheduling of LINS.

Lines 1 to 4 are also an initialization process. Since the VANs are scheduled in the descending order of the

**Input:** conflict graph  $G = \langle V, E \rangle$ ,  $b$  of each PU in descending order  
**Output:** active time  $A_i$  for each  $v_i \in V$

- (1) **for** each  $v_i \in V$  **do**
- (2)    $A_i = \emptyset$ ;
- (3)    $v_i$  length =  $b_i$ ;
- (4) **while**  $V \neq \emptyset$  **do**
- (5)   Find out  $v_{\max} \in V$  with the largest length;
- (6)   Find out the earliest time  $t$  in which  $v_{\max}$  does not interfere with other scheduled VANs;
- (7)    $a_j$  begin =  $t$ ;
- (8)    $a_j$  length =  $v_{\max}$  length;
- (9)    $A_i = A_i \cup a_j$ ;
- (10)    $V = V - \{v_{\max}\}$ ;

ALGORITHM 2: LINS.

bandwidth requirement, the VAN with the largest demanded bandwidth is fetched in Line 5, and it is allocated a service time as early as possible from Lines 6 to 10.

The computation complexity of LINS is  $O(n)$ , which is linear.

**4.3. Case Study.** Suppose four VANs coexist in a WiFi network, and the conflict graph of them is demonstrated in Figure 2. Each node in the conflict graph represents a VAN. If a line connects two nodes, it means the corresponding VANs interfere with each other.

The demanded bandwidths of the four primary users are 0.3, 0.3, 0.2, and 0.2, respectively. The scheduling schemes under MISS and LINS are shown in Figure 3.

For MISS, a maximum independent set containing User 1 and User 3 is found firstly, according to the interference relation. In this maximum independent set, User 3 has the least demanded bandwidth. Hence, the two users are allocated an active interval with the length of 0.2, which is the demanded bandwidth of User 3. After this interval, the bandwidth requirement of User 3 has been satisfied, and it is removed from the conflict graph. In the next round of the loop, another maximum independent set containing User 1 and User 4 is found. Since the remainder unsatisfied bandwidth requirement of User 1 is 0.1, which is lower than the demanded bandwidth of User 4, the two users are allocated an active interval with the length of 0.1. After that, User 1 is removed from the conflict graph because the total length of the allocated active intervals equals its demanded bandwidth. In the next round, a maximum independent set containing only User 2 is found, and it is allocated an interval with the length of 0.3. At last, User 4 is the only node left in the conflict graph, and it is allocated an interval to satisfy its remainder demanded bandwidth. A feature of MISS is that the demanded bandwidth of a user may be divided into multiple separated intervals, like User 4 in this case.

For LINS, the primary users are allocated active intervals in the descending order of the demanded bandwidth, and the interval is not divided into separated parts. In this case, User 1 is allocated an interval with the length of 0.3 first. Next, User 2 is allocated the following interval with the

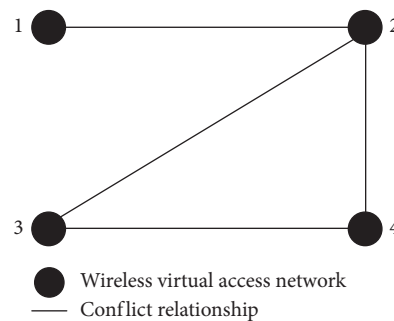


FIGURE 2: An example of the conflict graph.

length of 0.3, due to the fact that it has a conflict relation with User 1. When User 3 is considered, since it does not interfere with User 1, it is allocated a concurrent interval with User 1. At last, User 4 is allocated an interval after User 2's interval, because an earlier continuous interval cannot be found.

Besides the bandwidth allocated to primary users, the residual bandwidth can be used by secondary users. In this case, the residual bandwidth of MISS is 0.3, while the residual bandwidth of LINS is 0.2. Hence, more bandwidth is residual under the scheduling of MISS, and the WiFi network has a higher utilization.

It is revealed from the case that MISS has a more fine-grained scheduling than LINS, but LINS has higher efficiency.

## 5. Performance Evaluation

A simulator is developed, and extensive simulations are carried out to evaluate the performance of the proposed VAN scheduling algorithms.

**5.1. Experiment Setup.** A WiFi network with 25 APs and 50 primary users is generated in the simulator. The APs are deployed as a  $5 \times 5$  grid in a  $500 \text{ m} \times 500 \text{ m}$  rectangular area. The primary users distribute randomly in the area, following a uniform distribution. Both the transmission range and the interference range of the wireless node follow a disk model, and the transmission radius is 100 m and the interference

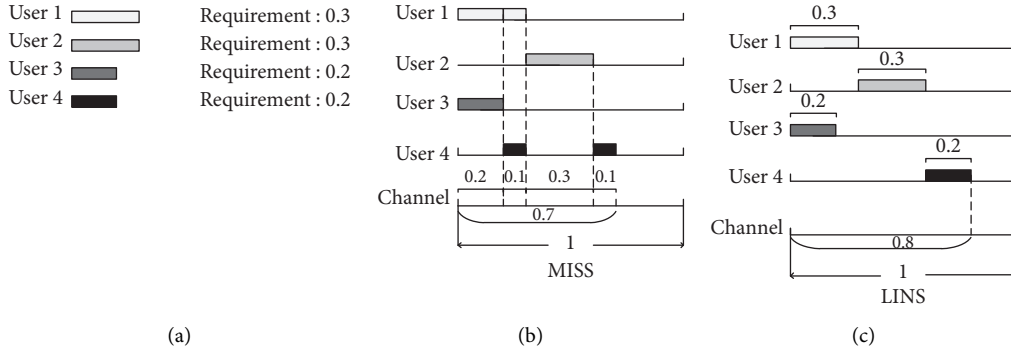


FIGURE 3: Scheduling comparison of MISS and LINS.

radius is 150 m. The default parameters in the simulations are listed in Table 1.

A benchmark is introduced as a comparison with MISS and LINS, which is called SUM. The SUM is the sum of the required bandwidths of all primary users without considering the possible overlaps between them as follows:

$$\text{SUM} = \sum b_i. \quad (7)$$

To evaluate the performance from multiple perspectives, the following metrics are measured:

- (1) For the sake of fair comparison, optimized service time for all primary users to that of SUM ratio (OTS) is presented as a new metric (computed as the following equation):

$$\text{OTS} = \frac{O}{\text{SUM}}, \quad (8)$$

where  $O$  is the total service time for primary users under the scheduling of MISS or LINS. OTS is a ratio between 0 and 1, and it should be minimized.

- (2) The average time costs for scheduling (i.e., the time for program execution) should be minimized.
- (3) The residual bandwidth for secondary users should be maximized.

**5.2. Experiment Results.** When the number of primary users varies from 10 to 100, the OTSs of MISS and LINS are shown in Figure 4. The OTSs of both MISS and LINS range from 40% to 50% and do not change obviously with the increase in the number of primary users. In each case, the OTS of MISS is slightly lower than that of LINS (2% on average), which indicates that MISS outperform LINS only a little.

When changing the distances between the adjacent APs from 50 meters to 150 meters, the OTSs of MISS and LINS are shown in Figure 5. As the AP density decreases, the OTSs of the two scheduling algorithms decline slowly. This is because when the distance between APs increases, the transmission conflicts among AP groups reduce, and more concurrent transmissions are available. The OTSs of MISS are smaller than those of LINS, and the differences between MISS and LINS are still limited.

TABLE 1: Default parameters in the simulations.

Parameter	Default value
Number of APs	25
Number of primary users	50
Demanded bandwidth of primary users	0.02
Distance between adjacent APs (m)	100
Transmission radius of AP (m)	100
Interference radius of AP (m)	150

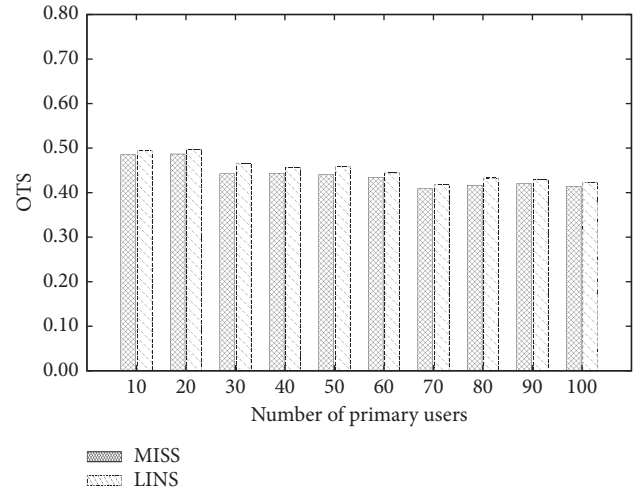


FIGURE 4: OTS of MISS and LINS when the number of primary users varies.

In Figure 6, the OTSs of MISS and LINS are displayed when the demanded access bandwidth scope of primary users changes. From the results, it can be seen that the demanded access bandwidth scope of primary users does not affect OTSs of the two scheduling algorithms.

From Figures 4–6, it is concluded that MISS has lower OTS than that of LINS when different parameters change. However, the differences between the two scheduling algorithms are quite small.

The average time costs for scheduling under different circumstances are also measured. In Figure 7, when the number of primary users increases from 10 to 100, the time cost of MISS surges exponentially, while the time cost of

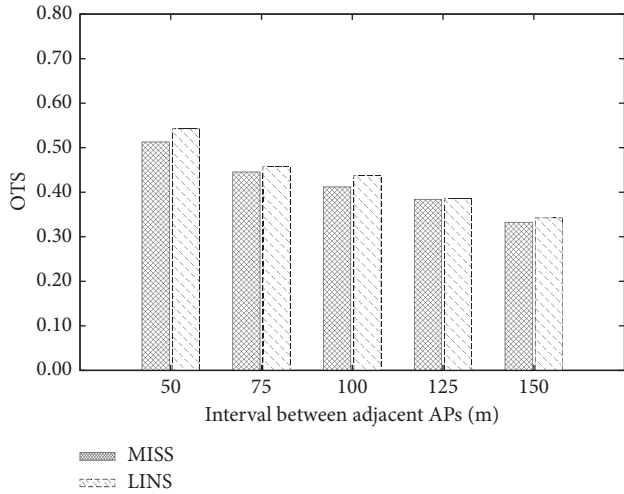


FIGURE 5: OTS of MISS and LINS when the distance between the adjacent APs varies.

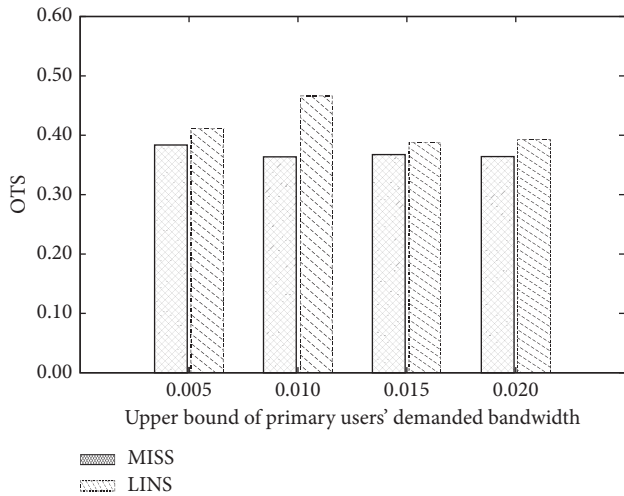


FIGURE 6: OTS of MISS and LINS when the demanded bandwidth of primary users varies.

LINS has a slight linear increment. The different computation complexities of the two scheduling algorithms lead to this result.

The time costs relative to AP densities are shown in Figure 8. The time cost of MISS still has obvious growth. In this case, however, the time cost of LINS even decreases slowly, because transmission conflicts reduce.

When the demanded bandwidth scope of the primary users becomes larger, the time cost of MISS has a slight increment, while the time cost of LINS keeps stable, as depicted in Figure 9. The differences between the two algorithms are still large.

From Figures 7–9, it is concluded that the time cost of MISS grows much faster than that of LINS. The difference is obvious especially when the scale of the WiFi network is large. The reason is that MISS has an exponential computing complexity, and LINS has a linear complexity. Since the two

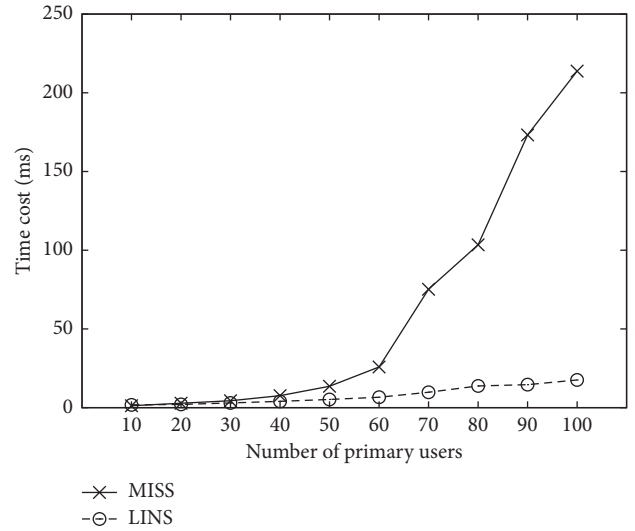


FIGURE 7: The average time cost of MISS and LINS when the number of primary users varies.

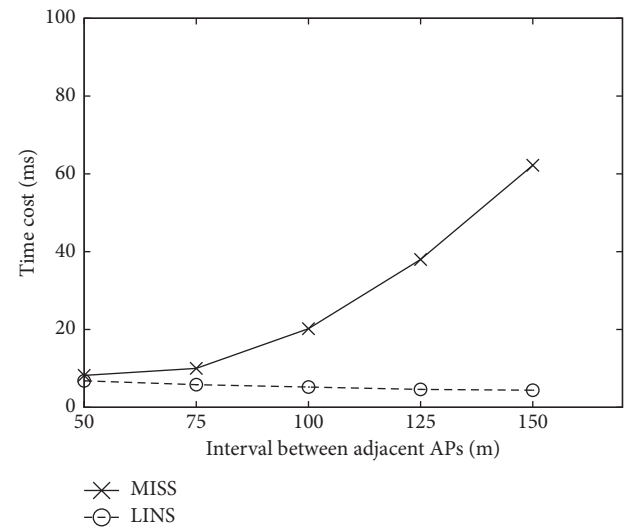


FIGURE 8: The average time cost of MISS and LINS when the distance between the adjacent APs varies.

scheduling algorithms have similar performance in OTS, LINS is more suitable for practical use. Hence, in the following simulations, only LINS is compared with the benchmark SUM.

The residual bandwidth is provided to secondary users, which should be maximized to improve network utilization. When the number of the primary users changes, the residual bandwidths of LINS and SUM are demonstrated in Figure 10. Thanks to the concurrent transmission scheduled by LINS, the residual bandwidth of LINS decreases much slower than that of SUM.

In Figure 11, the residual bandwidth of LINS increases slightly when the APs are deployed sparsely, while the AP density does not affect the residual bandwidth of SUM.

As shown in Figure 12, when the demanded bandwidths of the primary users increase, the residual bandwidth of

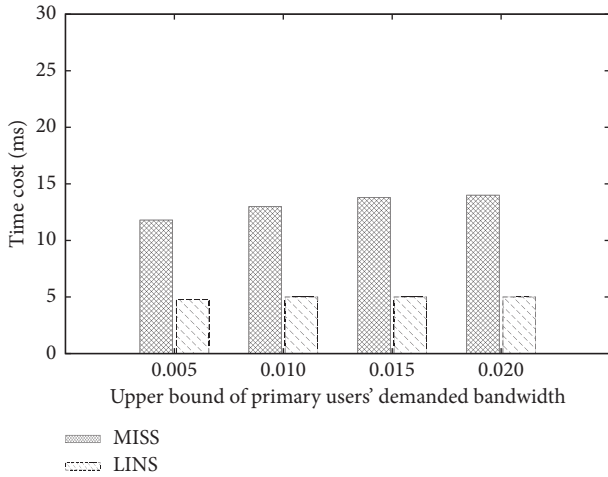


FIGURE 9: The average time cost of MISS and LINS when the demanded bandwidth of primary users varies.

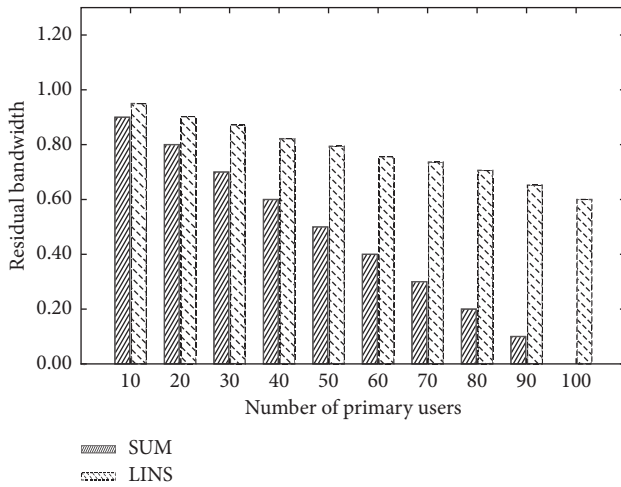


FIGURE 10: The residual bandwidth of LINS and SUM when the number of primary users varies.

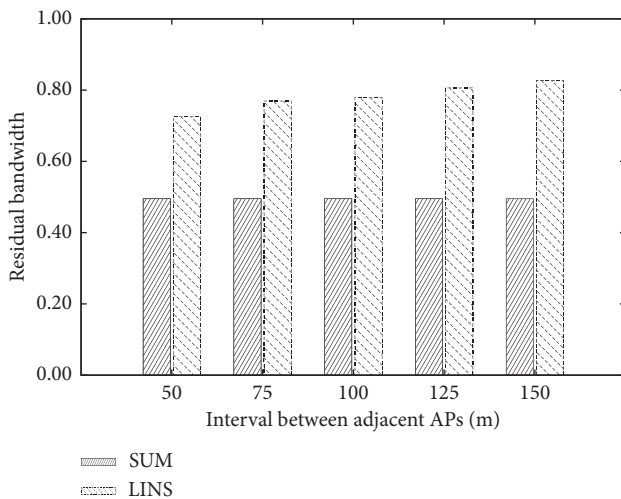


FIGURE 11: The residual bandwidth of LINS and SUM when the distance between the adjacent APs varies.

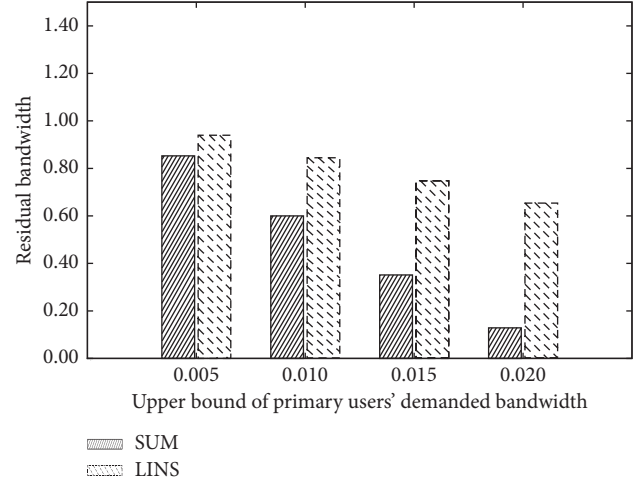


FIGURE 12: The residual bandwidth of LINS and SUM when the demanded bandwidth of primary users varies.

SUM decreases greatly. However, LINS has a slower rate of descent, owing to its scheduling.

From Figures 10–12, a conclusion can be drawn that the network utilization is improved due to transmission scheduling in WiFi networks.

## 6. Conclusion

To provide differentiated and guaranteed downlink bandwidth for devices in smart spaces, a framework of wireless virtual access network was designed in software-defined WiFi networks. Based on this framework, two scheduling algorithms, that is, MISS and LINS, were proposed to maximize the network utilization. The MISS algorithm employed the maximum independent set in the conflicted graph to increase concurrent transmissions of AP groups. The LINS algorithm scheduled the AP groups in a linear order, which had lower computation complexity than MISS. Extensive simulation experiments were conducted, and the results indicated that both algorithms scheduled wireless virtual access networks effectively. Nevertheless, the LINS algorithm was more suitable for practical use due to its high efficiency.

## Data Availability

The data used to support the findings of the study are available from the first author upon request (lvpin@gxu.edu.cn).

## Conflicts of Interest

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

## Acknowledgments

This research was funded by the Special Funds for Guangxi BaGui Scholars, National Natural Science Foundation of China (61402513 and 62062008), Guangxi Natural Science



Foundation (2019JJA170045, 2018JJA170194, 2018JJA170028, and 2016JJB170040), and Scientific Research Foundation of Guangxi University (XGZ150322).

## References

- [1] J. Sheth and B. Dezfouli, "Enhancing the energy-efficiency and timeliness of IoT communication in WiFi networks," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9085–9097, 2019.
- [2] Y. Sun, J. Chen, Y. Tang, and C. Yanjia, "Energy modeling of IoT mobile terminals on WiFi environmental impacts †," *Sensors*, vol. 18, no. 6, p. 1728, 2018.
- [3] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018.
- [4] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, "Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning," *Energy and Buildings*, vol. 177, pp. 12–22, 2018.
- [5] C. Del-Valle-Soto, L. J. Valdivia, R. Velázquez, L. Rizo-Dominguez, and J.-C. López-Pimentel, "Smart campus: an experimental performance comparison of collaborative and cooperative schemes for wireless sensor network," *Energies*, vol. 12, no. 16, p. 3135, 2019.
- [6] C. Liang and F. R. Yu, "Wireless network virtualization: a survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358–380, 2015.
- [7] T. M. Ho, N. H. Tran, S. M. A. Kazmi, Z. Han, and C. S. Hong, "Wireless network virtualization with non-orthogonal multiple access," in *Proceedings of the NOMS 2018—2018 IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Taipei, Taiwan, April 2018.
- [8] P. Lv, X. Wang, and M. Xu, "Virtual access network embedding in wireless mesh networks," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1362–1378, 2012.
- [9] Y. Amir, C. Danilov, R. Musualoiu-Elefteri, and N. Rivera, "The SMesh wireless mesh network," *ACM Transactions on Computer Systems (TOCS)*, vol. 28, no. 3, p. 6, 2010.
- [10] H. Wen, P. K. Tiwary, and T. Le-Ngoc, *Wireless Virtualization*, Springer, Berlin, Germany, 2013.
- [11] M. Yang, Y. Li, D. Jin, L. Zeng, X. Wu, and A. V. Vasilakos, "Software-defined and virtualized future mobile and wireless networks: a survey," *Mobile Networks and Applications*, vol. 20, no. 1, pp. 4–18, 2015.
- [12] N. Zhang, P. Yang, S. Zhang et al., "Software defined networking enabled wireless network virtualization: challenges and solutions," *IEEE Network*, vol. 31, no. 5, pp. 42–49, 2017.
- [13] D. B. Rawat, "Fusion of software defined networking, edge computing, and blockchain technology for wireless network virtualization," *IEEE Communications Magazine*, vol. 57, no. 10, pp. 50–55, 2019.
- [14] K. M. Park and C. K. Kim, "A framework for virtual network embedding in wireless networks," in *Proceedings of the 4th International Conference on Future Internet Technologies*, Seoul, Republic of Korea, June 2009.
- [15] P. Lv, X. Wang, X. Xue, and M. Xu, "SWIMMING: seamless and efficient WiFi-based internet access from moving vehicles," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 1085–1097, 2015.
- [16] P. Lv, X. Wang, Y. Yang, and M. Xu, "Network virtualization for smart grid communications," *IEEE Systems Journal*, vol. 8, no. 2, pp. 471–482, 2013.
- [17] A. Banchs, P. Serrano, P. Patras, and M. Natkaniec, "Providing throughput and fairness guarantees in virtualized WLANs through control theory," *Mobile Networks and Applications*, vol. 17, no. 4, pp. 435–446, 2012.
- [18] A. Checco and D. J. Leith, "Fair virtualization of 802.11 networks," *IEEE/ACM Transactions on Networking*, vol. 23, no. 1, pp. 148–160, 2015.
- [19] Y. Zhao, D. Guo, J. Xu et al., "CATS: Cooperative allocation of tasks and scheduling of sampling intervals for maximizing data sharing in WSNs," *ACM Transactions on Sensor Networks (TOSN)*, vol. 12, no. 4, pp. 1–26, 2016.

## Research Article

# Development of Hepatitis Disease Detection System by Exploiting Sparsity in Linear Support Vector Machine to Improve Strength of AdaBoost Ensemble Model

Wasif Akbar <sup>1</sup>, Wei-ping Wu,<sup>1</sup> Sehrish Saleem,<sup>2</sup> Muhammad Farhan,<sup>3</sup>  
Muhammad Asim Saleem,<sup>4</sup> Ashir Javeed,<sup>4</sup> and Liaqat Ali <sup>5,6</sup>

<sup>1</sup>School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China

<sup>2</sup>Department of Computer Science, MNS University of Engineering and Technology Multan, Multan, Pakistan

<sup>3</sup>Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan

<sup>4</sup>School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China

<sup>5</sup>School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China

<sup>6</sup>Department of Electrical Engineering, University of Science and Technology, Bannu, Pakistan

Correspondence should be addressed to Wasif Akbar; [sewasif@hotmail.com](mailto:sewasif@hotmail.com)

Received 19 March 2020; Revised 22 June 2020; Accepted 9 October 2020; Published 3 November 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Wasif Akbar 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.

Hepatitis disease is a deadliest disease. The management and diagnosis of hepatitis disease is expensive and requires high level of human expertise which poses challenges for the health care system in underdeveloped and developing countries. Hence, development of automated methods for accurate prediction of hepatitis disease is inevitable. In this paper, we develop a diagnostic system which hybridizes a linear support vector machine (SVM) model with adaptive boosting (AdaBoost) model. We exploit sparsity in linear SVM that is caused by  $L_1$  regularization. The sparse  $L_1$ -regularized SVM is capable of eliminating redundant or irrelevant features from feature space. After filtering features through the sparse linear SVM, the output of the SVM is applied to the AdaBoost ensemble model which is used for classification purposes. Two types of numerical experiments are performed on the clinical features of hepatitis disease collected from UCI machine learning repository. In the first experiment, only conventional AdaBoost model is used, while in the second experiment, a feature vector is applied to the sparse linear SVM before its application to the AdaBoost model. Simulation results demonstrate that the strength of a conventional AdaBoost model is enhanced by 6.39% by the proposed method, and its time complexity is also reduced. In addition, the proposed method shows better performance than many previously developed methods for hepatitis disease prediction.

## 1. Introduction

Hepatitis is considered a major chronic liver disease worldwide. The liver is considered to be the heaviest and one of the largest organs of the human body [1]. The liver is one of the key organs of a human body responsible for different functions. These functions include bile secretion, protein formation, and elimination of toxins from body. Hence, inflammation of liver (caused by hepatitis) results in

dysfunction of the liver, and consequently, the health of the subject is deteriorated. The symptoms of hepatitis are different in different patients, with some subjects showing no signs. Well-known symptoms include yellowish eyes and skin, abdominal pain, poor appetite, and tiredness [2, 3]. Hepatitis can be acute or chronic depending on duration. If it lasts for less than six months, it is acute; however, if it lasts for more than six months, it is chronic [4]. It has been reported that hepatitis results in more than a million deaths

each year. Diagnosis of hepatitis through conventional methods is a difficult job and requires expensive medical tests [5]. Additionally, the diagnosis of such disease through intelligent system reduces the cost and also examines the patient in shorter time. Hence, development of intelligent diagnostic systems for such type of disease prediction is very important.

In the past, numerous hybrid models for disease detection have been developed by different researchers. These include automated systems for Parkinson's disease prediction [6–8], mortality prediction [9, 10], cancer detection [11, 12], and heart disease [6, 13, 14]. These models are developed by hybridizing data mining models (for feature preprocessing) such as principal component analysis (PCA) and Fisher discriminant analysis (FDA) with machine learning models such as decision trees, logistic regression, support vector machine (SVM), Naive Bayes, neural network models, ensembles of neural networks,  $K$ -nearest neighbors, deep neural networks, and optimized and stacked SVMs [15–24]. For example, Adamczak developed different automated models for hepatitis prediction. These models include MLP + BP, RBF (Tooldiag), and FSM without rotation and achieved a prediction accuracy of 77.4%, 79%, and 88.5%, respectively [25]. In another study conducted by Passi, MLO was developed for hepatitis which resulted in hepatitis prediction of 79.70% [26, 27]. Stern and Dobnikar developed AIS, LDA, and FDA models which achieved the hepatitis prediction accuracy of 82%, 84.5%, and 86.40%, respectively [27]. Nilashi et al. developed KNN, ANFIS, NN, and SVM and achieved hepatitis prediction accuracy of 71.41%, 79.67%, 78.31%, and 81.17%, respectively [28]. Recently, Polat and Gunes discussed the hybridization of the feature extraction through the principal component analysis model with classification through artificial immune recognition system for the prediction of hepatitis disease [1, 29].

In this paper, we develop a hybrid intelligent diagnostic system. To improve the strength of AdaBoost predictive model, we propose to use  $L_1$ -penalized linear SVM. The  $L_1$  penalty makes the linear SVM sparse, thus making it capable of eliminating redundant features by making their coefficients zero through sparse solutions. After elimination of redundant features through the sparse linear SVM, the remaining features are supplied to the AdaBoost model for classification. In order to analyze the impact of the sparse linear SVM on the AdaBoost model, we performed two types of numerical experiments. In the first experiment, we developed the conventional AdaBoost model, while in the second experiment, we constructed a learning system by stacking the sparse SVM with the AdaBoost model. The performance of both the models, developed in the two experiments, was evaluated using an online hepatitis disease data. Experimental results demonstrated that the sparse linear SVM enhances the accuracy of conventional AdaBoost (for the hepatitis disease prediction based on the collected clinical features). Additionally, the sparse linear SVM also reduces AdaBoost model's complexity as the optimal subset of features contains less number of features.

The rest of the manuscript is organized as follows. Datasets, the proposed sparse linear SVM, and AdaBoost-

based learning system are elaborated in Section 2. Section 3 discusses various schemes for validation as well as multiple metrics for evaluation used in the manuscript. Section 4 discusses experimental setup and obtained results, whereas the last section concludes the paper.

## 2. Materials and Methods

*2.1. Dataset Description.* The hepatitis dataset consists of 155 samples, and each sample contains 19 features. Details about the 19 commonly used features for the hepatitis dataset are given in Table 1. The label of the dataset is binary, i.e., it can have a value of 1 or 2, where 1 means the sample belongs to a patient who died, while 2 means the sample is that of a subject who survived. There are 32 samples having label 1 and 123 samples having the label value of 2, i.e., the dataset contains 123 samples belonging to healthy class and 32 samples belonging to patient class. In machine learning, we split the data into two parts, namely, training and testing. The training part is used to train the model, and its performance is checked by testing the trained model on the testing data. In this study, the dataset is divided into training and testing datasets using 70–30 data partitioning. Hence, out of the 155 samples, 108 samples are used for training purposes, and the remaining 47 samples are used for testing purposes. Out of the 108 training samples, 23 samples belong to the patient class, and 85 patients belong to healthy class. On the other hand, out of the 47 testing samples, 7 samples belong to the patient group, and 38 samples belong to the healthy group. It can be noticed that lower class distribution of the patient class is a limitation of the dataset.

*2.2. Proposed Method.* As discussed above, in this paper, we exploit the sparsity in linear SVM to improve the strength of machine learning models, namely,  $k$ -nearest neighbours (KNN), Gaussian Naive Bayes (GNB), linear discriminant analysis (LDA), and AdaBoost ensemble model. Initially,  $L_1$ -penalized linear SVM is used to generate sparse features, i.e., to process the full set of features, null the redundant features, and yield a subset of features containing relevant features only. The generated subset of features by sparse linear SVM is supplied to machine learning models for classification purposes. The sparsity of the linear SVM is controlled by its hyperparameter  $\lambda$ . Hence, for distinct values of  $\lambda$ , various distinct features will be nullified resulting in different subsets of features. Thus, for achieving better hepatitis prediction accuracies, it is necessary to develop a sparse linear SVM that would nullify the most redundant or irrelevant features and generate a subset of the most relevant features. This can be accomplished by tuning the hyperparameter  $\lambda$ . In order to better comprehend the functioning of the proposed learning system, it is pertinent to briefly discuss the  $L_1$ -penalized linear SVM model and its formulation. The formulation is as follows.

Support vector machines (SVMs) are considered powerful learning methods and have been widely used in different biomedical- and health informatics-related problems [30]. During the training process, SVM tries to construct an

TABLE 1: Details of the 19 hepatitis features.

Feature no	Feature code	Feature description	Values
1	$D_1$	Age	10, 20, 30, . . . , 70, 80
2	$D_2$	Sex	Male, female
3	$D_3$	Steroid	1, 2
4	$D_4$	Antivirals	1, 2
5	$D_5$	Fatigue	1, 2
6	$D_6$	Malaise	1, 2
7	$D_7$	Anorexia	1, 2
8	$D_8$	Liver big	1, 2
9	$D_9$	Liver firm	1, 2
10	$D_{10}$	Spleen palpable	1, 2
11	$D_{11}$	Spiders	1, 2
12	$D_{12}$	Ascites	1, 2
13	$D_{13}$	Varices	1, 2
14	$D_{14}$	Bilirubin	0.39, 0.8, 1.2, 2.0, 3.0, 4.0
15	$D_{15}$	Alkaline phosphatase	33, 80, 120, 160, 200, 250
16	$D_{16}$	SGOT	100, 200, 300, 400, 500
17	$D_{17}$	Albumin	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
18	$D_{18}$	Protime	10, 20, 30, 40, . . . , 80, 90
19	$D_{19}$	Histology	1, 2

optimal hyperplane that can better differentiate the data points of the two classes (in case of binary classification) [31]. The major reason that motivates machine learning researchers to use SVM for their problems is that SVMs have powerful generalization capabilities to unseen data and they depend on very small number of hyperparameters [32].

Considering a dataset  $D_S$  with  $S$  instances  $D = \{(p_i, q_i) | p_i \in R^Q, q_i \in \{-1, 1\}\}_{i=1}^S$ , where  $p_i$  stands for  $i^{\text{th}}$  instance,  $Q$  represents the dimension of the original feature space of hepatitis data, and  $q_i$  denotes the class labels, i.e., presence or absence of hepatitis disease. The value is 19 for the hepatitis dataset considered in this paper. The SVM model determines a hyperplane calculated by  $g(x) = \beta^T * x + \delta$ , where  $\delta$  represents the bias and  $\beta$  denotes the weight vector. Based on the training data, the hyperplane  $g(x)$  of SVM augments the margin, whereas it curtails the classification error [33]. The sum of the distances between the closest negative and closest positive instances is called margin. In other words, the hyperplane augments the margin distance  $2/\|\beta\|_2^2$ .

SVM uses a set of slack variables denoted by  $\theta_i$ ,  $i = 1, \dots, S$  and a penalty parameter, i.e.,  $\lambda$ , and attempts to maximize  $\|\beta\|_2^2$  and minimize the errors of misclassification [34]. This fact is formulated as follows:

$$\min_{\beta, \delta, \theta} \underbrace{\frac{1}{2}\|\beta\|_2^2}_{\text{Regularizer}} + \lambda \underbrace{\sum_{i=1}^S \theta_i}_{\text{Error loss}}, \quad (1)$$

subject to  $\begin{cases} y_i(\beta x_i + \delta) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, \dots, S \end{cases}$ , where  $\theta$  is the slack variable

that calibrates the degree of misclassification and Euclidean norm or  $L_2$ -norm is the penalty term. A varied version of SVM was introduced by Bradley and Mangasarian which replaces the Euclidean norm, i.e.,  $L_2$ -norm with  $L_1$ -penalty function [35]. The  $L_1$ -penalized SVM produces sparse solutions and has the feature selection property due to its competence of overthrowing irrelevant or noisy features

automatically and hence can be used for feature selection. The formulation of  $L_1$ -penalized SVM is given as follows:

$$\min_{\beta, \beta, \xi} \underbrace{\|\beta\|_1}_{\text{Regularizer}} + \lambda \underbrace{\sum_{i=1}^S \theta_i}_{\text{Error loss}}, \quad (2)$$

$$\text{subject to } \begin{cases} y_i(\beta x_i + \delta) \geq 1 - \theta_i \\ \theta_i \geq 0, i = 1, \dots, S. \end{cases}$$

From the above formulas, it can be seen that, for different settings of the hyperparameter of the  $L_1$  SVM, i.e.,  $\lambda$ , different features will be nulled; consequently, a different subset of features will be produced [36]. The goal is to tune the value of  $\lambda$  in such a way to produce a subset of features which will show best performance in terms of hepatitis disease prediction accuracies. This is done by using exhaustive search methodology. After production of the features' subset, its application to AdaBoost machine learning models is carried out. The AdaBoost model is used for classification task.

AdaBoost (also known as adaptive boosting classifier) is an ensemble learning model. It utilizes boosting approach to construct a metaclassifier by combining the strengths of base classifiers, i.e., weak estimators. The boosting operation helps convert the weak estimators into a stronger or boosted model. During the process of boosting, weighted sum of the base learners or estimators is evaluated to produce the final output of the boosted model. This fact is reflected in the following formulation:

$$G(x) = \text{sign} \left( \sum_{m=1}^M \alpha_m B_m(x) \right), \quad (3)$$

where the  $m^{\text{th}}$  base classifier is denoted by  $B_m$  and  $\alpha_m$  denotes the weight of the  $m^{\text{th}}$  classifier or estimator. To implement the AdaBoost model, we used scikit-learn python API [37]. In the following discussion,  $E$  denotes the total

number of classifiers or estimators used for constructing the eventual AdaBoost model.

The primary objective of this paper is to investigate and exploit the sparsity in the linear  $L_1$ -regularized SVM to further improve the strength of the AdaBoost model. To meet this objective, we develop a cascade of the  $L_1$  linear sparse SVM and AdaBoost model. The full feature set is supplied at the input of  $L_1$  SVM which produces different subset of features based on the value of its hyperparameter  $\lambda$ . Performance of the subset of features is evaluated by their application to AdaBoost model. Thus, in the initial stages, we need to discretize the  $\lambda$  hyperparameter. After discretization of  $\lambda$ , we will have to search the optimal value of  $\lambda$  that will produce optimal subset of features which will show best classification performance. The whole process of the proposed method is shown in the Figure 1. From the figure, it can be seen that initially, a subset of features is generated by utilizing a specific value of  $\lambda$ . The subset of features is given to the AdaBoost model which is trained using one value of  $E$ . For the subset of features, performance is evaluated under optimal  $E$ . Furthermore, another subset of features is generated by utilizing another discrete value of  $\lambda$ , and again the AdaBoost model is trained and evaluated under optimal value of  $E$ . The process is repeated until all the subset of features are evaluated and tested. At the end, the optimal subset of features is selected based on the performance.

### 3. Evaluation of the Proposed Method

In literature, different researchers have utilized various metrics for performance evaluation of their proposed methods. However, for a more realistic evaluation of the performance of our proposed method, we utilized the following five evaluation metrics known as accuracy (ACC), specificity (Spec.), sensitivity (Sen.), and Matthews correlation coefficient (MCC). Accuracy gives information about the total number of correctly classified subjects (whether healthy or patients). Specificity conveys information about the number of healthy subjects which are classified correctly. Similarly, sensitivity represents the percentage of subjects which are classified correctly. MCC is used to measure the quality of binary classification. The basic formulas for these metrics are given as follows:

$$\begin{aligned}
 \text{ACC} &= \frac{tp + tn}{tp + fp + tn + fn}, \\
 \text{Sen} &= \frac{tp}{tp + fn}, \\
 \text{Spec} &= \frac{tn}{tn + fp}, \\
 \text{MCC} &= \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}
 \end{aligned} \tag{4}$$

## 4. Results and Discussion

In this section, the experimental setting and the obtained results are analyzed and discussed. All the experiments (including conventional machine learning-based experiments and the proposed method-based experiments) are performed using Python software (scikit-learn). The experiments were simulated using Intel Core i5 processor with 8 GB RAM and 64-bit operating systems. For the purpose of comparison, we performed two types of experiments. First, the conventional AdaBoost model is developed for the prediction of hepatitis disease. Second, the proposed hybrid model is developed to predict hepatitis disease based on the filtered set of features.

*4.1. Simulation of Conventional AdaBoost Model on Hepatitis Data.* In this experiment, we develop the conventional AdaBoost model for the hepatitis disease data. The model is trained using 70% of the dataset and tested on the remaining 30% of the data. An exhaustive grid search algorithm is used to search the optimized version of the AdaBoost model. The results on both optimal hyperparameters and nonoptimal hyperparameters are given in Table 2. It is evident from the table that best performance of 82.97% accuracy, 11.11% sensitivity, 100% specificity, and MCC of 0.302 is obtained at optimal hyperparameter, i.e.,  $E = 3$ .

*4.2. Simulation of the Proposed Method Using the Sparse Linear SVM and AdaBoost Model on Hepatitis Data.* In this experiment, the proposed learning system is developed by using both the models, i.e., sparse linear SVM and AdaBoost model. The simulation results are reported in Table 3. As can be seen in the table, different values of  $\lambda$  for the sparse SVM generate different subsets of features with different sizes. For subset of features with sizes from  $N = 1-10$ , no improvement in the performance is observed. However, from  $N = 10$  onwards, we see changes in performance of the system. It is evident from the table that best performance of 89.36% is obtained at  $N = 16$ , i.e., with subset of features having only 16 features. However, the best performance on full feature set, i.e., on conventional AdaBoost is 82.97% which is shown in the last row of the table. Hence, it can be observed that coupling the conventional AdaBoost model with sparse linear SVM model improves the performance by 6.39%.

To statistically analyze the results on the testing data, we utilize confusion matrix. As discussed above, the dataset is divided into training and testing datasets using 70-30 data partitioning. Hence, out of the 155 samples, 108 samples are used for training purposes, and the remaining 47 samples are used for testing purposes. Out of the 108 training samples, 23 samples belong to the patient class, and 85 patients belong to healthy class. On the other hand, out of the 47 testing samples, 7 samples belong to the patient group, and 38 samples belong to the healthy group. The predicted results of the proposed  $L_1$  SVM-AdaBoost model are depicted statistically in the confusion matrix in Figure 2.

To further show that the coupling of the sparse linear SVM with conventional AdaBoost model enhances the

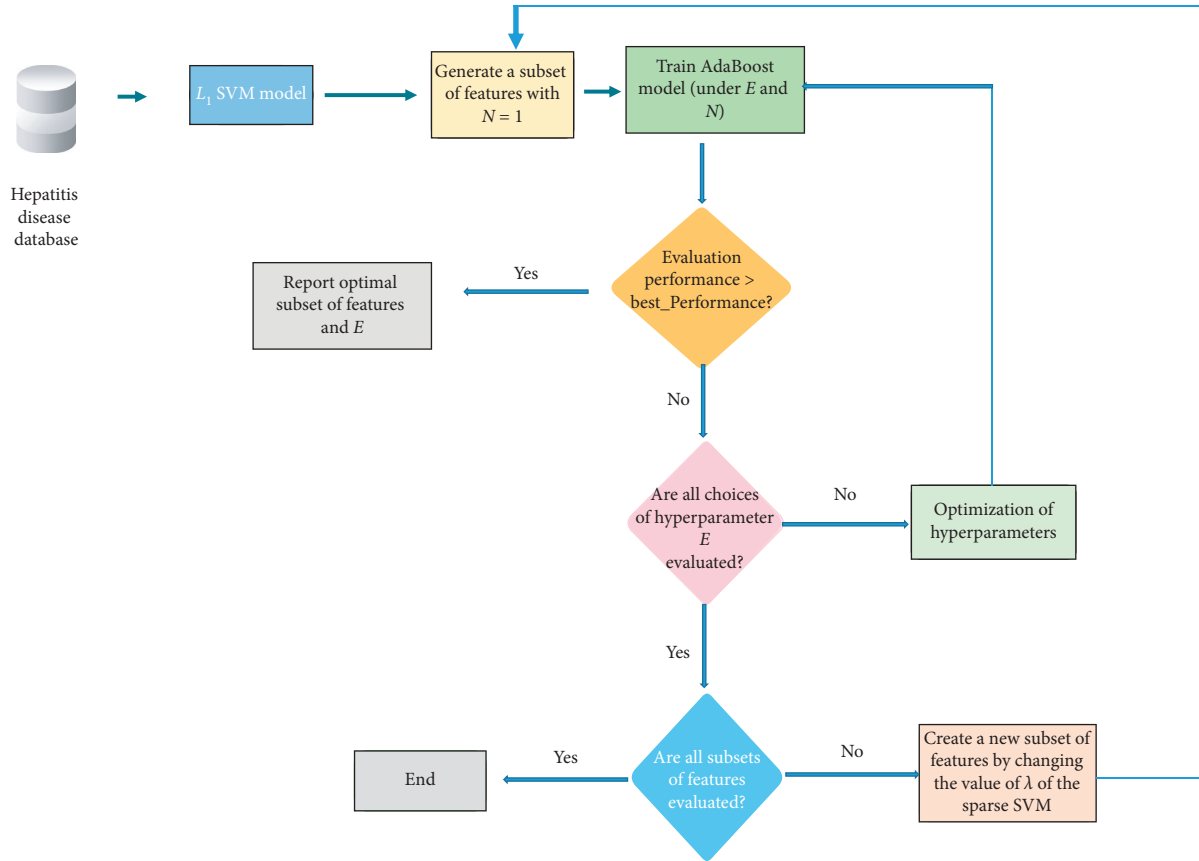


FIGURE 1: Block diagram of the proposed diagnostic system.  $E$ : number of estimators used by the AdaBoost model,  $N$ : size of subset of features, and  $\lambda$ : hyperparameter of the  $L_1$  SVM model that controls the sparsity.

TABLE 2: Performance of the conventional AdaBoost model on HF data.

$E$	$Acc_{test}$	$Acc_{train}$ (%)	Sens. (%)	Spec. (%)	MCC
<b>3</b>	<b>82.97</b>	<b>85.18</b>	<b>11.11</b>	<b>100.0</b>	<b>0.302</b>
10	76.59	93.51	11.11	92.10	0.045
12	74.46	96.29	11.11	89.47	0.007
14	74.46	97.22	11.11	89.47	0.007

Bold values indicate optimal performance.

TABLE 3: Performance of the proposed sparse SVM and AdaBoost-based learning system at optimal hyperparameters of the two models on hepatitis disease data.

$N$	$\lambda$	$E$	$Acc_{test}$	$Acc_{train}$ (%)	Sens. (%)	Spec. (%)	MCC
1	0.01	1	80.85	86.11	22.22	94.73	0.239
2	0.015	1	80.85	86.11	22.22	94.73	0.239
3	0.02	1	80.85	86.11	22.22	94.73	0.239
4	0.04	1	80.85	86.11	22.22	94.73	0.239
5	0.06	1	80.85	86.11	22.22	94.73	0.239
6	0.065	1	80.85	86.11	22.22	94.73	0.239
7	0.07	1	80.85	86.11	22.22	94.73	0.239
8	0.085	1	80.85	86.11	22.22	94.73	0.239
9	0.088	1	80.85	86.11	22.22	94.73	0.239
10	0.09	1	80.85	86.11	22.22	94.73	0.239
11	0.1	8	82.97	90.74	22.22	97.36	0.315
<b>16</b>	<b>0.3</b>	<b>75</b>	<b>89.36</b>	<b>100.0</b>	<b>44.44</b>	<b>100</b>	<b>0.626</b>
17	0.9	36	87.23	100.0	33.33	100.0	0.536
18	3	3	82.97	85.18	11.11	100.0	0.302
19	—	3	82.97	85.18	11.11	100.0	0.302

Bold values indicate optimal performance.

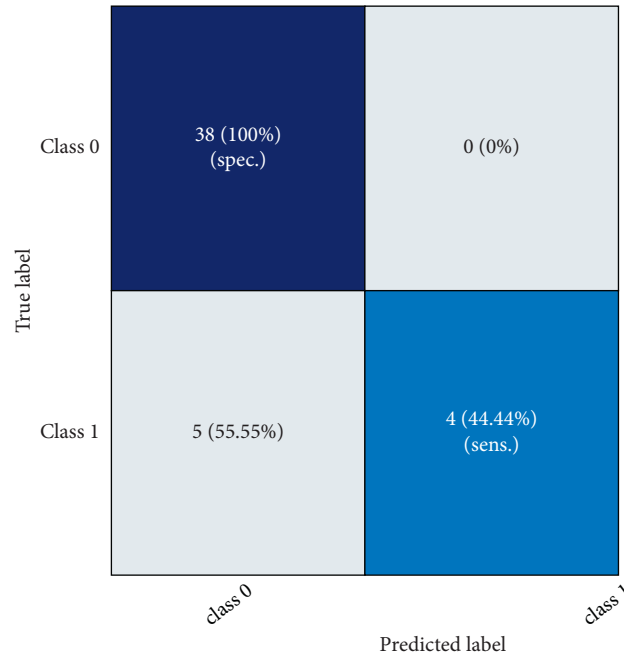


FIGURE 2: Graphical depiction of statistics of the obtained results on the testing dataset in terms of confusion matrix. Spec: specificity; Sens: sensitivity.

performance of conventional AdaBoost model, we use AUC. The AUC in case of conventional AdaBoost model is 0.587, while AUC in case of the proposed method is 0.649. Hence, the ROC charts further validate the fact that the coupling of the sparse linear SVM enhances the performance of AdaBoost for hepatitis disease data.

**4.3. Comparison of the Proposed Method with Some Other Proposed Methods Applied to Hepatitis Data.** The above discussion validates that the learning system proposed in this paper significantly augments the strength of the conventional AdaBoost model. In this section, the effectiveness of the learning system thus developed is further validated by carrying out a comparison of its performance with some of the well-known models presented in previous studies. The prediction accuracies and brief details about the models are given in Table 4. It is evident that our proposed method promises better performance upon 23 other machine learning models.

By analyzing Table 4, it can be seen that previous methods have exploited various machine learning-based methods to improve the hepatitis disease prediction accuracy. For example, Stern and Dobnikar developed methods based on discriminant analysis (including linear discriminant analysis and quadratic discriminant analysis) and could achieve a classification accuracy of 85.8% with quadratic discriminant analysis. Similarly, Ozyildirim and Yildirim developed a number of models for searching out optimum model with better classification accuracy. They obtained the highest classification accuracy of 83.75% using radial basis function (RBF). Moreover, if we analyze the results tabulated in Table 4, the previous methods have carried out analysis of their proposed method by only

considering classification accuracy. In this paper, we analyzed the results of the proposed hybrid method with a number of metrics and proved the robustness of the proposed method from two key metrics, i.e., classification accuracy and area under the curve (AUC).

**4.4. Limitations of the Study.** Although this paper demonstrated the effectiveness of exploitation of sparsity in feature space to improve the performance of the machine learning models, the main limitation is lower sensitivity rate. This is due to the low representation of the patient class in the dataset. The main limitation of the hepatitis disease dataset is its imbalanced nature. The dataset has uneven class distribution, i.e., out of 155 samples, 123 samples belong to the healthy class, and 32 samples belong to the patient class. Recent research pointed out that machine learning models trained under such imbalanced classes show biased performance against the minority class (i.e., the models show very poor performance on the minority class) [40]. On the other hand, the models are biased towards the majority class, i.e., the models will show very good performance on the majority class. In case of the hepatitis disease dataset, the minority class is the patient class, and the majority class is the healthy class. From the results, it can be seen that the majority class has 100% detection accuracy (i.e. 100% specificity) while the minority class has poor detection accuracy, i.e., 44%. In future studies, we need to collect balance datasets, i.e., having the same representation for both the classes. Machine learning models trained under such balanced scenario are supposed to show better sensitivity. Moreover, the exhaustive search method for hyperparameters optimization is time-consuming. In future, application of metaheuristic algorithms [41, 42] should be explored.

TABLE 4: Comparison of the proposed method with some well-known methods proposed for hepatitis disease in terms of prediction accuracy [25, 28, 38, 39].

Model or method number	Model or method	Study or authors	Acc. (%)
1	K-nearest neighbours (KNN)	Nilashi et al.	71.41
2	Neural network	Nilashi et al.	78.31
3	ANaFIS	Nilashi et al.	79.67
4	SVM	Nilashi et al.	81.17
5	ASI	Stern and Dobnikar	82.0
6	Multilayer perceptron + backpropagation	Adamczak	77.4
7	Linear discriminant analysis (LDA)	Stern and Dobnikar	86.4
8	Multilayer perceptron (MLP)	Ozyildirim, yildirim	74.37
9	Radial basis function (Tooldiag)	Adamczak	79.0
10	1NN	Stern and Dobnikar	85.3
11	Radial basis function (RBF)	Ozyildirim, yildirim	83.75
12	15NN, stand. Euclidean	Grudzinski	89.0
13	FSM with rotations	Adamczak	89.7
14	FSM without rotations	Adamczak	88.5
15	Multilayer perceptron with backpropagation	Stern and Dobnikar	82.1
16	Quadratic discriminant analysis	Stern and Dobnikar	85.8
17	(NB and semi-NB), i.e., Naive Bayes and semi-NB	Stern and Dobnikar	86.3
18	Fisher discriminant analysis (FDA)	Stern and Dobnikar	84.5
19	LVQ	Stern and Dobnikar	83.2
20	GRNN	Ozyildirim, yildirim	80.0
21	ASR	Stern and Dobnikar	85.0
22	IncNet	Norbert jankowski	86.0
23	Classification and regression tree (decision tree)	Stern and Dobnikar	82.7
24	LFC	Stern and Dobnikar	81.9
25	$L_1$ -SVM-AdaBoost	The proposed method	89.36

## 5. Conclusion and Future Work

This work developed an automatic hepatitis disease detection system by using machine learning methods. The AdaBoost model was developed for the hepatitis disease prediction. To improve the classification strength of the AdaBoost model, sparsity in the linear SVM model was exploited. The SVM model eliminated redundant or irrelevant features and thus improved the prediction accuracy of the AdaBoost model. It was also shown that the proposed sparse linear SVM also proves helpful in decreasing the time complexity of the AdaBoost model. Moreover, as evident by the simulation results, our proposed method surpassed many previously published methods in terms of hepatitis disease prediction accuracy. Given the experimental quantitative figures and results, it can thus be safely concluded that the proposed methodology can also be exploited to improve performance of other machine learning models and thus can help to make quality decisions in various other disease detection problems as well.

As discussed above, although the proposed method can be used as a tool to improve the performance of machine learning models, the obtained accuracy still needs considerable amount of improvement. Thus, in future studies, more robust cascaded models should be developed by using deep learning approaches for classification. Additionally, the low rate of sensitivity that is caused by lower class representation of the patient class in the dataset is also a limitation of the study that should be considered as an open challenge for the future work. In future studies, extended hepatitis disease datasets should be collected that will have balanced class distribution.

## Data Availability

All the data used in this study are available at the UCI Machine Learning Repository.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] K. Polat and S. Güneş, "Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system," *Applied Mathematics and Computation*, vol. 189, no. 2, pp. 1282–1291, 2007.
- [2] E. Dogantekin, A. Dogantekin, and D. Avci, "Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11282–11286, 2009.
- [3] Y. F. Liaw and C. M. Chu, "Hepatitis b virus infection," *The Lancet*, vol. 373, no. 9663, pp. 582–592, 2009.
- [4] B. Rehmann and M. Nascimbeni, "Immunology of hepatitis b virus and hepatitis c virus infection," *Nature Reviews Immunology*, vol. 5, no. 3, pp. 215–229, 2005.
- [5] K. Polat and S. Güneş, "Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation," *Digital Signal Processing*, vol. 16, no. 6, pp. 889–901, 2006.
- [6] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by



- simultaneous sample and feature selection,” *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.
- [7] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, “Automated detection of Parkinson’s disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.
  - [8] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, “Reliable Parkinson’s disease detection by analyzing hand-written drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model,” *IEEE Access*, vol. 7, pp. 116480–116489, 2019.
  - [9] F. S. Ahmad, L. Ali, H. A. Khattak et al., “A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRS),” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2020.
  - [10] F. S. Ahmed, L. Ali, B. A. Joseph, A. Ikram, R. Ul Mustafa, and S. A. C. Bukhari, “A statistically rigorous deep neural network approach to predict mortality in trauma patients admitted to the intensive care unit,” *The Journal of Trauma and Acute Care Surgery*, vol. 89, no. 4, pp. 736–742, 2020.
  - [11] T. Meraj, A. Hassan, S. Zahoor et al., “Lungs nodule detection using semantic segmentation and classification with optimal features,” *Neural Computing and Applications*, vol. 1, 2019.
  - [12] L. Ali, I. Wajahat, N. A. Golilarz, F. Keshtkar, and S. A. Chan Bukhari, “LDA-GA-SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine,” *Neural Computing and Applications*, 2020.
  - [13] L. Ali and S. Bukhari, “An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction,” *IRBM*, 2020.
  - [14] L. Ali, S. U. Khan, N. A. Golilarz et al., “A feature-driven decision support system for heart failure prediction based on statistical model and Gaussian naive bayes,” *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6314328, 2019.
  - [15] X. Tian, Y. Chong, Y. Huang et al., “Using machine learning algorithms to predict hepatitis b surface antigen seroclearance,” *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6915850, 2019.
  - [16] N. K. Kumar and D. Vigneswari, “Hepatitis-infectious disease prediction using classification algorithms,” *Research Journal of Pharmacy and Technology*, vol. 12, no. 8, pp. 3720–3725, 2019.
  - [17] V. K. Yarasuri, G. K. Indukuri, and A. K. Nair, “Prediction of hepatitis disease using machine learning technique,” in *Proceedings of the Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 265–269, IEEE, Palladam, India, 2019.
  - [18] G. Ahmad, M. A. Khan, S. Abbas, A. Athar, B. S. Khan, and M. S. Aslam, “Automated diagnosis of hepatitis b using multilayer mamdani fuzzy inference system,” *Journal of Healthcare Engineering*, vol. 2019, Article ID 6361318, 2019.
  - [19] M. P. McRae, B. Bozkurt, C. M. Ballantyne et al., “Cardiac scorecard: a diagnostic multivariate index assay system for predicting a spectrum of cardiovascular disease,” *Expert Systems with Applications*, vol. 54, pp. 136–147, 2016.
  - [20] O. Altay, T. Gurgenc, M. Ulas, and C. Özel, “Prediction of wear loss quantities of ferro-alloy coating using different machine learning algorithms,” *Friction*, vol. 8, no. 1, pp. 107–114, 2020.
  - [21] G. Manogaran, R. Varatharajan, and M. K. Priyan, “Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system,” *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4379–4399, 2018.
  - [22] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, “Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network,” *Knowledge-Based Systems*, vol. 132, pp. 62–71, 2017.
  - [23] A. D. Dolatabadi, S. E. Z. Khadem, and B. M. Asl, “Automated diagnosis of coronary artery disease (cad) patients using optimized SVM,” *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 117–126, 2017.
  - [24] M. Abdar, M. Zomorodi-Moghadam, R. Das, and I.-H. Ting, “Performance analysis of classification algorithms on early detection of liver disease,” *Expert Systems with Applications*, vol. 67, pp. 239–251, 2017.
  - [25] D. Çalişir and E. Dogantekin, “A new intelligent hepatitis diagnosis system: PCA-LSSVM,” *Expert Systems with Applications*, vol. 38, no. 8, pp. 10705–10708, 2011.
  - [26] P. Luukka, “Similarity classifier using similarities based on modified probabilistic equivalence relations,” *Knowledge-Based Systems*, vol. 22, no. 1, pp. 57–62, 2009.
  - [27] Y. Kaya and M. Uyar, “A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease,” *Applied Soft Computing*, vol. 13, no. 8, pp. 3429–3438, 2013.
  - [28] M. Nilashi, H. Ahmadi, L. Shahmoradi, O. Ibrahim, and E. Akbari, “A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique,” *Journal of Infection and Public Health*, vol. 12, no. 1, pp. 13–20, 2019.
  - [29] K. Polat and S. Güneş, “An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease,” *Digital Signal Processing*, vol. 17, no. 4, pp. 702–710, 2007.
  - [30] L. Ali, A. Niamat, J. A. Khan et al., “An optimized stacked support vector machines based expert system for the effective prediction of heart failure,” *IEEE Access*, vol. 7, pp. 54007–54014, 2019.
  - [31] S. A. Naghibi, K. Ahmadi, and A. Daneshi, “Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping,” *Water Resources Management*, vol. 31, no. 9, pp. 2761–2775, 2017.
  - [32] S. Maldonado, J. Pérez, R. Weber, and M. Labbé, “Feature selection for support vector machines via mixed integer linear programming,” *Information Sciences*, vol. 279, pp. 163–175, 2014.
  - [33] X. Yuan, Q. Tan, X. Lei, Y. Yuan, and X. Wu, “Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine,” *Energy*, vol. 129, pp. 122–137, 2017.
  - [34] H. S. Jang, K. Y. Bae, H. S. Park, and D. K. Sung, “Solar power prediction based on satellite images and support vector machine,” *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1255–1263, 2016.
  - [35] P. S. Bradley and O. L. Mangasarian, “Feature selection via concave minimization and support vector machines,” in *Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998)*, Madison, WI, USA, 2020.

- [36] J. Zhu and H. Zou, "Variable selection for the linear support vector machine," *Studies in Computational Intelligence Book Series*, vol. 35, pp. 35–39, Springer, Berlin, Germany, 2017.
- [37] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "SCIKIT-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [38] H. L. Chen, D. Y. Liu, B. Yang, J. Liu, and G. Wang, "A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11796–11803, 2011.
- [39] J. S. Sartakhti, M. H. Zangoeei, and K. Mozafari, "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA)," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 570–579, 2012.
- [40] Y. Sun, A. K. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 4, pp. 687–719, 2009.
- [41] S. U. Khan, M. Rahim, and L. Ali, "Correction of array failure using grey wolf optimizer hybridized with an interior point algorithm," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 9, pp. 1191–1202, 2018.
- [42] N. A. Golilarz, H. Gao, R. Kumar, L. Ali, Y. Fu, and C. Li, "Adaptive wavelet based MRI brain image de-noising," *Frontiers in Neuroscience*, vol. 14, 2020.

## Research Article

# A Lightweight Location-Aware Fog Framework (LAFF) for QoS in Internet of Things Paradigm

**Qaisar Shaheen** <sup>1,2</sup> **Muhammad Shiraz**,<sup>3</sup> **Muhammad Usman Hashmi**,<sup>2</sup>  
**Danish Mahmood**,<sup>4</sup> **Zhu zhiyu**,<sup>1</sup> and **Rizwan Akhtar** <sup>1</sup>

<sup>1</sup>*School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, China*

<sup>2</sup>*Department of Computer Science, Superior College, Lahore, Pakistan*

<sup>3</sup>*Department of Computer Science, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan*

<sup>4</sup>*Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad Campus, Islamabad, Pakistan*

Correspondence should be addressed to Rizwan Akhtar; rizwan@just.edu.cn

Received 18 March 2020; Revised 13 August 2020; Accepted 29 August 2020; Published 16 September 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Qaisar Shaheen 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.

Realization of Internet of Things (IoT) has revolutionized the scope of connectivity and reachability ubiquitously. Under the umbrella of IoT, every object which is smart enough to communicate with other object leads to the enormous data generation of varying sizes and nature. Cloud computing (CC) employs centralized data centres for the provisioning of remote services and resources. However, for the reason of being far away from client devices, CC has their own limitations especially for time and resource critical applications. The remote and centralized characteristics of CC often result in creating bottle necks, being latent, and hence deteriorate the quality of service (QoS) in the provisioning of services. Here, the concept of fog computing (FC) emerges that tends to leverage CC and end devices for data congestion and processing locally in a distributed and decentralized way. However, addressing latency and bottleneck issues for time critical applications are still challenging. In this work, a lightweight framework is proposed which employs the concept of fog head node that keeps track of other fog nodes in terms of user registrations and location awareness. The proposed lightweight location-aware fog framework (LAFF) persistently satisfies QoS by providing an accurate location-aware algorithm. A comparative analysis is also presented to analyse network usage, service time, latency, and RAM and CPU utilization. The comparison results depicts that the LAFF reduces latency, network use, and service time by 11.01%, 7.51%, and 14.8%, respectively, in contrast to the state-of-the-art frameworks. Moreover, considering RAM and CPU utilization, the proposed framework supersedes IFAM and TPFC targeting IoT applications. The RAM consumption and CPU utilization are reduced by 8.41% and 16.23% as compared with IFAM and TPFC, respectively, making the framework lightweight. Hence, the proposed LAFF improves QoS while accessing remote computational servers for the outsourced applications in fog computing.

## 1. Introduction

The concept of Internet of Things (IoT), supported by computational intelligence, has revolutionized in almost all domains of life. With every passing day, many new applications and domains in IoT and computational intelligence are emerging to help mankind in one way or other. On the other hand, providing such applications to general public has opened new horizons of business as well. In IoT, such

businesses and applications mostly rely on the sensory data that has to be gathered for effective decision making. For fusion or manipulating big data (that may be some streaming data or in shape of batches), there are some requirements, i.e., distributed processing capability, effective communication and uncompromised network so that decision making process may yield better accuracy. Clouds being service providers tend to solve these problems. However, for the reason of being faraway from client

devices, they have their own limitations for time critical applications. Hence to reduce such complexities, the models of fog or edge computing are employed. Basic infrastructure for such environment comprises of *things* (computing devices) which have computing, communicating, and storing capability. Based on current trends, it is expected that by 2025 such smart environments will incorporate over 1 trillion IoT devices with 50% increased demand for latency sensitive applications [1]. Fog computing (FC) refers to a hierarchically distributed computing paradigm that bridges cloud data centers and IoT devices. The fog environment offers both infrastructure and a platform to run diversified software services. At different hierarchical levels of the fog environment, the physical devices are commonly called fog nodes. This technology overcomes the limitations of cloud computation by enabling data acquisition, processing, and storage at decentralized and locally available fog nodes [2]. The idea of this model is initially described by CISCO [3]. Figure 1 shows the general architecture of FC.

However, ensuring rich user experience (QoS) is the main concern to be addressed specially for time-sensitive applications such as health care IoT [4], web-based gaming [5], and video streaming applications [6]. The large distance between users and end devices increases the number of routers/hops which results in higher latency rate and network usage. Hence, real-time provisioning of services is obstructed and QoS is decreased while leveraging remote fog nodes for the outsourced applications.

In this work, we propose a lightweight location-aware fog framework (LAFF) which employs the concept of fog head node that keeps track of other fog nodes in terms of user registrations and location. The proposed LAFF persistently improves QoS by employing location-aware algorithm. LAFF addresses the issues of high latency, service time, and network usage in distributed data on fog server in order to improve the QoS. The location-aware algorithm involves user registration on fog head. The user/actuator is responded by analyzing their requested data types. Data types are divided into multimedia data (MMD) and textual data (TD). QoS through LAFF is compared with other contemporary frameworks [7, 8] to validate performance of the proposed framework. The significant contributions of this research are include the following:

- (i) A fog-based lightweight framework is devised to provide better QoS to users
- (ii) A location-aware algorithm is developed that enables latency reduction, service time reduction, and minimal usage of network resources
- (iii) Resources utilization (RAM and CPU) is reduced to make the framework lightweight

The rest of the paper is organized as follows. The literature review is presented in Section 2. Section 3 discusses the lightweight location-aware fog framework (LAFF), location-aware algorithm, architecture, and analytical model. Section 4 focuses on the experimental setup of the framework. Section 5 presents the evaluation of the LAFF. Section

6 details the results of the simulations and discussion. The concluding remarks are conducted in Section 7.

## 2. Literature Review

In a simplified structure, FC is characterized by a geographically distributed computing design, prepared with heterogeneous devices connected at the edge of the network. The authors in [9, 10] highlight the advantages gained from FC. An algorithm is developed and implemented in [11] which is based on local computing. Through this algorithm, workload of cloud and fog processing is reduced. However, the proposed algorithm only works with star topology. In [12], a new layer is proposed, the fog layer (computing) of resources which is closer to the edge of the network to provide location awareness. A Fog-2-Fog (F2F) coordinated effort model is proposed in [13] that presents offloading approach amongst fog nodes, as per their load and handling capacities by Fog Resource Management Scheme (FRMC). In [14], the idea of resource allocation in a fog environment is introduced. The authors present a three-layer architecture that includes cloud, fog, and the user to divide the workload between the cloud and fog nodes. However, proposed architecture is only for homogenous environment. Moreover, scalability and associated challenges are not considered in the study.

Fan and Ansari [15] discussed the problem of load balancing in fog network through a distributed technique that assigns IoT devices to appropriate fog nodes and reduce the latency. In this technique, a fog node periodically broadcasts the computing and traffic estimated load. An FC framework is devised in [16] considering the medical field. Resource management is tackled by considering fog association, placement of VM, and task distribution. In [17], the workload placement algorithm is devised in tier edge cloud network to improve the response time of all tasks. The algorithm allocates computing resources between different tiers of fog node for completing assigned task. The idea of distributing the workload of a fog server receiving high traffic from IoT is presented in [18]. Two load balancing algorithms (task distributing and task grasping) are developed in [19] for large-scale FC. Through this structure, load balancing overhead is reduced when the scale of fog increased to get benefits of centralized and decentralized computing.

Puthal et al. [20] focus on developing an efficient dynamic load-balancing algorithm with an authentication method for edge data centers. Tasks were assigned to an underutilized edge data center by applying the breadth-first search (BFS) method. Each data center is modeled using the current load and the maximum capacity used to compute the current load. The authentication method allowed the load-balancing algorithm to find an authentic data center. IoT resource provisioning issue is discussed in [21], and a solution is proposed to overcome this problem. The model aims to boost fog resources and the minimization of system delay. The work in [21] is extended in [14] where QoS measurements and the deadlines for the provisioning of each kind of resources are considered.

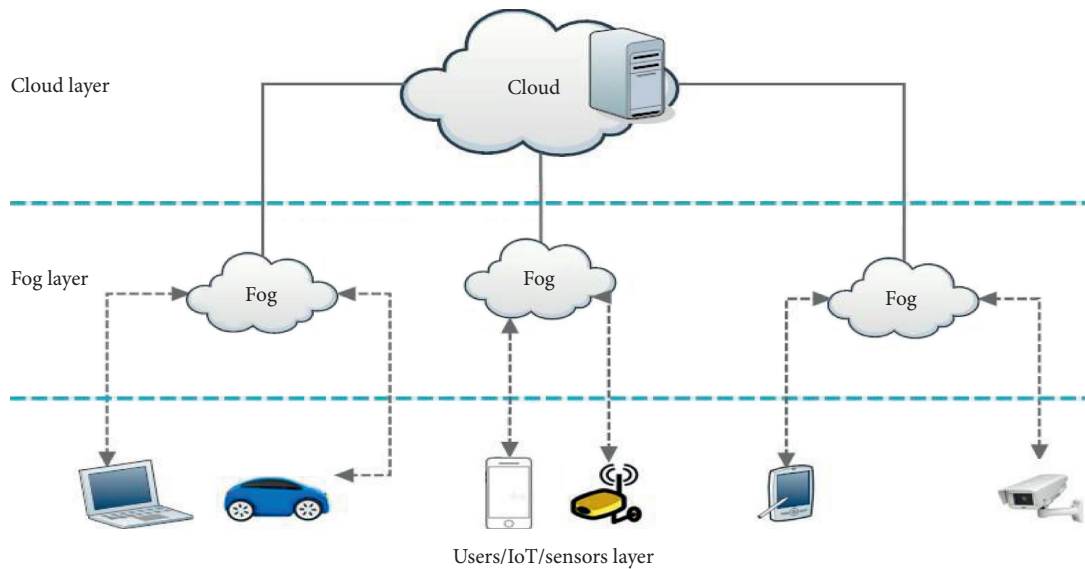


FIGURE 1: Architecture of fog computing.

In [22], a framework named FOGPLAN for QoS-aware dynamic fog service provisioning (QDFSP) is introduced. In order to meet low latency and QoS requirements of applications, QDFSP dynamically deploys application services on fog nodes or the release of application services that have previously been deployed on fog nodes. However, different characteristics of wireless and wired fog nodes are not considered. Also, the framework is neither location aware nor fulfils the real-time requirements of IoT tasks. Lin and Shen [23] designed a fog-based lightweight system to develop cloud gaming with high QoS. This system is a three-layer model, including cloud, fog, and devices (e.g., desktops/smartphone players). A set of supernodes were considered, which were near to end users and are connected to the cloud. The QoS requirements are achieved through reducing latency and bandwidth consumption.

A service management technique is introduced in [24] as iHome for smart house in the cloud. The paper proposed a service oriented architecture (SOA) to monitor home applications with real-time responses. The performance of services in terms of CPU and RAM in iHome is evaluated. The results show that the real-time responses can be returned under heavy burden of loading. The proposed system is tested under a limited number of physical appliances in a modular approach. However, many other important influential factors like cost and energy consumption are not addressed. Also, the system does not consider the user management and network condition. The proposed framework FATEH in [25] uses a three-layered architecture to improve QoS parameters. The first layer contains IoT devices and an agent node to collect data, and the gathered data are then submitted to the next layer. The third layer consists of fog manager to efficiently process the request on smart fog node. The processing and storage of less sensitive data are done at the third layer. The data coming from the fog manager are also processed at the third layer. The drawback of this system is that it does not

take into account the network condition and user management.

An algorithm for task management in fog infrastructure is proposed in [26] aiming to focus on task scheduling at the fog layer while minimizing the response time dependent on resources requested by these tasks. In any case, explicit QoS prerequisites have not been considered in their methodology. Zeng et al. proposed an algorithm in [27] that works with a unified scheme for mobile IPV6 and suggests scheduling and handle user mobility. The issue of resource sharing among the fog nodes to execute computational requests was discussed, while they especially focused on fog-enabled little cells in cellular systems. In [28], Kim and Chung target the shaping clusters of small cells, where each cluster represents a collection of little cells that offer resources for offloading mobile devices from their remaining workload. The aim of this work is to reduce latency for each user through clusters shaping, bandwidth allocation, and computational resources.

Location-based services (LBSs) [29] become increasingly popular in recent years due to recent advancements in mobile computing. LBS refers to service provisioning through location-based information of users, i.e., the geographic position.

In [30], website performance optimization is automated by fogging at the edge servers. This idea explains the significance of edge location by giving dynamic and customizable optimization dependent on local network and the conditions of user's devices. WiCloud [31] is developed as mobile-edge computing platform with OpenStack to improve location awareness and to manage inter-mobile-edge communication and data acquisition for an innovative service.

Providing an acceptable level of QoS is an important issue in FC [32]. To design an efficient fog-based system, various QoS factors are considered. Extracting from the literature, eleven factors of QoS are defined, i.e., latency,

security, service time, availability, cost, energy consumption, resource utilization, reliability, execution time, deadline, and scalability [33, 34]. Moreover, latency is investigated as one of the important factors of QoS.

A framework is required to ensure QoS provisioning without burdening a single resource and provide service near to the edge focusing abovementioned performance metrics. This framework needs to be more useful to reduce latency, service time, and network use through user and location management by considering the network condition. A lightweight LAFF is devised through this study, and the framework possesses following characteristics.

LAFF has taken into account various IoT data requirements (multimedia data and textual data). Major emphasis in proposed framework is given to location awareness, i.e., knowing the exact location of the users/actuators. LAFF registers users on fog heads and employs  $K^*$  heuristic algorithm [35, 36] to find the shortest path between user and fog node. Moreover, the algorithm also takes the decision of fog head selection considering the requested data type.

The proposed work is compared with IFAM (intelligent FC analytical model) [7] and TPFC (task placement on fog computing) [8]. In [7], an analytical model and reinforcement learning algorithm in an FC environment is introduced. This model aims to reduce the latency among healthcare IoT, cloud servers, and end users. This paper proposes a novel multitier fog processing system that provides IoT services. However, in this work, the author did not consider user's location and network condition. The other drawback of this research is that user's request for normal data are transferred to the cloud to respond. The LAFF is better in terms that it considers user's location and network conditions. The framework also transfers both type of data, MMD and TD, to fog to fulfill user's request. In [8], a context-aware information-based approach ideally uses virtual resources accessible on the system edges to improve the presentation of IoT benefits in terms of response time, cost, and energy decrease. The approach utilizes context-aware information including network conditions location of IoT devices and service type to provide resources to IoT applications. However, the increase in the number of fog nodes and services causes an exponential increase in time for problem solving.

### 3. Location-Aware Fog Framework (LAFF)

In the proposed LAFF, location awareness under fog computing umbrella is introduced to reduce the latency, service time, and network usage along with minimal resources utilization. LAFF employs a location-aware algorithm that has ability to trace user's exact location through fog head. Fog head is the controller of data center of all fog nodes. The idea of fog head is used in fog computing technology [38]. The fog head node is not only limited to search for current nodes but also for new nodes ( $F_{\text{head}} \rightarrow F_{\text{MMD}} + F_{\text{TD}} + F_{\text{others}}$ ).  $F_{\text{head}}$  represents fog head node,  $F_{\text{MMD}}$  refers to the fog multimedia data node,  $F_{\text{TD}}$  is fog textual-data node, and  $F_{\text{others}}$  are  $n$ th new fog nodes. The

search radius of fog head ( $F_{\text{head}}$ ) is extended to  $n$ th new nodes as the framework is developed by keeping the idea of scalability as well. After accessing the user's exact location, fog head dedicates a nearest fog node in response to the user's request considering requested data type. If any nearest fog node is hard to reach, then the  $k^*$  algorithm is used to find the shortest path from user/actuator to fog node by estimating the coordinates [35, 36]. This dedicated node serves the user without any interruption. This framework also registers users (user management) and determines the requested data type. TD requests include text-based information, images, etc. (fog-TD servers handle these data types). MMD requests include videos, movies, etc. (fog-MMD servers handle these data types). The lightweight LAFF reduces the latency  $L_{\text{id}}$ , service time  $f$ , and network use  $u_{\text{nw}}$ . Figure 2 shows a detailed view of the LAFF.

#### 3.1. Components of LAFF

**3.1.1. Cloud Layer.** The top layer of lightweight LAFF is a cloud layer which coordinates with lower layers for data collection and storage for future use. The cloud layer can be used for data processing and storage for a large amount of the data for longer duration. If the fog head fails to provide services to the user then cloud facilitates the users. Cloud layer components are as follows.

**3.1.2. Cloud.** Cloud is placed at higher layer of the lightweight LAFF. Cloud facilitates the fog layer in terms of storing data for later use and high processing when needed. Cloud servers are the centralized hosts. Cloud possesses all the necessary software needed to run, and it can also work as an independent unit. Cloud layer plays a supervisory role to handle communication and data storage. Cloud storage has many distributed resources acting as one unit. This distribution of data makes the cloud very fault tolerant. In this work, cloud is connected to the fog head to communicate with all fog nodes. Cloud communicates to fog head for all necessary communication. Cloud agent is responsible to manage communication between fog head and cloud.

The fog layer is the middle layer of the lightweight LAFF which aims to provide the processing facility of the data near to the edge. The following sections explain the modules of the fog layer:

**3.1.3. Fog Head.** Fog heads are fixed and predetermined physically with respect to geographical region and have larger hardware resources. Fog head is deployed between the fog nodes and the cloud and is responsible to communicate with cloud and all fog nodes. Fog head works according to the devised algorithm to access user's location and to identify requested data type. Users are registered at fog head. Fog head knows the exact location of all fog nodes. Tasks are assigned to the nodes considering the requested data type. Fog head is also responsible to manage and maintain the information on hardware level. Fog head has the following

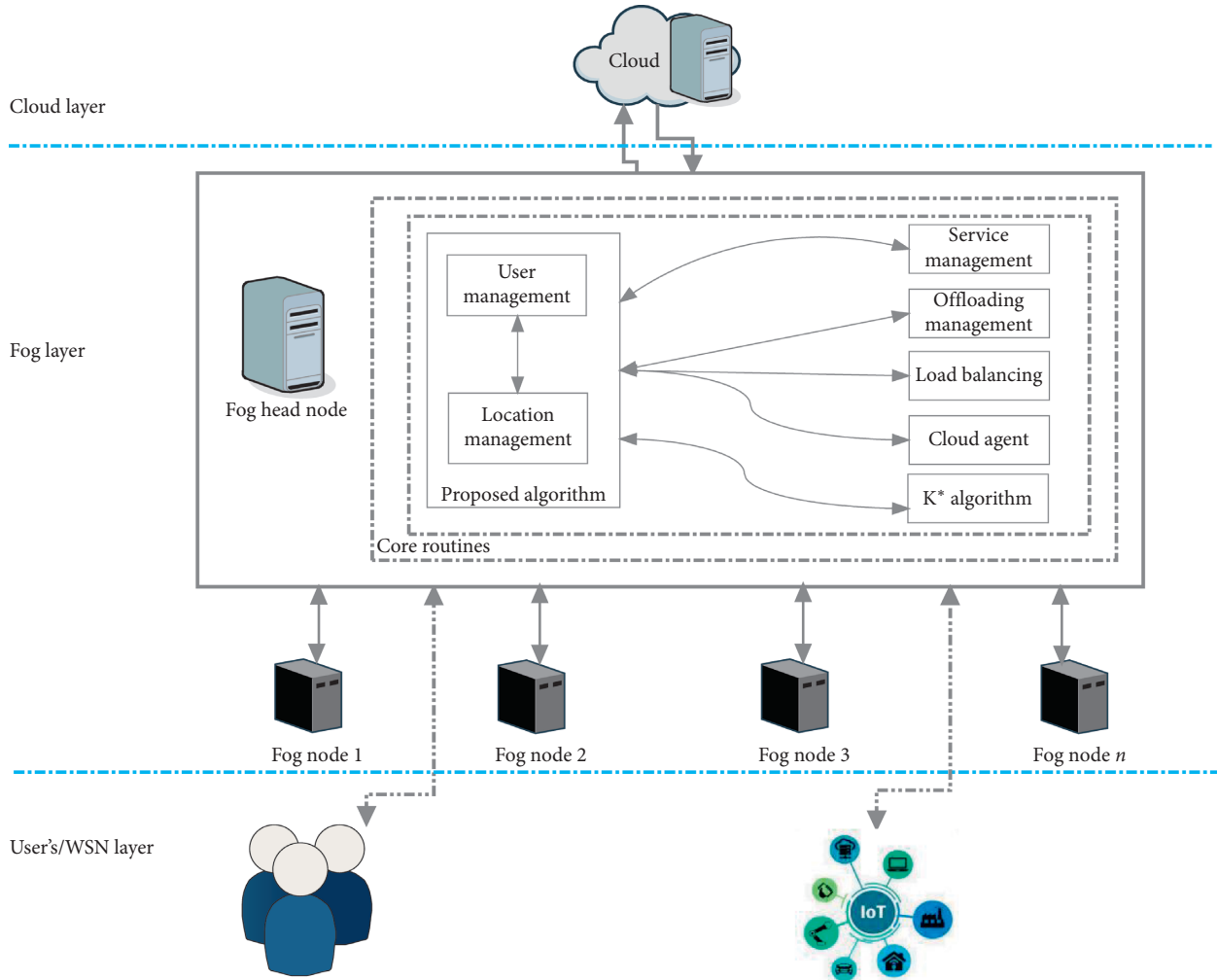


FIGURE 2: A detailed view of lightweight LAFF.

helping modules, and the proposed algorithm calls these modules as per their requirement.

**3.1.4. User Management Module.** Registration or management of the users and to store their details for future use is the responsibility of the user management module. The registered users are stored in Hashmap against specific identifiers for fast track communication. The advantage of using Hashmap is that it is not synchronized and hence saves additional usage of network and service time. The user management module communicates with the location management module to update/get user's location to provide services.

**3.1.5. Location Management Module.** The location management module manages the location of the users. The location is configured with coordinates. The  $x$  and  $y$  are range coordinate variables that are used for finding the shortest path in  $K^*$  search. The coordinates from 10 to 50 identify the location of existing users while other coordinates (coord1 and coord2) contain new users and  $n$  represents the coordinate value range. The mathematical representation of Geo function is described in the following equation:

$$\forall x \cup y \exists \text{Geo}(\text{coord1}, \text{coord2}) \Delta \text{coord} \theta < n \cap \text{coord} \theta > n, \quad (1)$$

where  $10 \leq n \leq 50$ .

On each request of the user, the location management module is accessed to match/update the location management table.

**3.1.6. Service Management Module.** The service management module (SMM) provides services to the fog layer. SMM registers services and coordinates with fog nodes to provide services to users. It manages the fog node service delivery assurance. SMM monitors all the resources of the nodes and fog head.

**3.1.7. Offloading Management Module.** The offloading management module offloads the task from a fog node and assigns to other nodes to provide dedicated services to assure QoS. Through this module, the framework enables to offload the task from a fog node and dedicate to the user.

**3.1.8. Load Balancing Module.** The load balancing module distributes the fog layer traffic among different fog nodes. Through this module, the framework becomes more responsive and available for users.

**3.1.9. Cloud Agent Module.** The cloud agent module facilitates the fog head and cloud to communicate with each other for storing and updating data on the cloud. The cloud agent module works as a broker between the fog layer and the cloud layer.

**3.1.10. Fog Nodes.** Fog nodes work as a server of the geographic area in which the fog node is deployed. Fog nodes process data near to the edge to reduce the burden of the cloud. Through the fog nodes, the lightweight LAFF provides better QoS by reducing latency and service time to accomplish the request.

**3.1.11.  $K^*$  Algorithm Module.** Through the proposed algorithm, user's location is accessed and nearest fog node is assigned to the user to fulfill the request. If this nearest fog node is hard to reach due to any abnormality, this algorithm uses a heuristic search algorithm  $k^*$  [35], which is used to find shortest path between users and fog node. The vertices in this case are added between register and unregistered users. The advantage of using  $K^*$  algorithm is that it only uses the executed portion of the graph. It reduces the network usage by only working on a required portion instead of communicating to whole weighted graph. The complexity of  $K^*$  algorithm is  $O(n \log n + r_u + u_u)$ , where  $n$  is the number of vertices. In [39], Mishra et al. used the same  $k^*$  algorithm to find shortest path between source and destination.

### 3.2. Proposed Algorithm.

The LAFF algorithm is provided in Algorithm 1.

**3.3. Features of LAFF.** To minimize the service delay, fog head communicates to fog nodes and queries are processed on a short distance; in this way, the service latency is minimized. If queries are not communicated through fog nodes and transferred to the upper layer like fog head and cloud, then the service delays are at a larger value. The latency  $L_{ld}$  is calculated by dividing available time  $T_{available}$  with total time  $T_{total}$  under product of 100. The following formula is used for calculating latency:

$$L_{ld} = \frac{T_{available}}{T_{total}} * 100 \text{ (milliseconds)}. \quad (2)$$

IoT service delay-minimizing policy: policy adopted in this regard is to implement a minimum delay tolerance system. The values are considered to be very low as compared with that of other systems' latency. Latency, network use, and service time are reduced by using equation (1).

If a fog node is hard to reach, the LAFF uses  $k^*$  algorithm to find shortest path between users and the IoT devices. The

path is selected from a pool of fog nodes ( $F_1-F_n$ ) to have idle space for processing in order to provide better QoS. The list of fog nodes  $F = \{F_1 + F_2 + F_3 + \dots + F_n\}$  and users  $U = \{U_1 + U_2 + U_3 + \dots + U_n\}$  having tasks  $T$  for updating the Cloud  $C$  is represented by the following equation:

$$U_n \prod (F_n, T) \longrightarrow C. \quad (3)$$

Equation (3) represents the  $n$  array product from completion of task to update the cloud.

The remaining components of the proposed paradigm are mathematically defined in the analytical model. In the analytical model, we have discussed the mapping between the components of different layers. In the simulation, we implemented the analytical model in iFogSim.

**3.4. Analytical Model.** A set (S)  $T_{IoT}$  for all sensors  $S \{S_1, S_2, S_3, \dots, S_n\}$  and actuators  $A \{A_1, A_2, A_3, \dots, A_n\}$  under a tuples load  $\alpha$  with transmission time  $L'$  is defined. Events  $E \{E_1, E_2, E_3, \dots, E_n\}$  happen at sensors  $\{S\}$ , where  $n$  is the  $n$ th mapped sensor to an event  $E$ . Equations (4) and (5) represent the events that happened at fog °F and cloud °C through sensors:

$${}^\circ\text{F}: T_{\text{iot}}\{S_n, E_n\} * \alpha \longrightarrow A_n, \quad (4)$$

$${}^\circ\text{C}\#\text{G}: T_{\text{iot}}\{S_n, E_n\} * \alpha \longrightarrow A_n. \quad (5)$$

Within the increase in the number of hops amongst sensors, the latency, network usage, and service time also increase. The mapping of a sensor to a fog node is described in equation (6) that expresses the relationship between transmissions. Here,  $i$  is the IoT device number,  $j$  represents the column of devices where IoT device are mapped, and  $M$  is the mapping.  $L$  is the load (MMD or TD load),  $S$  is the sensor, and  $F$  is the fog node:

$$\mathbf{M}(i, \mathbf{1}): \Sigma_{i=1}^n, \quad j = \mathbf{1}L' * \mathbf{S}i \longrightarrow {}^\circ\text{F}. \quad (6)$$

The latency  $L_{td}$  is computed using equation (2).

The service time  $f$  is expressed in terms of time taken by a service provider SP  $\{SP_1, SP_2, SP_3, \dots, SP_n\}$  by providing a service  $\check{T}$  to a user(s)  $\acute{u} \{U_1, U_2, U_3, \dots, U_n\}$ . The mapping relation is explained in

$$f: \mathbf{SP} \longrightarrow \acute{u} \prod L' * \check{T}. \quad (7)$$

To calculate the service time  $f$  in simulation environment, the following equation is used:

$$f = C_{\text{ins}}(T_{\text{ms}}) - T_k(S_t) \text{ (ms)}, \quad (8)$$

where  $C_{\text{ins}}(T_{\text{ms}})$  represents the time in milliseconds fetched by calendar instance and  $T_k(S_t)$  is the simulation time stored by timekeeper class. The simulation time is the amount of time spent in processing the  $K^*$  search, allocating nodes, processing requests of users, and updating cloud related to processing. In order to calculate network usage  $\acute{u}_{\text{nw}}$ , the tuple  $T_{\text{ud}}$  captured by network usage monitor  $M_{\text{nu}}$  are added to the total bandwidth used  $B_u$  in transmission and then divided by maximum simulation time  $ST_{\text{max}}$ . the following equation is used for calculation:



```

Inputs: tasks  $T$ , start services  $S$ , user  $u$ , Geo (coord1, coord2) gets integer based coordinates.
Output: assign nearest Fog-MMD or Fog-TD to the user.
start;
submit tasks;
place operators;
start services;
while allusers do
  getlocation;
  if coord > 10 coord < 50 then
    existing reg user;
  else
    reg as new user;
  end
  if reguser then
    if multimedia data then
      if clocation == plocation then
        if hard to find then
          start  $K^*$  search;
          calculate tasks on nodes ( $F_n * T_n$ );
          offload data from nearest fognode ( $F * T - 1/T$ );
          allocate fog-MMD;
           $F(u, T)$ ;
        else
          find nearest fog node;
          Search ( $F_1 \rightarrow F_n$ );
        end
      else
        register location;
        Geo (coord1, coord2);
      end
    else
      transfer to fog-TD;
       $F_1(u, T)$ ;
    end
  else
    unreg user;
  end
  find idle fog node;
   $u_u(F_1 \rightarrow F_n)$ ;
  send to cloud;
   $u_u(C, T)$ ;
  repeat;
end

```

ALGORITHM 1: LAFF.

$$\dot{u}_{nw} = M_{nu}(T_{ud}) + \left( \frac{B_u}{ST_{max}} \right) (\text{Kbps}). \quad (9)$$

#### 4. Experimental Setup

This lightweight LAFF is developed by conducting extensive simulation in CloudSim [40] and iFogSim simulators [41]. CloudSim is responsible for the simulation and events handling at Cloud. iFogSim handles events at Fog devices. This also minimizes the latency as servers become near to the edge of devices [42]. Following are the important steps and parameters which are needed to execute simulation.

The calendar is initialized to keep the current instance to conclude at the end when the simulation starts. In the end,

the simulation variable is initialized by tracing flag to “false” so that the detail log which is not relevant to the simulation is not shown. The fog broker is initialized based on the data center broker. Considering the requirements of the clients related to QoS, the data center broker class coordinates between users and cloud service. A fog broker helps users to create tuples on the fog. Tuples are extended from cloudlets class to model tasks in CloudSim and iFogSim.

The cloud and fog data centers have their own characteristics. In real case, the characteristics of fog device are less powerful and have less storage than cloud data center. The capacity function of cloud  $\zeta(l + d)^n$  and fog  $F(l)^n$  for load  $l$  and new expected data  $d$  is represented in equations (10) and (11):

$$\mathcal{C}(l+d)^n = \sum_{k=0}^n \binom{n}{k} u^k d^{n-k}, \quad (10)$$

$$F(l)^n = \sum_{k=0}^n \binom{n}{k} u^k l^{n-k}, \quad (11)$$

where  $n$  is the number of total requests and  $k$  represents the capacity of responses that is sent against the requests  $n$ . The response  $k$  is always sent against request  $n$ . The  $\mathcal{C}(l+d)^n$  function equation (10) shows that the cloud has more storage than fog devices (equation (11)).

**4.1. Cloud Data Centers.** Cloud data centers (CDCs) are the centralized hosts and play a supervisory role to handle communication and data storage. Cloud storage has many distributed resources acting as one unit.

**4.2. Fog Data Centers.** Fog data centers store data for further processing and communication with users.

**4.3. Location Manager Data Centers.** Location management data store information regarding user's location.

**4.4. Fog Head.** Fog head knows the location of all fog nodes and communicates between the cloud and all nodes. Fog head is also responsible for managing and maintaining the information on the hardware level. Characteristics of fog data center, location manager, proxy server, fog head, Fog-TD, and Fog-MMD are given in Table 1.

**4.5. Gateway Devices.** These gateway devices are part of the fog layer and communicate with proxy server and cloud devices. Table 2 represents characteristics of gateway devices.

**4.6. Sensor Devices.** Sensor devices are created for scenarios which produce the data with following characteristics (Table 3).

**4.7. Sensors and Actuators.** As the actual device model is based on sensor devices, generating a huge amount of data that need to be processed, each device involves a sensor and an actuator attached to it. The purpose of the sensor is to "sense" the data which are identified by the selector module of the server.

**4.8. Module to Module Interaction.** Tuples are sent from one module to the other in order to interact with each other. The tuples which are sent up to the fog or cloud for processing are identified as TuplesUp and tuples that are sent downward from one module to the other are TupleDown. Also, tuples are mapped to modules using the tuple mapping techniques defined in iFogSim. The network usage is calculated on the basis of tuple flow. The network usage  $\mu^n$  is defined in terms of  $\mu^{\text{fog}}$  (fog network length) and  $\mu^{\text{cloud}}$  (cloud network

length) by dividing a tuple size  $T^L$  with simulation total time  $st$  as presented in the following equation:

$$\mu^n \longrightarrow \mu^{\text{fog}} + \mu^{\text{cloud}} \Pi \frac{T^L}{st}. \quad (12)$$

## 5. Evaluation of the LAFF

The fog-based approach of the LAFF is shown in Figure 3: Initially, the normal flow of the system is as follows:

User -> UserIdentifier -> ServiceHandler -> FogHead -> proxyServer -> Fog (MMD or TD) -> Cloud-server

The fog head handles user's requests. Through the location management module, user's location is traced and a fog head is deployed there to respond. If the location manager is not idle, then the proxy server can be formed. Fog head asks user identifier to identify the type of requested data. Requests may be for MMD or TD. After the fog head determines the type of the requested data, it allocates the required fog nodes to the users. The specific fog node facilitates the user accordingly. The fog-MMD node is loaded with very powerful processing capability, whereas a low spec is configured on fog-TD node. Table 1 represents the specifications of both fog-TD and fog-MMD. The proposed algorithm makes this work so unique and distinguishing.

The flow after initial one is given below:

User -> UserIdentifier -> ServiceHandler -> FogHead -> Fog (MMD or TD) -> User

If fog head fails to identify the relative fog service provider, the request is then transferred to the Cloud-server to facilitate the user as represented in Figure 4. The lightweight LAFF is a fault-tolerant framework due to the cloud's availability in case fog head fails to fulfill the request.

**5.1. Data Configuration.** A data set with tuple size 3000, bandwidth 1000, and network length 500 is implemented in the below mentioned configurations. The tuple in iFogsim represents the term data row, where there are sequences of bytes in such data rows.

Simulation runs on iFogSim for different configurations. The configurations are presented in Table 4.

The results of the abovementioned configurations are shown below.

**5.1.1. Use Case.** To prove the significance of the proposed algorithm, a use case is described.

**5.1.2. Actors.** Jeena, thief, and users (police vans) were the actors

**5.1.3. Preconditions.** Registered user with known location and requested MMD.

TABLE 1: Characteristics of devices.

Name of device	*MIPS	RAM	Uploading bandwidth	Downloading bandwidth	Level
Fog-data-center	20000	10 GB	10 Gbits	10 Gbits	1 (cloud child)
Location-manager	2000	1 GB	10 Gbits	10 Gbits	1
Proxy server	2000	2 GB	10 Gbits	10 Gbits	1
Fog-head	20000	8 GB	10 Gbits	10 Gbits	2
Fog-TD	20000	2 GB	10 Gbits	10 Gbits	3 (fog head child)
Fog-MMD	20000	4 GB	10 Gbits	10 Gbits	3 (fog head child)

\*MIPS = million instructions per second.

TABLE 2: Characteristics of gateway devices.

Million instructions per second	RAM	Uploading bandwidth	Downloading bandwidth	Level
1000 Mips	1 GB	10 Gbits	10 Gbits	4

TABLE 3: Characteristics of sensor devices.

Million instructions per second	RAM	Uploading bandwidth	Downloading bandwidth	Level
1000 Mips	1 GB	10 Gbits	256 Mbits	4

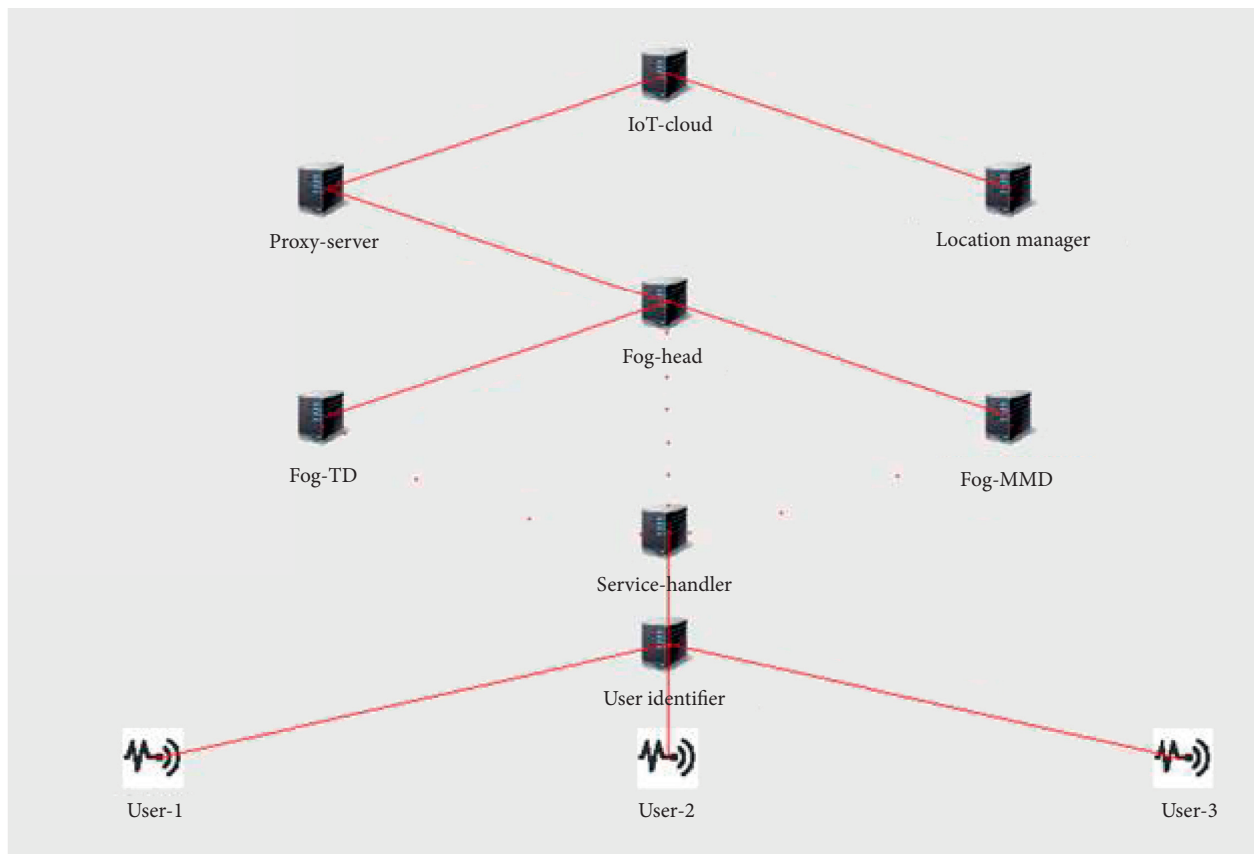


FIGURE 3: Topology of the LAFF.

5.1.4. *Postconditions.* A user is able to request fog framework to access CCTV cameras to get live streaming.

5.1.5. *Scenario.* Jeena is walking through a street; a thief snatched her bag and ran away. Jeena called the police and

complained about the thief. The police man asked Jeena’s location where she is now and to which direction the thief has gone. Jeena provides the police officer his desired information. The police officer started tracking the thief through CCTV cameras to get live streaming of the thief and also informed the police vans of the area where the thief is

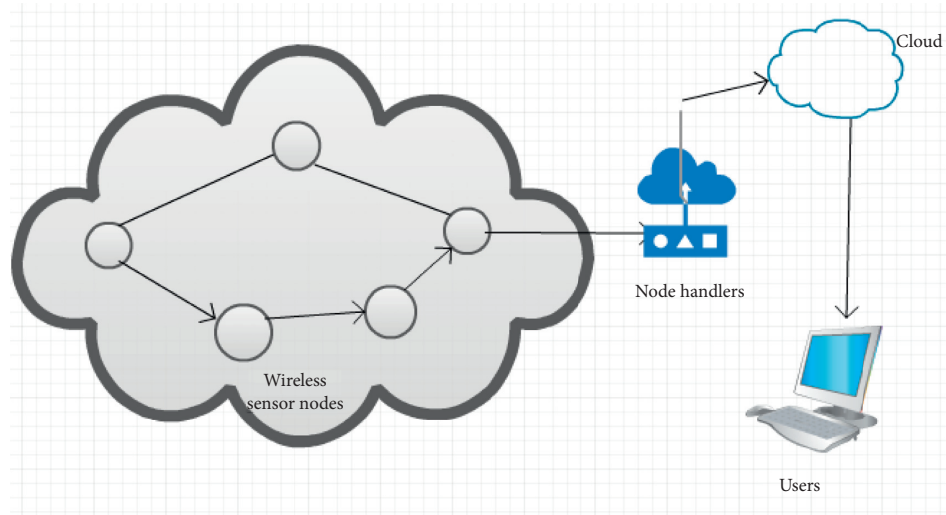


FIGURE 4: Flow of the LAFF in case when fog head fails to fulfil user's request.

TABLE 4: Data configuration.

Configuration	No. of fog nodes	No. of users
1	2	10
2	3	16
3	4	20
4	5	24
5	6	28
6	7	36
7	8	48
8	9	52
9	12	60
10	15	90

traced. The police vans caught the thief through accessing the exact location of the thief.

However, live streaming is a heavy task to run and requires a lot of computational power which requires a framework with low latency, service time, and network use to assure QoS. In this case, a nearest fog node will be assigned to the police vans so that they can trace the thief without any data loss and interruption.

## 6. Results and Discussion

The lightweight LAFF is compared with two other fog-based frameworks: IFAM (intelligent FC analytical model) and TPFC (task placement on fog computing) [7, 8]. The primary motivation behind this evaluation is to confirm the adequacy of the LAFF in terms of reducing latency, service time, and network use to facilitate users by providing better QoS. LAFF is a lightweight framework as it consumes less computational resources. RAM utilization and CPU consumption of a framework can increase the burden on resources. Since most of the fog nodes are not abundant in resources, execution of heavyweight software systems can cause significant computing overhead on them. Therefore, it is required to deploy lightweight frameworks in fog

computing environments. The framework that consumes less RAM and CPU consumption is considered lighter than the other frameworks [43]. Ten configurations are employed with varying numbers of devices and nodes so that consistent patterns could be extracted.

**6.1. Latency.** Security applications are very time-sensitive. Results cannot be delayed. For instance, if we come to know that a terrorist is going to blast a bomb somewhere, finding the terrorist's locations on time is a crucial and time-sensitive task. Delay cannot be afforded as it can lead to very negative consequences. This delay is calculated by implementing a control loop. Latency is calculated by using a module to module latency, and then average of them is taken; latency is much higher when IFAM and TPFC modules are executed as depicted in Figure 5. This comparison is done in the established scenario for the LAFF. The results depicted that the lightweight LAFF reduced the average latency by 11.01% when compared with that of both the frameworks. The agenda not only stops at reducing the latency but also reduces the network usage and service time in order to provide better QoS and consistent data.

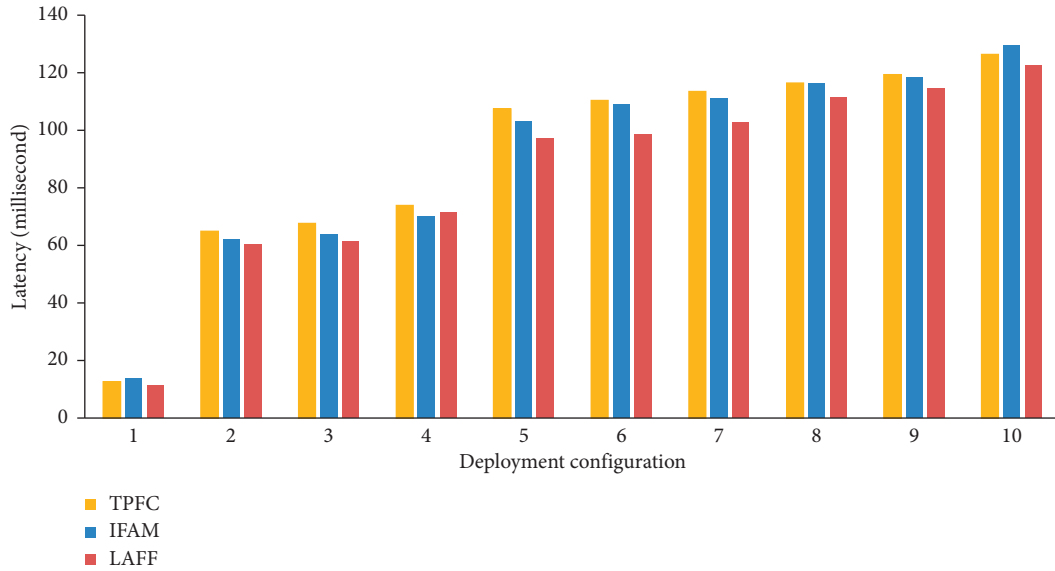


FIGURE 5: Latency comparison.

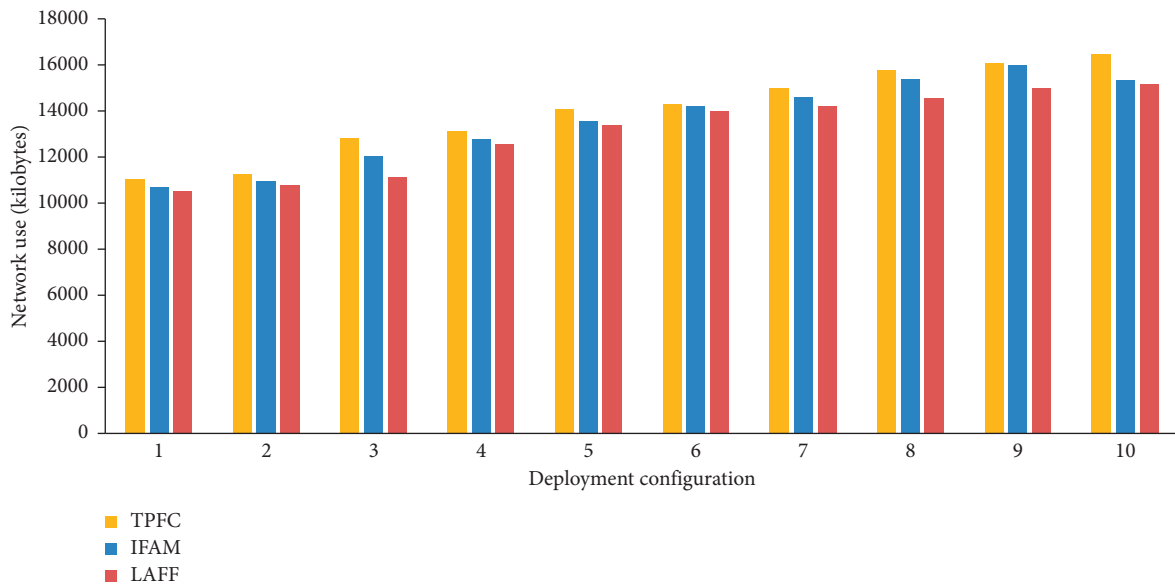


FIGURE 6: Network usage comparison.

6.2. *Network Usage.* This parameter is characterized by the utilization of system resources in terms of data sent and received from the network interfaces. The network usage should be kept at minimum for better performance. LAFF reduced the network traffic and consumption in terms of resource utilization. The results depict that the network utilization of the LAFF is reduced by average 7.51% as compared with that of the IFAM and TPFC as shown in Figure 6. The comparison of the LAFF with TPFC and IFAM showed that the LAFF provides better QoS.

6.3. *Service Time.* Service time is the most important parameter in sense of QoS. Service time is the amount of time

spent to provide services to a user by a service provider. The service providers are the small hosts integrated with fog nodes and cloud in order to use storage and transmissions. The service time comparison is shown in Figure 7. It shows that the average amount of time is 14.8% lesser than that of the TPFC and IFAM.

6.4. *RAM Consumption.* RAM is one of the most important components of the fog node. If the framework consumes more RAM, the RAM system will crash and become unresponsive. To prove that the proposed framework is lightweight, RAM consumption of the framework is compared with that of TPFC and IFAM. Figure 8 shows the RAM

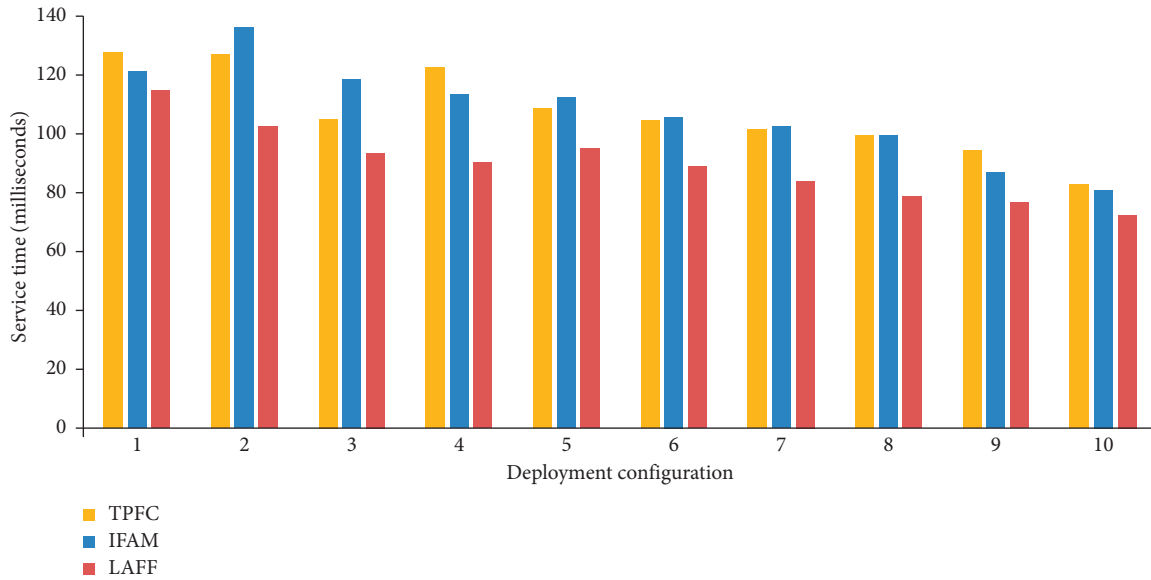


FIGURE 7: Service time comparison.

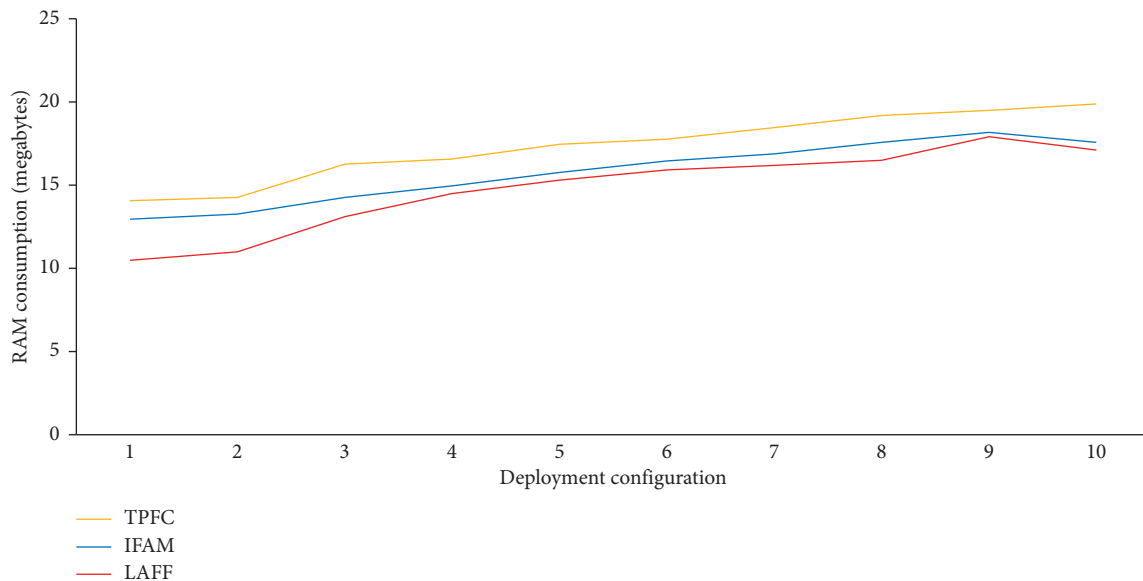


FIGURE 8: Ram consumption.

consumption for the data transmission and processing in fog nodes. The results showed that the proposed framework's RAM consumption is average 8.41% less than the both compared frameworks.

**6.5. CPU Utilization.** CPU utilization is the amount of work handled by a CPU of the fog node. The time taken between the start and the completion of a given task executed on a fog node is referred to as CPU utilization

and measured in milliseconds. In this study, we do not include the time taken for separating and combining tasks before and after their scheduling. A task is composed of a set of instructions. We assume that each instruction requires one clock cycle to be executed. In the proposed framework, offloading module helps to minimize the CPU utilization, therefore increasing the fog node performance. The results in Figure 9 show that the proposed framework's CPU utilization is average 16.23% less than the both compared frameworks.

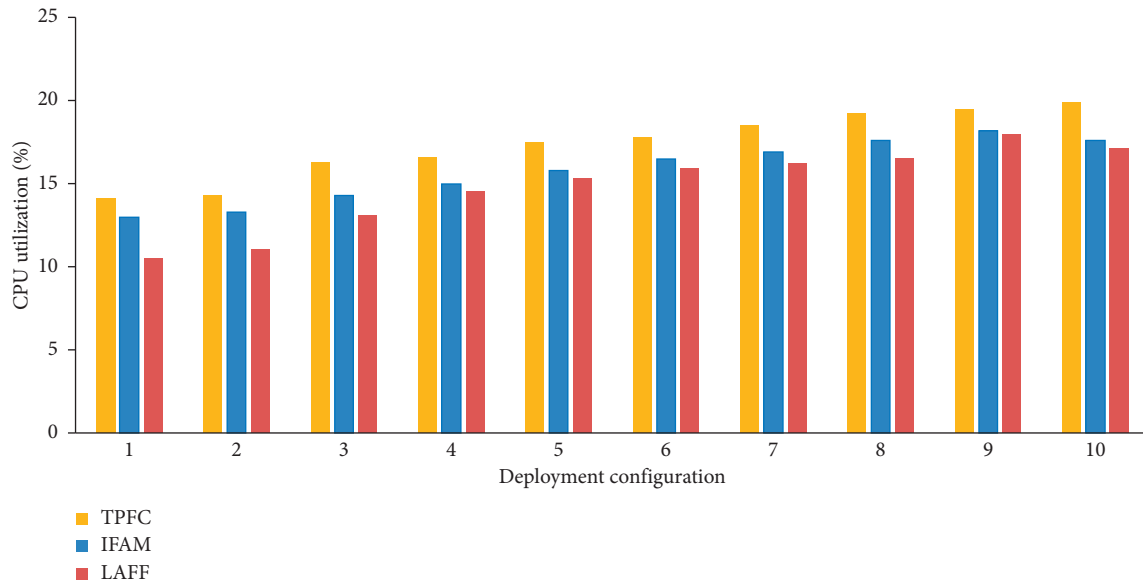


FIGURE 9: CPU utilization.

## 7. Conclusion and Future Work

The access to data and content is more smoother and faster when accelerated with location awareness. Responsiveness and consistency are increased if the latency is minimized and bottleneck issues are catered. LAFF is designed as a location-aware algorithm which ensures rich user experiences and provides better QoS by reducing the network utilization, service time, and latency. It is examined that the LAFF lessens the average latency, network use, and service time by 11.01%, 7.51%, and 14.8%, respectively, as compared with those of IFAM and TPFC. Similarly, resource utilization in terms of RAM and CPU is reduced by average 8.41% and 16.23% as compared with that of TPFC and IFAM making LAFF comparatively a lightweight framework. Location-aware feature is significant in defense and intelligence areas. Hence, the proposed LAFF improves QoS while accessing remote computational servers for the outsourced applications in fog computing. For future work, it is suggested that a module for predictive analysis must be integrated in cloud which will be able to predict the user's requests by analyzing the user's location and previous requests time. We also plan to develop optimization mechanisms such as in [44] to determine the optimal distribution and configuration of fog nodes while taking into consideration the computational resources with a backup plan to provide the backup in case of system failures such as in [45] through introducing learning methods.

### Data Availability

The data are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the research project of the Natural Science Foundation of China (NSFC) under grant no. 61671222.

## References

- [1] McKinsey & Company, *The Internet of Things: How to Capture the Value of IoT*, McKinsey & Company, New York, NY, USA, 2018, <https://www.mckinsey.com/featured-insights/internet-of-things/our-insights/the-internet-of-things-how-to-capture-the-value-of-iot>.
- [2] National Intelligence Council, *Disruptive Civil Technologies: Six Technologies with Potential Impacts on US Interests Out to 2025*, CreateSpace, Scotts Valley, CA, USA, 2008.
- [3] CISCO, *Fog Computing and the Internet of Things: Extend the Cloud to where the Things Are*, CISCO, San Jose, CA, USA, 2015.
- [4] A. H. Sodhro, Z. Luo, A. K. Sangaiah, and S. W. Baik, "Mobile edge computing based QoS optimization in medical healthcare applications," *International Journal of Information Management*, vol. 45, pp. 308–318, 2019.
- [5] S. Yan, M. Peng, M. A. Abana, and W. Wang, "An evolutionary game for user access mode selection in fog radio access networks," *IEEE Access*, vol. 5, pp. 2200–2210, 2017.
- [6] S. F. Hassan and R. Fareed, "Video streaming processing using fog computing," in *Proceedings of the 2018 International Conference on Advanced Science and Engineering (ICOASE)*, pp. 140–144, Duhok, Iraq, 2018.
- [7] S. Shukla, M. F. Hassan, M. K. Khan, L. T. Jung, and A. Awang, "An analytical model to minimize the latency in healthcare internet-of-things in fog computing environment," *PLoS One*, vol. 14, no. 11, Article ID e0224934, 2019.
- [8] M.-Q. Tran, D. T. Nguyen, V. A. Le, D. H. Nguyen, and T. V. Pham, "Task placement on fog computing made efficient for IoT application provision," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 6215454, 17 pages, 2019.

- [9] A. Munir, P. Kansakar, and S. U. Khan, "IFCIoT: integrated fog cloud IoT: a novel architectural paradigm for the future Internet of things," *IEEE Consumer Electronics Magazine*, vol. 6, no. 3, pp. 74–82, 2017.
- [10] G. R. kumar, N. Saikiran, N. Saikiran, and A. Sathish, "FOG: a novel approach for adapting IoT/ToE in cloud environment," *International Journal of Engineering Trends and Technology*, vol. 42, no. 4, pp. 189–192, 2016.
- [11] V. Mihai, C. Dragana, Grigore Stamatescu, and D. Popescu, "Loretta Ichim wireless sensor network architecture based on fog computing," in *Proceedings of the 2018 5th International Conference on Control, Decision and Information Technologies, (CoDIT'18)*, Thessaloniki, Greece, April 2018.
- [12] M. Malensek, S. L. Pallickara, and S. Pallickara, "HERMES: federating fog and cloud domains to support query evaluations in continuous sensing environments," *IEEE Cloud Computing*, vol. 4, no. 2, pp. 54–62, 2017.
- [13] M. Al-khafajiy, T. Baker, H. Al-Libawy, Z. Maamar, M. Aloqaily, and Y. Jararweh, "Improving fog computing performance via fog-2-fog collaboration," *Future Generation Computer Systems*, vol. 100, pp. 266–280, 2019.
- [14] S. Agarwal, S. Yadav, and A. K. Yadav, "An efficient architecture and algorithm for resource provisioning in fog computing," *International Journal of Information Engineering and Electronic Business*, vol. 8, no. 1, pp. 48–61, 2016.
- [15] Q. Fan and N. Ansari, "Towards workload balancing in fog computing empowered IoT," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 253–262, 2018.
- [16] L. Gu, D. Zeng, S. Guo, A. Barnawi, and Y. Xiang, "Cost efficient resource management in fog computing supported medical cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 1, pp. 108–119, 2017.
- [17] W. Tian, J. Zeng, Y. Lai et al., "Data collection from WSNs to the cloud based on mobile fog elements," *Future Generation Computer Systems*, vol. 105, pp. 864–872, 2017.
- [18] I. M. Al-Joboury and E. H. Al-Hemiary, "IoT-F2CDM-LB: IoT based fog- to-cloud and data-in-motion architectures with load balancing," *EAI Endorsed Transactions on Internet of Things*, vol. 4, no. 13, 2018.
- [19] C. Li, H. Zhuang, Q. Wang, and X. Zhou, "SSLB: self-similarity-based load balancing for large-scale fog computing," *Arabian Journal for Science and Engineering*, vol. 43, no. 12, pp. 7487–7498, 2018.
- [20] D. Puthal, M. S. Obaidat, P. Nanda, M. Prasad, S. P. Mohanty, and A. Y. Zomaya, "Secure and sustainable load balancing of edge data centers in fog computing," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 60–65, 2018.
- [21] A. V. Dastjerdi, H. Gupta, R. N. Calheiros, S. K. Ghosh, and R. Buyya, "Fog computing: principles, architectures, and applications," in *Internet of Things*, pp. 61–75, Elsevier, Amsterdam, Netherlands, 2016.
- [22] A. Yousefpour, A. Patil, G. Ishigaki et al., "FOGPLAN: a lightweight QoS-aware dynamic fog service provisioning framework," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5080–5096, 2019.
- [23] Y. Lin and H. Shen, "CloudFog: leveraging fog to extend cloud gaming for thin-client MMOG with high quality of service," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 2, pp. 431–445, 2017.
- [24] G. Myrizzakis and E. G. M. Petrakis, "iHome: smart home management as a service in the cloud and the fog," in *Advanced Information Networking and Applications*, pp. 1181–1192, Springer International Publishing, Berlin, Germany, 2020.
- [25] S. Prabhdeep and K. Rajbir, "Design and develop quality of service framework using fog computing for smart city applications," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 1S, 2019.
- [26] Skarlat, O. Nardelli, M. Schulte, and S. Dustdar, "Towards QoS-aware fog service placement," in *Proceedings of the 1st International Conference on Fog and Edge Computing (ICFEC)*, Madrid, Spain, May 2017.
- [27] D. Zeng, L. Gu, S. Guo, Z. Cheng, and S. Yu, "Joint optimization of task scheduling and image placement in fog computing supported software-defined embedded system," *IEEE Transactions on Computers*, vol. 65, no. 12, pp. 3702–3712, 2016.
- [28] W.-S. Kim and S.-H. Chung, "User incentive model and its optimization scheme in user-participatory fog computing environment," *Computer Networks*, vol. 145, pp. 76–88, 2018.
- [29] M. K. Tefera, X. Yang, and Q. T. Sun, "A survey of system architectures, privacy preservation, and main research challenges on location-based services," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 6, pp. 3199–3218, 2019.
- [30] J. Zhu, D. S. Chan, M. S. Prabhu, P. Natarajan, H. Hu, and F. Bonomi, "Improving web sites performance using edge servers in fog computing architecture," in *Proceedings of the 2013 IEEE 7th International Symposium on Service Oriented System Engineering (SOSE)*, pp. 320–323, San Francisco, CA, USA, March 2013.
- [31] H. Li, G. Shou, Y. Hu, and Z. Guo, "Mobile edge computing: progress and challenges," in *Proceedings of the 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, pp. 83–84, Oxford, UK, March 2016.
- [32] Y. Jiang, Z. Huang, and D. H. K. Tsang, "Challenges and solutions in fog computing orchestration," *IEEE Network*, vol. 32, no. 3, pp. 122–129, 2017.
- [33] M. Haghi Kashani, A. M. Rahmani, and N. Jafari Navimipour, "Quality of service-aware approaches in fog computing," *International Journal of Communication Systems*, vol. 33, p. e4340, 2020.
- [34] B. K. Dar, M. A. Shah, S. U. Islam, C. Maple, S. Mussadiq, and S. Khan, "Delay-aware accident detection and response system using fog computing," *IEEE Access*, vol. 7, pp. 70975–70985, 2019.
- [35] E. M. Tordera, "What is a fog node a tutorial on current concepts towards a common definition," 2016, <https://arxiv.org/abs/1611.09193>.
- [36] A. Husain and L. Stefan, "K\*: a heuristic search algorithm for finding the k shortest paths," *Artificial Intelligence*, vol. 175, no. 18, pp. 2129–2154, 2011.
- [37] V. Santhanam and D. B. Shanmugam, "Integrating wireless sensor networks with cloud computing and emerging IT platforms using middleware services," *International Research Journal of Engineering and Technology*, vol. 5, no. 8, 2018.
- [38] Z. Ning, J. Huang, and X. Wang, "Vehicular fog computing: enabling real-time traffic management for smart cities," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 87–93, 2019.
- [39] M. Mishra, S. K. Roy, and A. Mukherjee, "An energy-aware multi-sensor geo-fog paradigm for mission critical applications," *Journal of Ambient Intelligence and Humanized Computing*, vol. 41, 2019.
- [40] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of



- resource provisioning algorithms,” *Software: Practice and Experience*, vol. 41, no. 1, pp. 23–50, 2011.
- [41] H. Gupta, A. Vahid Dastjerdi, S. K. Ghosh, and R. Buyya, “iFogSim: a toolkit for modeling and simulation of resource management techniques in the Internet of things, edge and fog computing environments,” *Software: Practice and Experience*, vol. 47, no. 9, pp. 1275–1296, 2017.
- [42] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: vision and challenges,” *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.
- [43] S. Tuli, R. Mahmud, S. Tuli, and R. Buyya, “FogBus: a blockchain-based lightweight framework for edge and fog computing,” *Journal of Systems and Software*, vol. 154, pp. 22–36, 2019.
- [44] M. Taneja and A. Davy, “Resource aware placement of IoT application modules in fog-cloud computing paradigm,” in *Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, IEEE, Lisbon, Portugal, pp. 1222–1228, 2017.
- [45] M. K. Saroa and R. Aron, “Fog computing and its role in development of smart applications,” in *Proceedings of the 2018 IEEE International Conference on Parallel & Distributed Processing with Applications*, Melbourne, Australia, December 2018.

## Research Article

# Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification

Ashir Javeed,<sup>1</sup> Sanam Shahla Rizvi ,<sup>2</sup> Shijie Zhou,<sup>1</sup> Rabia Riaz,<sup>3</sup> Shafqat Ullah Khan,<sup>4</sup> and Se Jin Kwon <sup>5</sup>

<sup>1</sup>School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China

<sup>2</sup>Raptor Interactive (Pty) Ltd., Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa

<sup>3</sup>Department of CS&IT, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

<sup>4</sup>Department of Electronics, University of Buner, Buner, Pakistan

<sup>5</sup>Department of Computer Engineering, Kangwon National University, Samcheok 25806, Republic of Korea

Correspondence should be addressed to Se Jin Kwon; [sjkwon@kangwon.ac.kr](mailto:sjkwon@kangwon.ac.kr)

Received 16 March 2020; Revised 4 May 2020; Accepted 29 July 2020; Published 26 August 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Ashir Javeed 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.

Diagnosis of heart disease is a difficult job, and researchers have designed various intelligent diagnostic systems for improved heart disease diagnosis. However, low heart disease prediction accuracy is still a problem in these systems. For better heart risk prediction accuracy, we propose a feature selection method that uses a floating window with adaptive size for feature elimination (FWAFE). After the feature elimination, two kinds of classification frameworks are utilized, i.e., artificial neural network (ANN) and deep neural network (DNN). Thus, two types of hybrid diagnostic systems are proposed in this paper, i.e., FWAFE-ANN and FWAFE-DNN. Experiments are performed to assess the effectiveness of the proposed methods on a dataset collected from Cleveland online heart disease database. The strength of the proposed methods is appraised against accuracy, sensitivity, specificity, Matthews correlation coefficient (MCC), and receiver operating characteristics (ROC) curve. Experimental outcomes confirm that the proposed models outperformed eighteen other proposed methods in the past, which attained accuracies in the range of 50.00–91.83%. Moreover, the performance of the proposed models is impressive as compared with that of the other state-of-the-art machine learning techniques for heart disease diagnosis. Furthermore, the proposed systems can help the physicians to make accurate decisions while diagnosing heart disease.

## 1. Introduction

The heart is a vital organ in the human body that is liable for blood circulation. The heart is responsible for oxygen and energy supply to all organs of the body including itself. Heart disease causes the abnormal blood circulation in the body that might be fatal for human life. Hence, if the heart stops its normal functionality, the whole system will be dead. From the literature, various risk factors are identified that cause the heart disease. The risk factors of heart diseases are classified into two major types such as the risk factors that can alter,

e.g., smoking and physical exercise, and the risk factors that cannot alter, e.g., gender, age, and patient's family history [1]. The diagnosis of heart through conventional medical methods is quite difficult, complex, time consuming, and costly. Therefore, the diagnosis of heart disease is worst in developing countries due to lack of state-of-the-art examination tools and medical experts [2, 3]. Additionally, the invasive medical procedure for examination of heart failure is formed on various tests suggested by physicians, after studying the medical history of the patient and analyzing the relevant symptoms [4]. Angiography is considered as the

gold standard among the medical tests for diagnosis of heart failure. Heart disease cases are affirmed through angiography as it is the best practice for diagnosis of heart disease. Moreover, angiography has side effects as well as higher cost for diagnosis of heart disease and demands extraordinary technical expertise [5, 6]. Therefore, machine learning and data mining techniques are needed to design the expert systems for resolving the problems of angiography.

To address the abovementioned problems, researchers have designed different noninvasive diagnosis systems by exploiting machine learning based predictive models. These models include logistic regression, naive Bayes,  $k$ -nearest neighbor (KNN), decision tree, support vector machine (SVM), artificial neural network (ANN), and ensembles of ANN for heart failure disease classification [1, 7–18]. Robert Detrano utilized logistic regression for heart failure risk prediction and attained classification accuracy of 77%. Newton Cheung utilized various predictive models consisting C4.5, naive Bayes, BNND, and BNNF algorithms. The accuracies of proposed algorithms were 81.11%, 81.48%, 81.11%, and 80.95%, respectively, for precise classification of patients and healthy subjects. A. Khemphila and V. Boonijjing proposed a classification technique based on multilayer perceptron (MLP) in addition to back-propagation learning algorithm and biomedical test values for diagnosing the heart disease through a feature selection algorithm. Information gain is utilized to filter features through elimination of the features which do not contribute for precise results. Total number of thirteen features is reduced to eight by using a feature selection algorithm. For the classification, ANN is used as a classifier. The accuracy of training dataset was 89.56%, while for data validation, the accuracy of 80.99% was reported.

Recently, Paul et al. proposed a fuzzy decision support system (FDSS) in order to detect the heart disease [19]. They proposed a genetic algorithm based on FDSS that has five key components such as preprocessing of the dataset, effective features selection through diverse methods, weighted fuzzy rules that are set up through genetic algorithm, generated fuzzy knowledge used to build FDSS, and heart disease prediction. The proposed system obtained the accuracy of 80%. Verma et al. proposed a hybrid model for coronary artery disease (CAD) diagnosis [20]. The proposed method consists of jeopardizing factor identifiers adopting a correlation based subset (CFS) selection with particle swarm optimization (PSO) search model and  $K$ -means. Supervised learning algorithms such as multilayer perceptron (MLP), multinomial logistic regression (MLR), fuzzy unordered rule induction algorithm (FURIA), and C4.5 are then utilized to design CAD cases. The accuracy of the proposed approach was 88.4%. The proposed model enhanced the efficiency of classification techniques from 8.3% to 11.4% of Cleveland dataset. Shah et al. proposed a technique based on the feature extraction for reducing feature dimensions [21]. The proposed approach used probabilistic principal component analysis (PPCA). Projection dimensions are extracted through PPCA that compliments high covariance and also helps to eliminate feature dimension. Parallel analysis (PA) helps in the selection of projection vectors. The feature

subset of reduce feature vector is input to the radial basis function (RBF) kernel-based support vector machines (SVMs). Two types of classification are categories into heart patient (HP) and normal subject (NS) through RBF-based SVM serves. The proposed model is tested against accuracy, specificity, and sensitivity on the datasets of UCI, i.e., Cleveland. The accuracy of the proposed model for Cleveland dataset was 82.18%, 85.82%, and 91.30%, respectively.

Most recently, Dwivedi tests the performance of different machine learning methods for the prediction of heart disease. The highest classification accuracy of 85% was reported based on logistic regression [22]. Amin et al. evaluate the different data mining methods and identify the significant features for predicting heart disease [23]. Predictive models were built from different combinations of features and well-known classification methods, e.g., LR, SVM, and  $K$ -NN. From experimental results, it was studied that the best performance of the data mining technique for classification accuracy was 87.4% for the heart disease prediction. Özşen and Güneş proposed an expert system developed from an artificial immune system (AIS) and achieved accuracy of 87% [24]. An expert system was proposed by Özşen and Güneş based on the artificial immune system (AIS). The accuracy of 87% was reported for the developed expert system. Polat et al. developed another similar system and obtained 84.5% accuracy [25]. Das et al. utilized a neural network ensemble model with the purpose of improving classification accuracy. His ensemble model obtained the classification accuracy of 89.01% [1]. Recently, Samuel et al. proposed a diagnostic system developed from ANN and Fuzzy AHP. The prediction accuracy of 91.10% was reported from the ANN and Fuzzy AHP diagnosis system [4].

As clear from the literature survey, ANN-based diagnostic systems have shown better performance on the heart disease data. Hence, we also attempt to design a diagnostic system based on neural network for heart disease detection. The development of various noninvasive diagnostic systems for heart disease detection motivates us to design an expert diagnostic system based on neural networks. From the empirical result, it is analyzed that proposed model shows promising performance. Hence, it can be used in clinics to make accurate decisions while diagnosing heart failure.

## 2. Materials and Methods

In previous studies, researchers used feature sets without eliminating irrelevant or noisy features. In this study, we propose a novel feature elimination method for removing noisy or irrelevant feature vectors and thus selected an optimal subset of feature vectors before feeding them to ANN or DNN. The proposed algorithm uses a window with adaptive size. The window size is initialized from one and is placed at the first feature of the feature vector. The feature or features to which the window points are eliminated while the remaining features constitute the subset of features that are supplied to the neural network for classification. To find the optimal configuration of the neural network for the subset of features, grid search algorithm is used. It is noteworthy that the previous studies utilized conventional ANN with only

one hidden layer for heart failure detection problem. However, in this study, we found out that deep neural networks with more than one hidden layers and trained with new learning algorithms show better performance. Additionally, this study evaluates the feasibility of features selection algorithm at the input level of DNN. The working of the proposed diagnostic system is clearly shown in Algorithm 1 and Figure 1.

*2.1. Dataset Description.* For this research, an online repository of machine learning and data mining from University of California, Irvine (UCI), for the heart disease dataset was used that is known as a Cleveland heart disease database. Data were gathered from the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation by Dr. Robert Detrano [26]. The dataset is comprised of 303 subjects. Furthermore, the number of subjects having missing values in the dataset is 6. In the dataset, 297 subjects have complete data values out of 303 subjects. Hence, the number of subjects that have complete data values is used for experiments. Moreover, each subject in dataset has 76 raw features. In the previous work, the researchers mostly used 13 prominent features out of 76 raw features of each subject for the diagnosis of heart disease. Therefore, mostly used 13 features for diagnosis of heart failure is considered for this study. Table 1 depicts the most commonly used 13 features of heart disease.

*2.2. The Proposed Method.* The proposed diagnostic system has two main components that are hybridized as one black-box model. The main reason for hybridizing the two components into one block is that they work in connection with each other. The first component of the system is a feature selection module, while the second component is a predictive model. Feature selection methods use data mining concepts to improve the performance of the machine learning models [27, 28]. The feature selection module uses a search strategy to find out the optimal subset of features which are applied to the DNN that acts as a predictive model. The feature selection module uses a window that scans the feature vector. The working of the proposed method can be depicted from the algorithm.

Initially, the size of the window is set to 1. And, the window is placed at the left most side of the feature vector with size  $n$ , i.e., having  $n$  number of features. Hence, initially, the feature is eliminated from the feature vector on that the window is placed and the remaining features constitute the subset of features which are supplied to the DNN for classification. The performance of the subset of features is saved. In the next step, the window floats towards the right direction. Again, that feature is eliminated on which the window is placed and the remaining features constitute feature subset whose performance is checked by the DNN model. The same process is repeated until the window reaches the last feature, i.e., the  $n^{\text{th}}$  feature. With this, the first round of window floating is completed. It is important to note that the features subset size is  $n - 1$  in the first round.

In the next round, the window size is updated to 2. Hence, in this round, the window points towards the two features at a

time. Again, the window starts the floating process from the left most side of the feature vector and eliminates the first two features. The remaining  $n - 2$  features constitute the features subset that is applied as the input to the DNN model for classification, and the results are compared with the best performance achieved on the previous subset of features. If the performance is better than the previous best performance, the best performance and optimal subset of features is updated. In the next iteration, the window floats towards the right direction and those two features are eliminated on which the window is placed. The remaining features constitute the subset of features which are applied to DNN. The same process is repeated until the window reaches to the right most side of the feature vector. This marks the end of the second round. In the third round, the window size is made 3 and the same process is repeated that was carried out for the first two rounds. Finally, at the  $n - 1$  round, the window size is made  $n - 1$ . In this round, the window can float just once towards the right. And then, the whole process is ended. Finally, the subset of features that give us the best results is declared as the optimal subset of features. The whole process of features selection through adaptive floating window is clearly illustrated in Figure 2. Each time a subset of features is supplied to the DNN, the DNN architecture is optimized using the grid search algorithm. The performance of a DNN is highly dependent on its architecture [29]. Inappropriate DNN architecture will result in poor performance although there are chances that the DNN is applied with an optimal subset of features. The main reason for such a poor performance is that, if the DNN architecture selected for the classification is with insufficient capacity, then it will result in underfitting [30, 31]. In such a case, the DNN will show poor performance on both data, i.e., training data and testing data. However, if the DNN architecture has excessive capacity, it will overfit to the training data; thus, it will show better performance on the training data but poor performance on the testing data. Hence, we need to search optimal architecture of DNN that will show good performance on both testing and training data. To understand the relationship between DNN architecture and the capacity of DNN, we need to understand the formulation of DNN. The neural network is formulated as follows:

Neural networks are generated by the computational system based on mathematical models that simulate the human brain. The key element in the neural network model is known as perceptron or a node [32]. Nodes are shaped into groups which are called layers. Artificial neurons work on the same principal which is followed by the biological neuron. As an artificial neuron receives one or more inputs from the adjoined neurons, it then processes the information and transfers the output to the next perceptron. Artificial neurons are connected through a link that is known as weights. The input information  $\chi_i$  is weighted either positive or negative during the computation of output. An internal threshold value  $\rho$  and weights are assigned for the solution of a problem under consideration. On every node, the result is calculated by multiplying the input values  $\chi_n$  and associated weight  $w_n$  that is fine-tuned by the threshold value  $\rho$ . The output is then calculated through an activation function or transfer function ( $\alpha$ ) and is given in the following equation:

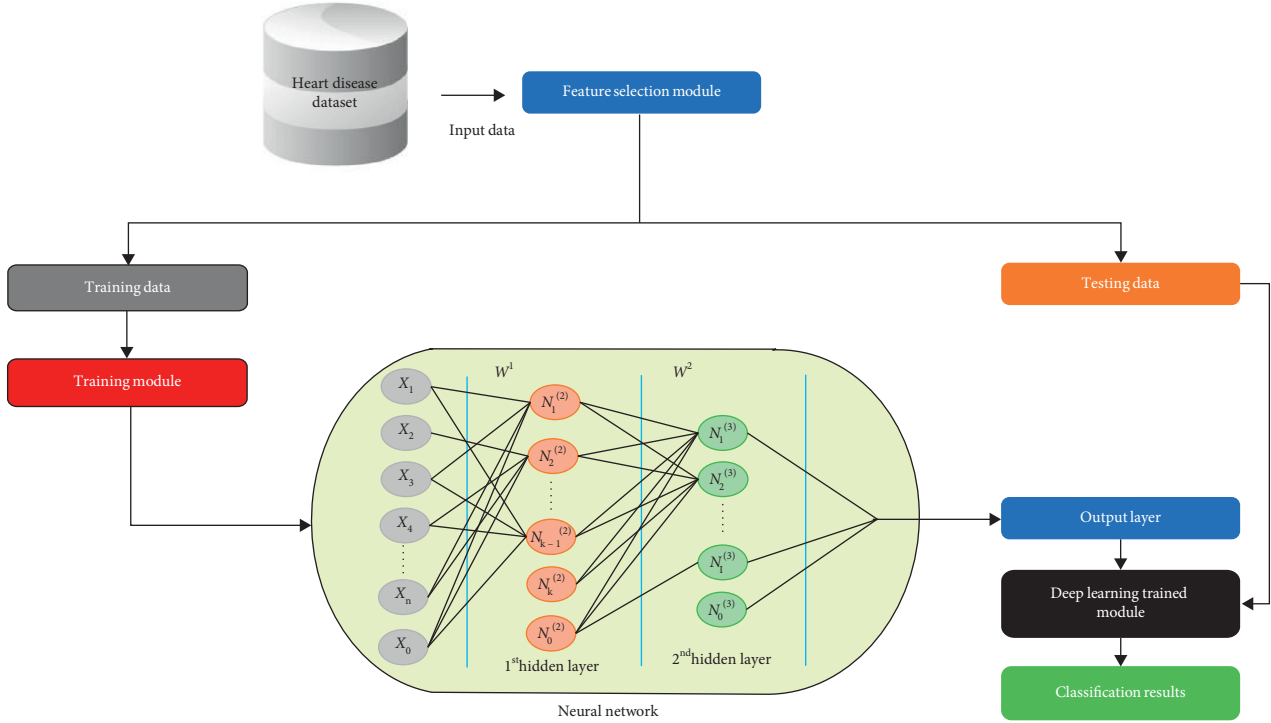


FIGURE 1: Block diagram of the newly proposed method.

**Input:**  $\{ N_F: \text{Features size}, \beta: \text{Hyperparameters} \}$   
**Output:**  $\{ F_s: \text{Optimal Subset of features}, \beta_o: \text{Optimal} \}$

- (1)  $W; 1, 2, 3, \dots, F$ , where  $W$  is the window size
- (2)  $\text{Best\_Acc} = 0$
- (3) Initialize
- (4) **for**  $1 = W_{\max}$
- (5)   **for**  $1 = \beta_{\max}$
- (6)     Acc using  $F_s$  and  $\beta$
- (7)     **if**  $(\text{Acc} > \text{Best\_Acc})$
- Begin if**
- $\text{Best\_Acc} = \text{Acc}$
- Note down  $F_s$  and  $\beta$  as  $\beta_o$
- End if**
- (9)     **END for**
- (10) **END for**
- (11) Display  $F_s, \beta_o$  and  $\text{Best\_Acc}$

ALGORITHM 1: FWAFE-DNN.

$$\delta_i = \alpha \left( \sum \omega_n \cdot \chi_n - \varrho \right). \quad (1)$$

The transfer can be linear or nonlinear. In the case of nonlinear function tangent, hyperbolic or radial basis form is applied. The sigmoid function,  $\alpha(\delta_i)$ , is done at the following layer as an output value (equation (2)).  $\gamma$  is related to the shape of the sigmoid function. The increase in parameter  $\gamma$  value strengthened the nonlinearity of the sigmoid function:

$$\alpha(\delta_i) = \frac{1}{[1 + \exp(-\gamma\delta_i)]}. \quad (2)$$

The neural network is obtained by connecting the artificial neurons. If the constructed neural network model has

only one hidden layer, we name it ANN [17]. However, if the constructed neural network model has more than one hidden layer, we name it DNN [17].

### 3. Validation Scheme and Evaluation Metrics

**3.1. Validation Scheme.** In earlier works, the performance of the expert diagnosis systems has been evaluated through holdout validation schemes. The dataset has to be partitioned into two parts: one is for training purpose, while another is used for testing. In the past, researchers have been using various train-test split percentages of data partitioning. Furthermore, Das et al. in [1] and Paul et al. in [33] used

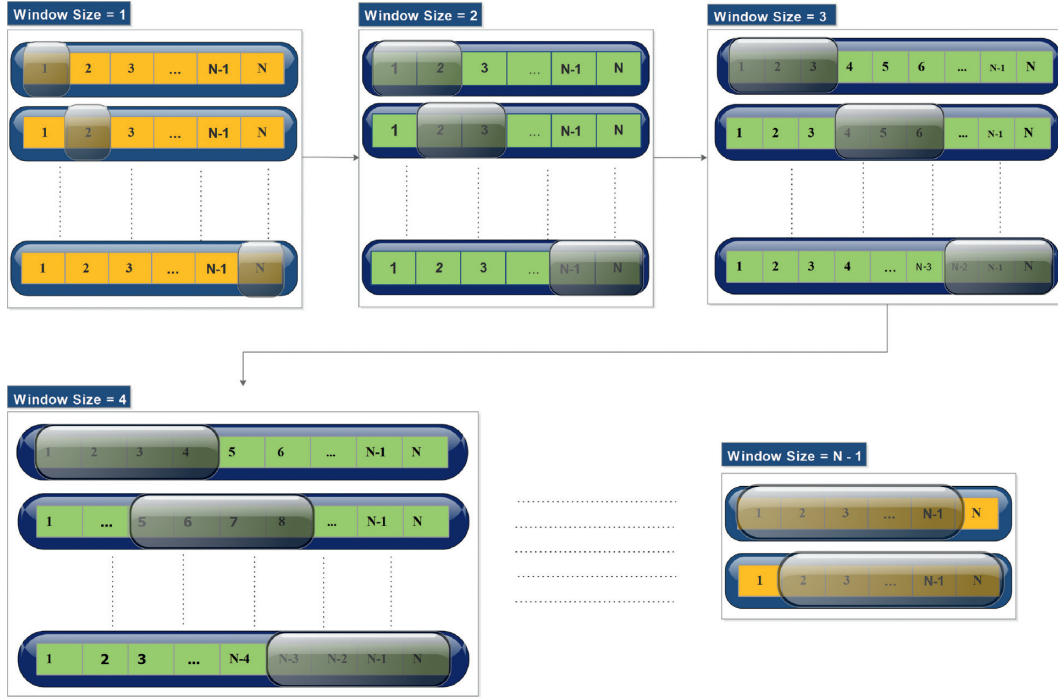


FIGURE 2: Feature selection by eliminating features using floating window with adaptive size.

holdout validation schemes in their research. They have partitioned the dataset into 70%–30% ratio, where 70% of the data set is utilized for the training purpose of the predictive model while 30% of the dataset is utilized for testing the performance of the predictive model. Therefore, we also utilized the same criteria of data partition for train-test purpose.

**3.2. Evaluation Metrics.** Various evaluation metrics such as specificity, sensitivity, accuracy, and Matthews correlation coefficient (MCC) are utilized for evaluating the performance and efficiency of the proposed model. The percentage of the precisely classified subjects is known as accuracy. Sensitivity is the accurate classification of the patients, whereas specificity is the absolute classification of healthy subjects. All the evaluation metrics are formulated in equations (3)–(6):

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

where TP stands for true positives, TN describes true negatives, FP shows false positives, and FN stands for false negatives.

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad (4)$$

$$\text{specificity} = \frac{TN}{TN + FP}, \quad (5)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

The characteristic of binary classification is assessed using MCC for machine learning and statistics. The value of MCC is ranging between  $-1$  and  $1$ . The  $-1$  value of MCC denotes the total conflict between prediction and observation, whereas  $1$  shows the exact prediction, while  $0$  describes the classification as random prediction. Moreover, in this study, another evaluation metric, namely, the receiver operating characteristic (ROC) curve was also exploited. The ROC curve is a well-known metrics that is used to statistically evaluate the quality of a predictive model. The ROC curve provides area under the curve (AUC); thus, a model is considered a better model if its AUC is high.

## 4. Experimental Results and Discussion

In this session, two kinds of diagnostic systems are proposed. Moreover, experiments are done to test the performance of the proposed diagnostic system. In the first experiment, FWAFE-ANN is developed and stimulated, while in the second experiment, FWAFE-DNN is utilized. In the first experiment, the FWAFE algorithm is used to construct a subset of features. Furthermore, a subset of features is applied to ANN that is used as a predictive model. In the second experiment, FWAFE is used to construct a subset of features, whereas DNN is utilized for classification. All the experiments were simulated by using Python programming software package.

**4.1. Experiment No. 1: Feature Selection by FWAFE and Classification by ANN.** In this experiment, at the first stage, FWAFE is used, while in the second stage, ANN is used. The feature selection module eliminates noisy and irrelevant features by exploiting a search strategy, whereas the second model is deployed as a predictive model. The proposed

TABLE 1: Types of features of the dataset.

Feature no.	Feature description	Feature code	(Mean $\pm$ std) <sub>Healthy</sub>	(Mean $\pm$ std) <sub>Patients</sub>
1	Age (AGE)	FC <sub>1</sub>	52.64 $\pm$ 9.52	56.84 $\pm$ 7.42
2	Sex (SEX)	FC <sub>2</sub>	0.55 $\pm$ 0.49	0.81 $\pm$ 0.385
3	Chest pain type (CPT)	FC <sub>3</sub>	2.79 $\pm$ 0.92	3.60 $\pm$ 0.79
4	Resting blood pressure (RBP)	FC <sub>4</sub>	129.17 $\pm$ 16.32	134.85 $\pm$ 18.69
5	Serum cholesterol (SCH)	FC <sub>5</sub>	243.49 $\pm$ 53.58	250.73 $\pm$ 49.83
6	Fasting blood sugar (FBS)	FC <sub>6</sub>	0.14 $\pm$ 0.35	0.15 $\pm$ 0.35
7	Resting electrocardiographic results (RES)	FC <sub>7</sub>	0.84 $\pm$ 0.98	1.14 $\pm$ 0.97
8	Maximum heart rate achieved (MHR)	FC <sub>8</sub>	158.59 $\pm$ 18.98	138.89 $\pm$ 22.74
9	Exercise induced angina (EIA)	FC <sub>9</sub>	0.14 $\pm$ 0.35	0.54 $\pm$ 0.49
10	Old peak (OPK)	FC <sub>10</sub>	0.59 $\pm$ 0.78	1.64 $\pm$ 1.29
11	Peak exercise slope (PES)	FC <sub>11</sub>	1.41 $\pm$ 0.59	1.83 $\pm$ 0.56
12	Number of major vessels colored by fluoroscopy (VCA)	FC <sub>12</sub>	0.27 $\pm$ 0.63	1.13 $\pm$ 1.01
13	Thallium scan (THA)	FC <sub>13</sub>	3.78 $\pm$ 1.55	5.90 $\pm$ 1.70

diagnostic system achieves accuracy of 91.11% using only a subset of features. The optimal subset of features is obtained for  $n = 6$ ,  $n = 7$ , and  $n = 11$  where  $n$  stands for the size of the feature subset. The simulation results are reported in Table 2. In the table, the last record displays a case where all the features are used, i.e., no feature selection is performed. It can be noticed that the best accuracy of 90% is achieved after optimizing the architecture of ANN by a grid search algorithm using all the features. Thus, it is evidently coherent that the proposed model is competent as it presents us better performance with the least number of features. The best performance of the proposed model is observed at 11 features for the peak training accuracy. Additionally, the feature selection module increases the performance of the optimized ANN by 1.11%. Moreover,  $F_e$  denotes the features that are eliminated from the features, space during the feature selection process. The results from distinct subsets of features and diverse hyperparameters are displayed in Table 2.

**4.2. Experiment No. 2: Feature Selection by FWAFE and Classification by DNN.** In this experiment, at the first stage, FWAFE is used, while at the second stage, DNN is implied. The feature selection module eliminates noisy and irrelevant features by exploiting a search strategy, whereas the second model is utilized as a predictive model. The proposed diagnostic system achieves an accuracy of 93.33% using only a subset of features. The optimal subset of features is obtained for  $n = 11$  which includes FC<sub>1</sub>, FC<sub>2</sub>, FC<sub>3</sub>, FC<sub>4</sub>, FC<sub>7</sub>, FC<sub>8</sub>, FC<sub>9</sub>, FC<sub>10</sub>, FC<sub>11</sub>, FC<sub>12</sub>, and FC<sub>13</sub>, i.e., by eliminating feature number 5 and 6. The experimental outcomes are displayed in Table 3. To validate the effectiveness of the proposed feature selection method, i.e., FWAFE, the experiment is performed using the DNN model on full features without using the feature selection module. The DNN architecture was optimized using grid search algorithm. The best accuracy of 90% was obtained using neural network with four layers. The size of 1<sup>st</sup> layer is equivalent to the number of features, 2<sup>nd</sup> layer consists 50 neurons, and 3<sup>th</sup> layer contains 2 neurons and output layer has only one neuron. In Table 3, the last row represents a case, whereas all features are utilized. Hence, it is evidently clear that the feature selection module boots the performance of DNN by 3.33%. Moreover, FWAFE-DNN

TABLE 2: Results of different subsets of features for the heart disease dataset.

$n$	$F_e$	$N$	Acc <sub>train</sub>	Acc <sub>test</sub> (%)	Sens. (%)	Spec. (%)	MCC
12	4	3	83.09	90.00	85.36	93.87	0.799
12	11	8	87.92	90.00	85.36	93.87	0.799
11	(5, 6)	4	82.60	91.11	85.36	95.91	0.822
11	(11, 12)	4	81.15	90.00	87.80	91.83	0.798
10	(4 to 6)	7	85.99	90.00	82.92	95.91	0.801
7	(3 to 8)	49	95.16	91.11	87.80	93.87	0.820
6	(6 to 12)	6	81.15	91.11	95.12	87.75	0.825
13	—	1	85.02	90.00	87.80	91.83	0.798

shows better performance than FWAFE-ANN. The results at distinct subsets of features on various hyperparameters are shown in Table 3. The ROC charts are utilized to analyze the performance of the proposed model. A method whose ROC chart has maximum area beneath the curve is considered the best. The ROC chart whose points are in the upper left corner is considered to be the best. Figure 3(a) shows the ROC chart of the proposed FWAFE-ANN diagnostic system, while Figure 3(b) denotes the ROC chart of the ANN-based diagnostic system. From the figure, it is evidently vivid that the feature selection module increases the performance of the ANN model owing to more area beneath the curve. Similarly, Figure 4(a) represents the ROC chart of the proposed FWAFE-DNN diagnostic system, while Figure 4(b) depicts the ROC chart of DNN-based diagnostic system. From the figure, it is clearly observed that the feature selection module also increases the performance of the DNN model.

**4.3. Experiment No. 3: Results of Other State-of-the-Art Machine Learning Models.** In this segment, a comparative analysis is done with other state-of-the-art machine learning models on biomedical datasets against our proposed model. The classifier selected for comparison are random forest (RF) classifier, randomized decision tree classifier, Adaboost ensemble classifier, SVM with radial basis function (RBF) kernel, and linear support vector machine (SVM). Table 4

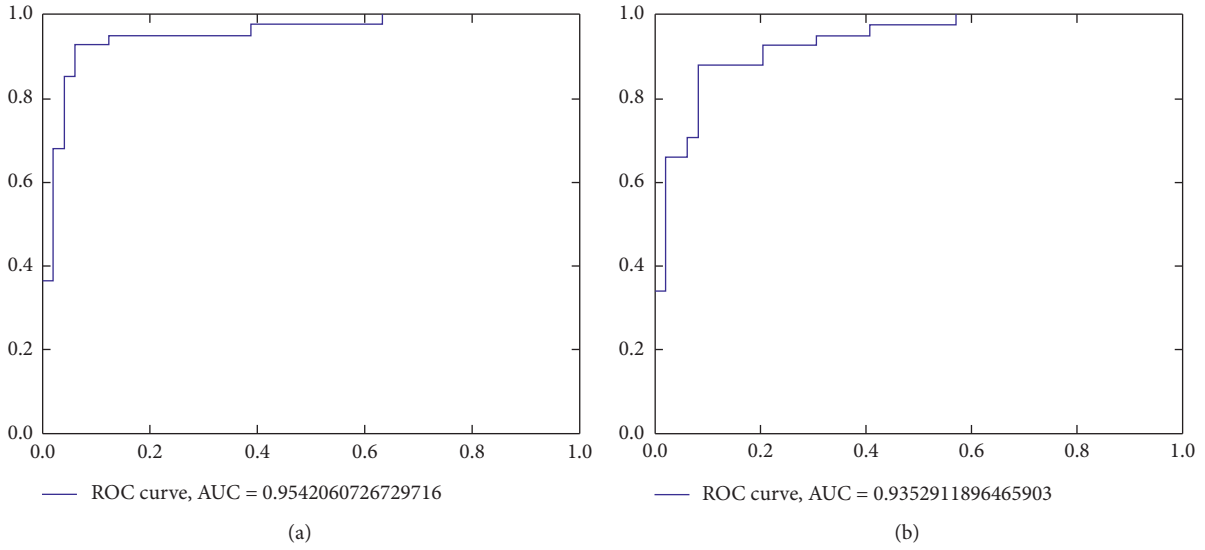


FIGURE 3: The ROC chart of the proposed ANN-based method. The ROC chart of the (a) proposed FWAFE-ANN diagnostic system and (b) conventional ANN-based diagnostic system.

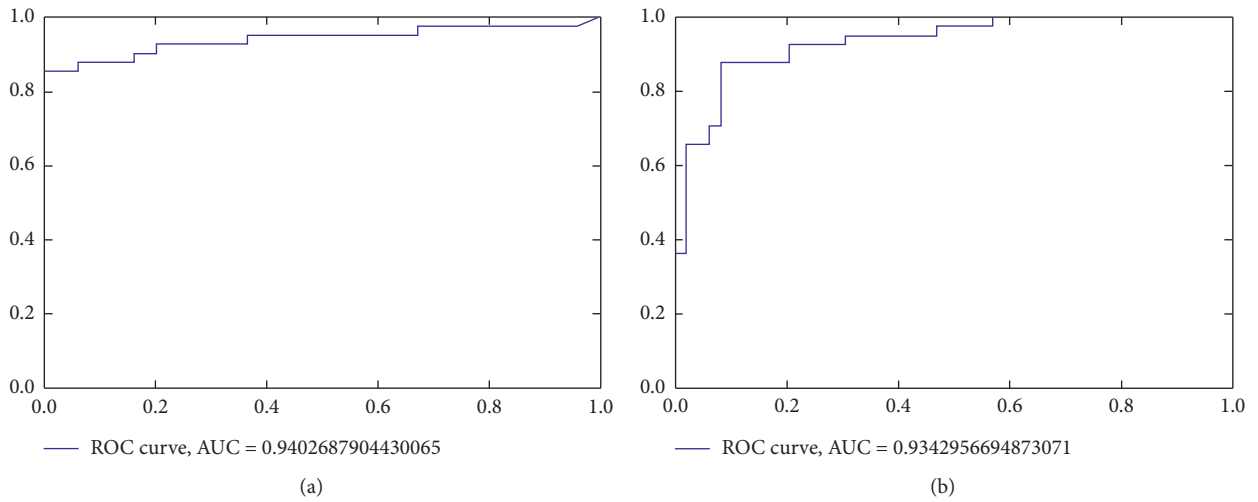


FIGURE 4: The ROC chart of the proposed DNN-based method. The ROC chart of the (a) proposed FWAFE-DNN diagnostic system and (b) conventional DNN-based diagnostic system.

TABLE 3: Results of different subsets of features for the heart disease dataset.

$n$	$F_e$	$N_1$	$N_2$	$Acc_{test}$	$Acc_{train} (\%)$	Sen. (%)	Spec. (%)	MCC
12	1	2	19	85.99	90.00	85.36	93.87	0.799
12	2	2	64	85.55	90.00	82.92	95.91	0.801
12	4	3	4	85.50	92.22	82.92	100	0.851
12	5	2	2	83.09	91.11	87.80	93.87	0.820
12	6	2	35	84.05	91.11	85.36	95.91	0.822
12	7	6	8	85.99	90.00	90.24	89.79	0.799
12	10	5	66	87.43	91.11	92.68	89.79	0.822
12	11	3	25	88.40	91.11	85.36	95.91	0.822
12	12	2	12	84.54	91.11	90.24	91.83	0.820
11	(5, 6)	50	2	83.57	93.33	85.36	100	0.872
6	(3 to 9)	16	42	93.23	92.22	90.24	93.87	0.843
13	—	2	4	85.02	90.00	87.80	91.83	0.798



TABLE 4: Performance of various predictive models on the heart disease dataset.

Model	Hyperparameters	Acc <sub>test</sub>	Acc <sub>train</sub>	Spec.	Sens.	MCC
Adaboost	$N_e = 25$	85.55	90.33	87.75	82.92	0.708
Adaboost	$N_e = 50$	86.66	94.68	87.75	85.36	0.731
Adaboost	$N_e = 75$	84.44	94.68	85.71	82.92	0.686
Adaboost	$N_e = 100$	82.22	97.58	81.63	82.92	0.643
Random forest	$N_e = 10$	81.11	98.06	87.75	73.17	0.619
Extra tree	$N_e = 10$	68.88	100.0	65.30	73.17	0.383
SVM (linear)	$C = 0.055$	90.00	84.05	93.87	85.36	0.799
SVM (RBF)	$C = 5, G = 0.2$	90.00	84.54	93.87	85.36	0.799
Proposed	$(50, 2)$	93.33	84.05	100	85.36	0.872

TABLE 5: Classification accuracies of the proposed method and other methods in the literature that used heart disease dataset.

Study (year)	Method	Accuracy (%)
ToolDiag, RA [35]	IB1-4	50.00
WEKA, RA [35]	InductH	58.50
ToolDiag, RA [35]	RBF	60.00
WEKA, RA [35]	FOIL	64.00
ToolDiag, RA [35]	MLP + BP	65.60
WEKA, RA [35]	T2	68.10
WEKA, RA [35]	1R	71.40
WEKA, RA [35]	IB1c	74.00
WEKA, RA [35]	$K^*$	76.70
Robert Detrano [35]	Logistic regression	77.00
Paul et al. [19]	FDSS	80
Khemphila and Boonijing (2011) [36]	MLP-backpropagation	80.99
Newton Cheung (2001) [37]	BNNF	80.96
Newton Cheung (2001) [37]	C4.5	81.11
Newton Cheung (2001) [37]	Naïve Bayes	81.48
Newton Cheung (2001) [37]	BNND	81.11
WEKA, RA [35]	Naïve Bayes	83.60
Šter and Dobnikar [38]	Fisher discriminant analysis	84.2
Šter and Dobnikar [38]	Linear discriminant analysis	84.5
Šter and Dobnikar [38]	Naïve Bayes	82.5–83.4
Polat et al. (2005) [25]	AIRS	84.50
Dwivedi (2018) [22]	LR	85
Ozsen et al. (2005) [39]	Kernel functions with AIS	85.93
Kahramanli and Allahverdi (2008) [40]	Hybrid neural network system	86.8
Polat et al. (2006) [10]	Fuzzy-AIRS-knn based system	87.00
Amin et al.(2019) [23]	LR, SVM, $K$ -NN	87.4
Verma et al. (2016) [20]	CFS-PSO	88.4
Das et al. (2009) [1]	Neural network ensembles	89.01
Jankowski and Kadirkamanathan (1997) [41]	IncNet	90.00
Ali and Bukhari (2019) [12]	Mutual information + DNN	90
Kumar (2011) [42]	ANFIS	91.18
Shah et al. (2017) [21]	PPCA	91.30
Samuel et al. (2017) [4]	ANN-fuzzy-AHP	91.10
Kumar (2012) [43]	Fuzzy resolution mechanism	91.83
Ali et al. (2019) [34]	Stacked SVMs	92.22
Proposed method (2020)	Feature selection based on FWAFE + ANN	91.11
Proposed method (2020)	Feature selection based on FWAFE + DNN	93.33

denotes the results of the abovementioned models. The performance of each model with hyperparameters values is also depicted from Table 4. For Adaboost classifier, the hyperparameter  $N_e$  represents the maximum number of estimators at which the boosting is terminated. For the RF classifier, the hyperparameter  $N_e$  denotes the number of trees in the forest. The ensemble model based on

randomized decision trees used average for improving the prediction accuracy. In case of SVM, the width of the Gaussian kernel is denoted by  $G$  and soft margin constant is denoted by  $C$ . Lastly, the number of neurons in 1<sup>st</sup> hidden layer of DNN is denoted in the last record of the table  $(50, 2)$ , i.e.,  $N_1 = 50$ , and the number of neurons in 2<sup>nd</sup> hidden layer of the DNN, i.e.,  $N_2 = 2$ . Moreover, the performance of the

proposed model is evidently precised and then the various state-of-the-art ensemble models as well as SVM model and are depicted from Table 4.

**4.4. Comparative Study with Previously Reported Methods.** In this section, experimental results of the proposed method are compared with those of the other methods discussed in the literature. The performance comparison is based on the prediction accuracy. Hence, Table 5 tabulates the prediction accuracies of our proposed method and other previously proposed methods in the literature. From the experimental outcomes, it is evident that the proposed hybrid method shows promising performance on heart disease, while the main limitation of the proposed method is its high time complexity.

From Table 5, it can be seen that many studies proposed numerous methods for automated detection of HF. For example, Ali et al. developed a two stage system using linear SVM at the first stage for feature selection and linear discriminant analysis model for classification at the second stage and obtained 90% accuracy. In another study, Verma et al. in [20] utilized the correlation-based feature subset (CFS) for feature selection and particle swarm optimization (PSO) algorithm for  $k$ -means clustering. Their method produced an accuracy of 88.4%. Saqlain et al in [21] proposed probabilistic principle component analysis and obtained an accuracy of 91.30%. Ali et al. in [34] proposed a novel hybrid method for improving the heart disease prediction accuracy. Their proposed method utilized linear SVM for feature selection and another SVM (with linear and nonlinear kernels) for classification. Their proposed method produced 92.22% heart disease detection accuracy. Hence, based on comparison with these methods, it is clear that our proposed method is a step forward in improving heart disease detection accuracy.

## 5. Conclusions

In this paper, an effort has been made to design a two stage diagnostic system that can improve the prediction accuracy of heart risk failure prediction. Two types of systems were developed. Both systems used same feature selection method, while the first system used ANN for classification and the second system used DNN for classification. A classification accuracy of 91.11% was achieved with the ANN-based system, while an accuracy of 93.33% was obtained with the DNN-based diagnostic system. It was also observed that the proposed diagnostic system shows better performance than other state-of-the-art machine learning models. From the experimental results, it can be safely concluded that the proposed system can help the physicians to make accurate decision while diagnosing heart disease.

## Data Availability

All the data used in this study are available at UCI machine learning repository.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2017R1D1A3B04031440 and by the Natural National Science Foundation of China under grant 61472066, Sichuan Science and Technology Program (Nos. 2018GZ0180, 2018GZ0085, 2017GZDZX0001, 281 2017GZDZX0002, 2018GZDZX0006, and 2018FZ0097).

## References

- [1] 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.
- [2] H. Yang and J. M. Garibaldi, "Automatic detection of protected health information from clinic narratives," *Journal of Biomedical Informatics*, vol. 58, pp. S30–S38, 2015.
- [3] L. A. Allen, L. W. Stevenson, K. L. Grady et al., "Decision making in advanced heart failure," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.
- [4] O. W. Samuel, G. M. Asogbon, A. K. Sangaiyah, P. Fang, and G. Li, "An integrated decision support system based on ann and fuzzy\_ahp for heart failure risk prediction," *Expert Systems with Applications*, vol. 68, pp. 163–172, 2017.
- [5] L. Ali, S. U. Khan, N. A. Golilarz et al., "A feature-driven decision support system for heart failure prediction based on statistical model and Gaussian naive bayes," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6314328, 8 pages, 2019.
- [6] L. Ali, S. U. Khan, M. Anwar, and M. Asif, "Early detection of heart failure by reducing the time complexity of the machine learning based predictive model," in *Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, IEEE, Swat, Pakistan, July 2019.
- [7] P. K. Anooj, "Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules," *Prediction of King Saud University—Computer and Information Sciences*, vol. 24, no. 1, pp. 27–40, 2012.
- [8] İ. Babaoglu, O. Findik, and E. Ülker, "A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3177–3183, 2010.
- [9] S. B. Patil and Y. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," *European Journal of Scientific Research*, vol. 31, no. 4, pp. 642–656, 2009.
- [10] K. Polat, S. Şahan, and S. Güneş, "Automatic detection of heart disease using an artificial immune recognition system (airs) with fuzzy resource allocation mechanism and  $k$ -nn (nearest neighbour) based weighting preprocessing," *Expert Systems with Applications*, vol. 32, no. 2, pp. 625–631, 2007.
- [11] S. U. Khan, M. K. A. Rahim, and L. Ali, "Correction of array failure using grey wolf optimizer hybridized with an interior point algorithm," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 9, pp. 1191–1202, 2018.

- [12] L. Ali and S. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *IRBM*, .
- [13] N. A. Golilarz, A. Addeh, H. Gao et al., "A new automatic method for control chart patterns recognition based on convnet and harris hawks meta heuristic optimization algorithm," *IEEE Access*, vol. 7, pp. 149398–149405, 2019.
- [14] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart diseases diagnosis using neural networks arbitration," *International Journal of Intelligent Systems and Applications*, vol. 7, no. 12, p. 72, 2015.
- [15] G. Manogaran, R. Varatharajan, and M. Priyan, "Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4379–4399, 2018.
- [16] L. Ali, S. U. Khan, M. Arshad, S. Ali, and M. Anwar, "A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease," in *Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, IEEE, Swat, Pakistan, July 2019.
- [17] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34938–34945, 2019.
- [18] F. S. Ahmed, L. Ali, B. A. Joseph, A. Ikram, R. Ul-Mustafa, and S. A. C. Bukhari, "A statistically rigorous deep neural network (DNN) approach to predict mortality in trauma patients admitted to the intensive care unit," *Journal of Trauma and Acute Care Surgery*, 2020.
- [19] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. Akhand, "Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease," in *Proceedings of the 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 145–150, IEEE, Dhaka, Bangladesh, May 2016.
- [20] L. Verma, S. Srivastava, and P. Negi, "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data," *Journal of Medical Systems*, vol. 40, no. 7, p. 178, 2016.
- [21] S. M. S. Shah, S. Batoool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, "Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis," *Physica A: Statistical Mechanics and Its Applications*, vol. 482, pp. 796–807, 2017.
- [22] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," *Neural Computing and Applications*, vol. 29, no. 10, pp. 685–693, 2018.
- [23] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics and Informatics*, vol. 36, pp. 82–93, 2019.
- [24] S. Özşen and S. Güneş, "Attribute weighting via genetic algorithms for attribute weighted artificial immune system (awais) and its application to heart disease and liver disorders problems," *Expert Systems with Applications*, vol. 36, no. 1, pp. 386–392, 2009.
- [25] K. Polat, S. Sahan, H. Kodaz, and S. Günes, "A new classification method to diagnosis heart disease: supervised artificial immune system (airs)," in *Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN)*, Izmir, Turkey, June 2005.
- [26] D. Dua and C. Graff, *UCI Machine Learning Repository*, 2017.
- [27] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, "Reliable Parkinson's disease detection by analyzing hand-written drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model," *IEEE Access*, vol. 7, pp. 116480–116489, 2019.
- [28] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.
- [29] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.
- [30] T. Meraj, A. Hassan, S. Zahoor et al., "Lungs nodule detection using semantic segmentation and classification with optimal features," *Neural Computing and Applications*, .
- [31] L. Ali, I. Wajahat, N. A. Golilarz, F. Keshkar, and S. A. C. Bukhari, "LDA-GA-SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Computing and Applications*, pp. 1–10, 2020.
- [32] F. S. Ahmad, L. Ali, Raza-Ul-Mustafa et al., "A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs)," *Journal of Ambient Intelligence and Humanized Computing*, 2020.
- [33] A. K. Paul, P. C. Shill, M. R. I. Rabin, and K. Murase, "Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease," *Applied Intelligence*, vol. 48, no. 7, pp. 1739–1756, 2017.
- [34] L. Ali, A. Niamat, J. A. Khan et al., "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019.
- [35] Datasets Used for Classification: Comparison of Results (2007), <http://www.is.umk.pl/projects/datasets.html>.
- [36] A. Khemphila and V. Boonjing, "Heart disease classification using neural network and feature selection," in *Proceedings of the 2011 21st International Conference on Systems Engineering*, pp. 406–409, IEEE, Las Vegas, NV, USA, October 2011.
- [37] N. Cheung, *Machine learning techniques for medical analysis. school of information technology and electrical engineering*, Ph.D. thesis, B. Sc. Thesis, University of Queensland, Brisbane, Australia, 2001.
- [38] B. Šter and A. Dobnikar, "Neural networks in medical diagnosis: comparison with other methods," in *Proceedings of the International Conference on Engineering Applications of Neural Networks*, pp. 427–430, London, UK, June 1996.
- [39] S. Ozsen, S. Gunes, S. Kara, and F. Latifoglu, "Use of kernel functions in artificial immune systems for the nonlinear classification problems," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 621–628, 2009.
- [40] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 82–89, 2008.
- [41] N. Jankowski and V. Kadirkamanathan, "Statistical control of rbf-like networks for classification," in *Proceedings of the International Conference on Artificial Neural Networks*, Springer, Lausanne, Switzerland, pp. 385–390, October 1997.
- [42] A. S. Kumar, "Adaptive neuro-fuzzy inference system for heart disease diagnosis," in *Proceedings of the International*

*Conference on Information System, Computer Engineering & Application (ICISCEA 2011)*, pp. 91–99, Singapore, 2011.

- [43] A. S. Kurnar, “Diagnosis of heart disease using fuzzy resolution mechanism,” *Journal of Artificial Intelligence*, vol. 5, no. 1, pp. 47–55, 2012.

## Research Article

# Evaluation of the Challenges in the Internet of Medical Things with Multicriteria Decision Making (AHP and TOPSIS) to Overcome Its Obstruction under Fuzzy Environment

Muhammad Imran Tariq <sup>1</sup>, Natash Ali Mian,<sup>2</sup> Abid Sohail,<sup>3</sup> Tahir Alyas,<sup>4</sup> and Rehan Ahmad<sup>5</sup>

<sup>1</sup>Department of Computer Science and Information Technology, Superior University, Lahore, Pakistan

<sup>2</sup>School of Computer and Information Technology, Beaconhouse National University, Lahore, Pakistan

<sup>3</sup>Department of Computer Sciences, COMSATS University Islamabad, Lahore, Campus, Pakistan

<sup>4</sup>Department of Computer Science, Lahore Garrison University, Lahore, Pakistan

<sup>5</sup>Department of Computer Science, The University of Lahore, Lahore, Pakistan

Correspondence should be addressed to Muhammad Imran Tariq; [imrantariqbutt@yahoo.com](mailto:imrantariqbutt@yahoo.com)

Received 10 March 2020; Revised 13 May 2020; Accepted 11 July 2020; Published 26 August 2020

Academic Editor: Sungchang Lee

Copyright © 2020 Muhammad Imran Tariq 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 exponential speed of advancement of innovation has expanded the needs of all users to avail all their information on the Internet 24/7. The Internet of things (IoT) enables smart objects to develop a significant building block in the development of the pervasive framework. The messaging between objects with one another means the least work and least expense for the enterprise. The industry that intends to implement the Internet of medical things (IoMT) in its organizations is still facing difficulties. Recognition and solving of these challenges are a time-consuming task and also need significant expenses if not adequately evaluated and prioritized. The application of the Internet of things is covered in almost every area, including medical/healthcare. In this research, the authors investigated the factors dealing with the Internet of medical things. The outcome of this study is to prioritize the level of significance of the elements causing these challenges, evaluated through fuzzy logic and multicriteria decision-making (MCDM) techniques like Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Analytic Hierarchy Process (AHP). It would be beneficial for enterprises to save time and revenue. The main criteria, as well as subcriteria, were determined after due consultation with the Internet of medical things experts. In this study, our goals are to figure out which criteria/factors create hurdles in the adoption of the Internet of medical things. Through the investigation, we figured out 20 criteria ought to be given more importance/preference by the industry that is in the transition phase of the Internet of medical things adoption. The enterprise, with the help of this study, will be enabled to accelerate that adoption by limiting time and fiscal misfortune.

## 1. Introduction

Internet of things (IoT) is probably the most sizzling innovation in the period of digital transformation, connecting and ensuring the availability of every device over the Internet. It is the most significant innovation behind smart homes, driving vehicles, smart utility billing, and intelligent urban communities [1]. Be that as it may, there are so many fundamental challenges for the eventual fate of the IoT. The adaptability of IoT devices is quickly expanding in the course

of the most recent couple of years [2]. As indicated by the survey organization, namely, Gartner, there will be more than 26 billion connected IoT devices around the globe by 2020. While IoT devices made possible viable communication between devices, computerized things, and also ensured saving of time and cost, it also has various advantages [3–5].

Numerous organizations are sorting out themselves to concentrate on integrating IoT and the availability of their future items and services. For the IoT business to flourish,

there are three classifications of difficulties to survive, and this is valid for any new pattern in innovation, not just IoT: innovation, business, and society [6].

The effect of IoT has been reformed in all areas of life; however, its impact on the medical system has been substantial because of its front-line transition. The job of IoT turns out to be increasingly prevailing when it is bolstered by the highlights of mobile computing [7]. This thing broadens the usefulness of IoT in the medical environment by getting tremendous help in the type of mobile health [8].

The unexpected ascent in the populace has given rise to numerous challenges in medical administrations and services, and eventually, it has led to a shortage of clinical assets [2, 9]. The medical organization does not have the expertise to resolve medical technology-related issues and challenges and provide an effective solution keeping in view the available resources [10]. IoT and mobile communication offer excellent solutions to the healthcare industry due to its less cost and easy to use features. The central theme of the medical IoT is to provide luxurious services to the users with very minimum price and best quality of service [11].

The main goal of the IoT is to give network services to the available healthcare resources and trustable, efficient, and medical services to the old patient. The IoT enabled medical and healthcare facilities, which consist of sensors, wireless networks for transferring data to a server, and also cloud computing to forward the same data over the Internet [12]. Furthermore, the Internet of medical things system also focuses on providing patient monitoring, treatment suggestions, and many more.

It is an undeniable fact that soon many Internet of things applications will be introduced in the market and many new smart objects will be available to connect with each other. The selection of Internet of medical things challenges depends upon various factors like specifications of the smart objects, cost, legal, and security issues; thus, it is required to compare a large number of criteria and subcriteria to alternatives, which need massive efforts and time. Second, the Internet of things firms and suppliers also write up their white and technical papers to inform the IoT users about the Internet of medical things challenges. Third, many blog writers who write up their survey reports on the Internet of medical things challenges also ranked these challenges without considering exact criteria and subcriteria or applying any mathematical model to prioritize the Internet of medical things challenges. Finally, many factors are required to be taken into consideration during the decision process. In the light of above, prioritization and evaluation of Internet of medical things challenges are complex multicriteria decision-making (MCDM) problems [13, 14]. It is pertinent to add here that due to a large number of Internet of medical things challenges, high complexity and computational power are required to be reduced.

To evaluate and choose the most challenging position for the Internet of medical things, the recommendations of the experts are the most appropriate solutions. In this study, the authors formulated an efficient and effective expert opinion system using MCDM methodology to evaluate and prioritize the Internet of medical things challenges. To simplify the

proposed solution and to reduce the complexities in the existing solutions, the authors proposed a hybrid multicriteria decision-making approach incorporating TOPSIS and AHP. The AHP is used to develop local and global weights of the criteria and subcriteria, and the TOPSIS is used to prioritize the alternatives. The authors, after a systematic literature review, claimed that this study is the first approach that they used to evaluate the Internet of medical things challenges. We properly designed and proposed hybrid AHP and TOPSIS model-based expert's opinion systems in the context of the Internet of medical things. This study is also the first to study existing criteria and develop its approaches after considering the pros and cons of the existing methods. After validation of the proposed framework with existing frameworks, the results proved that the proposed framework is better to rank the challenges of the Internet of Medical Things (IoMT).

The rest of the article is organized as follows. Section 2 is regarding literature review on the Internet of medical things challenges, MCDM techniques, and the most critical Internet of medical things challenges that the medical industry is currently facing. Section 3 describes identifying the main criteria methods, subcriteria, and alternatives, whereas Section 4 discusses the proposed research methodology used in this paper to resolve the problem. Section 5 is related to the simulation of the proposed methodology and results. In Section 6, a comparison of the proposed technique with existing technique is presented, and finally, Section 7 concludes the article.

## 2. Literature Review

The idea of IoT has a wide range of definitions in innovative dimensions. This is because analysts and industry have given significant importance to the IoT, keeping in view their requirements and business interests. Generally, the idea of IoT depends upon three methodologies: the Internet-based methods, the significance-based approach, and the object-based method. Internet of things is penetrating in our lives in the shape of intelligent objects like applications that can communicate on the system, have single IP addresses, are based on relevant communication protocols and procedures, and can sense changes in the environment like heat and radiations. A network connected with interrelated devices and internet, receive sensor's data as input, process on it, and send information to the desired nodes without human interaction known as the Internet of Things.

During the literature review, the authors of [6, 15, 16] gave information on benefits and challenges in the IoT and also informed the many issues like security and cost. Dizdarević et al. [17], Conti et al. [18], Farahani et al. [19], and Stojkoska and Trivodaliev [20] also discussed the various types of the IoT challenges like security, privacy, vendor lock-in, malicious insider, complexity in integration, competing standards, ubiquitous connectivity, law and regulations, business policies and procedures, the volume of data, and data analysis.

The Internet of things also connected with other technologies like cloud computing [21] and deep learning [22].

During a systematic literature review, it is observed that security and privacy are still a big challenge for the Internet of medical things. Several studies were already conducted in the near past to deal with the Internet of medical things security challenges [23–26] and as well as to deal with security challenges of cloud computing [27–32].

The IoT is the complex network developed through smart devices that are connected in different ways, producing data and information for communication and exchanging information within the network. The IoT is formulated with objects, tools, sensors, computers, desktop computers, and handheld devices, and it also helps the system to artificially think, feel, and speak.

Although there are several global definitions of the IoT, the core idea of this concept is the connectivity of different smart objects, which are automatically able to produce, communicate, and use information with very little human intervention to achieve common goals. Many studies were conducted regarding the challenges of the IoT, benefits or opportunities, and prioritization of the IoT challenges, but none of the studies were conducted till now for the prioritization of the Internet of medical things challenges. Ullah et al. [33] using fuzzy ANP conducted a study to prioritize the IoT challenges in the context of Iran. Uslu et al. studied the challenges of the IoT and prioritized these challenges by using AHP and ANP methodologies of the MCDM [34]. Mashal et al. examined the selection of the best application of IoT through MCDM and AHP [31].

Similarly, Mashal et al. [35] used the fuzzy analytic hierarchy process model for the analysis of the IoT. Kao et al. [36] evaluated the problems of the IoT in manufacturing industries using multicriteria decision making. Liu et al. [37] used hybrid MCDM tools to notice consumer adoption for mobile healthcare services. Abdel-Basset et al. [38] used MCDM to develop a medical decision support system. Shin et al. [39] conducted their study on the sustainability of the Internet of medical things and its integrated acceptance. A thorough investigation and research were carried out by Alsubaei et al. on the security and privacy of the Internet of medical things, and they also conducted risk assessment [40]. The selection of a stable IoT platform is also a challenge for the industry, and it is also solved through MCDM [41]. Another study was carried out by using MCDM for the zone head selection of the IoT-based WSN [42]. The authors used the fuzzy analytic hierarchy process and TOPSIS for the evaluation and ranking of ISO/IEC 27001:2013 information security controls [43].

However, after searching and investigating, the authors could not find a single study on the prioritization of the Internet of medical things challenges, especially using a combination of AHP and TOPSIS. It is pertinent to add here that such prioritization will help the industry to address critical challenges first to save time and reduce costs.

It would be practically difficult to cover the broad scope of the Internet of medical things challenges in a single article. Therefore, after critically reviewing the literature, Internet of medical things-related problems were dug out from high impact factor scientific papers. The authors, after intensive review, prepared the list of the Internet of medical things

challenges and, after due consultations with the Internet of medical things experts, categorized the challenges into five distinct categories and also presented the hierarchical structure. It is worthwhile to mention here that the experts confirmed the adequacy of the difficulties, and the validity of the criteria was verified.

*2.1. Multicriteria Decision Making.* Multicriteria decision making refers to discover the best option among the different alternatives having different characteristics, usually conflicting and based on decision matrix. It addresses and helps in the decision-making process when there are complex decision criteria. MCDM defines specific criteria and subcriteria, evaluates, selects, and prioritizes the alternatives. Many MCDM methods have been studied during the literature review. It is a very sophisticated and easy decision-making tool that provides quantitative and qualitative factors.

Multicriteria decision making has various methods, as shown in Figure 1. Many of these methods are already used for solving problem IoT-related problems like Analytic Hierarchy Process (AHP), Analytic Network Process (ANP), TOPSIS, Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE), Grey Relational Analysis (GRA), Goal Programming (GP), Value Analysis (VA), Value Engineering (VE), ELimination Et Choix Traduisant la REalité (ELECTRE), and Simple Additive Weighting (SAW).

In this article, the authors selected a combination of AHP and TOPSIS to evaluate the Internet of medical things challenges, develop criteria and subcriteria, formulate a comparison matrix, calculate local weights of the criteria and global weights of the subcriteria, and finally develop a decision matrix to rank the alternatives.

*2.2. Analytical Hierarchy Process (AHP).* Renowned mathematician Thomas Saaty formulated AHP to complicated decisions [44] and facilitated all kinds of industries to decide their priorities in all areas among different alternatives [43, 45, 46]. It is a powerful technique to solve unstructured, complex, and complicated problems. In AHP, the complex problem is always breakdown into small problems and organized into hierarchical levels. Each level of the hierarchical structure represents several criteria, subcriteria, and alternatives [47]. Each alternative is synthesized to finally rank/prioritize them to get the best solution. The AHP is a multicriteria analysis methodology based on the weighting process, and each criterion represents its importance. AHP is applied in various areas like education, engineering, healthcare, and especially in the financial areas. In AHP, initially, the importance of each criterion is set through pairwise comparison in which each criterion is compared with the rest of the criteria by assigning values. The first version of the AHP has various shortcomings [48]. Mr. Chen and Yang while highlighting deficiencies in AHP, pointed out that it is only used for crisp information, an unbalanced scale is used for judgment, it also not able to handle the uncertainty associated with human opinions, generate imprecise ranking, and final

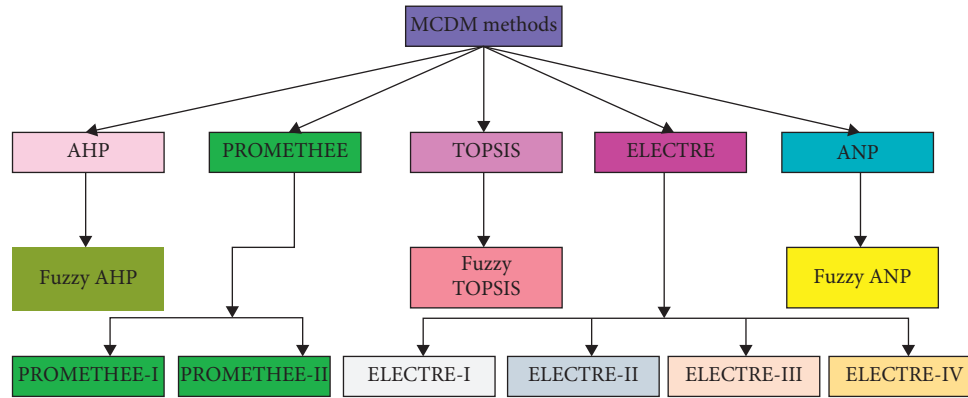


FIGURE 1: Multicriteria decision-making methods.

selection of alternatives are only based on experts' opinion. To remove the discrepancies in the AHP, various researchers integrated AHP with fuzzy set theory. In the year 1987, Buckley used fuzzy trapezoidal numbers to calculate the weights [49]. After applying the fuzzy analytic hierarchy process, the biasness of the evaluators can also be minimized, reliability can be increased, and it also becomes more validated [50].

**2.3. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS).** The TOPSIS technique was initially formulated in the year 1981 by Yoon and Hwang [51]. The main aim of the method is to evaluate the alternatives based on specific criteria, get opinions of the experts/decision makers to develop decision matrix, and select an alternative that has the shortest distance from the positive ideal solution and an alternative that has the farthest distance from the negative ideal solution. In a positive ideal solution, we give more weightage to the benefit criteria and less weightage to the cost criteria, whereas on the other hand, in a negative ideal solution, the cost criteria have more weightage and benefit criteria have less weightage. The upshot is the positive ideal solution provides the best values based on the criteria and vice versa. The reader who is more interested in studying the TOPSIS may read a broad survey [52]. According to Chen's method [53], the technique of fuzzy TOPSIS is like classical TOPSIS. Various studies have already been conducted by using fuzzy logic with TOPSIS for the IoT [54, 55].

### 3. Identifying the Criteria, Subcriteria, and Alternatives

In the AHP, the purpose of the criteria is to help in the selection problem. All the alternatives have to satisfy an independent set of criteria. The authors developed the criteria and their subcriteria based on the opinions of the Internet of medical things experts and proposed five criteria: (C1) security and privacy criterion, which represents the most influential factors regarding security and privacy; (C2) data criterion, which represents the Internet of medical things challenges related to data; (C3) technology criterion, which represents the technology-related Internet of medical things

challenges; (C4) legal criterion, which is used to highlight the legal issues of the IoMT; and (C5) cost criterion, which identifies the cost related to Internet of medical things challenges. The proposed framework is about the evolution of the Internet of medical things challenges as criteria, and three medial IoT scenarios, as alternatives, were developed and organized into four main layers, as shown in Figure 2. The first layer of the framework is to define the goal of the study, and the second layer consists of five significant forces, namely, security and privacy, data, technology, legal, and cost. In the third layer, twenty subcriteria are placed against each significant force. Every subcriterion is adequately connected with its relevant criteria. The fourth layer of Figure 2 is about alternatives, and we placed three alternatives for prioritization purposes based on criteria and subcriteria.

The analytic hierarchical process deliberated many assessment criteria and helped to select the best alternative [45, 56, 57]. Figure 2 shows a hierarchy structure based on the AHP method for the evaluation of the Internet of medical things challenges.

Figure 2 explains the methods of AHP, criteria, subcriteria, and alternatives in a graphical view. In the AHP, the first stage is to establish the goals of the study that is to evaluate and select the most critical Internet of medical things challenges. After formulating goals, criteria for evaluation and prioritization are required to be developed with the consultations of the experts/decision makers. In Figure 2, the authors defined five criteria and twenty subcriteria with the discussion of experts. Every criterion has been assigned a weight that designates the position of the criteria. The planned criteria are applied to alternatives to choose the best alternative, among other ambiguous alternatives, and also rank the alternatives. Figure 2 shows three Internet of medical things challenges as alternatives. The most critical Internet of medical things challenge and ranking of the IoMT challenges should be done according to 20 subcriteria.

### 4. Fuzzy Analytic Hierarchy Process and Fuzzy TOPSIS Methodology

In this study, the Internet of medical things- (IoMT-) related challenges in the medium-sized industry in Pakistan were assessed. The investigation's point was to decide



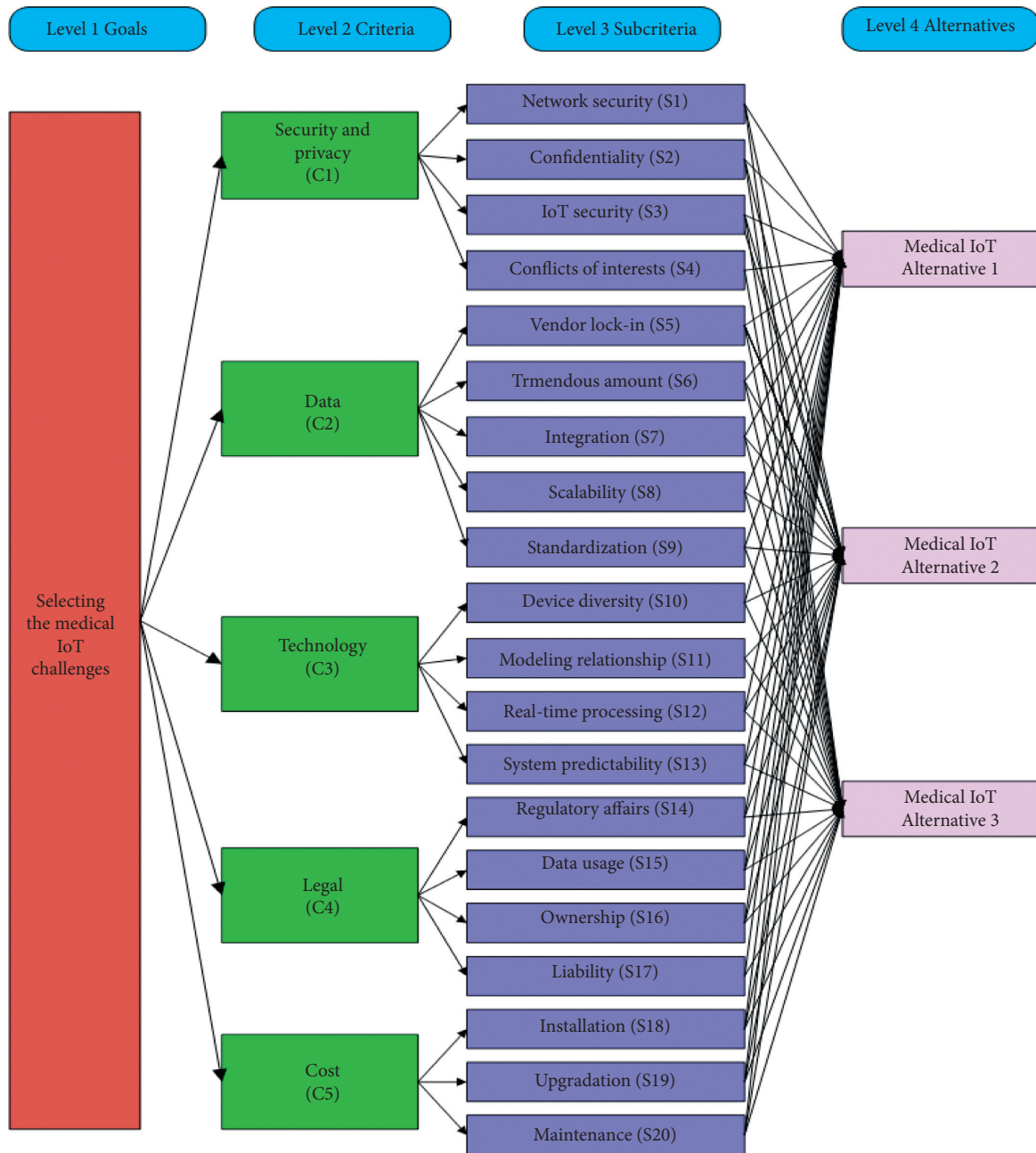


FIGURE 2: Hierarchy structure of AHP.

the Internet of medical things challenges in the adoption and deployment by studying and developing the importance of criteria. The criteria that ought to get specific consideration were underlined. Criteria were developed based on consultations with IoMT experts and relevant industry stakeholders. In this study, we considered small and medium-sized enterprises (SMEs) that are willing to adopt IoMT services as a case study. To evaluate the criteria, specialists were consulted for their expert opinion. As per the resources of the organization, three decision makers were asked. These decision makers were individuals who researched on the Internet of medical things. The Decision Makers have also studied the literature related to IoMT and also have experience to deal with such challenges.

The limited number of decision makers is the obstruction of this study. To cater to this prompting lacking assessments, the authors of this article get the help from the literature review to authenticate the accuracy of the evaluation. Three decision makers used fuzzy-based AHP and TOPSIS techniques for the assessment in the area of Internet of medical things. Fuzzy-based pairwise comparison matrices were assessed independently by the decision makers. These evaluations were then aggregated by using fuzzy aggregation equations.

The proposed hybrid methodology consists of four phases, as shown in Figure 3. Each stage consists of the next stage to get the outcome. Based on Figure 3, the following steps were taken to evaluate and rank the IoMT challenges:

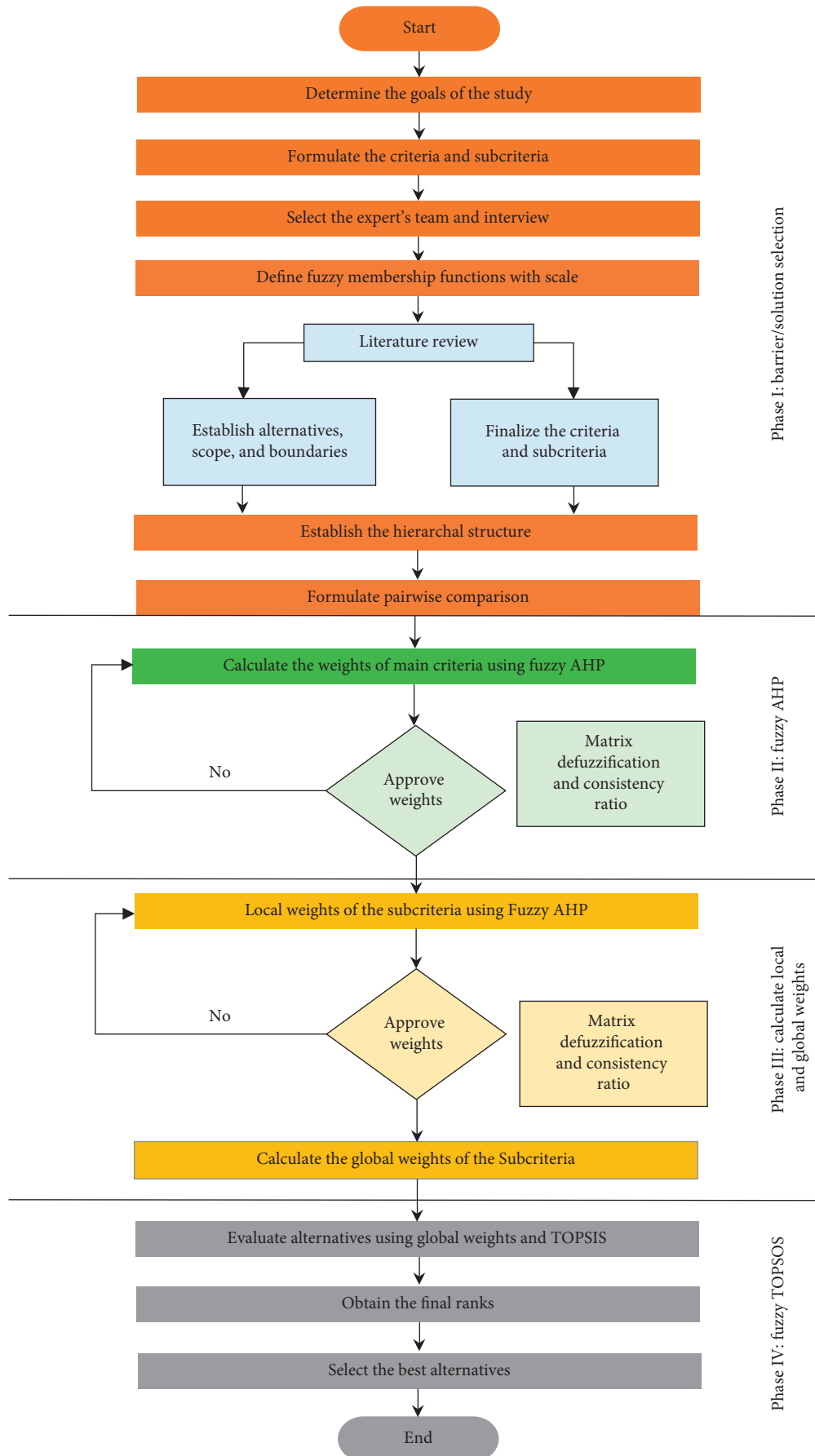


FIGURE 3: Fuzzy analytic hierarchy process and TOPSIS methodology.

- (1) Define the goals and objective of the study.
- (2) Contact decision makers/experts and hold a meeting with them to formulate criteria and sub-criteria and select alternative.
- (3) For pairwise comparison, the express scale consists of fuzzy membership functions.
- (4) Conduct systematic literature review (SLR).
- (5) Finalize criteria, subcriteria, and alternatives with the Internet of medical things experts.
- (6) Define scope and boundaries of the Internet of medical things.
- (7) Prepare matrix based on hierarchal structure.
- (8) Establish a pairwise comparison matrix.
- (9) Initiate the fuzzy defuzzification process and check the CR. If  $CR < 0.1$  and weights of the criteria are according to pairwise comparison, then move forward; otherwise, go back to Step 7.
- (10) Compute the weight of criteria using the fuzzy analytic hierarchy process technique.
- (11) Formulate the pairwise comparison matrix of each subcriterion based on the opinions of decision makers. Check the consistency ratio of the sub-criteria matrix and move forward to calculate the local weights of the subcriteria; otherwise, repeat previous steps to rectify the error.
- (12) Calculate the local weights of the subcriteria using fuzzy analytic hierarchy process.
- (13) Calculate the global weights of the subcriteria by multiplying weights of the criteria with local weights of the subcriteria.
- (14) Develop decision matrices using fuzzy TOPSIS.
- (15) Aggregate the options of the decision makers by applying fuzzy aggregation equations.
- (16) Normalize the decision matrix and assign the global weights to the decision matrix.
- (17) Compute final ranks through fuzzy-based TOPSIS techniques.
- (18) Select the best alternative/Internet of medical things challenges.

Figure 3 reflects the proposed methodology for the prioritization of the Internet of medical things challenges. The core advantages of this integrated and hybrid methodology are that the limitation of the AHP is covered with the introduction of the fuzzy logic for the weighting of criteria and the best-ranking technique, i.e., TOPSIS is used along with fuzzy logic to evaluate and prioritize the best alternative.

**4.1. Phase I: Identification of Medical IoT Challenge.** A detailed systematic literature review was conducted on the Internet of medical things and examined. We also investigated the same nature of the problem addressed in the literature review; the methodology used by the author of the study, criteria, and method adopted by the authors were also

determined. During the systematic literature review, we observed that each area of science has its importance which is still under development. Although the idea of IoT is not new, it has dynamic characteristics due to rapid change in growth, especially in the area of healthcare. Initially, we conducted this study in collaboration with small and medium-sized enterprises. The criteria and subcriteria of the Internet of medical things were developed by the authors of this article and duly examined and evaluated by the Internet of medical things experts having industry experience, and the fuzzy analytic hierarchy process and fuzzy TOPSIS methods were introduced. The decision makers/experts before evaluation first examined the organizational structure. The decision makers, thereafter, studied the concepts of the literature review of this article and the Internet of medical things. The experts, keeping in view the existing study on the related topics, prepared the goals of the survey and amended and finalized their own proposed criteria and subcriteria of the Internet of medical things challenges. The validity and authenticity of the criteria were discussed in detail with experts and industry stakeholders, and finally, the evaluation process was initiated. The goals of the study, main criteria, subcriteria, and alternatives are shown in Figure 2.

The next step is the formulation of the pairwise comparison matrix and transformation of the real numbers into fuzzy numbers. To accomplish this, triangular fuzzy numbers (TFNs) are used, as shown in Figure 4.

Although there are many other methods like trapezoidal fuzzy numbers (TrFNs), due to full acceptance of the TFNs, we used this in AHP and TOPSIS. Equation (1) is used to formulate FTMF, and  $(l, m, u)$  are used for notation.

$$\mu(\bar{M}) = \begin{cases} 0, & x < l, \\ (x-l)/(m-l), & l \leq x \leq m, \\ (u-x)/(u-m), & m \leq x \leq u, \\ 0, & x > u. \end{cases} \quad (1)$$

Use equation (2) to construct the pairwise comparison matrix, which has comparison in pairs, and select the appropriate linguistic values for alternatives regarding criteria. The goal is to define relative priorities for each element. The value  $a_{ij}$  demonstrates the relative significance of criterion  $i$  ( $c_i$ ) in comparison with criterion  $j$  ( $c_j$ ) in Saaty's scale.

$$A = [a_{ij}] = \begin{matrix} & \begin{matrix} c_1 & c_2 & \dots & c_n \end{matrix} \\ \begin{matrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{matrix} & \begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{21} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{n1} & 1/a_{n2} & \dots & 1 \end{bmatrix} \end{matrix} \quad (2)$$

Let  $X = \{x_1, x_2, x_3, \dots, x_n\}$  be an object set and  $G = \{g_1, g_2, g_3, \dots, g_n\}$  be the goal setting. Therefore, we can calculate  $m$  extent analysis value for every object using the following equation:

$$M_{gi}^1, M_{gi}^2, \dots, M_{gi}^m, \quad i = 1, 2, 3, \dots, n. \quad (3)$$

**4.2. Phase II: Fuzzy Analytic Hierarchy Process.** To calculate the weights of the criteria, first, we have to figure the fuzzy synthetic extent concerning  $i$ th object by using the following equation:

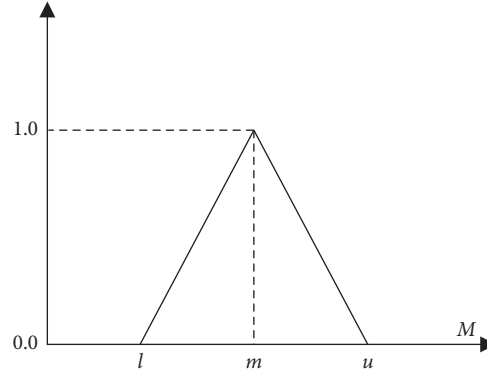


FIGURE 4: Fuzzy triangular number.

$$S_i = \sum_{j=1}^m M_{gi}^j \oplus \left[ \sum_{i=1}^n \sum_{j=1}^m M_{gi}^j \right]^{-1}. \quad (4)$$

To obtain  $\sum_{j=1}^m M_{gi}^j$ , the fuzzy adding formula will be applied to  $m$  extent analysis through the following equation:

$$\sum_{j=1}^m M_{gi}^j = \left( \sum_{j=1}^n l_j, \sum_{j=1}^n m_j, \sum_{j=1}^n u_i \right). \quad (5)$$

To obtain  $[\sum_{i=1}^n X_i \sum_{j=1}^m M_{gi}^j]^{-1}$ , fuzzy logic-based addition function set in equation (6) will be applied, and we also compute the inverse of equation (6) by using equation (7).

$$\sum_{i=1}^n \sum_{j=1}^m M_{gi}^j = \left( \sum_{j=1}^n l_j, \sum_{j=1}^n m_j, \sum_{j=1}^n u_i \right), \quad (6)$$

$$\left[ \sum_{i=1}^n \sum_{j=1}^m M_{gi}^j \right]^{-1} = \left( \frac{1}{\sum_{j=1}^n u_i}, \frac{1}{\sum_{j=1}^n m_i}, \frac{1}{\sum_{j=1}^n l_i} \right). \quad (7)$$

Calculate degree to the possibility in a situation where  $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ , and it is defined as equations (8) to (10).

$$V(M_2 \geq M_1) = \sup_{y \geq x} [\min(\mu_{M_1}(x), \mu_{M_2}(y))], \quad (8)$$

$$V(M_2 \geq M_1) = \text{hgt}(M_1 \cap M_2) = \mu_{M_2}(d), \quad (9)$$

$$V(M_2 \geq M_1) = \text{hgt}(M_1 \cap M_2) = \frac{(l_1 - u_2)}{(m_2 - u_2)} - (m_1 - l_1), \quad (10)$$

where  $(d)$  is an ordinate linked with intersection point  $(D)$ , and the highest intersection point between TFNs is shown in Figure 5.

To calculate the lowest degree of possibility  $M_2 \geq M_1$ , fuzzy values  $M_i = (1, 2, \dots, k)$  are required to be computed as shown in the following equation:

$$V(M \geq M_1, M_2, \dots, M_k) = \min V(M \geq M_i), \quad (i = 1, 2, \dots, k). \quad (11)$$

Assuming that  $d'(A_i) = \min V(S_i \geq S_k)$  for  $K = 1, 2, \dots, n$ , finally weights of main criteria shall be computed using the following equation:

$$W' = (d'(A_1), d'(A_2), \dots, d'(A_n)), \quad (12)$$

where  $A_i = (i = 1, 2, \dots, n)$  and  $n$  are elements.

It is imperative to normalize the matrix;  $W'$  represents priority weights and is calculated using the following equation:

$$W = (d(A_1), d(A_2), \dots, d(A_n))^T, \quad (13)$$

where  $W$  is not a fuzzy number.

AHP calculates the consistency index (CI) to validate the results of the comparison matrix. Equation (14) helps to

calculate inconsistency in the decisions of experts, whereas  $\lambda_{\max}$  is the principal eigenvalue of the decision matrix.

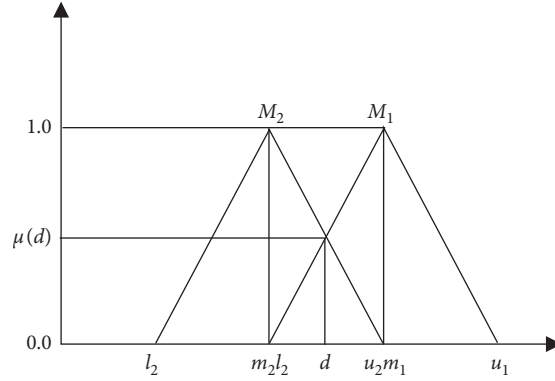
$$CI = \frac{\lambda_{\max} - n}{n - 1}. \quad (14)$$

Equation (15) assists in calculating the consistency ratio (CR) to authenticate the final weights.

$$CR = \frac{CI}{RI}. \quad (15)$$

The values of the random index (RI) are available in [58].

**4.3. Phase III: Calculating Local and Global Weights.** After calculating the weights of the criteria, the next step is to calculate the local weights of the subcriteria. For each subcriterion, the pairwise comparison matrix will be


 FIGURE 5: The intersection between  $M_1$  and  $M_2$ .

developed, and equation (1) to equation (15) will be used to compute the local weights of the subcriteria. The numerical example is shown in all steps in Section 5.

Global weights of the criteria are also required to be calculated by multiplying the values of the weights with local weights of the subcriteria.

The complete steps using the fuzzy analytic hierarchy process to calculate the weights, local weights, and global weights of the criteria are explained in Figure 6.

**4.4. Phase IV: Fuzzy TOPSIS.** The fourth stage of the study is to introduce a fuzzy TOPIS technique to evaluate the alternatives based on the weights of the criteria determined through the fuzzy analytic hierarchy process.

Equation (16) is used to develop a decision matrix of the TOPSIS method. Linguistic variables are used in this process.

$$E = [e_{ij}]_{n \times m} = \begin{matrix} A_1 \\ \vdots \\ A_m \end{matrix} \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{bmatrix}_{m \times n} \quad (16)$$

Thereafter, the linguistic variables are required to be converted into TFNs so that related fuzzy operations may be applied.

In TOPSIS, we have taken the expert opinion of three decision makers individually. Equations (17) to (20) are related to the aggregation of experts' preference values.

$$\tilde{X}_{ij} = (a_{ij}, b_{ij}, c_{ij}), \quad (17)$$

$$a_{ij} = \min_K \{a_{ij}^k\}, \quad (18)$$

$$b_{ij} = \frac{1}{k} \sum_{k=1}^k b_{ij}^k, \quad (19)$$

$$c_{ij} = \max_K \{a_{ij}^k\}. \quad (20)$$

Normalize the fuzzy decision matrix using equation (21), and it is denoted as  $\tilde{B}$ :

$$\tilde{B} = [r_{ij}]_{m \times n}. \quad (21)$$

In order to compute beneficial criteria, equation (22) is used.

$$\tilde{r}_{ij} = \left( \frac{a_{ij}}{c_j^*}, \frac{b_{ij}}{c_j^*}, \frac{c_{ij}}{c_j^*} \right), \quad (22)$$

$$c_j^* = \max\{c_{ij}\}.$$

To compute non-bifacial criteria/criteria, equation (23) is used.

$$\tilde{r}_{ij} = \left( \frac{a_j^-}{c_{ij}}, \frac{a_j^-}{b_{ij}}, \frac{a_j^-}{a_{ij}} \right), \quad (23)$$

$$c_j^* = \min\{a_{ij}\}.$$

Equation (24) is used to compute weighted normalized fuzzy decision matrix:

$$\tilde{v}_{ij} = \tilde{r}_{ij} \times w_j. \quad (24)$$

Equation (25) is related to calculating fuzzy positive ideal solution (FPIS), and equation (26) is associated with fuzzy negative ideal solution (FNIS).

$$V_i^+ = \left\{ \max_{1 \leq i \leq n} \left( \{v_{ij}\}_{i=1}^n \mid j \in J^+ \right), \min_{1 \leq i \leq n} \left( \{v_{ij}\}_{i=1}^n \mid j \in J_- \right) \mid j = \{v_{ij} \mid j = 1, 2, \dots, n\} \right\}, \quad (25)$$

$$V_i^- = \left\{ \min_{1 \leq i \leq n} \left( \{v_{ij}\}_{i=1}^n \mid j \in J^+ \right), \max_{1 \leq i \leq n} \left( \{v_{ij}\}_{i=1}^n \mid j \in J_- \right) \mid j = \{v_{ij} \mid j = 1, 2, \dots, n\} \right\}. \quad (26)$$

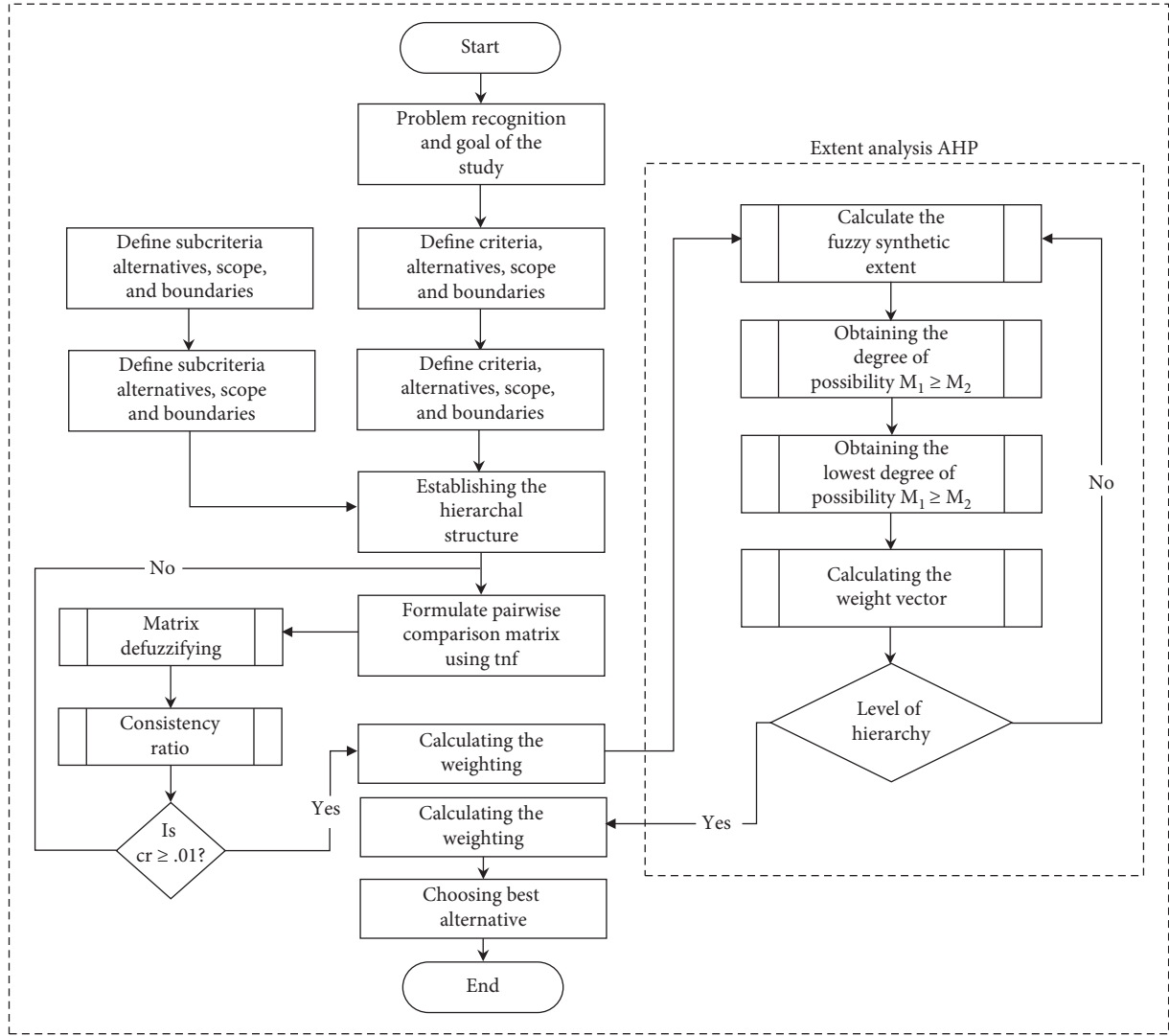


FIGURE 6: Fuzzy analytic hierarchy process methodology.

Calculate the distance of alternatives from fuzzy positive ideal solution using equation (27) and fuzzy negative ideal solution through equation (28).

$$d_i^+ = \sqrt{\sum_{j=1}^m (V_{ij} - V_i^+)^2}, \quad (27)$$

where  $(i = 1, 2, \dots, n)$ .

$$d_i^- = \sqrt{\sum_{j=1}^m (V_{ij} - V_i^-)^2}, \quad (28)$$

where  $(i = 1, 2, \dots, n)$ .

The final relative proximity  $P_i$  is calculated using the following equation:

$$P_i = \frac{d_i^-}{d_i^+ + d_i^-}. \quad (29)$$

## 5. Results and Discussion

The organization deploying the Internet of medical things may use the proposed methodology for the evaluation of real-world Internet of medical things-related challenges. In this Internet of medical things challenge problem, there are five criteria, twenty subcriteria, and four alternatives. The hierarchical structure to choose the most critical Internet of medical things challenge is shown in Figure 2. The proposed organization size is a small and medium-sized enterprise. In this study, detailed interviews were conducted with three Internet of medical things experts and one industry stakeholder to identify weight coefficients. Before gathering data from the experts, we defined the linguistic scale based on the fuzzy triangular numbers and discussed the same scale with the experts. The authors contacted many IoMT experts, but the majority of the experts refused to assist the authors due to their personal and official limitations. After the approval of the linguistic scale, we also discussed the criteria and subcriteria with the experts.

It is pertinent to mention here that due to personal access to the IoMT experts, one of the authors of this article visited each expert and discussed the aforementioned things. It is also paramount to add here that to protect their privacy, the name of the experts and their organizations are not mentioned in this article. After finalizing the proposed criteria and linguistic scale, the experts were given both things with the guideline to insert their input in the pairwise matrix. After the development of the pairwise matrix, the authors applied the proposed fuzzy analytic hierarchy process methodology, and the results are shown in Table 1. The authors then identified the four alternatives as a sample to evaluate and prioritize them. The experts then gave their opinion on each alternative by considering the TOPSIS parameters and ranked each alternative. The aggregated matrix is shown in Table 2.

Fuzzy analytic hierarchy process and fuzzy TOPSIS techniques were integrated to obtain the best results and to rectify the shortcomings in both methodologies. The step by step numerical example of the proposed method for the Internet of medical things challenges is given as follows.

*Step 1.* The first step is to define the objectives and goals of the study. Here in this example, the purposes of the research are to evaluate and prioritize the Internet of medical things challenges and find the most critical problems to save cost and take measures in due course of time.

*Step 2.* The most important and critical stage of this study is to finalize the criteria to evaluate and rank the alternatives. We already explained the details about the formulation of the criteria in the previous section.

*Step 3.* During the literature review, we studied many linguistic scales for the evaluation of alternatives, and these scales were developed, keeping in view the nature of the problem. In this study, we agreed to use TFNs and Saaty's scale [59], as shown in Table 3. The range of the scale is 0 to 11. Figure 7 represents the fuzzy analytic hierarchy process scale opted for evaluation, and Table 3 expresses the scale for pairwise comparison matrix.

*Step 4.* In the next level, prepare a comparison matrix table using given above linguistic scale, as shown in Table 4.

*Step 5.* By using equations (2) and (3), develop a fuzzy pairwise comparison matrix based on the decision maker's opinion and linguistic scale, as shown in Table 5.

*Step 6.* Use equation (4) to calculate the weights of the main criteria by synthesizing values.

$$S_{C1} = (7.00, 13.00, 19.00) * (0.017, 0.023, 0.035) = (0.116, 0.301, 0.666).$$

$$S_{C2} = (6.14, 10.20, 14.33) * (0.017, 0.023, 0.035) = (0.102, 0.236, 0.502).$$

$$S_{C3} = (6.68, 8.87, 11.67) * (0.017, 0.023, 0.035) = (0.111, 0.205, 0.409).$$

$$S_{C4} = (2.51, 4.81, 8.20) * (0.017, 0.023, 0.035) = (0.042, 0.111, 0.287).$$

$$S_{C5} = (6.20, 6.33, 7) * (0.017, 0.023, 0.035) = (0.103, 0.147, 0.245).$$

*Step 7.* Equations (9) and (10) were used to calculate the degree of possibility.

$$V(S_{C1} \geq S_{C2}) = 1, V(S_{C1} \geq S_{C3}) = 1, V(S_{C1} \geq S_{C4}) = 1, V(S_{C1} \geq S_{C5}) = 1.$$

$$V(S_{C2} \geq S_{C1}) = 0.856, V(S_{C2} \geq S_{C3}) = 1, V(S_{C2} \geq S_{C4}) = 1, V(S_{C2} \geq S_{C5}) = 1.$$

$$V(S_{C3} \geq S_{C1}) = 0.754, V(S_{C3} \geq S_{C2}) = 0.909, V(S_{C3} \geq S_{C4}) = 1, V(S_{C3} \geq S_{C5}) = 1.$$

$$V(S_{C4} \geq S_{C1}) = 0.474, V(S_{C4} \geq S_{C2}) = 0.598, V(S_{C4} \geq S_{C3}) = 0.653, V(S_{C4} \geq S_{C5}) = 0.839.$$

$$V(S_{C5} \geq S_{C1}) = 0.456, V(S_{C5} \geq S_{C2}) = 0.616, V(S_{C5} \geq S_{C3}) = 0.696, V(S_{C5} \geq S_{C4}) = 1.$$

*Step 8.* To compute priority weights, equation (11) was considered.

$$d'(C_1) = \min(1, 1, 1, 1, 1) = 1.$$

$$d'(C_2) = \min(0.856, 1, 1, 1, 1) = 0.856.$$

$$d'(C_3) = \min(0.754, 0.909, 1, 1) = 0.754.$$

$$d'(C_4) = \min(0.474, 0.598, 0.653, 0.839) = 0.474.$$

$$d'(C_5) = \min(0.456, 0.616, 0.696, 1) = 0.456.$$

The priority weights are  $W' = (1, 0.856, 0.754, 0.474, 0.456)$ .

*Step 9.* Equations (12) and (13) were used to normalize the priority weights of the criteria. The weights of each criterion are shown in Table 1. The contribution of the main criteria for the evaluation of the Internet of medical things challenges is shown in Figure 8.

*Step 10.* Compute the value of  $\lambda_{\max} = 7.629$  and then calculate the consistency index (CI) using equation (14) to validate the methodology, and 0.105 means the comparison matrix and expert's opinions are well. Similarly, equation (15) was used to calculate consistency ratio (CR), and it was 0.094, which was less than 0.1.

*Step 11.* The local weights of the subcriteria were calculated by using the fuzzy analytic hierarchy process equations (1) to (13). The local weights are shown in Table 6.

*Step 12.* The values of the consistency index (CI) and consistency ratio (CR) of each subcriterion were again calculated using equations (14) and (15), and the same is reflected in Table 7.

*Step 13.* The global weights of the main criteria are calculated by multiplying the weights of the criteria with the local weights of the subcriteria. The values of the global weights of

TABLE 1: Weights of each criteria.

(C1)	(C2)	(C3)	(C4)	(C5)
0.282	0.242	0.213	0.134	0.129

TABLE 2: Aggregated fuzzy-based decision matrix.

Subcriteria alternative	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Alternative 1	(5, 7, 9)	(3, 7, 9)	(3, 5, 7)	(3, 6, 9)	(1, 6, 9)	(1, 5, 9)	(3, 7, 9)	(3, 7, 9)	(3, 5, 7)	(3, 5, 7)
Alternative 2	(1, 6, 9)	(1, 6, 9)	(5, 7, 9)	(1, 6, 9)	(1, 4, 9)	(1, 3, 7)	(1, 6, 9)	(1, 6, 9)	(5, 7, 9)	(1, 6, 9)
Alternative 3	(5, 8, 9)	(1, 6, 9)	(5, 8, 9)	(5, 7, 9)	(5, 7, 9)	(1, 6, 9)	(3, 6, 9)	(5, 8, 9)	(5, 8, 9)	(5, 8, 9)
Alternative 4	(5, 8, 9)	(3, 5, 7)	(1, 2, 5)	(1, 4, 7)	(1, 6, 9)	(3, 6, 9)	(1, 5, 9)	(3, 5, 7)	(1, 2, 7)	(1, 3, 5)
Subcriteria alternative	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
Alternative 1	(3, 8, 9)	(1, 6, 9)	(1, 5, 9)	(5, 7, 9)	(1, 5, 9)	(3, 7, 9)	(3, 6, 9)	(3, 6, 9)	(3, 6, 7)	(3, 6, 9)
Alternative 2	(1, 4, 9)	(1, 4, 9)	(1, 4, 9)	(1, 5, 9)	(1, 4, 9)	(1, 6, 9)	(5, 7, 9)	(1, 5, 9)	(1, 6, 9)	(5, 7, 9)
Alternative 3	(5, 7, 9)	(1, 6, 9)	(3, 6, 9)	(1, 6, 9)	(3, 6, 9)	(3, 7, 9)	(5, 8, 9)	(5, 7, 9)	(5, 8, 9)	(5, 8, 9)
Alternative 4	(1, 4, 7)	(3, 6, 9)	(1, 6, 9)	(1, 4, 7)	(1, 6, 9)	(1, 4, 7)	(1, 4, 7)	(1, 3, 7)	(1, 4, 7)	(1, 4, 7)

TABLE 3: Fuzzy scale and numbers.

Scale (0–11)	Evaluation	Fuzzy scale	Reciprocal
1	Equally important (EI)	(1, 1, 1)	(1, 1, 1)
3	Moderately important (MI)	(1, 3, 5)	(1/5, 1/3, 1)
5	Strongly important (SI)	(3, 5, 7)	(1/7, 1/5, 1/3)
7	Very strongly important (VSI)	(5, 7, 9)	(1/9, 1/7, 1/5)
9	Extremely important (EI)	(7, 9, 11)	(1/11, 1/9, 1/7)

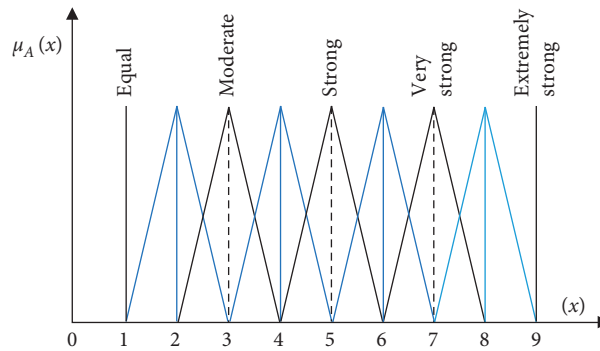


FIGURE 7: Linguistic scale based on triangular numbers.

TABLE 4: Comparison matrix based on linguistic scale.

Criteria	Security and privacy (C1)	Data (C2)	Technology (C3)	Legal (C4)	Cost (C5)
Security and privacy (C1)	1	5	3	3	1
Data (C2)	1/5	1	5	3	1
Technology (C3)	1/3	1/5	1	7	1/3
Legal (C4)	1/3	1/3	1/7	1	3
Cost (C5)	1	1	3	1/3	1

TABLE 5: Fuzzy-based pairwise comparison matrix.

Criteria	C1	C2	C3	C4	C5
C1	(1, 1, 1)	(3, 5, 7)	(1, 3, 5)	(1, 3, 5)	(1, 1, 1)
C2	(1/7, 1/5, 1/3)	(1, 1, 1)	(3, 5, 7)	(1, 3, 5)	(1, 1, 1)
C3	(1/5, 1/3, 1)	(1/7, 1/5, 1/3)	(1, 1, 1)	(5, 7, 9)	(1/3, 1/3, 1/3)
C4	(1/5, 1/3, 1)	(1/5, 1/3, 1)	(1/9, 1/7, 1/5)	(1, 1, 1)	(1, 3, 5)
C5	(1, 1, 1)	(1, 1, 1)	(3, 3, 3)	(1/5, 1/3, 1)	(1, 1, 1)



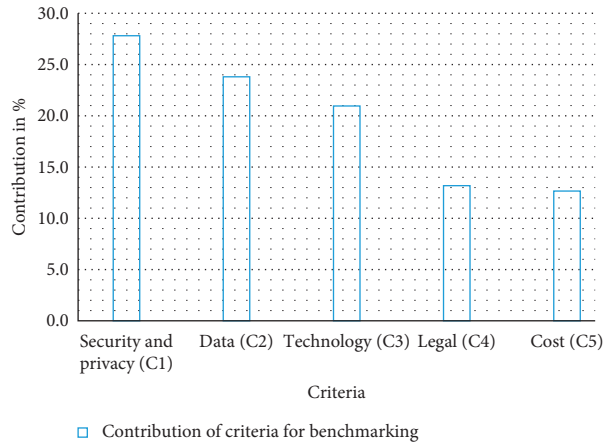


FIGURE 8: Contribution of the criteria for benchmarking Internet of medical things challenges.

TABLE 6: Local weights of the subcriteria.

Criteria	Subcriteria	Local weights of subcriteria
Security and privacy (C1)	Network security (S1)	0.2115
	Confidentiality (S2)	0.3786
	IoT security (S3)	0.2438
	Conflicts of interest (S4)	0.1661
Data (C2)	Vendor lock-in (S5)	0.2097
	Tremendous amount (S6)	0.2902
	Integration (S7)	0.2616
	Scalability (S8)	0.1746
	Standardization (S9)	0.0638
Technology (C3)	Device diversity (S10)	0.1718
	Modelling relationship (S11)	0.2405
	Real-time processing (S12)	0.3790
	System predictability (S13)	0.2088
Legal (C4)	Regulatory affairs (S14)	0.3769
	Data usage (S15)	0.3114
	Ownership (S16)	0.1895
	Liability (S17)	0.1221
Cost (C5)	Installation (S18)	0.4770
	Maintenance (S19)	0.3458
	Upgradation (S20)	0.1771

TABLE 7: Consistency index and consistency ratio of subcriteria.

Criteria	DM-1		DM-2		DM-3		DM-4		DM-5	
	CI	CR	CI	CR	CI	CR	CI	CR	CI	CR
Security and privacy	0.044	0.049	0.063	0.070	0.054	0.059	0.054	0.060	0.054	0.059
Data	0.089	0.079	0.034	0.031	0.054	0.048	0.078	0.070	0.056	0.050
Technology	0.067	0.074	0.054	0.060	0.051	0.057	0.044	0.049	0.054	0.059
Legal	0.088	0.098	0.066	0.074	0.023	0.025	0.031	0.034	0.017	0.019
Cost	0.002	0.004	0.038	0.066	0.029	0.050	0.038	0.066	0.029	0.050

the criteria are given in Table 8. The contribution of the global weights for benchmarking the Internet of medical things challenges is shown in Figure 9.

The criteria and weights were finalized, and these weights were used as weights of the TOPSIS decision matrix.

The robust features of fuzzy TOPSIS were used to evaluate and rank the alternatives.

*Step 14.* Formulate linguistic variables for the decision matrix. The authors developed the triangular fuzzy

TABLE 8: Global weights of the main criteria.

Criteria	Weights of the criteria	Subcriteria	Local weights of subcriteria	Global weights of the criteria
Security and privacy (C1)	0.282	Network security (S1)	0.2115	0.060
		Confidentiality (S2)	0.3786	0.107
		IoT security (S3)	0.2438	0.069
		Conflicts of interest (S4)	0.1661	0.047
Data (C2)	0.242	Vendor lock-in (S5)	0.2097	0.051
		Tremendous amount (S6)	0.2902	0.070
		Integration (S7)	0.2616	0.063
		Scalability (S8)	0.1746	0.042
		Standardization (S9)	0.0638	0.015
Technology (C3)	0.213	Device diversity (S10)	0.1718	0.037
		Modelling relationship (S11)	0.2405	0.051
		Real-time processing (S12)	0.3790	0.081
		System predictability (S13)	0.2088	0.044
Legal (C4)	0.34	Regulatory affairs (S14)	0.3769	0.128
		Data usage (S15)	0.3114	0.106
		Ownership (S16)	0.1895	0.064
		Liability (S17)	0.1221	0.042
Cost (C5)	0.129	Installation (S18)	0.4770	0.062
		Maintenance (S19)	0.3458	0.045
		Upgradation (S20)	0.1771	0.023

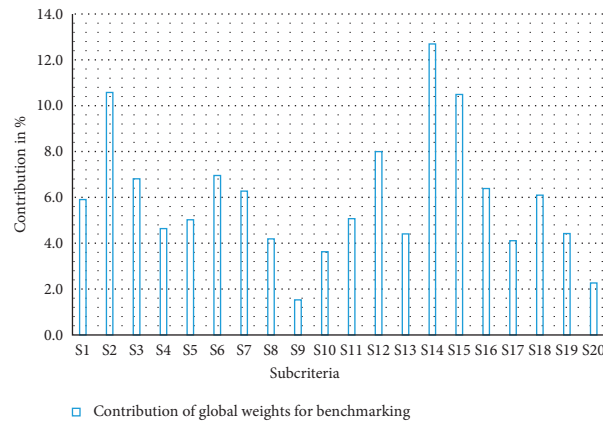


FIGURE 9: Contribution of the global weights for benchmarking Internet of medical things challenges.

number- (TFN-) based linguistic scale, as shown in Table 9.

*Step 15.* Establish a decision matrix in the light of the decision maker's rating. In this study, three decision makers participated who already gave their opinions in the fuzzy analytic hierarchy. For demonstration purposes, the authors selected 4 medical IoT challenges for the ranking/prioritization purpose. Equations (17) to (20) were used to aggregate the decisions. The aggregated decision matrix is shown in Table 2.

*Step 16.* Use equations (21) to (23) to normalize the decision matrix, as shown in Table 10. It is worthwhile and pertinent to add here that installation cost (S18), maintenance cost (S19), and upgradation cost (S20) are the nonbeneficial

TABLE 9: Linguistic variable ratings.

Linguistic variables	Assigned TFN
Very low	(1, 1, 3)
Low	(1, 3, 5)
Average	(3, 5, 7)
High	(5, 7, 9)
Very high	(7, 9, 9)

subcriteria and the remaining subcriteria are beneficial in said matrix.

*Step 17.* Compute weighted normalized matrix through equation (24). The output of  $(\tilde{v}_{ij})$  is shown in Table 11.

*Step 18.* Use equation (25) to compute  $(V_i^+)$  and equation (26) for  $(V_i^-)$ ; the results are shown in Table 12.

TABLE 10: Normalized fuzzy decision matrix.

Subcriteria alternative	S1	S2	S3	S4	S5
Alternative 1	(0.56, 0.78, 1)	(0.33, 0.78, 1)	(0.33, 0.56, 0.78)	(0.33, 0.70, 1)	(0.11, 0.63, 1)
Alternative 2	(0.11, 0.63, 1)	(0.11, 0.63, 1)	(0.56, 0.78, 1)	(0.11, 0.63, 1)	(0.11, 0.48, 0.78)
Alternative 3	(0.56, 0.86, 1)	(0.11, 0.70, 1)	(0.56, 0.86, 1)	(0.56, 0.78, 1)	(0.56, 0.78, 1)
Alternative 4	(0.56, 0.92, 1)	(0.33, 0.56, 0.78)	(0.11, 0.26, 0.56)	(0.11, 0.41, 0.78)	(0.11, 0.63, 1)
Subcriteria alternative	S6	S7	S8	S9	S10
Alternative 1	(0.11, 0.56, 1)	(0.33, 0.78, 1)	(0.33, 0.78, 1)	(0.33, 0.56, 0.78)	(0.33, 0.56, 0.78)
Alternative 2	(0.11, 0.33, 0.78)	(0.11, 0.63, 1)	(0.11, 0.63, 1)	(0.56, 0.78, 1)	(0.11, 0.63, 1)
Alternative 3	(0.11, 0.63, 1)	(0.33, 0.70, 1)	(0.56, 0.92, 1)	(0.56, 0.86, 1)	(0.56, 0.86, 1)
Alternative 4	(0.33, 0.63, 1)	(0.11, 0.56, 0.78)	(0.33, 0.56, 0.78)	(0.11, 0.26, 0.78)	(0.11, 0.33, 0.56)
Subcriteria alternative	S11	S12	S13	S14	S15
Alternative 1	(0.33, 0.86, 1)	(0.11, 0.63, 1)	(0.11, 0.56, 1)	(0.56, 0.78, 1)	(0.11, 0.56, 1)
Alternative 2	(0.11, 0.48, 1)	(0.11, 0.48, 1)	(0.11, 0.41, 1)	(0.11, 0.56, 1)	(0.11, 0.41, 1)
Alternative 3	(0.56, 0.78, 1)	(0.11, 0.70, 1)	(0.33, 0.70, 1)	(0.11, 0.63, 1)	(0.33, 0.70, 1)
Alternative 4	(0.11, 0.48, 0.78)	(0.33, 0.70, 1)	(0.11, 0.70, 1)	(0.11, 0.48, 0.78)	(0.11, 0.70, 1)
Subcriteria alternative	S16	S17	S18	S19	S20
Alternative 1	(0.33, 0.78, 1)	(0.33, 0.63, 1)	(0.11, 0.16, 0.33)	(0.11, 0.16, 0.33)	(0.11, 0.18, 0.33)
Alternative 2	(0.11, 0.70, 1)	(0.56, 0.78, 1)	(0.11, 0.20, 1)	(0.11, 0.16, 1)	(0.11, 0.14, 0.20)
Alternative 3	(0.33, 0.78, 1)	(0.56, 0.92, 1)	(0.33, 0.14, 0.20)	(0.11, 0.13, 0.20)	(0.11, 0.12, 0.20)
Alternative 4	(0.11, 0.48, 0.78)	(0.11, 0.41, 0.78)	(0.14, 0.33, 1.00)	(0.14, 0.23, 1)	(0.14, 0.27, 1)

TABLE 11: Weighted normalized fuzzy decision matrix.

Subcriteria alternative	S1	S2	S3	S4	S5
$W_j$	0.060	0.107	0.069	0.047	0.051
Alternative 1	(0.03, 0.05, 0.06)	(0.04, 0.08, 0.11)	(0.02, 0.04, 0.05)	(0.02, 0.03, 0.05)	(0.01, 0.05, 0.06)
Alternative 2	(0.01, 0.04, 0.06)	(0.01, 0.07, 0.11)	(0.04, 0.05, 0.07)	(0.01, 0.03, 0.05)	(0.01, 0.04, 0.06)
Alternative 3	(0.03, 0.05, 0.06)	(0.01, 0.07, 0.11)	(0.04, 0.06, 0.07)	(0.03, 0.04, 0.05)	(0.01, 0.04, 0.06)
Alternative 4	(0.03, 0.05, 0.06)	(0.04, 0.06, 0.08)	(0.01, 0.02, 0.04)	(0.01, 0.02, 0.04)	(0.02, 0.04, 0.06)
Subcriteria alternative	S6	S7	S8	S9	S10
$W_j$	0.070	0.063	0.042	0.015	0.037
Alternative 1	(0.01, 0.04, 0.07)	(0.02, 0.05, 0.06)	(0.02, 0.05, 0.06)	(0.02, 0.04, 0.05)	(0.02, 0.04, 0.05)
Alternative 2	(0.01, 0.02, 0.05)	(0.01, 0.04, 0.06)	(0.01, 0.04, 0.06)	(0.04, 0.05, 0.06)	(0.01, 0.04, 0.06)
Alternative 3	(0.01, 0.04, 0.07)	(0.02, 0.04, 0.06)	(0.04, 0.06, 0.06)	(0.04, 0.05, 0.06)	(0.04, 0.05, 0.06)
Alternative 4	(0.02, 0.04, 0.07)	(0.01, 0.04, 0.06)	(0.02, 0.04, 0.05)	(0.01, 0.02, 0.05)	(0.01, 0.02, 0.04)
Subcriteria alternative	S11	S12	S13	S14	S15
$W_j$	0.051	0.81	0.044	0.128	0.106
Alternative 1	(0.02, 0.05, 0.06)	(0.01, 0.04, 0.06)	(0.01, 0.04, 0.06)	(0.04, 0.05, 0.06)	(0.01, 0.04, 0.06)
Alternative 2	(0.01, 0.03, 0.06)	(0.01, 0.03, 0.06)	(0.01, 0.03, 0.06)	(0.01, 0.04, 0.06)	(0.01, 0.03, 0.06)
Alternative 3	(0.04, 0.05, 0.06)	(0.01, 0.04, 0.06)	(0.02, 0.04, 0.06)	(0.01, 0.04, 0.06)	(0.02, 0.04, 0.06)
Alternative 4	(0.01, 0.03, 0.05)	(0.02, 0.04, 0.06)	(0.01, 0.04, 0.06)	(0.01, 0.03, 0.05)	(0.01, 0.04, 0.06)
Subcriteria alternative	S16	S17	S18	S19	S20
$W_j$	0.064	0.042	0.062	0.045	0.023
Alternative 1	(0.02, 0.05, 0.06)	(0.02, 0.04, 0.06)	(0.01, 0.01, 0.02)	(0.01, 0.01, 0.02)	(0.01, 0.01, 0.02)
Alternative 2	(0.01, 0.04, 0.06)	(0.04, 0.05, 0.06)	(0.01, 0.01, 0.06)	(0.01, 0.01, 0.06)	(0.01, 0.01, 0.01)
Alternative 3	(0.02, 0.05, 0.06)	(0.04, 0.06, 0.06)	(0.01, 0.01, 0.01)	(0.01, 0.01, 0.01)	(0.01, 0.01, 0.01)
Alternative 4	(0.01, 0.03, 0.05)	(0.01, 0.03, 0.05)	(0.01, 0.02, 0.06)	(0.01, 0.01, 0.06)	(0.01, 0.02, 0.06)

Step 19. Use equation (27) to calculate the distance of every alternative from a positive ideal solution ( $d_i^+$ ) and negative ideal solution ( $d_i^-$ ) by using equation (28). The value to relative proximity ( $P_i$ ) is computed through equation (29), as shown in Table 13.

The fuzzy analytical hierarchy process and TOPSIS results indicate that security and privacy with a score of 0.282 (28.2%) and data with a score of 0.242 (24.2%) are the most influential factors in the decision-making process of the

IoMT challenges. The authors proposed 20 subcriteria of the five criteria. If we consider the local weights of the criteria and subcriteria given in Table 6 and also consider global weights of the subcriteria shown in Table 8, respectively, the first criterion, confidentiality, has the most weightage with the score of (0.3786, 0.107), followed by IoT security (0.2438, 0.069), network security (0.2115, 0.060), and conflicts of interest (0.1661, 0.047). The second criterion titled data has five subcriteria, and among these criteria, tremendous

TABLE 12: The fuzzy positive ideal solution and fuzzy negative ideal solution.

Subcriteria	S1	S2	S3	S4	S5
$(V_i^+)$	(0.03, 0.05, 0.06)	(0.04, 0.08, 0.11)	(0.04, 0.06, 0.07)	(0.03, 0.04, 0.05)	(0.03, 0.04, 0.05)
$(V_i^-)$	(0.01, 0.04, 0.06)	(0.01, 0.0, 0.06)	(0.01, 0.0, 0.04)	(0.01, 0.02, 0.04)	(0.01, 0.02, 0.05)
Subcriteria	S6	S7	S8	S9	S10
$(V_i^+)$	(0.02, 0.04, 0.07)	(0.02, 0.05, 0.06)	(0.04, 0.06, 0.06)	(0.04, 0.05, 0.06)	(0.04, 0.05, 0.06)
$(V_i^-)$	(0.01, 0.02, 0.05)	(0.01, 0.04, 0.06)	(0.01, 0.04, 0.05)	(0.01, 0.02, 0.05)	(0.01, 0.02, 0.04)
Subcriteria	S11	S12	S13	S14	S15
$(V_i^+)$	(0.04, 0.05, 0.06)	(0.02, 0.04, 0.06)	(0.02, 0.04, 0.06)	(0.04, 0.05, 0.06)	(0.02, 0.04, 0.06)
$(V_i^-)$	(0.01, 0.03, 0.05)	(0.01, 0.03, 0.06)	(0.01, 0.03, 0.06)	(0.01, 0.03, 0.05)	(0.01, 0.03, 0.06)
Subcriteria	S16	S17	S18	S19	S20
$(V_i^+)$	(0.02, 0.05, 0.06)	(0.04, 0.06, 0.06)	(0.01, 0.02, 0.06)	(0.01, 0.01, 0.06)	(0.01, 0.02, 0.06)
$(V_i^-)$	(0.01, 0.03, 0.05)	(0.01, 0.03, 0.05)	(0.01, 0.01, 0.01)	(0.01, 0.01, 0.01)	(0.01, 0.01, 0.01)

TABLE 13: Relative proximity and a final rank.

Alternative	$d_i^+$	$d_i^-$	$P_i$	Rank
Alternative 1	0.217	0.240	0.5249	2
Alternative 2	0.263	0.223	0.4591	3
Alternative 3	0.146	0.314	0.6825	1
Alternative 4	0.274	0.184	0.4021	4

TABLE 14: Comparison of fuzzy AHP and TOPSIS with other methodologies for IoMT challenges.

Criteria	[34]	[60]	[61]	[63]	[62]	Fuzzy AHP and TOPSIS technique of this article
Fuzzy logic	No	Yes	No	No	No	Yes
Pairwise comparison	No	No	Yes	Yes	No	Yes
Weighting of criteria	No	No	No	Yes	No	Yes
Complexity	Moderate	Moderate	Low	Moderate	Low	High
Consistency	No	No	Yes	No	No	Yes
Independence	No	Yes	Yes	Yes	Yes	Yes
Computational requirement	High	High	Low	Moderate	Low	High
Probability and possibility	No	No	Yes	No	No	Yes

amount (0.2902, 0.070) is the most influential factor in this category, followed by integration (0.2616, 0.063), vendor lock-in (0.2097, 0.051), scalability (0.1746, 0.042), and standardization (0.0638, 0.015). The most interesting criterion that every researcher used in their studies is technology, and it also has four subcriteria. The real-time processing subcriterion has taken the lead by scoring (0.3790, 0.081), followed by modelling relationship (0.2405, 0.051), system predictability (0.2088, 0.044), and device diversity (0.1718, 0.037). Legal is the fourth criterion that is also used by many authors as described in the introduction section; regulatory affairs has the highest score (0.3769, 0.128), followed by data usage (0.3114, 0.106), ownership (0.1895, 0.064), and liability (0.1221, 0.042). The last proposed criterion is cost, and it has only 3 subcriteria. Installation has greater score (0.4770, 0.062) among other scores ((0.3458, 0.045) and (0.1771, 0.023)).

In regard to subcriteria, the two most influential criteria are regulatory affairs, which has 0.128 (12.8%) value, and confidentiality, which has 0.107 (10.7%) value. In the light of final scores, we can say that Alternative 3 is the most challenging factor for the IoMT followed by Alternative 1; on

the other hand, Alternative 4 demonstrates the least challenge in the light of the expert decision.

## 6. Comparison of Fuzzy AHP and TOPSIS Techniques with Other Existing Techniques

The authors of this article studied and examined many criteria developed by different authors to rank the challenges of IoT, but none of the authors ranked the IoMT challenges by using hybrid techniques known as fuzzy logic and AHP TOPSIS. The authors who used multicriteria decision-making tools to evaluate the IoT challenges neither used the features of fuzzy logic nor used core functions of AHP and TOPSIS. Many authors used a general criterion to evaluate IoT challenges and still could not find proper criteria. In this study, the authors developed 5 criteria after studying highly cited papers on IoMT challenges and also developed 20 subcriteria after studying high impact factor journal papers.

The authors also critically studied each criterion and subcriteria that cover all the tasks and meet the characteristics of a good methodology. After a thorough examination

of the task that covers all approaches, we formulated a single methodology that covers all requirements.

The authors formulated simple criteria to evaluate the Internet of medical things challenges, and the same criteria have been implemented by using the proposed methodology in Section 5 just to show that the proposed methodology is producing better results as compared to other methodologies that were used in the past for the IoT challenges. The comparison of the IoT methodologies with the proposed methodology is given in Table 14.

Table 14 depicts many existing methodologies opted by different researchers to compare and prioritize IoT and select the most critical challenge. We compared the existing techniques to prove that the fuzzy AHP and TOPSIS technique is the most efficient technique to evaluate and prioritize the Internet of medical things challenges as it confronts ambiguity. This permits the formulation of criteria, associates subcriteria, establishes a pairwise comparison, normalizes the decision matrix, calculates local and global weights, and validates the results using the consistency index and consistency ratio.

## 7. Conclusions

Nowadays, smart objects communicate and interact with each other and play a significant role in human life. Industry stakeholders think that more usage of the Internet and interaction between smart objects will increase the number of opportunities and increase the level of competition. Healthcare enterprises that have the aim to introduce the Internet of things should know the current challenges, the significance of these challenges, and the methods to encounter these challenges with less effort, cost, and hardship.

Internet of medical things challenge selection decision became an essential operational and technical decision in a complex network environment. In this article, for the first time, a hybrid fuzzy multicriteria decision-making approach based on fuzzy logic, AHP, and TOPSIS is proposed for dealing with the Internet of medical things. Using fuzzy set theory with AHP to obtain the weights of the criteria of TOPSIS can minimize the ambiguities and doubts that are still roadblock in decision making about IoT challenges, especially for healthcare. We proposed a triangular fuzzy number-based methodology with AHP and computed weights of the criteria, local weights of the subcriteria, and global weight values used for TOPSIS. We also used linguistic variables and expert's opinion for pairwise comparison and decision matrix which made the final decision-making process easy, reliable, and realistic. The proposed methodology consists of four phases; each phase is independent and transforms its output to next step. In this study, small and medium-sized enterprises located in Pakistan were explored, focusing on challenges related to the Internet of medical things. The Internet of medical things challenges were identified as criteria, and experts of this field evaluated these challenges. The study comprises five criteria: security and privacy, data, technology, legal, and cost, which affected the challenges related to the Internet of medical things. There were twenty subcriteria and four alternatives. The significance of the criteria was computed by

using the fuzzy analytic hierarchy process and fuzzy TOPSIS methods. In light of the results, it was observed that the industry which was going to adopt the Internet of medical things should pay attention to security and privacy, data, technology, legal, and cost. When the global weights of the criteria were calculated through the fuzzy analytic hierarchy process as weights of the criteria of TOPSIS, the top four criteria are regulatory affairs with 12.8%, confidentiality with 10.7%, and data usage with 10.6%, followed by real-time processing with 8.1%.

For the further extension of this work, we will consider other decision-making methods for the selection of Internet of medical things challenges that can be employed. The comparison of current and previous studies is suggested. Furthermore, we will also consider other weight calculation methods used with TOPSIS like the entropy method, and least square programming methods can be applied. Moreover, different scenarios and criteria can be considered for future work.

## Data Availability

The Expert's opinion 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 article.

## Acknowledgments

This study was supported in part by the REMIND Project (The use of computational techniques to improve compliance to reminders within smart environments) from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 734355.

## References

- [1] P. Sethi and S. R. Sarangi, "Internet of Things: architectures, protocols, and applications," *Journal of Electrical and Computer Engineering*, vol. 2017, Article ID 9324035, 25 pages, 2017.
- [2] A. Abdelgawad and K. Yelamarthi, "Internet of Things (IoT) platform for structure health monitoring," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 6560797, 10 pages, 2017.
- [3] S. Jabbar, F. Ullah, S. Khalid, M. Khan, and K. Han, "Semantic interoperability in heterogeneous IoT infrastructure for healthcare," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9731806, 10 pages, 2017.
- [4] E. F. Jesus, V. R. Chicarino, C. V. de Albuquerque, and A. A. d. A. Rocha, "A survey of how to use blockchain to secure Internet of Things and the stalker attack," *Security and Communication Networks*, vol. 2018, Article ID 9675050, 27 pages, 2018.

- [5] P. P. Ray, "A survey on visual programming languages in Internet of Things," *Scientific Programming*, vol. 2017, Article ID 1231430, 6 pages, 2017.
- [6] S. Hameed, F. I. Khan, and B. Hameed, "Understanding security requirements and challenges in Internet of Things (IoT): a review," *Journal of Computer Networks and Communications*, vol. 2019, Article ID 9629381, 14 pages, 2019.
- [7] J.-X. Hu, C.-L. Chen, C.-L. Fan, and K. Wang, "An intelligent and secure health monitoring scheme using IoT sensor based on cloud computing," *Journal of Sensors*, vol. 2017, Article ID 3734764, 11 pages, 2017.
- [8] S. Nazir, Y. Ali, N. Ullah, and I. García-Magariño, "Internet of Things for healthcare using effects of mobile computing: a systematic literature review," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 5931315, 20 pages, 2019.
- [9] B. Risteska Stojkoska, K. Trivodaliev, and D. Davcev, "Internet of Things framework for home care systems," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 8323646, 10 pages, 2017.
- [10] S. M. Riazul Islam, D. Kwak, M. H. 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.
- [11] Y. Yuehong, Y. Zeng, X. Chen, and Y. Fan, "The Internet of Things in healthcare: an overview," *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, 2016.
- [12] M. A. Ferrag, L. A. Maglaras, H. Janicke, J. Jiang, and L. Shu, "Authentication protocols for Internet of Things: a comprehensive survey," *Security and Communication Networks*, vol. 2017, Article ID 6562953, 41 pages, 2017.
- [13] E. Triantaphyllou, "Multi-criteria decision making methods," in *Multi-Criteria Decision Making Methods: A Comparative Study*, Springer, Berlin, Germany, 2000.
- [14] E. K. Zavadskas, Z. Turskis, and S. Kildienė, "State of art surveys of overviews on MCDM/MADM methods," *Technological and Economic Development of Economy*, vol. 20, no. 1, pp. 165–179, 2014.
- [15] K. Kimani, V. Oduol, and K. Langat, "Cyber security challenges for IoT-based smart grid networks," *International Journal of Critical Infrastructure Protection*, vol. 25, pp. 36–49, 2019.
- [16] P. Bhojar, P. Sahare, S. B. Dhok, and R. B. Deshmukh, "Communication technologies and security challenges for Internet of Things: a comprehensive review," *AEU - International Journal of Electronics and Communications*, vol. 99, pp. 81–99, 2019.
- [17] J. Dizdarević, F. Carpio, A. Jukan, and X. Masip-Bruin, "A survey of communication protocols for Internet of Things and related challenges of fog and cloud computing integration," *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–29, 2019.
- [18] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, "Internet of Things security and forensics: challenges and opportunities," *Future Generation Computer Systems*, vol. 78, pp. 544–546, 2018.
- [19] B. Farahani, F. Firouzi, V. Chang, M. Badaroglu, N. Constant, and K. Mankodiya, "Towards fog-driven IoT eHealth: promises and challenges of IoT in medicine and healthcare," *Future Generation Computer Systems*, vol. 78, pp. 659–676, 2018.
- [20] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of Internet of Things for smart home: challenges and solutions," *Journal of Cleaner Production*, vol. 140, pp. 1454–1464, 2017.
- [21] S. A. Butt, M. I. Tariq, T. Jamal, A. Ali, J. L. Diaz Martinez, and E. De-La-Hoz-Franco, "Predictive variables for agile development merging cloud computing services," *IEEE Access*, vol. 7, pp. 99273–99282, 2019.
- [22] M. I. Tariq, S. Tayyaba, M. W. Ashraf, and V. E. Balas, "Deep learning techniques for optimizing medical big data," in *Deep Learning Techniques for Biomedical and Health Informatics*, pp. 187–211, Elsevier, Amsterdam, Netherlands, 2020.
- [23] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT security techniques based on machine learning: how do IoT devices use AI to enhance security?" *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 41–49, 2018.
- [24] K.-H. Yeh, "A secure IoT-based healthcare system with body sensor networks," *IEEE Access*, vol. 4, pp. 10288–10299, 2016.
- [25] A. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors*, vol. 19, no. 2, p. 326, 2019.
- [26] E. Al Alkeem, C. Y. Yeun, and M. J. Zemerly, "Security and privacy framework for ubiquitous healthcare IoT devices," in *Proceedings of the 2015 10th International Conference for Internet Technology and Secured Transactions, ICITST*, pp. 70–75, London, UK, 2015.
- [27] M. I. Tariq, "Agent based information security framework for hybrid cloud computing," *KSII Transactions on Internet & Information Systems*, vol. 13, no. 1, pp. 406–434, 2019.
- [28] M. I. Tariq, S. Tayyaba, M. U. Hashmi, M. W. Ashraf, and N. A. Mian, "Agent based information security threat management framework for hybrid cloud computing," *IJCSNS*, vol. 17, no. 12, pp. 57–66, 2017.
- [29] M. I. Tariq, S. Tayyaba, M. W. Ashraf, H. Rasheed, and F. Khan, *Risk Based NIST Effectiveness Analysis for Cloud Security*, Bahria University Journal of Information & Communication Technology, Islamabad, Pakistan, 2017.
- [30] M. I. Tariq, S. Tayyaba, H. Rasheed, and M. W. Ashraf, "Factors influencing the cloud computing adoption in higher education institutions of Punjab, Pakistan," in *Proceedings of the 2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*, pp. 179–184, Pakistan, 2017.
- [31] I. Mashal, O. Alsaryrah, T.-Y. Chung, and F.-C. Yuan, "A multi-criteria analysis for an Internet of Things application recommendation system," *Technology in Society*, vol. 60, p. 101216, 2020.
- [32] A. Iqbal, F. Ullah, H. Anwar et al., "Interoperable Internet-of-Things platform for smart home system using Web-of-objects and cloud," *Sustainable Cities and Society*, vol. 38, pp. 636–646, 2018.
- [33] A. K. Ullah, S. Ghafoori, A. Mohammadian, R. Mohammadkazemi, B. Mahbanooei, and R. Ghasemi, "A Fuzzy Analytic Network Process (FANP) approach for prioritizing Internet of Things challenges in Iran," *Technology in Society*, vol. 53, pp. 124–134, 2018.
- [34] B. Uslu, T. Eren, Ş. Gür, and E. Özcan, "Evaluation of the difficulties in the Internet of Things (IoT) with multi-criteria decision-making," *Processes*, vol. 7, no. 3, p. 164, 2019.
- [35] I. Mashal and O. Alsaryrah, "Fuzzy analytic hierarchy process model for multi-criteria analysis of Internet of Things," *Kybernetes*, 2019.
- [36] Y.-S. Kao, K. Nawata, and C.-Y. Huang, "Evaluating the performance of systemic innovation problems of the IoT in manufacturing industries by novel MCDM methods," *Sustainability*, vol. 11, no. 18, p. 4970, 2019.
- [37] Y. Liu, Y. Yang, Y. Liu, and G.-H. Tzeng, "Improving sustainable mobile health care promotion: a novel hybrid MCDM method," *Sustainability*, vol. 11, no. 3, p. 752, 2019.

- [38] M. Abdel-Basset, G. Manogaran, A. Gamal, and V. Chang, "A novel intelligent medical decision support model based on soft computing and IoT," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4160–4170, 2020.
- [39] D. Shin and Y. Hwang, "Integrated acceptance and sustainability evaluation of internet of medical things," *Internet Research*, vol. 27, no. 5, pp. 1227–1254, 2017.
- [40] F. Alsubaei, A. Abuhussein, and S. Shiva, "Security and privacy in the internet of medical things: taxonomy and risk assessment," in *Proceedings of the 2017 IEEE 42nd Conference on Local Computer Networks Workshops (LCN Workshops)*, pp. 112–120, Singapore, 2017.
- [41] Y. Kondratenko, G. Kondratenko, and I. Sidenko, "Multi-criteria decision making for selecting a rational IoT platform," in *Proceedings of the 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, pp. 147–152, Ukraine, Kyiv, May 2018.
- [42] H. Farman, B. Jan, H. Javed et al., "Multi-criteria based zone head selection in Internet of Things based wireless sensor networks," *Future Generation Computer Systems*, vol. 87, pp. 364–371, 2018.
- [43] M. I. Tariq, S. Tayyaba, N. A. Mian et al., "Combination of AHP and TOPSIS methods for the ranking of information security controls to overcome its obstructions under fuzzy environment," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 6075–6088, 2020.
- [44] T. L. Saaty, "What is the analytic hierarchy process?" in *Mathematical Models for Decision Support*, Springer, Berlin, Germany, 1988.
- [45] A. N. Patil, "FUZZY AHP Methodology and its sole applications," *International Journal of Management Research and Reviews*, vol. 8, no. 5, pp. 24–32, 2018.
- [46] M. I. Tariq, S. Ahmed, N. A. Memon et al., "Prioritization of information security controls through fuzzy AHP for cloud computing networks and wireless sensor networks," *Sensors*, vol. 20, no. 5, p. 1310, 2020.
- [47] A. Leśniak, D. Kubek, E. Plebankiewicz, K. Zima, and S. Belniak, "Fuzzy AHP application for supporting contractors' bidding decision," *Symmetry*, vol. 10, no. 11, p. 642, 2018.
- [48] C.-C. Yang and B.-S. Chen, "Key quality performance evaluation using fuzzy AHP," *Journal of the Chinese Institute of Industrial Engineers*, vol. 21, no. 6, pp. 543–550, 2004.
- [49] J. J. Buckley and V. Uppuluri, "Fuzzy hierarchical analysis," in *Uncertainty in Risk Assessment, Risk Management, and Decision Making*, pp. 389–401, Springer, Berlin, Germany, 1987.
- [50] K.-J. Zhu, Y. Jing, and D.-Y. Chang, "A discussion on extent analysis method and applications of fuzzy AHP," *European Journal of Operational Research*, vol. 116, no. 2, pp. 450–456, 1999.
- [51] C. L. Hwang and K. Yoon, "Methods for multiple attribute decision making," in *Multiple Attribute Decision Making*, pp. 58–191, Springer, Berlin, Germany, 1981.
- [52] M. Behzadian, S. Khanmohammadi Otahsara, M. Yazdani, and J. Ignatius, "A state-of-the-art survey of TOPSIS applications," *Expert Systems with Applications*, vol. 39, no. 17, pp. 13051–13069, 2012.
- [53] C.-T. Chen, "Extensions of the TOPSIS for group decision-making under fuzzy environment," *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1–9, 2000.
- [54] X. Guo, T. Zeng, Y. Wang, and J. Zhang, "Fuzzy TOPSIS approaches for assessing the intelligence level of IoT-based tourist attractions," *IEEE Access*, vol. 7, pp. 1195–1207, 2018.
- [55] N. Gupta, V. Sharma, and M. Kashyap, "A critical analysis of sensor based IoT architectures using fuzzy TOPSIS," in *Proceedings of the 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 21–27, Greater Noida, India, 2018.
- [56] Y.-C. Chou, H.-Y. Yen, V. T. Dang, and C.-C. Sun, "Assessing the human resource in science and technology for asian countries: application of fuzzy AHP and fuzzy TOPSIS," *Symmetry*, vol. 11, no. 2, p. 251, 2019.
- [57] M. Diouf and C. Kwak, "Fuzzy AHP, DEA, and managerial analysis for supplier selection and development; from the perspective of open innovation," *Sustainability*, vol. 10, no. 10, p. 3779, 2018.
- [58] H. A. Donegan and F. J. Dodd, "A note on saaty's random indexes," *Mathematical and Computer Modelling*, vol. 15, no. 10, pp. 135–137, 1991.
- [59] T. L. Saaty, "The analytic hierarchy process," McGraw-Hill, New York, NY, USA, 1980.
- [60] T. D. C. Frazão, D. G. G. Camilo, E. L. S. Cabral, and R. P. Souza, "Multicriteria decision analysis (MCDA) in health care: a systematic review of the main characteristics and methodological steps," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, p. 90, 2018.
- [61] P. T. M. Ly, W.-H. Lai, C.-W. Hsu, and F.-Y. Shih, "Fuzzy AHP analysis of internet of things (IoT) in enterprises," *Technological Forecasting and Social Change*, vol. 136, pp. 1–13, 2018.
- [62] D. Peng and Y. Ruan, "AHP-based QoS evaluation model in the internet of things," in *Proceedings of the 2012 13th International Conference on Parallel and Distributed Computing, Applications and Technologies*, pp. 578–581, USA, 2012.
- [63] R. A. Paramita and M. Dachyar, "The alternative selection for internet of things (IoT) implementation in medical rehabilitation," *International Journal of Advanced Science and Technology*, vol. 29, no. 7s, pp. 3632–3640, 2020.

## Research Article

# Personal Communication Technologies for Smart Spaces Density-Based Clustering for Content and Color Adaptive Tone Mapping

Maleeha Javed,<sup>1</sup> Hassan Dawood ,<sup>1</sup> Muhammad Murtaza Khan,<sup>2,3</sup> Ameen Banjar,<sup>4</sup>  
Riad Alharbey,<sup>4</sup> and Hussain Dawood <sup>5</sup>

<sup>1</sup>Department of Software Engineering, University of Engineering and Technology, Taxila, Pakistan

<sup>2</sup>Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 21589, Saudi Arabia

<sup>3</sup>School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

<sup>4</sup>Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Jeddah 21589, Saudi Arabia

<sup>5</sup>Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 21859, Saudi Arabia

Correspondence should be addressed to Hassan Dawood; [hasandawod@yahoo.com](mailto:hasandawod@yahoo.com)

Received 11 March 2020; Revised 25 June 2020; Accepted 31 July 2020; Published 17 August 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Maleeha Javed 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.

Tone mapping operators are designed to display high dynamic range (HDR) images on low dynamic range devices. Clustering-based content and color adaptive tone mapping algorithm aims to maintain the color information and local texture. However, fine details can still be lost in low dynamic range images. This paper presents an effective way of clustering-based content and color adaptive tone mapping algorithm by using fast search and find of density peak clustering. The suggested clustering method reduces the loss of local structure and allows better adaptation of color in images. The experiments are carried out to evaluate the effectiveness and performance of proposed technique with state-of-the-art clustering techniques. The objective and subjective evaluation results reveal that fast search and find of density peak preserves more textural information. Therefore, it is most suitable to be used for clustering-based content and color adaptive tone mapping algorithm.

## 1. Introduction

The demand for high dynamic range (HDR) images is rapidly increasing with the advent of sensors, display technology, and availability of multiple exposure photography. HDR images provide better visual quality as compared to their counterpart low dynamic range (LDR) images on account of their large ranged brightness levels and better preservation of color variations. This enables to store and visualize bright and dark objects in the same image which would normally require multiple images at different exposures to be in low dynamic range. Visualization of HDR

images on standard display devices (made for displaying LDR images) is challenging as they are incapable of displaying the high dynamic range of the image. To address this issue, tone mapping operators are used to convert HDR images to LDR images. However, the technology that directly displays HDR images is becoming more cost-effective. Recently, most of the available display devices have LDR, which is why HDR to LDR conversion still has practical applications.

For visualization of HDR images on LDR devices, Retinex theory has been widely used, which slices the image into its components: base layer and detail layer. Edge



preservation filtering techniques are used for the approximation of base layer in Retinex image decomposition. However, these methods have resulted in false coloring and halo artifacts because filtering is unable to capture complex local structure of the image. To overcome this problem, we need a method that can better adopt the local structure and color variation of HDR images in LDR images.

Furthermore, tone mapping operators are generally divided into two main categories: global tone mapping operators and local tone mapping operators. Global tone mapping operators use the same monotonic curve for dynamic range compression of the entire image [1, 2]. Initially, the focus of the research was on the design of global tone mapping operators. Drago et al. [3] proposed a tone mapping algorithm that used an adaptive logarithmic base for luminance compression while maintaining the image's details and contrast. Kim and Kautz [4] introduced a tone mapping operator based on the hypothesis that the human visual system adopts Gaussian distribution. Reinhard et al. [5] proposed the use of spontaneous dodging and burning for dynamic range compression. The terms "dodging" and "burning" originate from printing, where holding backlight from a region of print is called dodging and the addition of light to image is referred to as burning. This can be used to increase or decrease the brightness of a captured image effectively. Reinhard and Devlin [6] introduced the light adaptive tone mapping which provides satisfying visual effects. Khan et al. [7] introduced a lookup table-based approach, where they have utilized the histogram of luminance for tone mapping. To produce the low dynamic range image appropriate for various viewing conditions, Han et al. [8] modify Khan et al.'s [7] method by considering the impact of ambient light on HVS. Generally, global tone mapping operators are computationally efficient [9]. However, they are unable to preserve the local characteristics and dynamic range variations as they ignore local pixel intensity variations in an image [9, 10].

Local tone mapping operators solve the issue of global tone mapping operators by incorporating the ratio between neighboring pixels and compressing each pixel. Gu et al. [11] and Min et al. [12] proposed local tone mapping operators based on layer decomposition, where the input image is divided into base layer and detail layer. The layer-decomposition-based approach has also been utilized by authors in [13–16]. Durand and Dorsey [17] proposed a method that preserves the detail layer by encoding large-scale variations of the base layer. Debevec and Gibson [18] proposed a method that helped to preserve brightness and details of image by applying a local luminance adaption function for compression of dynamic range, followed by reinjecting details in the low dynamic range image. Image color appearance model (iCAM06) [19] employed tone mapping by using iCAM06 color appearance model by considering the viewing conditions to generate optimal results. However, iCAM06 [19] led to the poor visual quality due to introduction of halo artifact, color saturation, and gradient reversal at the edges. To eliminate these problems, the authors in [20] proposed an iCAM06 based model. In their proposed algorithm guided filter [20], color adaptive transformation

matrix and HPE primitives were altered to enhance the effectiveness of model.

Krawczyk et al. [21] introduced an algorithm based on anchoring theory that decomposed image luminance into patches and then calculated the lightness for each patch. Li et al. [22] proposed a symmetrical analysis-synthesis filter for reducing the intensity range. Jia and Zhang [23] used the guided image filter for tone mapping. Meylan et al. [24] suggested a method based on the characteristic of human retina to reduce dynamic range while increasing the local contrast of the image. Another retina inspired range compression algorithm is used [25] to overcome the halo artifacts, loss of details, different visualization across different displays, and color saturation.

Parraga and Otazu [26] developed a tone mapping method based on human color perception by dividing the image intensity into multiresolution contrast. Also, they had used a nonlinear saturation model for dynamic range reduction based on visual cortex behavior. Chen et al. [27] introduced Earth mover's distance to segment HDR image and then applied local tone mapping on each component of the image. This maintained local texture and balanced perceptual impression. To remove artifacts and contrast enhancement, Liang et al. [28] used nonlinear diffusion.  $l_2$ -based retinex model [29] was used for contrast enhancement. Ahn et al. [30] proposed a Retinex-based adaptive local tone mapping algorithm according to which the use of guided filter reduced halo artifacts. Recently, Liang et al. [31] proposed a hybrid  $l_1 - l_0$  norm-based layer decomposition model.  $l_1$  sparsity term was imposed on the base layer and  $l_0$  on the detail layer. In [32], the authors proposed the use of decomposed multiscale Retinex for information preservation in tone-mapped image. Shu and Wu [33] employed an optimal local tone mapping operator to avoid halo artifacts and double edges. Rana et al. [34] proposed a pixel-wise adaptive tone mapping operator based on support vector regression to overcome the problem of drastic illumination variation.

El Mezeni and Saranovac [35] introduced an enhanced local tone mapping operator (ELTM). ELTM decomposed an image into detail and base layers, where the base layer was compressed into both logarithmic and linear domain. Local tone mapping operator presented in [36] was operationally similar to HVS. Li and Zheng [37] presented a saliency-aware local tone mapping operator that preserved edges using guided filters. Ferradans et al. [38] proposed a two-stage algorithm incorporating both global tone mapping and local tone mapping. In the first step, they applied a global tone mapping operator based on the human visual system and followed by local method for contrast enhancement in the second stage. To make the local and global tone mapping algorithm more effective, Ambalathankandy et al. [39] used histogram equalization method implemented on FPGA with efficient resource usage. To resolve the overenhancement, a nonuniform quantization technique is proposed for CT image enhancement [40]. Recently, deep convolution neural network is also used for range compression [41]. The authors in [41] used the output of existing tone mapping operators as training set and therefore inherited the best properties of all

tone mapping operators. Hence, they performed robustly well for all testing images compared to current methods of tone mapping.

Li et al. [42] introduced a clustering-based content and color adaptive tone mapping algorithm that preserved the local structure and naturalness of image. They used  $K$ -means algorithm for the clustering of color structure. However,  $K$ -means is difficult to adopt for a specific problem due to prespecified number of clusters. In addition, the utilization of  $K$ -means for tone mapping has led to the loss of complex local detail in LDR images. Furthermore, the number of available clustering algorithms encourages the use of the most effective algorithm for clustering-based tone mapping algorithm. Therefore, to preserve the complex local details, we have adopted a better clustering technique.

In this paper, we have proposed a clustering-based tone mapping algorithm to overcome the problems of false coloring, halo artifacts, and loss of complex local structure. For that purpose, we have utilized the fast search and find of density peak for content and color adaptive tone mapping. This clustering technique automatically recognizes clusters regardless of dimensionality of data. The following study compares various available state-of-the-art clustering algorithms for tone mapping operation. The effectiveness of proposed technique is compared with different clustering algorithms through subjective and objective evaluation techniques. The experimental analysis suggests that the fast search and find of density peak provides visually appealing result without compromising the local structure of an image.

The rest of the paper is composed of the following sections. The second section provides a brief introduction to the clustering-based content and color adaptive tone mapping algorithm. The third section describes different applied clustering techniques. Furthermore, the fourth section discusses the experimentation conducted to evaluate the performance of different clustering techniques. Our overall work is concluded in the fifth section.

## 2. Clustering-Based Content and Color Adaptive Tone Mapping Algorithm

Li et al. [42] proposed clustering-based content and color adaptive tone mapping algorithm. Most of conventional tone mapping operators split an image into chrominance and luminance channels. Instead, Li et al. divided an image into overlapped color patches. The overall algorithm consists of training and testing phases. The algorithm [42] utilizes training phase to learn PCA transform matrix for each color structure cluster from HDR training images, while the testing phase is used to project the HDR test images and to find the closest match in the training set. In the first step, logarithmic transform is applied on each HDR image to enhance the contrast and brightness of low luminance values in a channel while compressing the higher ones. Afterward, each image is divided into overlapped color patches to avoid artifacts such as local graying out, hue shift, or color fringes [43]. Each patch is further divided into approximately uncorrelated components: color structure, color variation, and patch mean. Patches with varied intensity level may have

similar structures [42]. Therefore, patches are grouped together into different clusters based on color structure and then PCA transform matrix is obtained for each cluster. In the testing phase, for each patch color structure, similarity measure is calculated with cluster centers extracted from the training data and then relevant PCA matrix is retrieved. For tone mapping,  $S$  curve arctan function is applied on PCA projection matrix. The patch mean is compressed by a linear function and the color variation is controlled by an arctan function. The image is reconstructed by processed patches. At the end, postprocessing step is performed for contrast enhancement by clamping image at its maximum and minimum intensity value.

## 3. Compared Clustering Techniques

To explore the influence of different clustering algorithms,  $K$ -means is a reference clustering algorithm and the results are compared with Gaussian Mixture Model (GMM) [44], DBSCAN [45], and fast search and find of density peak (FSFDP) [46] clustering algorithms. Before presenting the results obtained by using each of these clustering schemes, a brief description of each clustering technique is presented with its strengths and shortcomings.

**3.1.  $K$ -Means.**  $K$ -means is an extensively used partition-based clustering technique proposed by MacQueen. Partitioning can be done based on minimizing the objective function known as square error [47] which can be calculated as follows:

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_i\|)^2, \quad (1)$$

where  $\|x_i - v_i\|$ , “ $c$ ,” and “ $c_i$ ” represented Euclidean distance, number of cluster centers, and the number of points in  $i^{th}$  cluster, respectively.  $K$ -means was considered by Li et al. [42] for the clustering of the color structure.  $K$ -means is a relatively simple and computationally efficient algorithm. However, it suffers from the restriction of requiring human interaction for selecting the number of clusters, which provides easy implementation and effective results. Results are influenced by the initialization and may affect the performance of the tone mapping algorithm. Furthermore,  $K$ -means is sensitive to the cluster with a single data point due to its square distance [48].

**3.2. Fast Search and Find of Density Peak.** Fast search and find of density peak (FSFDP) [46] assumes that the density of data at the center of a cluster is relatively high and the center of one cluster is far away from the center of another cluster. Based on this assumption, FSFDP calculates the number of clusters and their centers automatically. Moreover, it detects and removes outliers intuitively. Clusters of nonspherical shape are easy to be recognized using FSFDP. In FSFDP, the cutoff distance used for calculating the density of each data point has a great influence on the effectiveness of FSFDP. For

a data point  $i$ , FSFDP calculates local density,  $\rho_i$ , and distance,  $\delta_i$ , from its nearest center as follows:

$$\rho_i = \sum_j X(d_{ij} - d_c), \quad (2)$$

where  $d_{ij}$  denotes the distance from data point  $i$  to  $j$  and  $d_c$  is cutoff distance used to calculate the density of each data point. For assigning a data point  $i$  to the nearest cluster center, distance is calculated as

$$\delta_i = \begin{cases} \min_{j: \rho_j > \rho_i} (d_{ij}), & \text{if } \exists j, s, t \rho_j > \rho_i, \\ \max_{j: \rho_j > \rho_i} (d_{ij}), & \text{otherwise.} \end{cases} \quad (3)$$

Equations (2) and (3) represent that the cluster centers are points with greater distance “ $\delta$ ” from other cluster centers and high data density “ $\rho$ .” The fast search and find of density peak is further experimented on two scale implementations [42] and represented as two-scale fast search and find of density peak (TFSFDP).

**3.3. Gaussian Mixture Model.** Unlike  $K$ -means clustering, Gaussian Mixture Model (GMM) [44] allows clusters of different shapes. The cluster orientation is dependent on the Gaussian distribution. GMM sets the  $k$  number of Gaussians according to the data distribution. GMM learns the cluster belonging through the probability of each data point using its parameters such as variance  $\Sigma_k$ , mean  $\mu_k$ , and weight of the cluster  $\pi_k$ . Mathematically, it is computed by using

$$p(x) = \sum_{k=1}^K \pi_k X(x | \mu_k, \Sigma_k). \quad (4)$$

GMM allows cluster overlapping; however, it is computationally extensive. Therefore, GMM is not applicable for high-dimensional data. Moreover, GMM is dependent on Gaussian distribution (clusters).

**3.4. Density-Based Spatial Clustering of Applications with Noise.** Similar to FSFDP, Density-Based Spatial Clustering of Application with Noise (DBSCAN) [45] is a density-based clustering algorithm that does not require the number of clusters as an initial parameter. DBSCAN identifies the clusters of various shapes and sizes based on data connectivity. Furthermore, DBSCAN is widely adopted because of its efficient computation. However, the algorithm of DBSCAN depends on the initial radius “ $r$ ” of the cluster as follows:

$$N_r = \{i \in D | \text{dist}(i, j) \leq r\}. \quad (5)$$

If the value of the radius is set to be small, all points that belong to the sparse cluster are considered as noise. On the other hand, if the radius is set to be large, all points lie within a single cluster. Therefore, to obtain appropriate results, the algorithm is executed for a number of times with different values of “ $r$ .”

## 4. Experiments and Results

This section describes the details of the dataset and the evaluation criteria used in our experiments. As different clustering methods have different parameters, the details of these are discussed in parameter setting section. The focus is to elaborate the influence of different state-of-the-art clustering algorithms on clustering-based content and color adaptive tone mapping algorithm, which is examined in performance and effectiveness section.

**4.1. Dataset and Evaluation Criteria.** For the training phase, Kodak’s database [49] is used, which consists of a total of 24 images of size of  $512 \times 768 \times 3$ . All images are of true color (24 bits per pixel) and, for tone mapping, these images are used as standard set suit [49]. In the testing phase, we performed a series of experiments on Fun and Shi [50] HDR dataset to evaluate the influence of different clustering algorithms on clustering-based content and color adaptive tone mapping [42]. This dataset contains both indoor and outdoor HDR images of 105 scenes captured through Nikon D700 digital camera. Sample images from the dataset are shown in Figure 1.

We compared the results of different clustering algorithms both qualitatively and quantitatively. Tone mapping operators attempt to preserve desirable characteristics, including local structure and naturalness, and to avoid the halo artifacts while converting an HDR image into LDR images. The classical evaluation parameters for objective measurements like PSNR cannot be used for the evaluation of tone mapping due to the unavailability of reference LDR images. So, we used two metrics for quantitative evaluation of dynamic range reduction with different clustering techniques: tone-mapped image quality index (TMQI) [51] and feature similarity index for tone-mapped images (FSITM) [52]. TMQI [51] is used to compute the structural resemblance and indexes of naturalness, and FSITM [52] is to measure the local phase similarity between HDR and LDR images.

**4.2. Parameters Setting.** Appearance regeneration of an image is dependent not only on the intensity relationship but also on many local weighted attributes including brightness, contrast, gray level, and color relationship. Clustering-based content and color adaptive tone mapping [42] is a patch base model in which the size and shape of the patch depend upon the window used for the calculation of mean and color structure. In this study, we have used the default window with size of  $7 \times 7$  as defined by Li et al. [42]. We have used default values of the control parameter for color appearance, local structure, and luminance within the tone mapping process as used by [42].

For clustering algorithms, the control parameters are set with the most optimal values. The details of these parameters for each clustering method are as follows.

**4.2.1. For  $K$ -means.** For  $K$ -means clustering, the number of clusters is initially set to be 100.



FIGURE 1: HDR tone-mapped source images [50].

TABLE 1: Effect of cutoff distance on FSITM and TMQI parameters for an image of size  $1419 \times 2130 \times 3$ .

Objective evaluation	$d_c = 1.3838$	$d_c = 1.4141$	$d_c = 1.9192$	$d_c = 2.3465$	$d_c = 2.8979$	$d_c = 3.2348$	$d_c = 4$
FSITM	0.871	0.871	0.871	0.871	0.871	0.871	0.871
TMQI	0.881	0.881	0.881	0.881	0.881	0.881	0.881

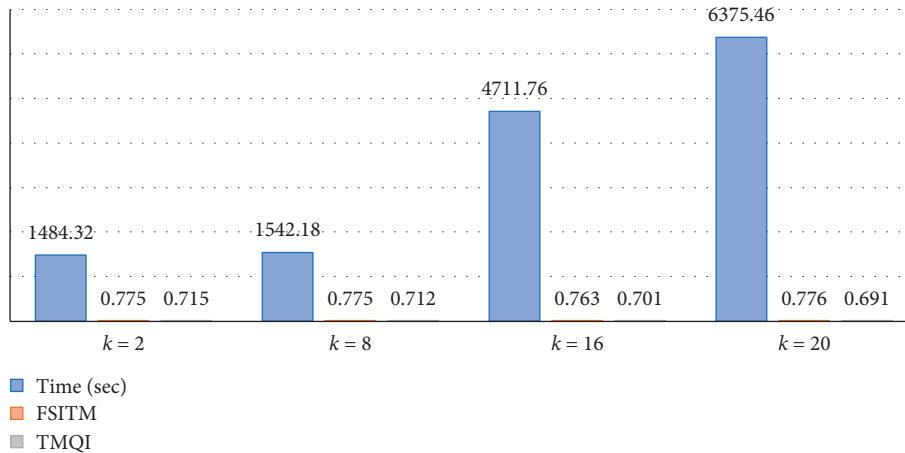


FIGURE 2: Effect of Gaussian distribution “ $k$ ” on time and FSITM and TMQI parameters.

4.2.2. *For Fast Search and Find of Density Peak.* For fast search and find of density peak, cut-off distance “ $d_c$ ” given in equation (2) is significant to be set. However, varying its default value that is “1.4141” has no difference in objective evaluation as shown in Table 1; therefore we set it as 1.4141.

4.2.3. *For Density-Based Spatial Clustering of Applications with Noise.* Furthermore, for DBSCAN, we execute the

algorithm several times to get to the appropriate value of “ $r$ ” as mentioned in equation (5). On the default value of “ $r$ ,” which is 10 for DBSCAN, the algorithm considers sparse datapoints as noise and therefore selected value of “ $r$ ” to be 0.2.

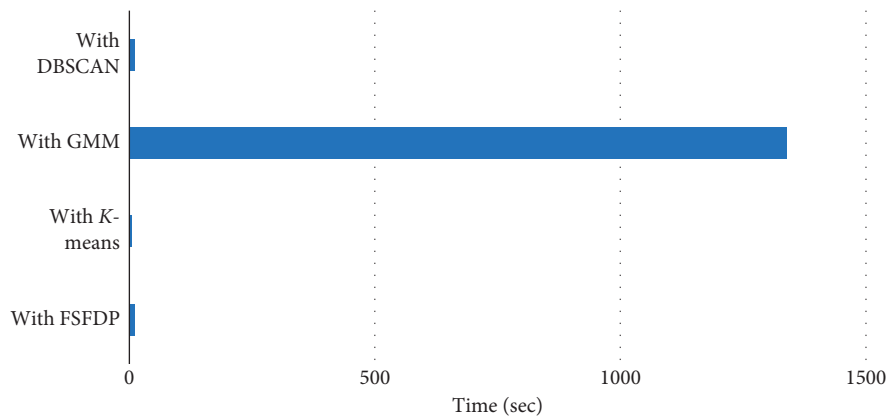
4.2.4. *For Gaussian Mixture Model.* For GMM, we used the default value of Gaussian distribution “ $k$ ” in equation (4) as (2).

TABLE 2: Tone-mapped image quality index score.

S. no.	With FSFDP	With TFSFDP	With K-means	With GMM	With DBSCAN
1	0.758	0.761	0.766	0.715	0.373
2	0.772	0.780	0.774	0.723	0.229
3	0.818	0.835	0.821	0.646	0.490
4	0.824	0.820	0.829	0.781	0.490
5	0.823	0.830	0.826	0.790	0.365
6	0.881	0.893	0.878	0.820	0.460
7	0.762	0.781	0.754	0.778	0.258
8	0.727	0.730	0.728	0.625	0.401
9	0.880	0.889	0.878	0.800	0.486
10	0.789	0.804	0.783	0.744	0.426
11	0.917	0.916	0.907	0.769	0.520
12	0.874	0.873	0.880	0.818	0.358
Avg.	0.819	0.826	0.819	0.683	0.405

TABLE 3: Feature similarity index for tone-mapped image score.

S. no.	With FSFDP	With TFSFDP	With K-means	With GMM	With DBSCAN
1	0.877	0.891	0.882	0.775	0.671
2	0.870	0.886	0.873	0.750	0.670
3	0.862	0.871	0.864	0.702	0.624
4	0.859	0.880	0.864	0.826	0.574
5	0.860	0.874	0.861	0.830	0.572
6	0.871	0.885	0.875	0.784	0.618
7	0.855	0.861	0.856	0.707	0.714
8	0.886	0.892	0.888	0.755	0.618
9	0.849	0.865	0.855	0.754	0.643
10	0.842	0.861	0.845	0.785	0.589
11	0.843	0.860	0.845	0.792	0.667
12	0.827	0.850	0.834	0.782	0.624
Avg.	0.858	0.873	0.862	0.770	0.632

FIGURE 3: Average execution time with different clustering algorithm applied on five images with size of  $535 \times 401 \times 3$ .

The experimentation shows that the different values of Gaussian distribution have none or little effect on objective evaluation. However, Gaussian distribution's larger values cause high computational cost. This is illustrated in Figure 2.

#### 4.3. Performance and Effectiveness

**4.3.1. Objective Evaluation.** Tables 2 and 3 present the results of different clustering techniques applied with the

method of [42] in terms of TMQI and FSITM, respectively. The best results are highlighted in both tables. Green, blue, and yellow numbers represent the first, second, and third best scores, respectively; likewise red number represents the worst score. The effectiveness of tone mapping operators depends upon the contrast, brightness, and local structure of HDR image scenes. So, a tone mapping operator cannot be equally effective for all HDR images [53]. Table 2 depicts that TFSFDP performs best for 9 images; however, considering

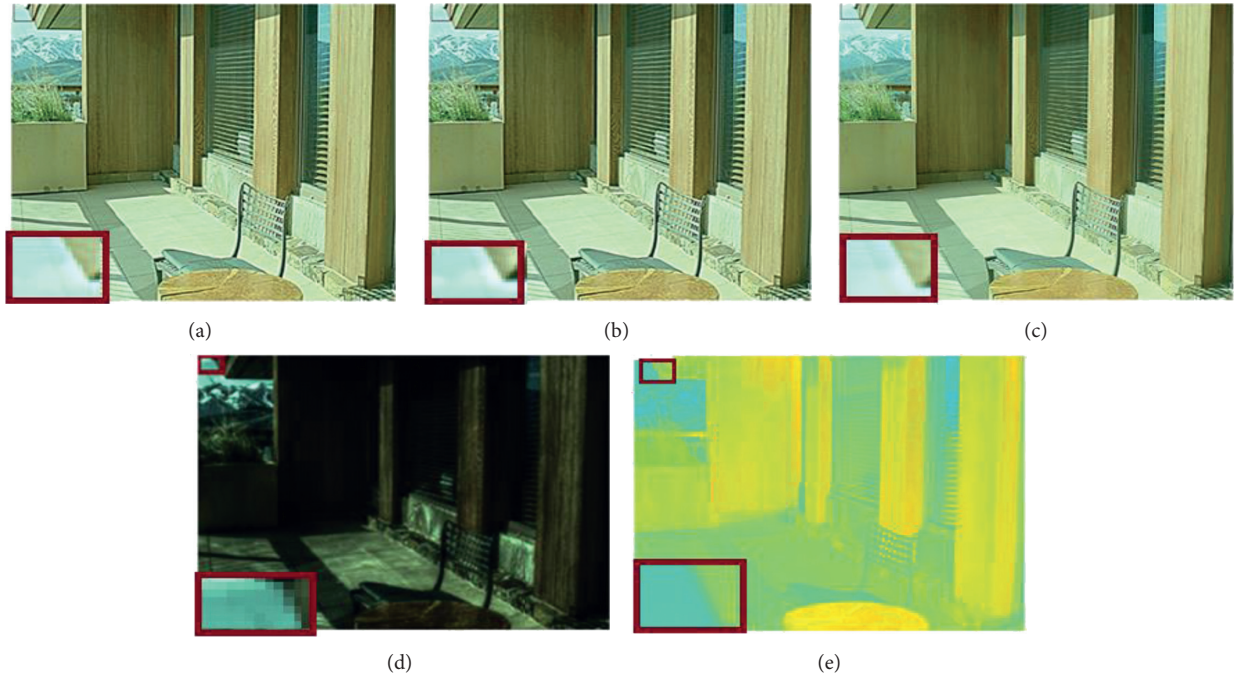


FIGURE 4: Visual comparison of outdoor tone-mapped images with (a) FSFDP, (b) TFSFDP, (c) *K*-means, (d) GMM, and (e) DBSCAN.

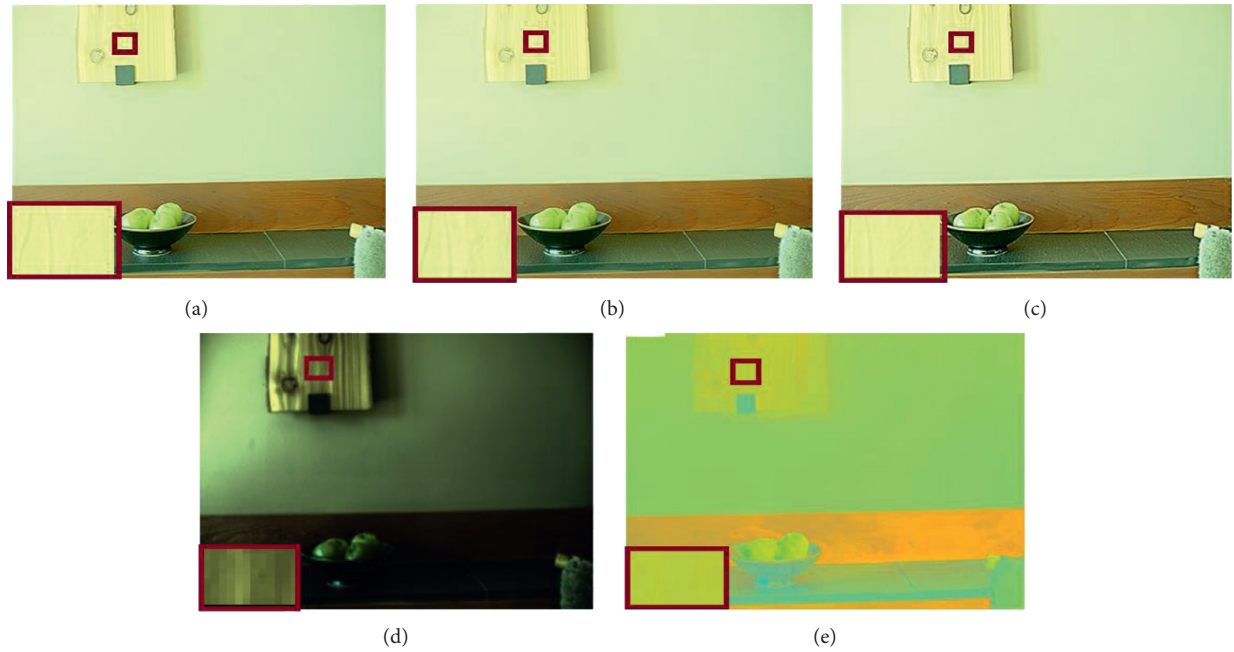


FIGURE 5: Visual comparison of indoor tone-mapped images with (a) FSFDP, (b) TFSFDP, (c) *K*-means, (d) GMM, and (e) DBSCAN.

the average of each method, *K*-means and FSFDP can be used alternatively. Table 3 has strengthened the results further, as FSITM shows that TFSFDP performs best, while *K*-means and FSFDP stood in the second and third positions, respectively. Furthermore, DBSCAN performs worst in terms of both FSITM and TMQI for every image. Based on the obtained results, DBSCAN is not recommended for tone mapping.

Performance of clustering-based tone mapping techniques is also compared in terms of their execution time. The execution times of different clustering-based tone mapping techniques are presented in Figure 3, which highlights that GMM has high computation cost irrespective of the provided TMQI and FSITM results. Therefore, it is not a preferred clustering algorithm to be used for tone mapping operation. In terms of fastest execution time, *K*-means outperforms the other

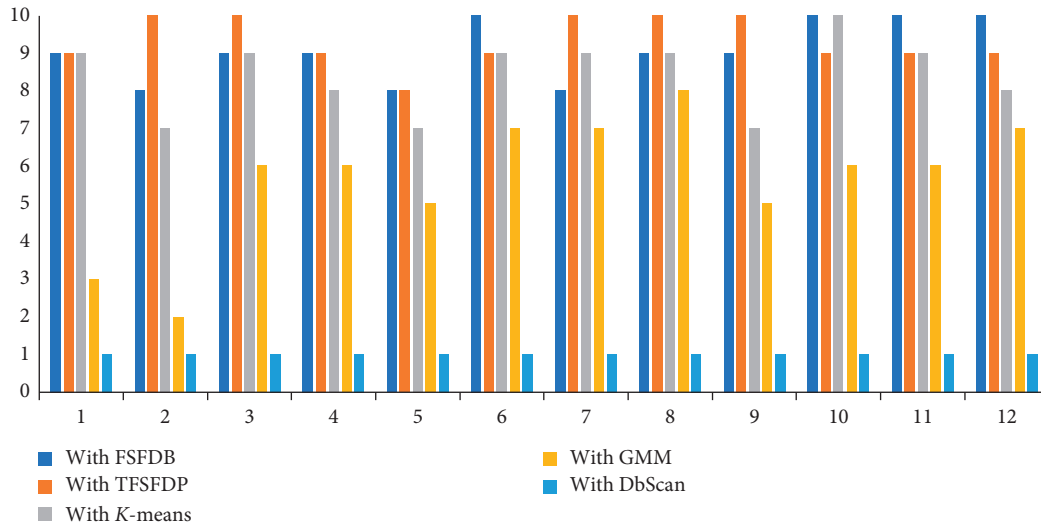


FIGURE 6: Average subjective evaluation score.

methods with least execution time of about “6.598 seconds,” whereas DBSCAN and FSFDP can also be considered for tone mapping operation as their execution times are about “11.31 seconds” and “12.218 seconds,” respectively.

**4.3.2. Subjective Evaluation Assessment.** Subjective evaluation is performed through visual assessment and scoring. Figures 4 and 5 exhibit a visual comparison of clustering-based techniques using tone-mapped indoor and outdoor scene images. FSFDP, TFSFDP, and K-means produced the visually appealing results as these clustering methods better preserve the local structure and color saturation as shown in Figures 4(a)–4(c) and Figures 5(a)–5(c). In case of FSFDP, minor details like clouds near the shaded corner shown in Figures 4(a) and 4(b) and wood pattern of cutting board as in Figures 5(a) and 5(b) remain better preserved than K-means that can be clearly illustrated from Figures 4(c) and 5(c). GMM suffers from information loss and unpleasant effect. In Figures 4(d) and 5(d), details on a glass of the window and shelf are not visible. In case of DBSCAN, colors are oversaturated and result in unnatural visual appearances of image as depicted in Figures 4(e) and 5(e).

Subjective evaluation of 12 images is performed by 8 volunteers currently working on image processing and machine learning. We used a monitor with a spatial resolution of  $2560 \times 1600$  to display images, side by side using the image viewer customizable window. We showed the resultant images of all methods to volunteers and asked them to score image from 1 to 10. “1” represents the worst score, while “10” is the highest score. Ratings are recorded by using score sheets provided to each volunteer during the subjective assessment. Figure 6 shows that, for most of the images, FSFDP has the highest averaged score.

## 5. Conclusion

Tone mapping is a complex field in which the conversion of HDR image to LDR image without information loss is

needed. The intention of this article is to present the best clustering technique among existing techniques for preserving complex local details lost in case of clustering-based content and color adaptive tone mapping method. For this purpose, the influence of different clustering techniques is examined. The effectiveness is measured in terms of subjective evaluation and FSITM and TMQI values. Experiments show that fast search and find of density peak results in more appealing results for tone mapping with acceptable computational cost. In the future, this model can be extended by using a feature other than color structure for clustering of patches. Besides, any other function in place of S shape arctan curve can be used for range compression.

## Data Availability

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

## Conflicts of Interest

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

## References

- [1] J. Tumblin and H. Rushmeier, “Tone reproduction for realistic images,” *IEEE Computer Graphics and Applications*, vol. 13, no. 6, pp. 42–48, 1993.
- [2] G. Ward, “A contrast-based scalefactor for luminance display,” in *Graphics Gems*, vol. 4, pp. 415–421, Elsevier, Amsterdam, Netherlands, 1994.
- [3] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, “Adaptive logarithmic mapping for displaying high contrast scenes,” *Computer Graphics Forum*, vol. 22, no. 3, pp. 419–426, 2003.
- [4] M. H. Kim and J. Kautz, “Consistent tone reproduction,” in *Proceedings of the Proceedings of Computer Graphics and Imaging*, Innsbruck, Austria, February 2008.

- [5] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 267–276, 2002.
- [6] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 13–24, 2005.
- [7] I. R. Khan, S. Rahardja, M. M. Khan, M. M. Movania, and F. Abed, "A tone-mapping technique based on histogram using a sensitivity model of the human visual system," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3469–3479, 2018.
- [8] J. Han, I. R. Khan, and S. Rahardja, "Lighting condition adaptive tone mapping method," in *Proceedings of the ACM SIGGRAPH 2018 Posters on—SIGGRAPH'18*, pp. 1-2, Vancouver, Canada, August 2018.
- [9] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," in *ACM Transactions on Graphics*, vol. 25, no. 3, , pp. 646–653, ACM, 2006.
- [10] D.-H. Lee, M. Fan, S.-W. Kim, M.-C. Kang, and S.-J. Ko, "High dynamic range image tone mapping based on asymmetric model of retinal adaptation," *Signal Processing: Image Communication*, vol. 68, pp. 120–128, 2018.
- [11] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 70–79, 2013.
- [12] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5638–5653, 2014.
- [13] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics (TOG)*, ACM, vol. 27, no. 3, , p. 67, 2008.
- [14] S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian filters," *Communications of the ACM*, vol. 58, no. 3, pp. 81–91, 2015.
- [15] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L 0 gradient minimization," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6, , p. 174, ACM, 2011.
- [16] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, p. 139, 2012.
- [17] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics (TOG)*, ACM, vol. 21, no. 3, pp. 257–266, 2002.
- [18] P. Debevec and S. Gibson, "A tone mapping algorithm for high contrast images," in *Proceedings of the 13th Eurographics Workshop on Rendering*, Pisa, Italy, June 2002.
- [19] J. Kuang, G. M. Johnson, and M. D. Fairchild, "iCAM06: a refined image appearance model for HDR image rendering," *Journal of Visual Communication and Image Representation*, vol. 18, no. 5, pp. 406–414, 2007.
- [20] S. Tong and Y. Yang, "A novel tone mapping algorithm," in *Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, pp. 427–431, IEEE, Chongqing, China, May 2019.
- [21] G. Krawczyk, K. Myszkowski, and H. P. Seidel, "Lightness perception in tone reproduction for high dynamic range images," *Computer Graphics Forum*, vol. 24, no. 3, pp. 635–645, 2005.
- [22] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 836–844, 2005.
- [23] Y. Jia and W. Zhang, "Efficient and adaptive tone mapping algorithm based on guided image filter," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 4, p. 2054012, 2020.
- [24] L. Meylan, D. Alleysson, and S. Süsstrunk, "Model of retinal local adaptation for the tone mapping of color filter array images," *Journal of the Optical Society of America A*, vol. 24, no. 9, pp. 2807–2816, 2007.
- [25] X.-S. Zhang, K.-F. Yang, J. Zhou, and Y.-J. Li, "Retina inspired tone mapping method for high dynamic range images," *Optics Express*, vol. 28, no. 5, pp. 5953–5964, 2020.
- [26] C. A. Parraga and X. Otazu, "Which tone-mapping operator is the best? A comparative study of perceptual quality," *Journal of the Optical Society of America A*, vol. 35, no. 4, pp. 626–638, 2018.
- [27] H. T. Chen, T. L. Liu, and C. S. Fuh, "Tone reproduction: a perspective from luminance-driven perceptual grouping," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 73–96, 2005.
- [28] Z. Liang, W. Liu, and R. Yao, "Contrast enhancement by nonlinear diffusion filtering," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 673–686, 2016.
- [29] M. K. Ng and W. Wang, "A total variation model for Retinex," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 345–365, 2011.
- [30] H. Ahn, B. Keum, D. Kim, and H. S. Lee, "Adaptive local tone mapping based on retinex for high dynamic range images," in *Proceedings of the 2013 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 153–156, IEEE, Las Vegas, NV, USA, January 2013.
- [31] Z. Liang, J. Xu, D. Zhang, Z. Cao, and L. Zhang, "A hybrid 11-10 layer decomposition model for tone mapping," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4758–4766, Salt Lake City, UT, USA, June 2018.
- [32] B. J. Lee and B. C. Song, "Local tone mapping using sub-band decomposed multi-scale retinex for high dynamic range images," in *Proceedings of the 2014 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 125–128, Las Vegas, NV, USA, January 2014.
- [33] X. Shu and X. Wu, "Locally adaptive rank-constrained optimal tone mapping," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 3, pp. 1–10, 2018.
- [34] A. Rana, G. Valenzise, and F. Dufaux, "Learning-based adaptive tone mapping for keypoint detection," in *Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 337–342, Hong Kong, China, July 2017.
- [35] D. M. El Mezeni and L. V. Saranovac, "Enhanced local tone mapping for detail preserving reproduction of high dynamic range images," *Journal of Visual Communication and Image Representation*, vol. 53, pp. 122–133, 2018.
- [36] S. A. Henley: "Rendering of high dynamic range images," U.S. Patent No. 7,480,421, U.S. Patent and Trademark Office, Washington, DC, USA, 2009.
- [37] Z. Li and J. Zheng, "Visual-saliency-based tone mapping for high dynamic range images," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 12, pp. 7076–7082, 2014.
- [38] S. Ferradans, M. Bertalmio, E. Provenzi, and V. Caselles, "An analysis of visual adaptation and contrast perception for tone mapping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 2002–2012, 2011.



- [39] P. Ambalathankandy, M. Ikebe, T. Yoshida et al., “An adaptive global and local tone mapping algorithm implemented on FPGA,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
- [40] A. Mehmood, I. R. Khan, D. Hassan, and D. Hussain, “Enhancement of CT images for visualization,” in *Proceedings of the ACM SIGGRAPH 2019 Posters*, vol. 83, pp. 1-2, New York, NY, USA, July 2019.
- [41] A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic, “Deep tone mapping operator for high dynamic range images,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1285–1298, 2019.
- [42] H. Li, X. Jia, and L. Zhang, “Clustering based content and color adaptive tone mapping,” *Computer Vision and Image Understanding*, vol. 168, pp. 37–49, 2018.
- [43] D. Ellis, *Tone Mapping for High Dynamic Range Cameras*, Department of Engineering Science Oxford University, Oxford, UK, 2008, [http://www.robots.ox.ac.uk/~dre/docs/ellis\\_d\\_stj\\_4yp.pdf](http://www.robots.ox.ac.uk/~dre/docs/ellis_d_stj_4yp.pdf).
- [44] Clustering with Gaussian mixture model—clustering with Gaussian mixture model—medium, 2018, <https://medium.com/clustering-with-gaussian-mixture-model/clustering-with-gaussian-mixture-model-c695b6cd60da>.
- [45] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” *Computer Science*, vol. 96, no. 34, pp. 226–231, 1996.
- [46] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks,” *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [47] S. A. Elavarasi, J. Akilandeswari, and B. Sathiyabhama, “A survey on partition clustering algorithms,” *International Journal of Enterprise Computing and Business Systems*, vol. 1, no. 1, 2011.
- [48] S. K. Popat and M. Emmanuel, “Review and comparative study of clustering techniques,” *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 805–812, 2014.
- [49] <http://r0k.us/graphics/kodak/>, 2018.
- [50] B. Funt and L. Shi, “HDR dataset,” 2018, [http://www.cs.sfu.ca/~colour/data/funt\\_hdr/](http://www.cs.sfu.ca/~colour/data/funt_hdr/).
- [51] H. Yeganeh and Z. Wang, “Objective quality assessment of tone-mapped images,” *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657–667, 2013.
- [52] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam, and M. Cheriet, “FSITM: a feature similarity index for tone-mapped images,” *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1026–1029, 2015.
- [53] K. Gu, S. Wang, G. Zhai et al., “Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure,” *IEEE Transactions on Multimedia*, vol. 18, no. 3, pp. 432–443, 2016.

## Research Article

# Impact of Node Density on the QoS Parameters of Routing Protocols in Opportunistic Networks for Smart Spaces

Puneet Garg <sup>1</sup>, Ashutosh Dixit <sup>1</sup>, Preeti Sethi,<sup>1</sup> and Plácido Rogerio Pinheiro<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, J. C. Bose University of Science & Technology YMCA, Faridabad 121006, India

<sup>2</sup>Applied Informatics, University of Fortaleza, Fortaleza, CE 60811-905, Brazil

Correspondence should be addressed to Puneet Garg; [puneetgarg.er@gmail.com](mailto:puneetgarg.er@gmail.com)

Received 27 May 2020; Revised 28 June 2020; Accepted 8 July 2020; Published 1 August 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Puneet Garg 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 need and importance of Smart Spaces have been potentially realized by the researchers due to its applicability in the current lifestyle. Opportunistic network, a successor of mobile ad hoc networks and a budding technology of network, is a best-suited technology for implementing Smart Spaces due to its wide range of applications in real-life scenarios ranging from building smart cities to interplanetary communication. There are numerous routing protocols which are available in opportunistic network, each having their pros and cons; however, no research till the time of listing has been done which can quantitatively demonstrate the maximum performance of these protocols and standardize the comparison of opportunistic routing protocols which has been a major cause of ambiguous performance evaluation studies. The work here presents a categorical view of the opportunistic routing protocol family and thereby compares and contrasts the various simulators suited for their simulation. Thereafter, the most popular protocols (selecting at least one protocol from each category) are compared based on node density on as many as 8 standard performance metrics using ONE simulator to observe their scalability, realism, and comparability. The work concludes by presenting the merits and demerits of each of the protocols discussed as well as specifying the best routing protocol among all the available protocols for Smart Spaces with maximum output. It is believed that the results achieved by the implemented methodology will help future researchers to choose appropriate routing protocol to delve into their research under different scenarios.

## 1. Introduction

In the era of consistently changing environment, communication devices are getting intelligent day by day and delivering rapid and robust connections. New applications are emerging with an advanced approach in wireless networking arena which is attracting new researchers in this field for further efforts. Due to the tremendous research in the wireless section, communication has become promising even in remote regions where previously, constructing a simple communication network was a huge challenge.

Owing to the pervasive applications of networking, existing technologies on wireless networking like vehicular ad hoc wireless network, wireless sensor networks, mobile ad hoc networks are observed to be insufficient in some instances such as interplanetary communication [1], Smart

Spaces [2], and social networks [3] to cope up with all the aspects and challenges concerned to the wireless networking. Some detected major difficulties with these technologies are connection failure and links discontinuation which degrades the overall performance of the network. To counter this challenge, researchers worked hard to create a new networking technology that led to the development of opportunistic networks (OppNets). According to Shu et al. [4], opportunistic networks (OppNets) are one of the categories of delay-tolerant networks [5], which support data communication through movement in nodes as it does not need any long-lasting links from sender to receiver nodes.

According to Kushwaha and Gupta [6], opportunistic network sare one of the rising advancements of the network system. In opportunistic networks, nodes can communicate with one another regardless of whether the route between

source to destination does not exist at that given moment. Opportunistic networks must be delay-tolerant (i.e., ready to tolerate bigger delays). Delay-tolerant network (DTN) utilizes the idea of “store-carry-forward” of data packets. DTNs can move data or set up a correspondence in a remote area or emergency condition where there is no network set up. DTNs have numerous applications like to provide smooth Internet arrangements in remote areas, in vehicular networks, noise observing, extraordinary terrestrial situations, and so on. It is in this manner promising to recognize viewpoints for reconciliation and integration of opportunistic network systems and advances into delay-tolerant networking.

OppNet is different than mobile ad hoc networks (MANETs) in the aspect of connectivity of participating nodes carrying data [7, 8]. The nodes participating in ad hoc networks for data communication remain connected constantly whether the nodes are in motion or static; on the other hand, nodes get to connect with other nodes in the OppNet when the communication is to be done between the nodes that make it a better approach in real-world applications. Therefore, conventionally defined protocols such as TCP/IP, DSR, AODV, and DSDV fail to function properly in opportunistic networks [9–11].

Rather, OppNet which is a type of delay-tolerant network is considered as the next generation of ad hoc networks which is further derived from standard wireless networks. Figure 1 clearly illustrates the evolution of opportunistic networks originating from the wireless networks domain through step-by-step growth. Every growth in each progressive step indicates the extension of personal communication networks towards solving further real-life problems which were a challenge earlier.

According to Nayyar et al. [12], the primary aim of developing opportunistic networks is to handle critical situations with effective manner such as disaster handling, war-field communications [13], satellite communications, flying warplanes/drones networking, underwater sensor networks [14], and forest surveillance. OppNets are highly useful where communication encounters high delays, no reliable connectivity, and high error rates. Nodes participating in OppNet are equipped with several attributes such as short communication range, high dynamic mobility, and low density. OppNet is designed specifically to connect almost every device which is capable of being connected through any wireless medium such as Bluetooth and Wi-Fi, etc. thereby making it a perfect choice for network designers all over the world. A general scenario of the opportunistic network connection is depicted in Figure 2.

*1.1. Advantages and Disadvantages of Opportunistic Networks.* According to Nayyar et al. [12], opportunistic networks are considered a strong option among the available networking technologies due to the following advantages:

- (1) OppNets can tolerate high delays if the destination or another intermediate node is not responding due to any reason during data communication

- (2) OppNet may allow data transfer with asymmetric data rates
- (3) OppNets can prevent data loss due to connection failure as it follows the store-carry-forward approach
- (4) OppNets can manage data communication even in continuous ups and downs with network state as it is specifically designed for operating under the situation of intermittent connection

Despite all the abovementioned advantages, there are certain challenges [15] in OppNet communication that needs to be dealt with. The various shortcomings of OppNet are as follows:

- (1) OppNet requires high buffer space as it stores the data to be forwarded which further increases its operational cost.
- (2) Due to intermittent connectivity, a node communicating in OppNet requires a large amount of energy as it may wait for a long time for forwarding the data it holds.
- (3) OppNet faces a challenge of security also like MANET because like MANET, nodes participating in OppNet forward the data towards destination via intermediate nodes. These intermediate nodes may be malicious sometime. Therefore, choosing a secure route between two communicating nodes is a challenge.

It is believed that the above-discussed challenges will soon be resolved by upcoming researchers through their continuous efforts to make OppNet better than its current version.

*1.2. Role of Opportunistic Networks in Smart Spaces.* In the era of Digital Connectivity, a large amount of population is equipped with smartphones that connect a person to the digital world via the Internet [16, 17]. Besides connecting the Internet, a smartphone comes with a different mode of connectivity with other devices such as Bluetooth and Wi-Fi [18]. According to Samaniego et al. [19], “Smart Spaces are common spaces that have capabilities to get data from the environment and apply knowledge to fulfill requirements of mobility, distribution, and context awareness of its inhabitants.” Smart Space is nothing but a virtual world full of information as per the interest of member nodes [2, 20]. According to Ismagilova et al. [21], the concept of Smart Spaces complements IoT technology specifically for designing smart cities.

These different modes of smartphone enable its user to make his/her private network as per the requirements as well as preferences of connected persons. These small and customizable networks are termed as Smart Spaces in the real world [22, 23]. The connected devices in such Smart Spaces are known as nodes in the networking terminology [2, 24]. For its smooth connectivity, opportunistic network is the best suited due to its inherent traits. The nodes in opportunistic networks use Wi-Fi or Bluetooth for interconnectivity and primarily initiate functioning with a single

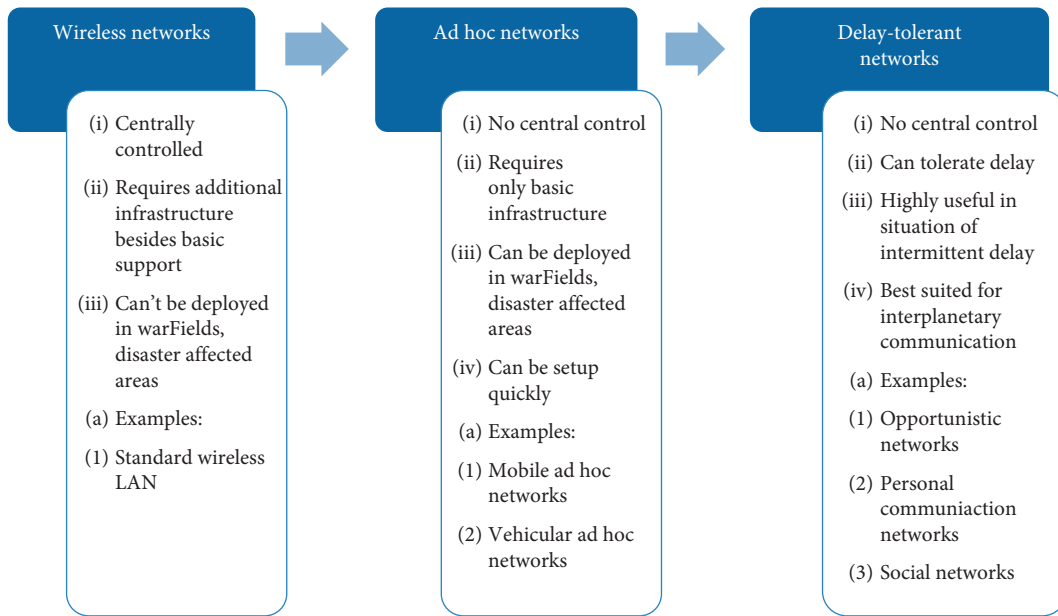


FIGURE 1: Evolution of opportunistic networks.

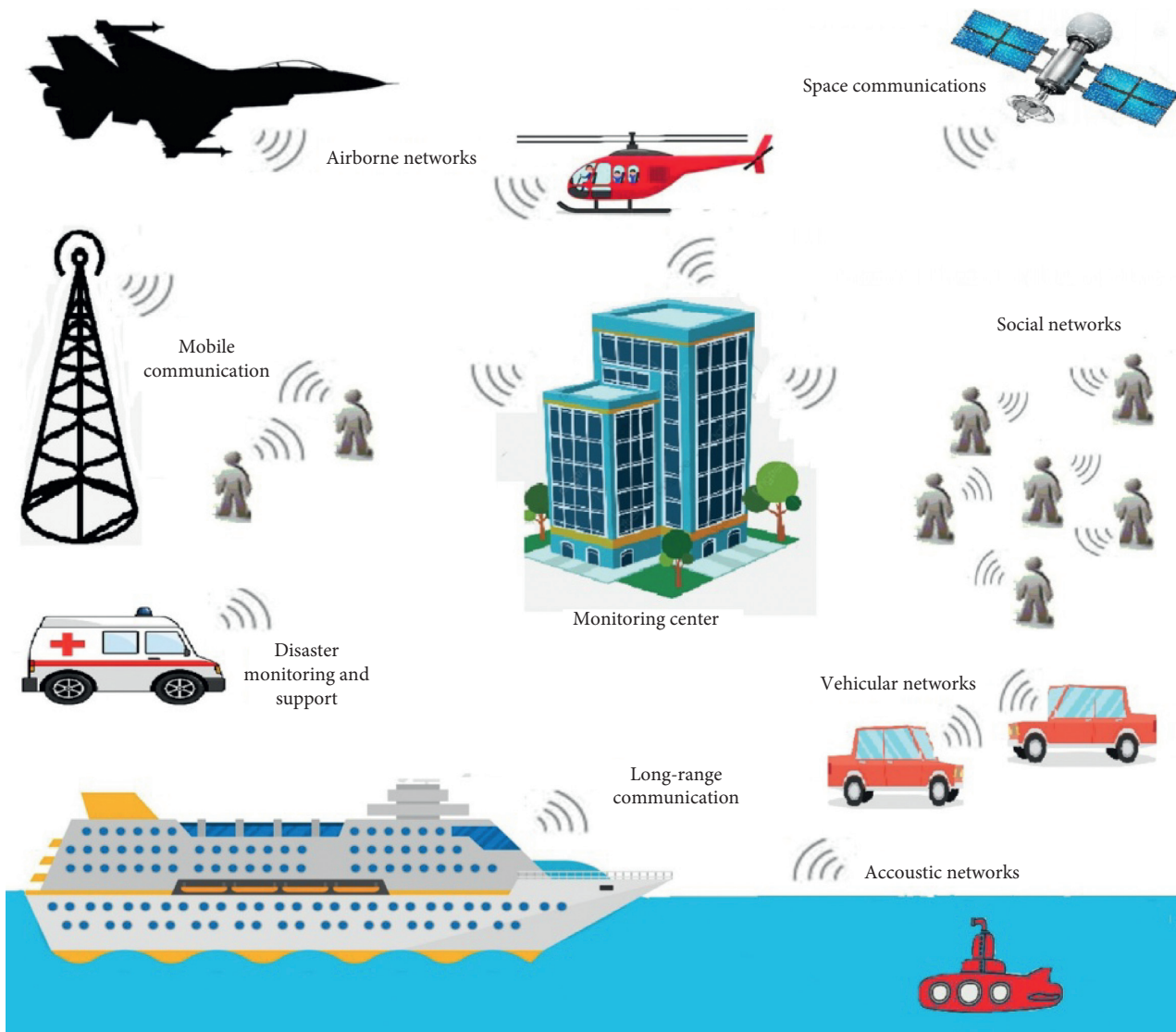


FIGURE 2: A general OppNet scenario.

node known as seed OppNet and expand further by implementing it among more member nodes that facilitate data forwarding in the network [25, 26].

Routing in OppNets relies upon contact opportunity between the nodes which is required due to their versatile nature. The most huge technique used in OppNet for routing movement is the store-carry-forward technique, where a message can be forwarded among intermediary nodes, and accordingly, the message is passed on to the destination node. The store-carry-forward technique is seen as a capable technique to ensure message delivery to destination nodes where message delivery may bear high delays. Thus, OppNets are a subclass of DTN where nodes must be outfitted with high buffer space to store messages for a strange timeframe to evade packet dropping.

Short-distance communication feature enabled node may help OppNet to gain large improvements and numerous scopes covering almost every industry such as information attacking, energy utilization, communication engineering, and information gathering. However, maintaining a stable network topology in OppNet is a cumbersome task; also, predicting the network topology is very difficult due to the frequent mobility among nodes and large communication range.

This paper initially presents the introduction of opportunistic networks followed by its role in building Smart Spaces and applications of OppNets which are presented in Section 2. Section 3 highlights various routing protocols associated with this class of networks. Section 4 elaborates numerous research simulators available in OppNets followed by the enlightenment of Java-based simulator ONE (Opportunistic Network Environment). Section 5 presents the analysis of the total nine standard routing protocols over standard QoS parameters. The paper concludes by submitting future work in this area.

## 2. Applications of Opportunistic Networks

Opportunistic networks have become ubiquitous nowadays. It has numerous applications in real-life scenarios covering almost all levels of modern communication requirements. Figure 3 demonstrates OppNet applications in the real world.

Some of the popular ones are described as follows:

- (a) LASSO: Saloni et al. [27] developed LASSO, a general-purpose smartphone-based application that uses the opportunistic networking feature using Bluetooth or smartphone for group monitoring. It has proved to be highly advantageous for the interconnectivity of a group of some persons roaming in a smart city and monitoring their locations to track if someone got missed. Its unique feature of decentralization device-to-device mode of operation makes it able to be used in any mobile scenario. Also, it does not need any pre-existence of any communication infrastructure. LASSO has performed well on small-scale implementation (i.e., 250 persons over a small geographic area) and it is being underdevelopment for further enhancements.
- (b) Shared wireless infostation model (SWIM): Small and Haas [28] proposed an infostation concept with the integration of opportunistic networking. It was experimented to observe the whale species by tying sensors on whale's back, thereby making them radiotagged whales. All sensor nodes are connected via opportunistic networking and data are forwarded in the same fashion as in OppNet and finally delivers to the infostation. This application has proved to be excellent to monitor whales' life closely.
- (c) Underwater communication networks: Detweiller et al. [29] experimented with a communication setup consisting of mobile sensor nodes with acrylic closure and other communications support hardware to establish underwater communication network. It is a quiet application of opportunistic networks as it can tolerate delay and respond accordingly to commensurate the real-life challenges in a typical sea environment. An experimental study proved it successful along with TDMA protocols with depths less than 100 meters for comprehension and demonstrating coral reefs. It can likewise support more prominent depths by supplanting acrylic enclosure with a glass/titanium enclosure.
- (d) ZebraNet: ZebraNet [30] is an OppNet-based project implemented by Princeton University under Mpala Research to track and monitor wild creatures in the forest of Kenya with the help of powerful sensors tied at animals' neck. Every sensor being used in it is enabled with wireless transceiver, CPU, and GPS. All sensors fitted on animal bodies interchange their sensed information in OppNet fashion and finally deliver to the desired station. It is focused to develop for monitoring the movement and speed of wild creatures in forests.
- (e) Composable Distributed Mobile Applications: Papadaki et al. [31] presented a system design that permits application developers to consider the future environment as a generic execution that opportunistically distributes and executes automatically the components of their applications. The concept of permitting mobile clients to utilize the resources present in the environment with the help of location-aware services relates this application to opportunistic networking and opportunistic computing. The primary aim of this system design is to hold a vision to a futuristic environment where clients do not require to search and use services already existing in the environment, rather, to use the environment to implement their custom applications. It has experimented successfully with the help of a prototype evaluation.
- (f) Saratoga: Wood et al. [1] presented Saratoga which is a light-weight protocol based on opportunistic networks. It was developed by Surrey Satellite

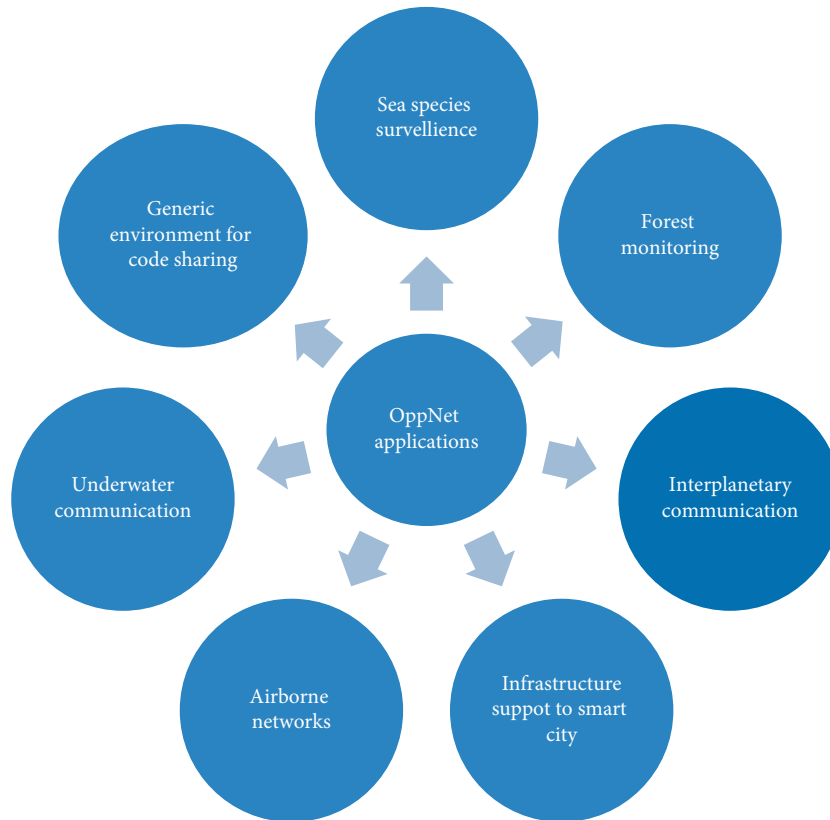


FIGURE 3: OppNet applications in the real world.

Technology Limited (SSTL) for file transfers of data recorded in image format by IP-based Disaster Monitoring Constellation (DMC) satellites revolving around earth from low orbit. Saratoga follows opportunistic routing as it only forwards the data packet when link connectivity is available which guarantees that the maximum possible data are transferred to the node during a 12-minute pass over a satellite ground station. Saratoga is fully operational for many years.

- (g) Underwater acoustic communication: underwater communication networks have been the prime area of research in recent years due to its various applications such as oil spills detection, disaster detection and avoidance, sea exploration, and detection of submarines. Menon and Prathap discussed [32] numerous opportunistic routing protocols developed especially for underwater acoustic communication. Two major categories of such protocols are pressure-based protocols and location-based protocols. Rahman et al. [33] proposed a routing algorithm named TORA (totally opportunistic routing algorithm) with a focus to overcome issues about underwater acoustic communication such as void nodes, horizontal transmission, high end-to-end delay, low throughput, and high battery drain. According to extensive simulation studies, TORA has been proved a better option over the existing algorithm up to a considerable extent.

These are some of the major applications opportunistic networks possess. But its scope has not been limited to the mentioned applications; rather, it has vast scope in airborne networks [34], space operations [35], backend support in smart cities [36, 37], and many other domains that are not discussed in this paper.

### 3. Routing Protocols in OppNets

Opportunistic networks contain a huge number of routing protocols. These protocols came into existence as a result of the rigorous efforts of several researchers done in the domain of opportunistic networking [3, 38–40].

According to Juyal et al. [41], numerous protocols can be categorized into various classes, viz., flooding-based routing protocols (e.g., Epidemic routing protocol and Spray-and-Wait routing protocol), forwarding-based routing protocol (e.g., Direct Delivery routing protocol and First Contact routing protocol), probability-based routing protocols (e.g., PROPHET and MaxProp), knowledge-based routing protocols (e.g., Epidemic Oracle routing protocol), social relationship-based routing protocols (e.g., FRESH routing protocol), and off-course hybrid routing protocols (e.g., RAPID protocol). The work presents the exhaustive survey of all these protocols in each category of routing protocols in opportunistic networks.

The taxonomy of routing protocols is depicted in Figure 4.

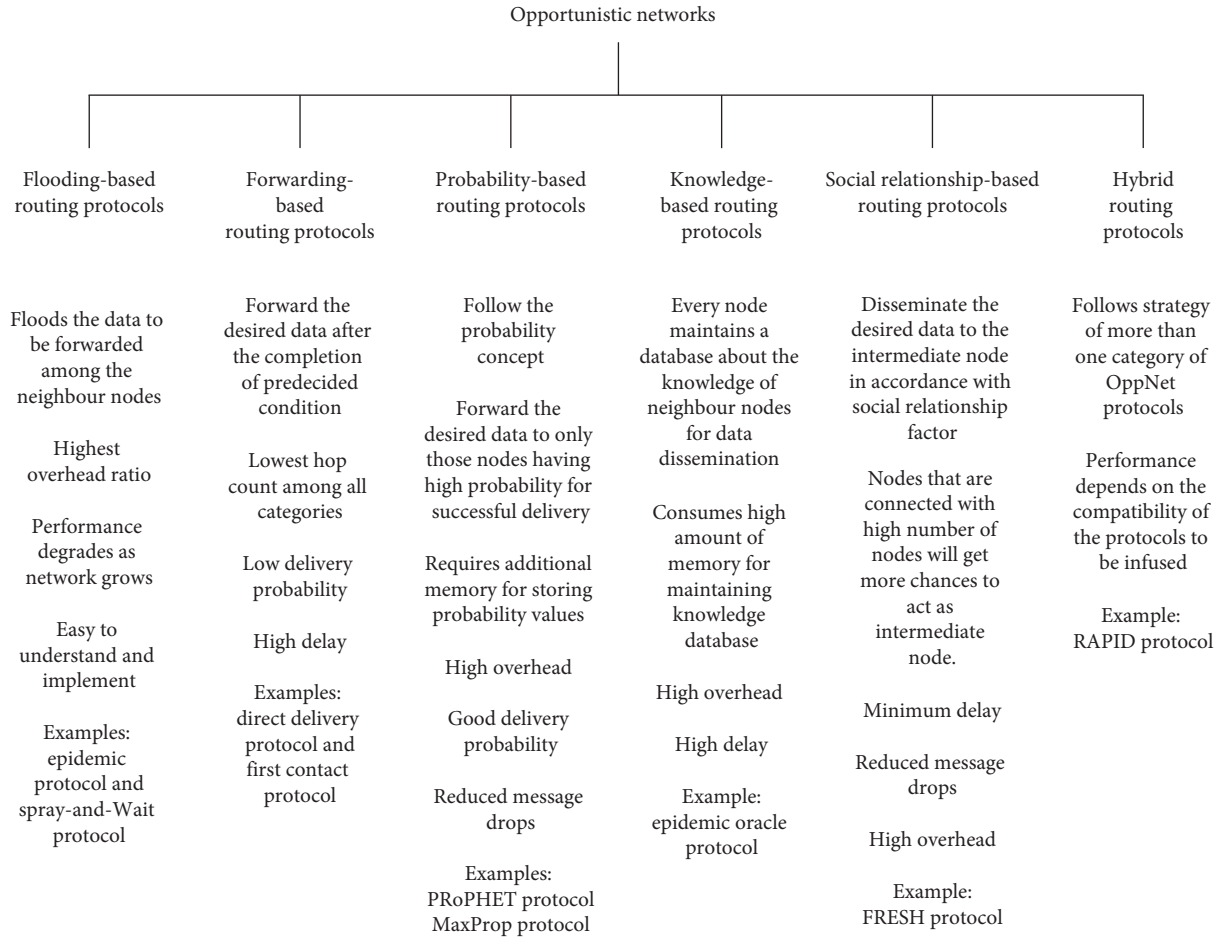


FIGURE 4: Taxonomy of routing protocols in opportunistic networks.

- (a) Epidemic routing protocol: the functioning of Epidemic protocol may be described as an epidemic disease spread in an area; therefore, any contact developed in such an infected area will spread it further. The only difference is that here, the disease is considered as a message containing data, and the area is considered as transmission range. Vahadat and Becker [42] introduced a message delivery technique, in which there is no connected path between source and destination. Initially, in mobile ad hoc networks, existing techniques available at that time were unable to counter this situation. Later on, it was adapted for opportunistic networks. Thus, epidemic routing was introduced where random pairwise interchanges of messages among participating nodes guarantee eventual message delivery. The Epidemic routing protocol is one of the oldest routing protocols in opportunistic networks. The Epidemic routing protocol is easy to understand and implement.
- (b) Direct Delivery routing protocol: Spyropoulos et al. [43] presented this protocol with an idea of single-copy routing in opportunistic networks. In this routing protocol, the message needs not to be forwarded via intermediate nodes rather; it is held by

the sender node itself. It waits for sending the message until it comes into contact with the destination node directly. It is the most simple and easy to understand among all the protocols currently available in opportunistic networks. The methodology adapted is not sufficiently reliable as the sender node may have to wait for infinite delay for destination node come into its contact. If this kind of situation happens, the entire message will be lost as the whole network would have only a single of that message. Spyropoulos et al. [43] defined the formula for calculating ED direct delivery as follows:

$$ED_{\text{DirectDelivery}} = 0.5N \left( 0.34 \log N - \frac{2^{K+1} - K - 2}{2^K - 1} \right), \quad (1)$$

where  $K \rightarrow$  transmission range of each node and  $N \rightarrow$  covered area.

- (c) Spray-and-Wait routing protocol: Spyropoulos et al. [44] proposed this scheme with a focus to improve the performance of the Epidemic routing protocol. It was found to be a better performer than Epidemic and other protocols that lie in the same category on

the ground of simulations as well as theory. With the increase of network size or connectivity level, Spray-and-Wait protocol also proved itself as a very scalable protocol. Also, Spray-and-Wait protocol generates less number of message replicas, thereby reducing the buffer space and relevant parameters. The functioning of this protocol can be divided into two phases, i.e., Spray phase and Wait phase. In the Spray phase, message replicas are broadcasted by source nodes among other nodes in its transmission range. All nodes accept the message from the source node and save it in their respective buffer. In the Wait phase, the nodes holding the message received from the source node wait for the opportunity to forward the message when another node comes into its contact [45].

The primary objective of Spray-and-Wait protocol is to reduce the number of message replicas to be forwarded like Epidemic routing protocol, thereby reducing the expected delay in data transmission. As per Spyropoulos et al. [44], the expected delay (ED) in Spray-and-Wait routing protocol is shown as follows:

$$ED_{\text{Spray-and-Wait}} = \frac{H_{n-1}}{(n-1)} ED_{dt}, \quad (2)$$

where  $H_n \rightarrow n^{\text{th}}$  harmonic number, viz.,  $H_n = \sum_{i=1}^n (1/i) = \Theta(\log n)$  and  $n \rightarrow$  total number of nodes.

After the comparison between the formulas of expected delay of Spray-and-Wait and Direct Delivery, it may be easily concluded that Spray-and-Wait protocol saves a considerable amount of time in data delivery.

- (d) First Contact routing protocol: Jain et al. [46] developed a scheme that requires limited additional knowledge about network topology, considerably less than the entire topology. This scheme is known as the First Contact routing protocol. In this protocol, the sender node disseminates the message to the very first node it encountered, and this node forwards it to the next first encountered node. The process continues until the message is received by the destination node. The encounters between the nodes are based on random walk search. The message will be retained in the buffer of the node if it does not find any other node through the encounter. It is found experimentally that First Contact routing protocol performs poor as it forwards the node based on random encounter and no topology or geographic condition is taken care of for message transmission towards the destination node. It is easy for implementation and may be used as a better option for multicast messages. Packet dropping and high delays are some of the problems that arise due to the First Contact routing protocol.

- (e) PROPHET routing protocol: Lindgren et al. [47] proposed a new routing algorithm named Probabilistic Routing Protocol using History of Encounters and Transitivity (PROPHET) in intermittently connected networks. It has similar functionality like Epidemic routing protocol, but the only difference is that in PROPHET, every node participating in opportunistic networks calculates a “probabilistic metric” also known as delivery predictability for each evaluated/known destination which empowers the source node to find out the accomplishment of message delivery.

Delivery predictability can be updated as per the following equation:

$$P_{(a,b)} = P_{(a,b)_{\text{old}}} + (1 - P_{(a,b)_{\text{old}}}) \times P_{\text{init}}, \quad (3)$$

where  $P_{\text{init}} \in (0, 1)$  and  $a, b \rightarrow$  nodes in the network. If a pair of two nodes does not have any experience of any cooperation in data forwarding due to any reason, their respective delivery predictability must be reduced as the time grows. Therefore, delivery predictability may be updated as per the *aging constant*  $\gamma$  [ $\gamma \in (0, 1)$ ] as mentioned in the following equation:

$$P_{(a,b)} = P_{(a,b)_{\text{old}}} \times \gamma^K, \quad (4)$$

where  $K \rightarrow$  time units.

Moreover, delivery predictability also follows the property of transitivity, i.e., if node A has high probability metric with node B and similarly node B has high probability metric with node C, then node A and node C would have also high probability metric due to the property of transitivity in PROPHET protocol even though node A and node C do not have any recent experience of any cooperation in data forwarding. The following equation illustrates the effect of transitivity on delivery predictability:

$$P_{(a,c)} = P_{(a,c)_{\text{old}}} + (1 - P_{(a,c)_{\text{old}}}) \times P_{(a,b)} \times P_{(b,c)} \times \beta, \quad (5)$$

where  $\beta$  is a scaling constant and  $\beta \in [0, 1]$ .

The computation of delivery predictability is done based on encountered nodes history or nodes visited history. At the point when two nodes came in contact with each other, summary vectors are interchanged containing delivery predictability. On the off chance that two nodes are encountered on a routine basis, they will have higher delivery predictability and those nodes which are having less predictability or never encountered have fewer changes of effective message delivery to the destination. The delivery predictability shifts from time to time. Simulation-based investigation explains that PROPHET protocol takes fewer message interchanges, low communication overhead,



and less delay and has a better packet delivery ratio when contrasted with epidemic routing.

- (f) MaxProp routing protocol: Burgess et al. [48] proposed MaxProp protocol. This protocol is based on prioritizing two kinds of schedules, i.e., schedule of messages to be dropped and schedule of messages to be transferred to other nodes. The main aim of designing this protocol is to improvise delivery rate and average latency. It functions by ranking the stored packets in nodes' memory on the ground of cost assigned.

The formula for calculating cost between source node  $a$  and destination node  $d$  is shown as follows:

$$c(a, a + 1, \dots, d) = \sum_{x=a}^{d-1} [1 - (f_{x+1}^x)], \quad (6)$$

where  $f_{x+1}^x \rightarrow$  probability of successful message transfer from node  $x$  towards  $x + 1$ .

MaxProp prevents duplication of packets and giving high priority to new packets. The priorities of the message are assigned based on the head start of a new message, hop count, previous intermediate nodes, and historical data. The functioning of MaxProp routing protocol starts from transmitting all the messages destined for the immediate neighbour in the network, after that, routing information is transmitted followed by acknowledgments of messages being delivered regardless of sender and

receiver nodes. At last, high priority has been given to those messages which are not delivered to the destined nodes for communication. MaxProp protocol is found better than Dijkstra, ME/DLE, and random routing protocols after experimental evaluation.

- (g) Epidemic Oracle routing protocol: Jain et al. [46] presented it and placed under the category of knowledge-based routing protocols as it maintains a history of all participating nodes in the entire opportunistic network. Epidemic Oracle routing protocol carries the message to be forwarded until there would be enough probability of delivering the message to the right destination. It has a knowledge database concerning future connectedness; therefore, it falls under the Knowledge-based routing protocols. Epidemic Oracle routing protocol delivers a better delivery probability than Epidemic protocol, PRoPHET protocol, and Direct Delivery protocol, but sometimes it may lead to prolonged delays in a case when there would be no sufficient probability of delivering the message to the right destination.

Like other routing protocols, one of the major objectives of this protocol is to achieve optimum delay. The expected delay (ED) in Epidemic Oracle routing protocol is described as follows:

$$ED_{\text{Epidemic Oracle}} = \min \left( \sum_{v \in V} \sum_{k \in K^v} \sum_{I_q \in I_E} (t_{q-1} - \omega(k)) \cdot \left( \sum_{e \in I^v} R_{e, I_q}^k - \sum_{e \in O^v} X_{e, I_q}^k \right) \right), \quad (7)$$

where  $\sum_{e \in I^v} R_{e, I_q}^k \rightarrow$  summation of data segments of message  $K$  which is received by node  $v$  in time duration of  $I_q$ ,  $\sum_{e \in O^v} X_{e, I_q}^k \rightarrow$  summation of all data segments of message  $K$  transmitted over edge  $e$  during the time duration of  $I \in I_E$ , and  $(t_{q-1} - \omega(k)) \rightarrow$  time duration consumed since the start of message transfer.

- (h) FRESH routing protocol: Dubois-Ferriere et al. [49] proposed FResher Encounter Search (FRESH) algorithm for path discovery in opportunistic networks in an efficient manner. In FRESH, participating nodes maintain a record of their most recent encounter times with other nodes. When a node requires forwarding the message to some another node, then it searches for any intermediate node with having the most recent encounters and forward the message to it. The intermediate node follows the same process up to when the message reaches the desired location.

Calculating search cost ( $C_s$ ) in finding a route between source  $s$  towards destination  $d$  can be

considered as a composition of several successive searches and the mathematical expression is given in the following equation:

$$C_s = \sum_{i=s}^{i=d} (\alpha |X_i - X_{i+1}|)^2, \quad (8)$$

where  $X_i \rightarrow$  positions of  $i^{\text{th}}$  node and  $\alpha \rightarrow$  radius of the search area.

FRESH protocol results in cheap route discovery by replacing a single whole network search to the series of small searches. It has been found experimentally that FRESH protocol reduces the flooding overhead to a considerable extent.

- (i) RAPID routing protocol: Balasubramanian et al. [50] presented the RAPID protocol to maximize the performance of specific performance parameters. It uses the utility function ( $U_i$ ) to assign utility value to every message on the ground of the average delay parameter. Primarily, it involves routing of packets

by replicating until a copy reaches to the destination node. At a transfer opportunity, it repeats a packet that locally brings about the most noteworthy increment in utility. By and large,  $U_i$  is characterized as the expected contribution of  $i$  to the given routing metric. For instance, the metric limit average delay is estimated by adding the delay of packets. In like manner, the utility of a packet is its normal delay. Along these lines, RAPID is a heuristic dependent on locally improving marginal utility, i.e., the normal increment in utility per unit resource utilized. RAPID imitates packets in diminishing requests of their marginal utility at each transfer opportunity.

The equation for calculating expected delay (ED) to deliver  $I$  is expressed as follows:

$$ED_{\text{RAPID}} = \left[ \sum_{j=1}^k \frac{1}{E(M_{x_j,z}) \cdot n_j(i)} \right]^{-1}, \quad (9)$$

where  $E(M_{x_j,z}) \rightarrow$  expected time of  $x_j$  to reach node  $z$  and  $n_j(i) \rightarrow$  the number of times each of the  $j$  nodes, respectively, required to contact the destination to deliver  $I$  directly.

It has been found through simulation that RAPID performs better than MaxProp, Spray-and-Wait, PROPHET on the ground of average delay, packet delivery ratio, and overall efficiency in opportunistic networks.

#### 4. Simulation Trends in Opportunistic Networks

Various researchers developed numerous simulators and made them available for simulation purposes. Some of the popular simulation tools are as shown in Table 1.

In addition to the abovementioned simulators, various custom-built simulators are also being an option for pursuing research in opportunistic networks. These simulators help to share original coding work for its reuse in the future. Few examples of such simulators are MONICA [57], E-ONE [58], and UDTNSim [59].

Kuppusamy et al. [15] surveyed the simulation trend followed by researchers focusing on opportunistic networks. The results reveal that there has been a substantial increase in the use of ONE simulator during the current decade. Figure 5 presents the contribution of different available simulators towards OppNet research.

Also, Kuppusamy et al. [15] have brought in to light the fact that the foremost reason for selecting ONE simulator as a major tool is that it is capable of supporting the maximum number of participating nodes during simulation among all discussed simulators (except custom-based simulators). However, it has some limitations regarding the underlay layers such as the MAC sublayer, but that can be ignored for the research work of this paper. Further, it is also found as the most accurate simulator which allows the researcher to get results with the maximum number of QoS parameters among all its counterparts.

Based on the abovementioned details, it may be stated that ONE (Opportunistic Network Environment) simulator is the most widely used simulator among researchers. Therefore, this paper uses ONE simulator for the implementation of the mentioned routing protocols aiming to cover a large group of researchers engaged in the opportunistic network research domain.

**4.1. ONE Simulator.** ONE (Opportunistic Network Environment) is a Java-based discrete event simulator whose main functions are node movement modeling, routing, message-handling, and internode contacts, Result collection and analysis are achieved through visualization and other postprocessing tools.

The results which are generated as a result of the simulation are generally logs of events that are further processed by external tools such as Graphviz for plotting graphs. Figure 6 illustrates the simulation environment of ONE simulator.

The popularity of ONE simulator is because it provides various tools to generate difficult mobility scenarios that are closer to real-life situations than any other available simulator in current time. Some of its features such as GPS Map data and Working Day Movement Model make it a better option to reality.

However, still, ONE simulator has some challenges; for example, the message generation process may perform better if group relationship and context information be added. Also, it must be mentioned here that several research groups are putting their efforts into enhancing supporting features in ONE simulator. Maybe, a newer version of ONE simulator will be added with some better features.

#### 5. Performance Evaluation of Routing Protocols for Smart Spaces

Rigorous review uncovers the fact that though numerous routing protocols have been proposed by various eminent researchers, yet none of them has quantitatively evaluated them. The authors in this work showed that there is an urgent need to do the same to determine which protocol is best suited in a given environment. Keeping this in mind, the authors have meticulously compared the numerous protocols of opportunistic networks.

In this section, nine different routing protocols are compared based on standard Quality-of-Service (QoS) parameters by varying the number of participating nodes. It is believed that this simulation comparison will describe the performance behavior of different protocols on the ground of node density [26, 38]. The main purpose of choosing node density as a primary factor is that it correlates to the real-life scenario of Smart Spaces very closely, for example, if we consider mobile handset device as participant node connected to OppNet via Bluetooth/Wi-Fi, then it would be around 50–60 nodes per square km in case of a park in opportunistic networks, but it can increase up to 500 or more in the situation of a conference hall. If we talk about interplanetary communication, then the number of

TABLE 1: Popular simulators used in opportunistic networks.

Simulator	Brief description
Adyton [51]	(i) C++-based event-driven simulator (ii) Supports numerous routing protocols and real-world contact traces (iii) Also provides several congestion control mechanisms and buffer management policies
MobEmu [52]	(i) Java-based free simulator (ii) Capable of executing a mobility model or replay a trace, while applying the desired routing or dissemination algorithm
Ns-3 [53]	(i) Python-based free simulator under the GNU GPLv2 license (ii) Also supports a real-time scheduler that facilitates several “simulation-in-the-loop” use cases for interacting with real systems
OMNet++ [54, 55]	(i) Free only for noncommercial organizations (ii) An extensible, component-based C++ simulation framework (iii) Runs basically on all platforms where a modern C++ compiler is available
ONE [56]	(i) Java-based free simulator (ii) Offers both keyboard and GUI interface for coding (iii) It allows researchers to design new protocols/architecture/framework in a very easy and defined way

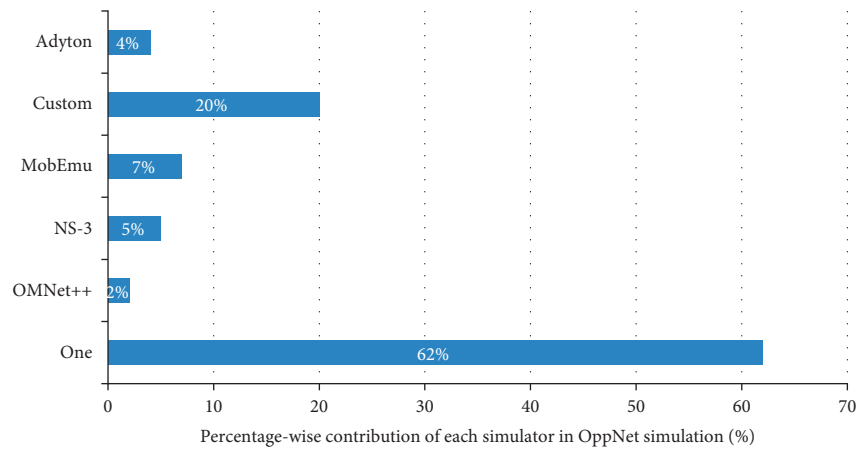


FIGURE 5: Contribution of available simulators towards OppNet research.

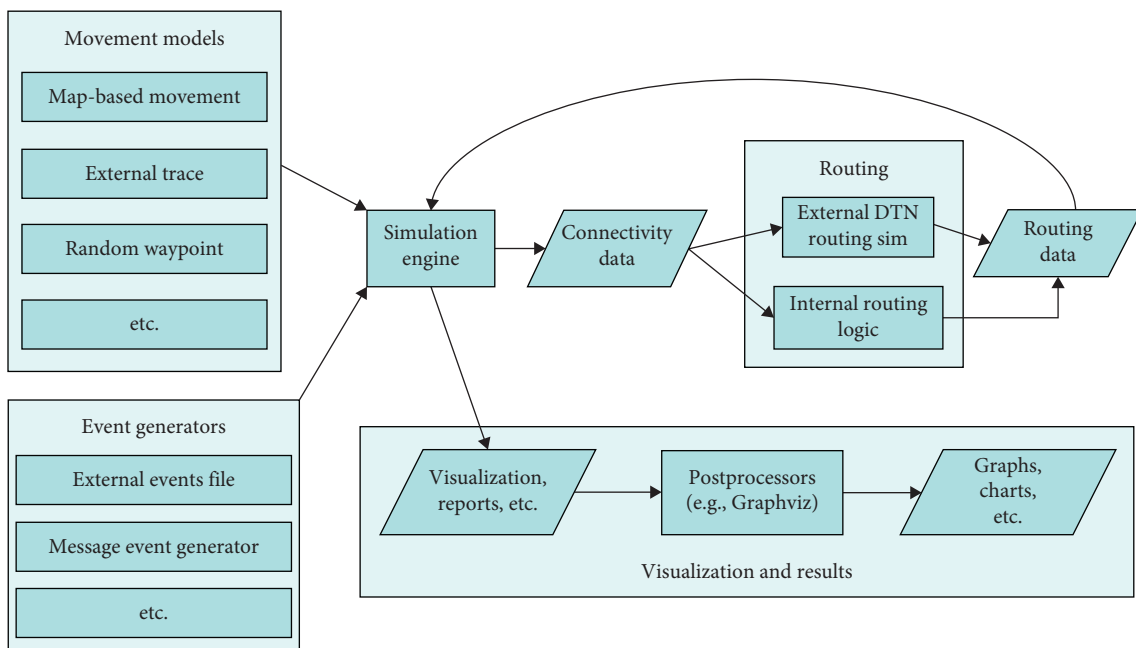


FIGURE 6: One simulation environment (source: Keränen et al. [56]).

participant nodes would be at most 1 or even less than 1 per square km. The ratio of the number of nodes per square km becomes 100–150 when a normal highway situation is considered. It may be increased up to 200–250 when a busy pedestrian path is taken as an example. There are many other real-life situations as well where node density differs as per the environment which directly affects the overall performance of Smart Spaces. Therefore, this paper aims to find the suitability of the protocol being used in different scenarios. The simulation comparison will result in the performance of different protocols in different node densities; it will help upcoming researchers to choose routing protocol accordingly for establishing different Smart Spaces.

*5.1. Common Parameters Used in Each Case.* Besides the variation in node density with a different routing protocol, several other parameters are kept constant to analyze the performance change only due to the change in several participating nodes. The details of these parameters are listed in Table 2.

*5.2. Quality-of-Service (QoS) Parameters Used.* The comparison needs some standard parameters, so that the performance comparison could explain which is better and which is worse [60–64]. Table 3 explains the various standard parameters that have been taken to decide the behavior of routing protocols for a different number of participating nodes.

- (1) Number of participating nodes: as already mentioned earlier, this paper aims to inspect the behavioral change of different routing protocols when several participating nodes vary to observe its suitability for setting up different Smart Spaces of different capabilities in different places for different purposes. The simulation study starts with several nodes 50 and it ends to 500 nodes in a particular simulation area ( $4500 \times 3400$  square meters). Therefore, this paper considers it as the primary QoS metric for judging the quality of different routing protocols under various situations concerning node density.
- (2) Message delivered: as per the message delivered is concerned, it is observed that FRESH routing protocol shows the best behavior for 50–250 nodes, but Spray-and-Wait protocol takes the lead as the participating nodes grow from 270 to 500. PRoPHET protocol also shows good performance after FRESH protocol for 100–250 nodes, but it lags when the number of nodes grows more than 250. Also, MaxProp starts giving good results when the number of nodes grows from 250 nodes. First Contact protocol behaves worse from starting values to the end, and also Direct Delivery protocol showed similar performance but somewhat better than First Contact protocol. Comparative performance may

also be viewed by the following graph depicted in Figure 7.

- (3) Message dropped: it may be easily analyzed that FRESH protocol shows best behavior, i.e., zero message dropped in every case of node density. Direct Delivery protocol also shows the same but only when the number of nodes becomes more than 150. First Contact protocol also delivers minimal message drop after the FRESH protocol and Direct Delivery protocol. Epidemic protocol gives the highest most message drops up to the node density of 350 and also in the case when the number of nodes is equal to 450. Besides Epidemic, PRoPHET protocol follows the higher drop rate and delivers the highest message drops in node density of 500. The comparative graph is available for reference as depicted in Figure 8.
- (4) Delivery probability: from the observations, it can be clearly stated that FRESH protocol shows the best performance up to the node density of 270. Spray-and-Wait protocol leads after the number of nodes is equal to 270 to 500, whereas the First Contact protocol and Direct Delivery protocol shows the worse performance on every case of the node number. The comparative graph is available for reference as depicted in Figure 9.
- (5) Overhead ratio: the Direct Delivery protocol exhibits zero overhead ratio. Spray-and-Wait protocol also delivers the second-most lowest overhead ratio after the Direct Delivery protocol in every case of node density. As far as other protocols are concerned, Epidemic Oracle Protocol delivers almost maximum overhead ratio up to the node density of 150, Epidemic protocol leads after that, and PRoPHET protocol also exhibits the high overhead ratio but somewhat less than Epidemic for the number of nodes higher than 150 and Epidemic Oracle for the number of nodes less than 150. FRESH protocol also delivers a very less overhead ration up to the 200 nodes in the simulation area, but it slightly changes its behavior when it grows from 200 towards 500 in node density. Refer to the graph as depicted in Figure 10.
- (6) Average latency: from the given graph as depicted in Figure 11, it may be easily concluded that Epidemic protocol gives the least average latency in the case of 50 nodes, but it surprisingly changes the behavior from least to higher values when many nodes grow from 50 to 100. Similarly, Direct Delivery protocol also behaves surprisingly. It exhibits considerable low average latency in case of node density of 50, 150, 250, and 350–450 as compared to the case when the number of nodes is 100, 200, 300, and 500. MaxProp and RAPID protocols consistently show a huge average latency in every case of node density. FRESH protocol shows the best performance in this

TABLE 2: Common parameters used in OppNet simulation.

Parameters	Values
Simulation area	4500 × 3400 sq. meters (15.3 square km)
Simulation time	10 hours
Movement model	Random waypoint movement
Time-to-live (per message)	240 minutes
Scenario update interval	0.1 second
Communication medium	Bluetooth and Wi-Fi (high speed)
Bluetooth interface speed	250 kbps
Bluetooth interface range	10 meters
Bluetooth interface scan interval	32 seconds
Wi-Fi interface speed	500 kbps
Wi-Fi interface range	10 meters
Wi-Fi interface interval	64 seconds
Node movement speed	From 0.5 m/s to 1.5 m/s
Transmission range	10 meters
Message size	From 500 KB to 1 MB
Warm-up period	1800 seconds
Buffer size	5 MB
Operating system	The mentioned research is carried out on MS Windows 10 platform, but the ONE simulator is a Java-based application; therefore, its performance is independent of the platform being used

TABLE 3: Performance metrics used in OppNet simulation.

S. no.	QoS parameters	Mathematical notations	Brief description
1	Number of participating nodes	$n$	Involves the total number of participating nodes in the network including those nodes also which are not participating in communication at any instance.
2	Message delivered	$M_{\text{delivered}} = \sum_{i=1}^{i=n} (M_{\text{created}_i} - M_{\text{dropped}_i})$	The total number of messages successfully delivered by the sender node to the destination node via the intermediate node. The number of intermediate nodes may vary following network topology and the distance between sender and receiver end.
3	Message dropped	$M_{\text{dropped}} = \sum_{i=1}^{i=n} (M_{\text{created}_i} - M_{\text{delivered}_i})$	The total number of messages lost due to any reason during communication between the sender node and receiver node in opportunistic networks.
4	Delivery probability	$P_{\text{delivery}} = ((\sum_{i=1}^{i=n} (M_{\text{created}_i} - M_{\text{dropped}_i})) / (\sum_{i=1}^{i=n} M_{\text{created}_i}))$	It is the probability of successful delivery of a data packet originated from source node directed towards destination node via intermediate nodes and is a pointer of how solid the network is as far as message delivery.
5	Overhead ratio	Overhead ratio = $((\sum_{i=1}^{i=n} M_{\text{relayed}_i} - \sum_{i=1}^{i=n} M_{\text{delivered}_i}) / (\sum_{i=1}^{i=n} M_{\text{delivered}_i}))$	The overhead ratio suggests the utilization of network resources and buffer space because of the utilization of different duplicates of a similar message to expand delivery possibilities.
6	Average latency	Latency <sub>avg</sub> = $((\sum_{i=1}^{i=n} T_{\text{successful delivery}_i}) / (\sum_{i=1}^{i=n} M_{\text{delivered}_i}))$	The average time is taken by a data to be completely disseminated from a source node to the destination node. Less average latency denotes a good characteristic of a good routing protocol.
7	Average hop count	Hop count <sub>avg</sub> = $\sum_{i=1}^{i=n} M_{\text{exchanged}_i} / n$	The average number of an intermediate number of nodes traveled by data from source to destination in a predefined duration of time.
8	Average buffer time	Buffer time <sub>avg</sub> = $(\sum_{i=1}^{i=n} T_{\text{delivered}_i} / n)$	It is the average time brought about by all messages that are delivered relinquished/stranded at the intermediate node buffers.

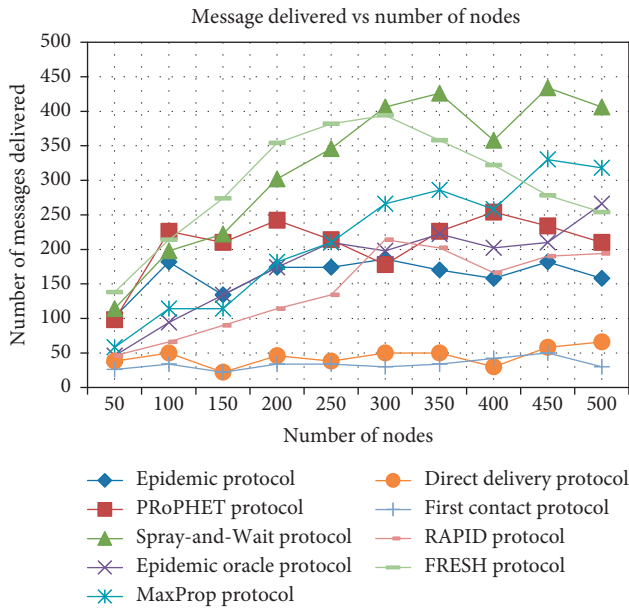


FIGURE 7: Protocols’ performance based on message delivered versus number of nodes.

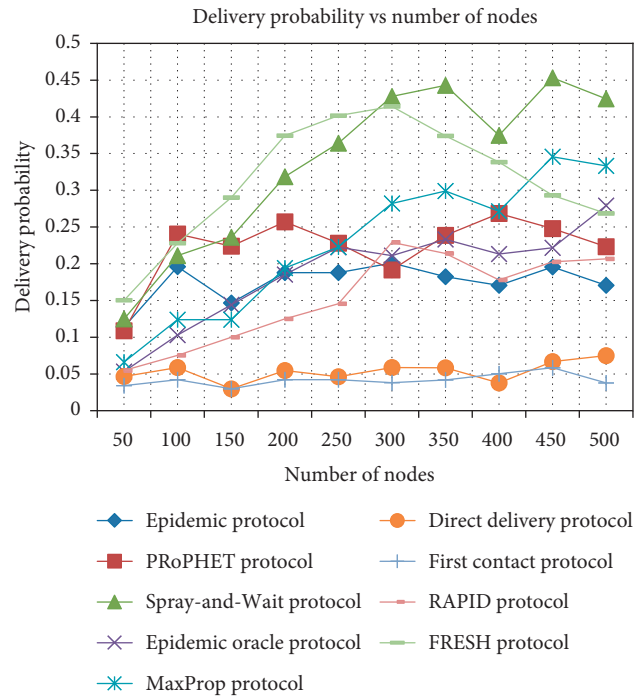


FIGURE 9: Protocols’ performance based on delivery probability versus number of nodes.

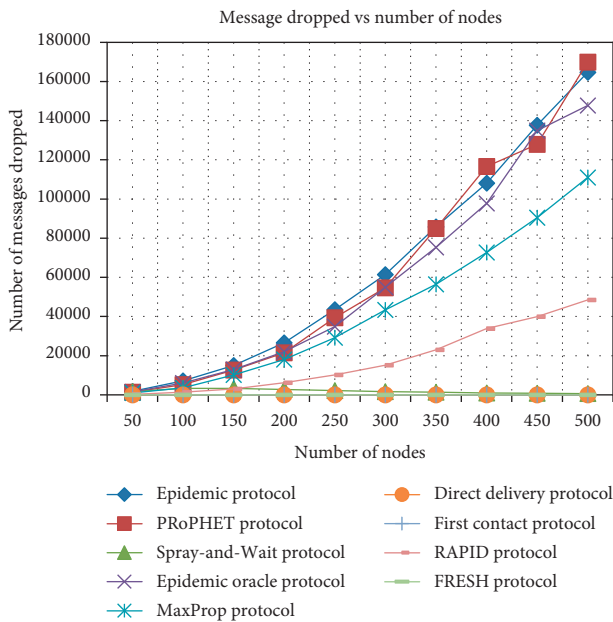


FIGURE 8: Protocols’ performance based on dropped message versus number of nodes.

parameter by exhibiting low average latency in each case of node density.

(7) Average hop count: as far as the average hop count is concerned, it may be summarized that Direct Delivery as per its methodology delivers messages only with one hop count. If this protocol is sided apart, then Spray-and-Wait and MaxProp protocols require the least average hop count to deliver a message in every case of node density. PRoPHET protocol also requires less average hop count but after Spray-

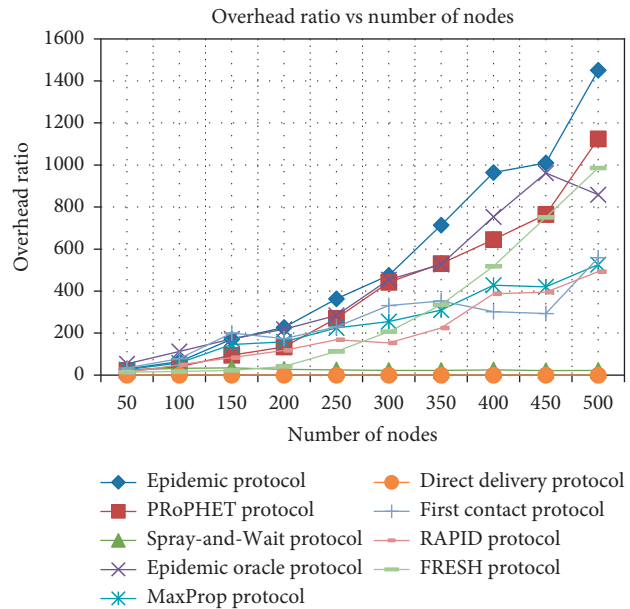


FIGURE 10: Protocols’ performance based on overhead ratio versus number of nodes.

and-Wait and MaxProp protocols. Epidemic Oracle protocol requires the highest average hop count except in the case of node density of 320–410. First Contact protocol also requires the highest average hop count for node density of 300–400, but its requirement surprisingly falls when the number of nodes is from 400 towards the higher side. FRESH protocol exhibits a consistent behavior in this

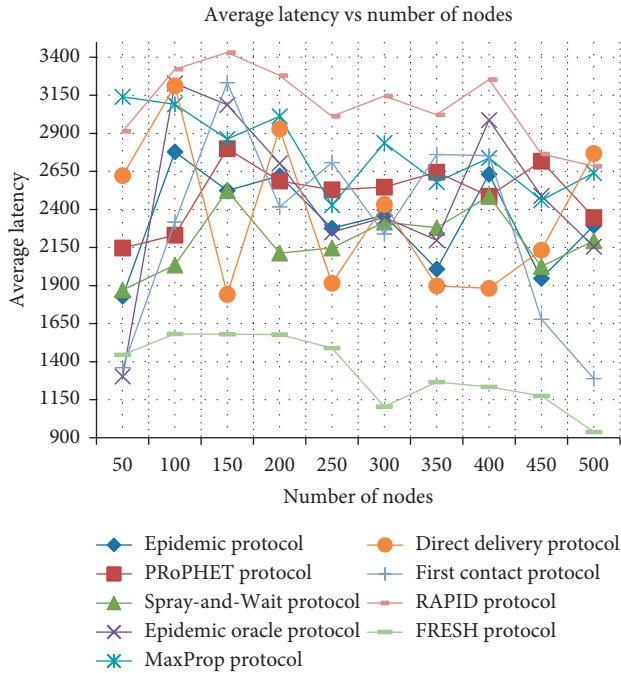


FIGURE 11: Protocols’ performance based on average latency versus number of nodes.

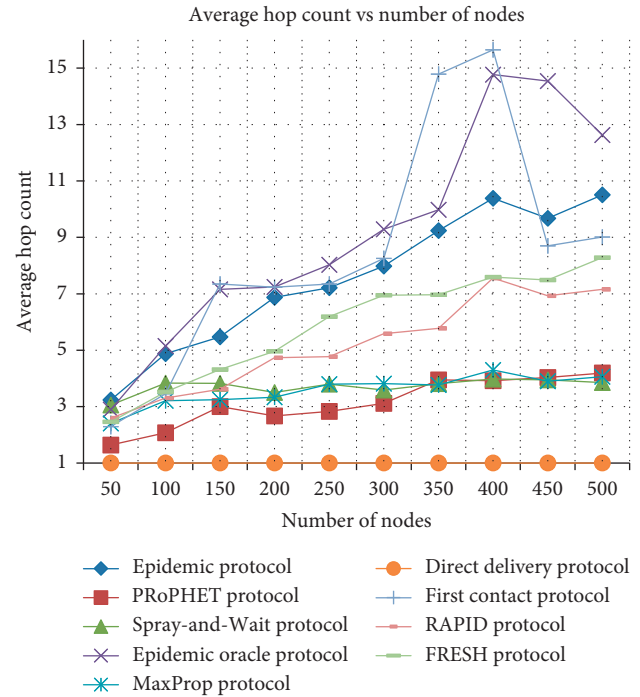


FIGURE 12: Protocols’ performance based on average hop count versus number of nodes.

parameter; it continually grows as the number of nodes grows. See the graph as depicted in Figure 12.

- (8) Average buffer time: as per the buffer time is concerned, it can be easily understood that the Direct Delivery protocol needs least most buffer time when the number of nodes is greater than 150. FRESH protocol also requires considerably less buffer time throughout every case of node density. Also, it has been observed that the requirement of average buffer time of almost every routing protocol (observed here) reduces as the number of nodes increases except Spray-and-Wait protocol; rather, it shows that it requires more buffer time as the number of nodes grows. See the comparative graph as depicted in Figure 13.

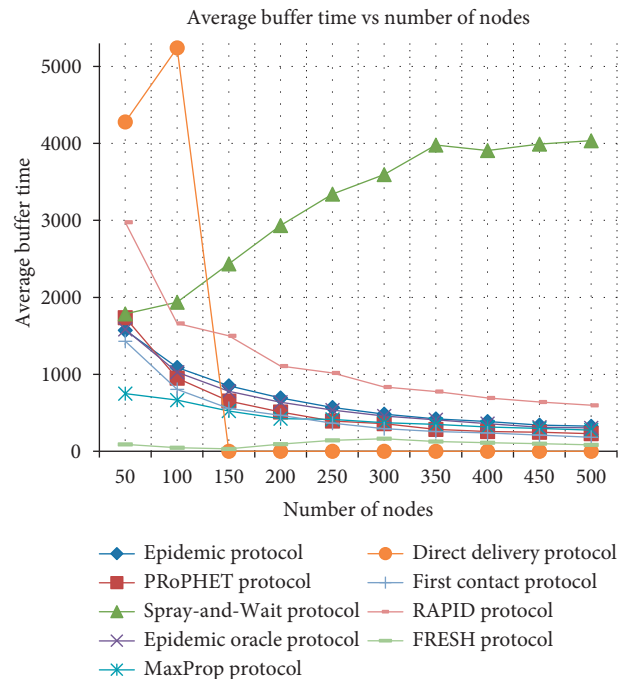


FIGURE 13: Protocols’ performance based on average buffer time versus number of nodes.

5.3. Analysis of Performance Evaluation. Total nine protocols (Epidemic protocol, PProPHET protocol, Spray-and-Wait protocol, Epidemic Oracle protocol, MaxProp protocol, Direct Delivery protocol, First Contact protocol, RAPID protocol, and FRESH protocol) among the six different categories of routing protocols of opportunistic networks (flooding-based routing protocols, forwarding-based routing protocols, probability-based routing protocols, knowledge-based routing protocols, social relationship-based routing protocols, and hybrid routing protocols) over eight different QoS parameters (number of participating nodes, message delivered, message dropped, delivery probability, overhead ratio, average latency, average hop count, and average buffer time) have been thoroughly explored in this

paper. Besides it, some protocols are observed superior in particular cases of node density concerning mentioned QoS parameters. A comprehensive evaluation of the suitability of routing protocols over node density has been prepared in Table 4 based on our simulation study.

TABLE 4: Analysis of routing protocols over node density.

Protocol	Extremely sparse environment (3–5 nodes per square km)	Sparse environment (6–15 nodes per square km)	Average environment (16–25 nodes per square km)	Populated environment (26–400 nodes per square km)	Dense environment (more than 400 nodes per square km)
Epidemic protocol	✓		✓		
Spray-and-Wait protocol	✓	✓	✓	✓	✓
Direct Delivery protocol	✓	✓	✓	✓	
First Contact protocol		✓	✓		
PRoPHET protocol		✓			
MaxProp protocol	✓		✓	✓	✓
Epidemic Oracle protocol	✓				
FRESH protocol	✓	✓	✓	✓	✓
RAPID protocol	✓	✓			

Furthermore, below are some findings taken after the simulation study of mentioned protocols in a different environment about node density over discussed QoS parameters:

- (1) Epidemic protocol: it is a simple protocol that replicates the maximum message to be forwarded. It provides the best delivery probability in case of node density of populated environment and its average buffer time reduces as the node density increases.
  - (2) Spray-and-Wait protocol: it must be termed as the finest routing protocol along with FRESH routing protocol among all discussed protocols in this paper. It delivers good performance and proved as a better option in all varieties of node density. It is also found better than FRESH protocol based on the criteria of overhead ratio, average hop count, message delivered, and delivery probability.
  - (3) Direct Delivery protocol: it is the simplest protocol OppNet can have. It is a good protocol up to the node density of 400 nodes per square km. It provides zero packet drop and average delivery probability throughout all node densities.
  - (4) First Contact protocol: First Contact protocol is suited well in node density ranging from 6 to 25 nodes per square kilometer. Packet dropping and high delays are some problems that arise in this protocol when node density increases.
  - (5) PRoPHET protocol: PRoPHET protocol functions of the basis of the history of encounters and transitivity. Due to this feature, it achieves high delivery probability as compared to many traditional protocols. Like Epidemic, its average buffer time reduces as the node density increases.
  - (6) MaxProp protocol: the primary objective of this protocol is to improvise delivery rate and average latency. It functions by ranking the stored packets in nodes' memory on the ground of cost assigned. It is proved a good choice in almost all categories of node densities in opportunistic networks.
  - (7) Epidemic Oracle protocol: it falls under knowledge-based routing protocols as it maintains a knowledge database of all participating nodes in the entire opportunistic networks which makes it a good option for Sparse Opportunistic Network Environment. But its performance degrades when the number of nodes increases due to the overhead of knowledge database.
  - (8) FRESH protocol: it is also the finest routing protocol along with Spray-and-Wait routing protocol in every case of node density discussed here. When compared to the Spray-and-Wait protocol, FRESH protocol is found better on the criteria of the number of messages dropped, average latency, and average buffer time.
  - (9) RAPID protocol: this protocol uses utility function to assign utility value to every message on the ground of average delay parameter which helps it to perform well in every situation, but it is proved a good choice only for the extremely sparse and sparse environment when compared to the other routing protocols discussed in this paper.
- It has been experimentally observed that two of the routing protocols (Spray-and-Wait and FRESH protocols) are found finest in every instance of node density. However, these routing protocols are leading as well as lagging to each other based on individual QoS parameters. The comparative



TABLE 5: Performance comparison of FRESH and Spray-and-Wait protocols on standard QoS metrics.

QoS metrics	Spray-and-Wait protocol	FRESH protocol
Message delivered	Best	May be improved
Message dropped	May be improved	Best
Delivery probability	Best	May be improved
Overhead ratio	Best	May be improved
Average latency	May be improved	Best
Average hop count	Best	May be improved
Average buffer time	May be improved	Best

performances of the two finest emerged protocols have been presented along with QoS metrics in Table 5.

From the observations made from Table 5, it can be summarized that Spray-and-Wait and FRESH protocols are exhibiting their best performance on different QoS metrics. However, due to the absence of common QoS parameters, it is hard to declare one protocol to be best among the discussed protocols on the mentioned criteria, but, for making a viable conclusion, one protocol must be declared as the best one.

To cope up with this problem, the importance of individual QoS metric must be evaluated based on its suitability for the best performance under Smart Space Environment. Literature survey [65, 66] reveals the fact that certain QoS metrics such as message dropped, average latency, and average buffer time are of significant importance for accurate and fast delivery with least additional storage requirement other than data packet to be transmitted which is an ideal condition for receiving best results in Smart Space Environment. In light of this fact, it can be stated that FRESH protocol is optimally suited to Smart Space Environment.

## 6. Conclusion and Future Scope

This paper in depth explores various routing protocols of opportunistic networks which can be used for establishing Smart Spaces. It also discusses in detail various simulation trends prevailing in the arena of opportunistic networks. The protocols are thereafter compared based on node density to determine the best-suited protocol for building Smart Spaces in the given simulation environment. The paper concludes with the fact that Spray-and-Wait outperforms the FRESH protocol by giving better results on the 5 standard QoS parameters. However, with the eye and muscle of Smart Space Environment, the paper also highlights the fact that certain QoS metrics such as message dropped, average latency, and average buffer Time are of significant importance for getting the best outcome in Smart Space Environment. Therefore, FRESH protocol must be considered as the best routing protocol suited for Smart Spaces Environment.

## 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 no conflicts of interest regarding the publication of this article.

## Acknowledgments

The simulation work of this paper was carried out in the Research Lab of J. C. Bose University of Science and Technology YMCA, Faridabad, Haryana, India. The authors are thankful to J. C. Bose University of Science and Technology YMCA, Faridabad, Haryana, India for permitting them to use the Lab Facilities. The authors are also thankful to the National Council for Scientific and Technological Development (CNPq) for the support received via grant no. 305805/2017-7.

## References

- [1] L. Wood, W. M. Eddy, W. Ivancic, J. McKim, and C. Jackson, "Saratoga: a delay-tolerant networking convergence layer with efficient link utilization," in *Proceedings of the 2007 International Workshop on Satellite and Space Communications*, pp. 168–172, IEEE, Salzburg, Austria, September 2007.
- [2] S.-L. Shaw and D. Sui, "Understanding the new human dynamics in smart spaces and places: toward a spartial framework," *Annals of the American Association of Geographers*, vol. 110, no. 2, pp. 339–348, 2020.
- [3] A. Socievole, A. Caputo, F. De Rango, and P. Fazio, "Routing in mobile opportunistic social networks with selfish nodes," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 6359806, 15 pages, 2019.
- [4] J. Shu, J. Xiong, L. Xu, X. Geng, and L. Liu, "An improved opportunistic sensor networks connectivity monitoring model based on network connectivity," in *China Conference on Wireless Sensor Networks*, pp. 64–76, Springer, Singapore, 2018.
- [5] A. Bujari, S. Gaito, D. Maggiorini, C. E. Palazzi, and C. Quadri, "Delay tolerant networking over the metropolitan public transportation," *Mobile Information Systems*, vol. 2016, Article ID 8434109, 14 pages, 2016.
- [6] V. Kushwaha and R. Gupta, "Delay tolerant networks: architecture, routing, congestion, and security issues," in *Handbook of Research on Cloud Computing and Big Data Applications in IoT*, pp. 448–480, IGI Global, Hershey, PA, USA, 2019.
- [7] S. Prabhavat, W. Narongkhachavana, T. Thongthavorn, and C. Phankaew, "Low overhead localized routing in mobile ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 9652481, 15 pages, 2019.
- [8] J. Molina-Gil, P. Caballero-Gil, and C. Caballero-Gil, "Comparative study of cooperation tools for mobile ad hoc networks," *Mobile Information Systems*, vol. 2016, Article ID 3435674, 9 pages, 2016.
- [9] S. H. Ahmed, A. K. Bashir, M. Elhoseny, W. Guibene, and S. H. Bouk, "Research on efficient data forwarding in vehicular networks," *Mobile Information Systems*, vol. 2019, Article ID 2353478, 2 pages, 2019.
- [10] H. Patel, D. Singh Rajput, G. Thippa Reddy, C. Iwendi, A. Kashif Bashir, and O. Jo, "A review on classification of imbalanced data for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 4, 2020.

- [11] A. K. Bashir, A. H. Akbar, S. A. Chaudhary, C. S. Hussain, and K. H. Kim, "Collaborative detection and agreement protocol for routing malfunctioning in wireless sensor networks," in *Proceedings of the 2006 8th International Conference Advanced Communication Technology*, vol. 1, pp. 327–332, IEEE, Phoenix Park, South Korea, February 2006.
- [12] A. Nayyar, R. S. Batth, D. B. Ha, and G. Sussendran, "Opportunistic networks: present scenario-a mirror review," *International Journal of Communication Networks and Information Security*, vol. 10, no. 1, pp. 223–241, 2018.
- [13] A. U. H. Yasar, H. Malik, and Z. Khan, "Mobile sensor networks applications and confidentiality," *Mobile Information Systems*, vol. 2015, Article ID 893438, 2 pages, 2015.
- [14] M. Cobos, F. Antonacci, A. Mouchtaris, and B. Lee, "Wireless acoustic sensor networks and applications," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 1085290, 3 pages, 2017.
- [15] V. Kuppusamy, U. Thanthrige, A. Udugama, and A. Förster, "Evaluating forwarding protocols in opportunistic networks: trends, advances, challenges and best practices," *Future Internet*, vol. 11, no. 5, p. 113, 2019.
- [16] S. Shelke and B. Aksanli, "Static and dynamic activity detection with ambient sensors in smart spaces," *Sensors*, vol. 19, no. 4, p. 804, 2019.
- [17] C. R. Lynch and V. J. Del Casino Jr., "Smart spaces, information processing, and the question of intelligence," *Annals of the American Association of Geographers*, vol. 110, no. 2, pp. 382–390, 2020.
- [18] F. Al-Turjman, "5G-enabled devices and smart-spaces in social-IoT: an overview," *Future Generation Computer Systems*, vol. 92, pp. 732–744, 2019.
- [19] M. Samaniego, C. Espana, and R. Deters, "Suspicious transactions in smart spaces," 2019, <https://arxiv.org/abs/1909.10644>.
- [20] F. Alrimawi, L. Pasquale, and B. Nuseibeh, "On the automated management of security incidents in smart spaces," *IEEE Access*, vol. 7, pp. 111513–111527, 2019.
- [21] E. Ismagilova, L. Hughes, N. Rana, and Y. Dwivedi, "Role of smart cities in creating sustainable cities and communities: a systematic literature review," in *International Working Conference on Transfer and Diffusion of IT*, pp. 311–324, Springer, Cham, Switzerland, 2019.
- [22] S. A. Al Ayyat, S. G. Aly, and K. A. Harras, "On integrating space syntax metrics with social-aware opportunistic forwarding," in *Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–7, IEEE, Marrakesh, Morocco, April 2019.
- [23] L. M. Funes, Systems and methods for smart spaces, U.S. Patent No. 10168677, U.S. Patent and Trademark Office, Washington, DC, USA, 2019.
- [24] M. Gelsomini, G. Leonardi, and F. Garzotto, "Embodied learning in immersive smart spaces," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, Honolulu, HI, USA, April 2020.
- [25] J. W. Lee and S. Helal, "Context awareness computing in smart spaces using stochastic analysis of sensor data," in *Advances in Intelligent Systems and Computing*, pp. 3–9, Springer, Cham, Switzerland, 2019.
- [26] S. E. Loudari and N. Benamar, "Effects of selfishness on the Energy consumption in opportunistic networks: a performance assessment," in *Proceedings of the 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, pp. 1–7, IEEE, Fez, Morocco, April 2019.
- [27] M. Saloni, C. Julien, A. L. Murphy, and G. P. Picco, "Lasso: a device-to-device group monitoring service for smart cities," in *Proceedings of the 2017 International Smart Cities Conference (ISC2)*, pp. 1–6, IEEE, Wuxi, China, September 2017.
- [28] T. Small and Z. J. Haas, "The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way)," in *Proceedings of the 4th ACM International Symposium on Mobile ad Hoc Networking & Computing*, ACM, Annapolis, MD, USA, pp. 233–244, June 2003.
- [29] C. Detweiler, I. Vasilescu, and D. Rus, "An underwater sensor network with dual communications, sensing, and mobility," in *Proceedings of the Oceans 2007-Europe*, pp. 1–6, IEEE, Aberdeen, UK, June 2007.
- [30] M. Martonosi, *The Princeton Zebrant Project: Sensor Networks for Wildlife Tracking*, Princeton University, Princeton, NJ, USA, 2004.
- [31] C. Papadaki, T. Kärkkäinen, and J. Ott, "Composable distributed mobile applications and services in opportunistic networks," in *Proceedings of the 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pp. 14–23, IEEE, Chania, Greece, June 2018.
- [32] V. G. Menon and P. J. Prathap, "Comparative analysis of opportunistic routing protocols for underwater acoustic sensor networks," in *Proceedings of the 2016 International Conference on Emerging Technological Trends (ICETT)*, pp. 1–5, IEEE, Kollam, India, October 2016.
- [33] Z. Rahman, F. Hashim, M. F. A. Rasid, and M. Othman, "Totally opportunistic routing algorithm (TORA) for underwater wireless sensor networks," *PLoS One*, vol. 13, no. 6, Article ID e0197087, 2018.
- [34] <https://www.airbornewirelessnetwork.com/index.asp#advantages>.
- [35] [https://www.nasa.gov/mission\\_pages/station/research/experiments/explorer/Investigation.html?id=717](https://www.nasa.gov/mission_pages/station/research/experiments/explorer/Investigation.html?id=717).
- [36] M. Zhou, Q. Liang, H. Wu, W. Meng, and K. Xu, "Wireless sensor networks for smart communications," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 4727385, 2 pages, 2018.
- [37] M. Xie, Y. Bai, Z. Hu, and C. Shen, "Weight-aware sensor deployment in wireless sensor networks for smart cities," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 5913836, 15 pages, 2018.
- [38] T. Gautam and A. Dev, "Opportunistic network routing protocols: challenges, implementation and evaluation," in *Proceedings of the 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 100–106, Noida, India, January 2019.
- [39] D. Chen, G. Navarro-Arribas, C. Pérez-Solà, and J. Borrell, "Message anonymity on predictable opportunistic networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 1–14, 2019.
- [40] T. E. Amah, M. Kamat, K. Abu Bakar, W. Moreira, A. Oliveira Jr., and M. A. Batista, "Preparing opportunistic networks for smart cities: collecting sensed data with minimal knowledge," *Journal of Parallel and Distributed Computing*, vol. 135, pp. 21–55, 2020.
- [41] V. Juyal, N. Pandey, and R. Saggarr, "Impact of varying buffer space for routing protocols in delay tolerant networks," in *Proceedings of the 2016 International Conference on Communication and Signal Processing (ICCSP)*, pp. 2152–2156, IEEE, Melmaruvathur, India, April 2016.
- [42] A. Vahdat and D. Becker, "Epidemic routing for partially connected Ad Hoc networks," 2000.

- [43] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Single-copy routing in intermittently connected mobile networks," in *Proceedings of the 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004*, pp. 235–244, IEEE, Santa Clara, CA, USA, October 2004.
- [44] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and wait: an efficient routing scheme for intermittently connected mobile networks," in *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking—WDTN'05*, ACM, Philadelphia, PA, USA, pp. 252–259, August 2005.
- [45] P. Dwivedi and R. S. Pippal, "Performance evaluation of spray and wait DTN routing protocol under different mobility models," *Journal of the Gujarat Research Society*, vol. 21, no. 16, pp. 186–192, 2019.
- [46] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 145–158, 2004.
- [47] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," in *Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc*, Annapolis, MD, USA, June 2003.
- [48] J. Burgess, B. Gallagher, D. D. Jensen, and B. N. Levine, "MaxProp: routing for vehicle-based disruption-tolerant networks," in *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, vol. 6, Barcelona, Spain, April 2006.
- [49] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age matters: efficient route discovery in mobile ad hoc networks using encounter ages," in *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing*, ACM, Annapolis, MD, USA, pp. 257–266, June 2003.
- [50] A. Balasubramanian, B. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 373–384, 2007.
- [51] N. Papanikos, D. G. Akestoridis, and E. Papapetrou, "Adyton: a network simulator for opportunistic networks," 2015.
- [52] R. I. Ciobanu, R. C. Marin, and C. Dobre, "Mobemu: a framework to support decentralized ad-hoc networking," in *Modeling and Simulation in HPC and Cloud Systems*, pp. 87–119, Springer, Cham, Switzerland, 2018.
- [53] G. F. Riley and T. R. Henderson, "The ns-3 network simulator," in *Modeling and Tools for Network Simulation*, pp. 15–34, Springer, Berlin, Heidelberg, 2010.
- [54] A. Varga, "The OMNeT++ discrete event simulation system," in *Proceedings of the European Simulation Multiconference*, Prague, Czech Republic, June 2001.
- [55] A. Udugama, A. Förster, J. Dede, and V. Kuppasamy, "Simulating opportunistic networks with omnet++," in *Recent Advances in Network Simulation*, pp. 425–449, Springer, Cham, Switzerland, 2019.
- [56] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, p. 55, Rome, Italy, March 2009.
- [57] P. Yuan and M. Song, "MONICA one simulator for mobile opportunistic," in *Proceedings of the 11th EAI International Conference on Mobile Multimedia Communications*, pp. 21–32, Qingdao, China, September 2018.
- [58] S. Saha, R. Verma, S. Saika, P. S. Paul, and S. Nandi, "e-ONE: enhanced ONE for simulating challenged network scenarios," *Journal of Networks*, vol. 9, no. 12, p. 3290, 2014.
- [59] S. Babu, G. Jain, and B. S. Manoj, "Urban delay tolerant network simulator (UDTNSim v0.1)," 2017, <https://arxiv.org/abs/1709.05645>.
- [60] S. B. M. Baskaran, G. Raja, A. K. Bashir, and M. Murata, "QoS-aware frequency-based 4G+ relative authentication model for next generation LTE and its dependent public safety networks," *IEEE Access*, vol. 5, pp. 21977–21991, 2017.
- [61] M. Cuka, D. Elmazi, K. Bylykbashi, E. Spaho, M. Ikeda, and L. Barolli, "Implementation and performance evaluation of two fuzzy-based systems for selection of IoT devices in opportunistic networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 2, pp. 519–529, 2019.
- [62] E. Spaho, K. Dhoska, K. Bylykbashi, L. Barolli, V. Kolici, and M. Takizawa, "Performance evaluation of routing protocols in DTNs considering different mobility models," in *Advances in Intelligent Systems and Computing*, pp. 205–214, Springer, Cham, Switzerland, 2019.
- [63] M. W. Kang, D. Y. Seo, and Y. W. Chung, "An efficient opportunistic routing protocol for ICN," in *Proceedings of the 6th ACM Conference on Information-Centric Networking*, pp. 159–160, Macao, China, September 2019.
- [64] J. Gandhi and Z. Narmawala, "Fair comparative analysis of opportunistic routing protocols: an empirical study," in *Data Communication and Networks*, pp. 285–294, Springer, Singapore, 2020.
- [65] H. Liu, H. Ning, Q. Mu et al., "A review of the smart world," *Future Generation Computer Systems*, vol. 96, pp. 678–691, 2019.
- [66] K. Rook, B. Witt, R. Bailey, J. Geigel, P. Hu, and A. Kothari, "A study of user intent in immersive smart spaces," in *Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 227–232, IEEE, Kyoto, Japan, March 2019.

## Research Article

# Your Knock Is My Command: Binary Hand Gesture Recognition on Smartphone with Accelerometer

Huixiang Zhang <sup>1</sup>, Wenteng Xu,<sup>1</sup> Chunlei Chen,<sup>2</sup> Liang Bai,<sup>3</sup> and Yonghui Zhang<sup>2</sup>

<sup>1</sup>School of Cyberspace Security, Northwestern Polytechnical University, Xi'an 710072, China

<sup>2</sup>School of Computer Engineering, Weifang University, Weifang 261061, China

<sup>3</sup>No. 203 Research Institute of China Ordnance, Xi'an 710065, China

Correspondence should be addressed to Huixiang Zhang; zhanghuixiang@nwpu.edu.cn

Received 11 March 2020; Revised 21 May 2020; Accepted 1 July 2020; Published 26 July 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Huixiang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motion-based hand gesture is an important scheme to allow users to invoke commands on their smartphones in an eyes-free manner. However, the existing scheme is facing some problems. On the one hand, the expression ability of one single gesture is limited. As a result, a gesture set consisting of multiple gestures is typically adopted to represent different commands. Users must memorize all gestures in order to make interaction successfully. On the other hand, the design of gestures needs to be complicated to express diverse intentions. However, complex gestures are difficult to learn and remember. In addition, complex gestures set a high recognition barrier to smart APPs. This leads to an imbalance problem. Different gestures have different recognition accuracy levels, which may result in instability of recognition precision in practical applications. To address these problems, this paper proposes a novel scheme using binary motion gestures. Only two simple gestures are required to express bit “0” and “1,” and rich information can be expressed through the permutation and combination of the two binary gestures. Firstly, four kinds of candidate binary gestures are evaluated for eyes-free interactions. Then, an online signal cutting and merging algorithm is designed to split accelerometer signals sequence into multiple separate gesture signal segments. Next, five algorithms, including Dynamic Time Warping (DTW), Naive Bayes, Decision Tree, Support Vector Machine (SVM), and Bidirectional Long Short-Term Memory (BLSTM) Network, are adopted to recognize these segments of knock gestures. The BLSTM achieves the top performance in terms of both recognition accuracy and recognition imbalance. Finally, an Android application is developed to illustrate the usability of the proposed binary gestures. As binary gestures are much simpler than traditional hand gestures, they are more efficient and user-friendly. Our scheme eliminates the imbalance problem and achieves high recognition accuracy.

## 1. Introduction

Eyes-free interaction is a method of controlling mobile devices without having to look at the device [1]. A variety of schemes have been developed to let users interact in an eyes-free manner. In [2], a digital calculator that operated with fingers on touch screens is developed. This method utilizes taps for digits input and uses swipes for other operations. Seventeen finger gestures are defined for arithmetic tasks. In [3], a nonvisual text entry method that uses the 6 bit Braille character encoding is presented. A signal is an input by touching the screen with several fingers where each finger represents one bit, either touching the screen or not. In addition to surface gestures, voice commands also provide a

solution [4]. Siri is one of the most prominent examples of a mobile voice interface. Another important way is to use a motion-based hand gesture [5]. To command a smartphone to execute a task, a user needs to perform a hand gesture with that phone in hand. The type of gesture is recognized through analysing data samples captured by motion sensors, such as accelerometers, gyroscopes, and orientation sensors.

Motion-based hand gestures enjoy several advantages. Firstly, users do not need to pay visual attention to the touchscreen because the physical location of the smartphone can be perceived via proprioception [6]. Secondly, hand-motion-gesture interaction puts forwards a few restrictions on the surrounding environment. For example, voice commands are prone to error in noisy environments [7], but

motion gestures can be performed as long as the hands of users are free. Finally, motion-based hand gestures can be designed in three-dimensional space. Compared to surface gestures, there remains larger design space for a variety of interactive tasks [8–10].

However, the scheme using motion-based hand gesture to command smartphones is facing three problems.

- (1) In order to represent different commands, a gesture set consisting of multiple gestures is required. For example, fourteen gestures are specified in the literature [5]; 11 gestures are proposed in the literature [11]. Users need to learn the set of hand gestures supported by a smartphone. They must memorize all gestures in order to make interaction successfully.
- (2) In order to distinguish these different gestures, hand gestures are defined not only in terms of the movement shape but also based on the motion kinematics [12]. Users are required to learn the features of gestures, in terms of movement shape and kinematics. It could be a daunting barrier to grasp details of such features. In addition, gestures with complex features set up a barrier to achieving high recognition accuracy.
- (3) The design of multiple gestures causes an uneven distribution of recognition accuracy levels among different gestures, which hinders the practical application of such design. For example, a deep feed-forward neural network is proposed to recognize 11 hand gestures in the literature [11]. They attained a minimum hit rate of 70.35% for Gesture 1 and a maximum hit rate of 100% for Gesture 10. As a result, the recognition accuracy levels of different gestures are dramatically different.

The root cause of the above problem is that multiple types of gestures are required to complete a specific interaction task with a phone. To address this problem, a novel interaction scheme using binary gestures is proposed in this paper. Only two kinds of hand gestures are needed to express binary bit “0” and “1.” Through the permutation and combination of the two binary gestures, a bit sequence is constructed. The application installed on smartphones can identify the bit sequence by analysing sensors’ signals. As the binary gestures are much simpler than traditional hand gestures, they are easy to learn and remember for users. High recognition accuracy can be achieved for both gestures. Thus, there will be no imbalance problem.

Taking the swiping movement gesture as an example, it is stipulated that users swipe of the smartphone horizontally to left and to right represent the bit “0” and “1,” respectively. By combining binary gestures, complex meanings can be expressed. For instance, if the user swipes the phone to the left four times in succession, it means that the command is “0000.” The permutation and combination of four binary gestures can represent up to 16 commands. We believe that it is easier for users to remember numbers than complex gestures.

It should be noted that we do not intend to design a set of gestures to meet the requirement of all kinds of interaction

tasks. We just provide an alternative for the eyes-free interaction scenarios. Its typical application scenarios include visually disabled users [13], distracted interaction [14], and covert operation [15].

The main work and contribution of the paper are summarized as follows.

- (1) A novel user-smartphone interaction scheme using binary gestures in an eyes-free manner is proposed.
- (2) An online signal cutting and merging algorithm is designed to extract the independent gesture signal segment from the binary gesture sequence. This online algorithm achieves an accuracy rate comparable to the offline SVM algorithm.
- (3) Five algorithms, including DTW, Naive Bayes, Decision Tree, SVM, and BLSTM, are adopted to recognize binary gestures, and BLSTM has reached a recognition accuracy of 98%.
- (4) A prototype application that uses binary gestures to send SMS messages is implemented on the Android platform.

The rest of this paper is organized as follows. The definition of binary gestures is introduced in Section 2. Section 3 describes the segmentation process of binary gesture sequences in detail. In Section 4, five algorithms are exploited to recognize a segmented knock gesture. Section 5 introduces a prototype application that uses binary gesture interaction. Finally, the work of this paper is concluded.

## 2. Definition of Binary Gestures

We exploit four categories of binary gesture according to a standard 3-axis coordinate system. In the standard 3-axis coordinate system, the  $x$ -axis is horizontal and points to the right. As illustrated in Figure 1, the  $y$ -axis is vertical and points up, and the  $z$ -axis points toward the outside of the screen face [27].

The definition of the four binary gestures is shown in Table 1. In the definition, the phone is supposed to hold in its portrait orientation by users’ two hands. The swipe, pitch, and flip gestures are performed along the  $z$ -axis,  $x$ -axis, and  $y$ -axis, respectively. For the knock gesture, the user holds the phone in one hand and taps on the screen with the index finger of the other hand.

A set of command encoded in binary is defined to represent the user’s interactive intention. A specific command is transformed to a gesture sequence consisting of single-actions and double actions. In each gesture category, a single-action is defined to represent the meaning of “0,” and a double-action is defined to represent the meaning of “1.” A double-action gesture includes two consecutive single-action gestures. Multiple gestures constitute a binary gesture sequence for interaction. Take knock gesture as an example; if a user wants to issue a 4-bit command “0101” to a smartphone, he is required to perform 4 knock actions in sequence. In other words, the user needs to perform “single-knock, double-knock, single-knock, double-knock” on the smartphone within a specified time range.

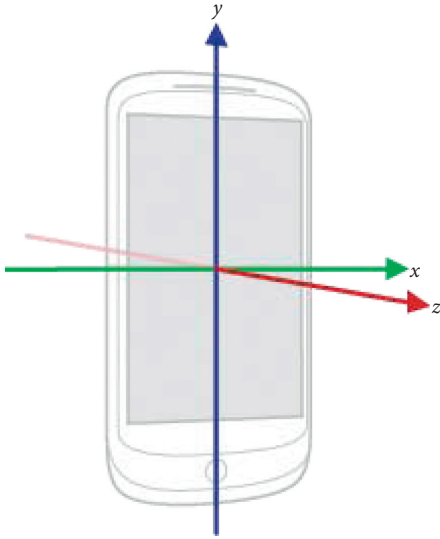


FIGURE 1: A standard 3-axis coordinate system [27].

TABLE 1: The definition of binary gestures.

Gesture category	Action	Meaning
Swipe (along the $z$ -axis)	Single	0
	Double	1
Pitch (along the $x$ -axis)	Single	0
	Double	1
Flip (along the $y$ -axis)	Single	0
	Double	1
Knock (on the screen)	Single	0
	Double	1

An accelerometer is very common on smartphones. It is a vital sensor to monitoring device motion, such as tilt, shake, rotation, and swing. In addition, it uses about 10 times less power than other motion sensors [16]. For the aforementioned reasons, we consider collecting accelerometer data to identify user gestures. The application installed in the smartphone analyses the acceleration sensor data to identify the binary bit sequence.

Figure 2 illustrates the collected 3-axis accelerometer data while performing two binary gestures in succession under different categories. The two successive gestures represent a bit sequence of “01.” The  $x$ ,  $y$ , and  $z$  curves correspond to the 3-axis accelerometer data. It can be seen from Figure 2(a), there is a lot of noise in the acquired accelerometer signal of the swipe gestures. It is difficult to distinguish the two swiping action gestures. In contrast, the pitch, flip, and knock gestures are easier to distinguish. The single and double actions of these gestures are mainly distinguished according to the number of crests or troughs. From Figure 2(b), it can be clearly seen that the single-pitch gesture has a significant trough in the  $z$ -axis and a significant crest in the  $y$ -axis, while the double-pitch gesture has two troughs and crests in the corresponding axis. In Figure 2(c), the waveform of flip gestures is similar to that of pitch

gestures, but the crests appear on the  $x$ -axis. For the knock gesture shown in Figure 2(d), the single-knock action has a significant crest, while double-knock action has two significant peaks. In summary, the pitch, flip, and knock gestures are considered in the following discussion.

In the next section, we will explain in detail how to identify the binary bit sequence passed by the user from the accelerometer signal.

### 3. Signal Segmentation

**3.1. Overall Process.** The overall processing flow is shown in Figure 3.

The 3-axis accelerometer signals are continuously acquired by an application installed on a smartphone. Before the start of each interaction, the phone is kept motionless for a period of time (more than 1 second). This motionless period is seen as a start signal of a gesture sequence. It is called the initial quiet period.

Firstly, the collected signals are preprocessed by synthesis and filtering. Then, the initial quiet period is detected. Once the start signal appears, an online bit cutting process is used to cut out independent gesture signal segments from a continuous signal stream. Next, the cut-out gesture signal segment is identified in its binary meaning. In an ideal state, a sequence composed of  $N$  binary gestures can be divided into  $N$  independent gestures signal segments. The final output is a  $N$ -bit binary sequence, which represents user’s command message.

#### 3.2. Signal Acquisition and Preprocessing

**3.2.1. Sampling Frequency.** In an Android smartphone, the sampling frequency of the various sensor is set in the system. There are four values that are available [17].

- ① SENSOR\_DELAY\_NORMAL, the sampling frequency is about 5 Hz.
- ② SENSOR\_DELAY\_UI, the sampling frequency is about 16 Hz.
- ③ SENSOR\_DELAY\_GAME, the sampling frequency is about 50 Hz.
- ④ SENSOR\_DELAY\_FASTEST, sample as fast as possible.

In the samples we collected, the duration of a single-knock gesture is about 0.2s-0.5s, which is equivalent to a gesture frequency of 2 Hz ~ 5 Hz. According to the Shannon sampling theorem, the sampling frequency of the signal should be no less than 10 Hz. If SENSOR\_DELAY\_FASTEST is used, the sampling frequency is much larger than 10 Hz, and too many samples are collected. This brings unnecessary overhead for subsequent calculations. The two frequencies of SENSOR\_DELAY\_UI and SENSOR\_DELAY\_GAME are more reasonable. Considering the accuracy of gesture recognition, we choose 50 Hz as the sampling frequency to obtain more sampling points.

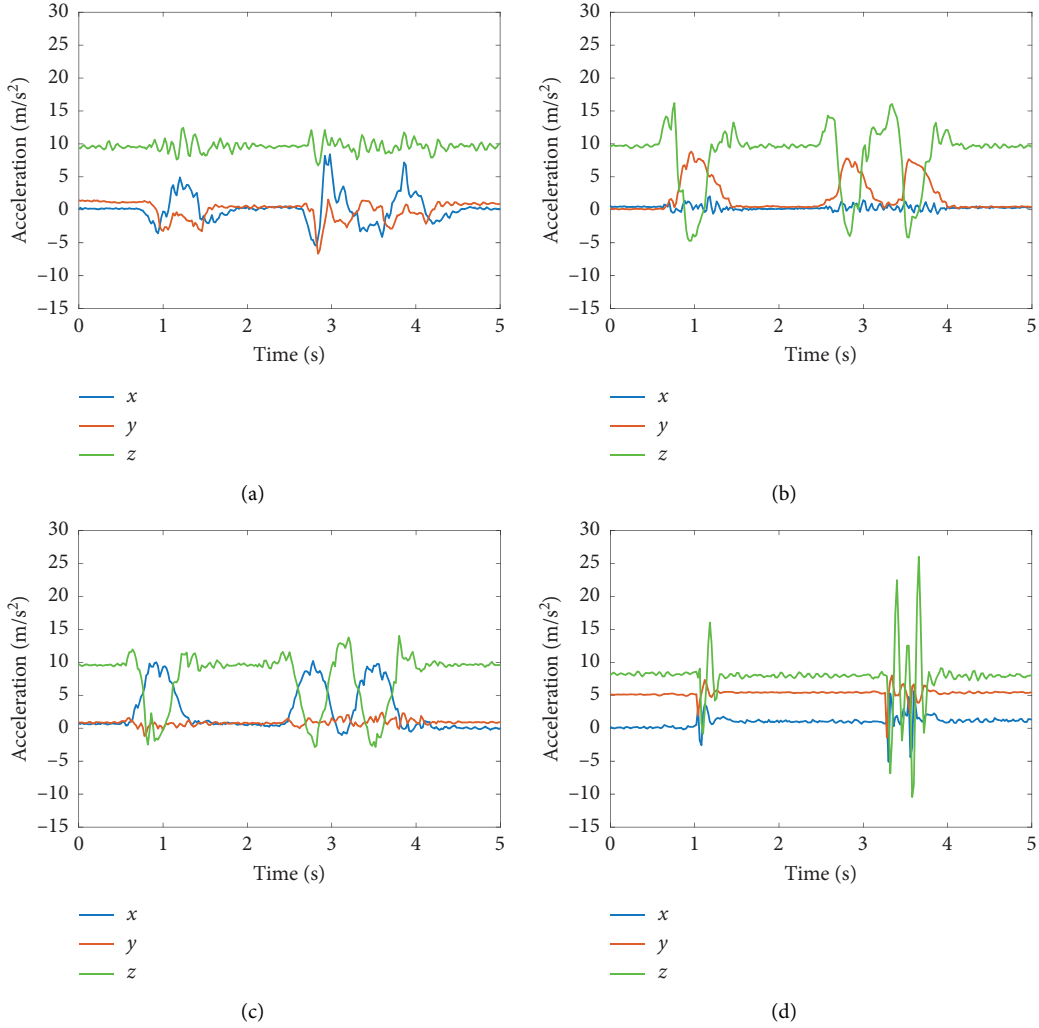


FIGURE 2: 3-axis accelerometer data for a bit sequence of “01” under different definitions. (a) Swipe. (b) Pitch. (c) Flip. (d) Knock.

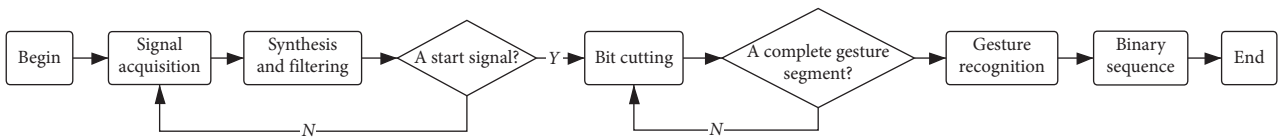


FIGURE 3: The overall processing flow.

**3.2.2. Signal Synthesis and Filtering.** To avoid the influence of the sensor’s own drift and gravity, we have performed vector synthesis on the 3-axis data [18]:

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2} - G, \quad (1)$$

where  $G$  represents the acceleration of gravity.  $A_x$ ,  $A_y$ , and  $A_z$  represent the accelerometer sampling value of the  $X$ -axis,  $Y$ -axis, and  $Z$ -axis, respectively.

In order to filter out abnormal points and noise in the collected data, a low-pass filter is performed as follows:

$$\bar{A}_i = \alpha A_i + (1 - \alpha) \bar{A}_{i-1}. \quad (2)$$

Here,  $A_i$  represents the  $i^{\text{th}}$  synthesized accelerometer sample and  $\bar{A}_i$  represents the value obtained after filtering. As the new sampling points are more significant for feature extraction and recognition, it is recommended to choose a large value of  $\alpha$  to retain a large proportion of sampled values.

**3.3. Bit Cutting Process.** The bit cutting process attempts to separate independent gesture signal segments from the continuously collected accelerometer signal stream. The bit cutting process operates in an online mode. Instead of acquiring the complete binary gesture sequence signals,

cutting and analysing operations run simultaneously. Figure 4 shows the complete flowchart of the bit cutting process.

The classic Sliding Window (SW) and Sliding Window and Bottom-up (SWAB) algorithms [19] are used to perform online signal segmentation. These algorithms cannot cut out a single complete binary gesture signal at one time. By contrast, such algorithms obtain a large number of short signal segments. Therefore, a merge algorithm is designed to combine these short signal segments into a complete binary gesture signal segment. The pseudocode of a bit cutting process is illustrated in Algorithm 1.

**3.3.1. Cutting Algorithm.** SW and SWAB are two kinds of online signal cutting algorithms used to extract physical signal segments from time-series signals. The SW algorithm read sample into a sliding window continuously then uses linear regression to fit a line for the samples in the window. At some points, the cumulative error is greater than a user-specified threshold (denoted as  $E_{\max}$ ), so the subsequence in the window is transformed into a segment. Then, the size of the sliding window is reduced to 0, and the process iterates until the entire time serial has been transformed into a piecewise linear approximation. The SWAB algorithm keeps a small buffer to gain a “semiglobal” view of the dataset for Bottom-Up. It scales linearly with the size of the dataset, requires only constant space, and produces high quality approximations of the data. That is beneficial to application in mobile devices.

The cumulative error  $E_{\text{cum}}$  of the linear approximation is calculated as follows:

$$E_{\text{cum}} = \sum_{i=1}^n \sqrt{(\bar{A}_i - \hat{A}_i)^2}. \quad (3)$$

Here,  $\hat{A}_i$  is the fitted value of the  $i^{\text{th}}$  data sample after signal synthesis and filtering, and  $n$  is the current window size. Whenever the window size changes, the cumulative error is recalculated.

Figure 5 shows a preprocessed accelerometer signal sequence generated by two consecutive knock gestures, which are a single-knock and thereafter a double-knock. As illustrated in Figure 5, there is a relatively calm interval between two adjacent knock gestures, such as the interval of 2.5 s–4.5 s. This kind of interval is called the quiet period. In contrast, the signal period with relatively strong fluctuations is called the fluctuation period, such as the interval of 1.5 s–2.5 s and the interval of 4.5 s–6.0 s. Those are the signal segments corresponding to the user’s knock gestures. Ideally, the quiet period and the fluctuation period alternate in the signal sequence of binary gestures.

After processing by the SW/SWAB, the signal sequence is cut into multiple short segments. As illustrated in Figure 5, these short segments are separated by blue vertical dashed lines. During the quiet period, there will be fitting errors due to small fluctuations. After a period of time, the cumulative error will eventually exceed the cutting threshold  $E_{\max}$ . Therefore, the signal in the quiet period will be cut into multiple sparse segments. During the fluctuation period, due to the relatively large fluctuation of the accelerometer signal,

the cumulative error will exceed the cutting threshold  $E_{\max}$  in a short time. Thus, the signal in the fluctuation period will be cut into multiple dense segments.

In order to extract a complete gesture, it is necessary to design a merge algorithm to combine multiple signal segments included in the fluctuation period. For the signal in Figure 5, two complete signal segments corresponding to the two knock gestures should be extracted after segments merging.

**3.3.2. Merge Algorithm.** For a segment, we can compute its average error  $E_{\text{avg}}$  as in the following equation:

$$E_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n \sqrt{(\bar{A}_i - \hat{A}_i)^2}. \quad (4)$$

Here,  $\bar{A}_i$  is the value of the  $i^{\text{th}}$  sample after signal synthesis and filtering,  $\hat{A}_i$  is the fitted value of the corresponding sample, and  $n$  is the number of samples in the segment. In particular, the average error of the initial quiet period is denoted as  $E_{\text{avgb}}$ .

Further, a characteristic  $p$  is defined to measure the fluctuation level of a segment. For the  $k$ th segment cut out by the SW/SWAB algorithm, its fluctuation characteristic  $p(k)$  is set according to the following equation:

$$p(k) = \begin{cases} 0, & E_{\text{avg}}(k) < \beta * E_{\text{avgb}}, \\ 1, & E_{\text{avg}}(k) > \beta * E_{\text{avgb}}. \end{cases} \quad (5)$$

A fluctuation characteristic of 0 indicates that the segment’s fluctuation is low and belongs to a quiet period. By contrast, a fluctuation characteristic of 1 indicates that the segment’s fluctuation is high and belongs to a fluctuation period.

$\beta$  is a coefficient used to balance  $E_{\text{avg}}(k)$  and  $E_{\text{avgb}}$ . In general, the average error of the segments included in a quiet period is slightly larger than  $E_{\text{avgb}}$ . Thus, the value of  $\beta$  should be greater than 1. However, if  $\beta$  is set to a large value, segments that belong to a fluctuation period would be marked as segments belonging to a quiet period incorrectly.

After the above processing, we can get a binary numerical sequence of the fluctuation characteristic, that is,  $P = [p(1), p(2), \dots, p(k)]$ . The merging algorithm processing flow is shown in Figure 6.

When the  $k^{\text{th}}$  segment is cut out ( $k \geq 3$ ), the merge operation is performed according to the fluctuation characteristics of the last three segments, i.e.,  $[p(k-2), p(k-1), p(k)]$ . There are three cases in which a merge operation can be performed:

- (1)  $p_{k-1}$  equals  $p_{k-2}$ , the  $(k-1)$ th and the  $(k-2)$ th segment are merged into a new segment, and the fluctuation characteristic of the new segment remains unchanged.
- (2) The sequence  $P$  matches  $[0, 1, 0]$ , and the size of the  $(k-1)$ th segment is less than  $C_{\min}$ ; it means that these three segments can be merged into a new segment with a fluctuation characteristic of 0.
- (3) The sequence  $P$  matches  $[1, 0, 1]$ , and the size of the  $(k-1)$ th segment is less than  $C_{\max}$ ; it means that



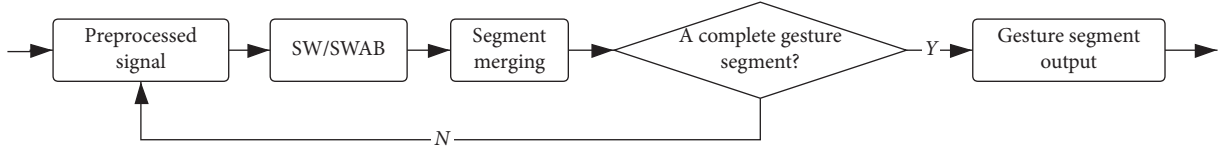


FIGURE 4: The flowchart of bit cutting process.

```

Input:  $\alpha$ , the coefficients of the low-pass filter
 $\beta$ , the coefficients to adjust the fluctuation characteristic
 $E_{\max}$ , the user-set maximum cumulative error threshold
 $E_{\text{bavg}}$ , the average error of the initial quiet period
Output: A complete gesture signal segment
/* initialization */;
 $i \leftarrow 0$ ,  $A \leftarrow []$ ;
 $k \leftarrow 0$ ,  $\text{Segment} \leftarrow []$ ,  $P \leftarrow []$ ;
while Get ith 3-axis acceleration sample:  $A_x, A_y, A_z$  do
  /* signal synthesis and filtering */;
   $A_i \leftarrow \sqrt{A_x^2 + A_y^2 + A_z^2} - G$ 
   $\bar{A}_i \leftarrow (1 - \alpha)\bar{A}_i + \alpha A_i$ ;
   $A[i] \leftarrow \bar{A}_i$ ;
   $i \leftarrow i + 1$ ;
  /* SW or SWAB */;
  start  $\leftarrow \text{Segment}[k - 1]$ , end  $\leftarrow i - 1$ ;
   $\hat{A} \leftarrow$  linear regression to fit a line for  $A[\text{start} : \text{end}]$ ;
  for  $j = \text{start} \rightarrow \text{end}$  do
     $E_{\text{cum}} \leftarrow E_{\text{cum}} + \sqrt{(\bar{A}_j - \hat{A}_j)^2}$ 
  end
  if  $E_{\text{cum}} > E_{\max}$  then
    /* a new segment is cut out by SW or SWAB */;
     $\text{Segment}[k] \leftarrow i - 1$ ;
     $E_{\text{avg}} \leftarrow 1/i E_{\text{cum}}$ ;
    /* set the fluctuation characteristic */;
    if  $E_{\text{avg}} < \beta E_{\text{bavg}}$  then
       $P[k] \leftarrow 0$ ;
    else
       $P[k] \leftarrow 1$ ;
    end
    /* process by merge Algorithm */;
     $\text{Segment}, P, k \leftarrow \text{merge}(\text{Segment}, P, k)$ ;
    /* check if a complete segment is cut-out */;
    if  $[P[0], P[1], P[2]] = [1, 0, 1]$  then
       $TS \leftarrow A[0 : \text{Segment}[0]]$ ;
      output  $TS$  for recognition;
    end
     $k \leftarrow k + 1$ ;
  end
end

```

ALGORITHM 1: The pseudocode of bit cutting process.

these three segments can be merged into a new segment with a fluctuation characteristic of 1.

If none of the above cases are met, the current round of merging operation ends, then waiting for a new segment is cut out by the SW/SWAB.

There are two important parameters in the merge process, i.e.,  $C_{\max}$  and  $C_{\min}$ . The size of the segment is actually the duration of the signal. In Case 3,  $C_{\max}$  indicates the maximum

duration of a quiet period allowed in a complete gesture signal. As a double-knock gesture is two consecutive single-knocks, there is usually a drop in the signal. The duration of the signal drop is about 100–300 ms in our experiments. Therefore,  $C_{\max}$  is set to 15 under a sample frequency of 50 Hz.

In Case 2,  $C_{\min}$  indicates the maximum duration of a fluctuation allowed in a quiet period.  $C_{\min}$  is affected by many factors, such as the use scenario and sensor accuracy. Therefore,  $C_{\min}$  is set to 3 conservatively in our experiments.

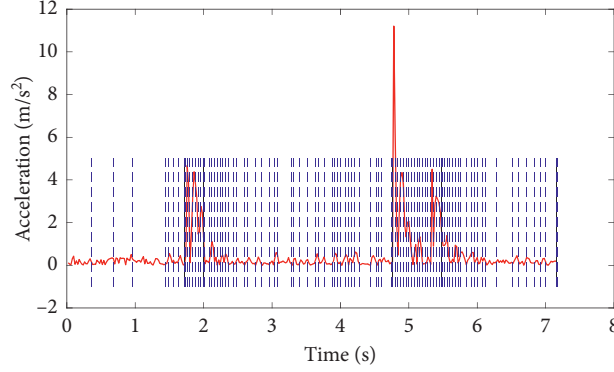


FIGURE 5: Cutting result of the SW/SWAB algorithm.

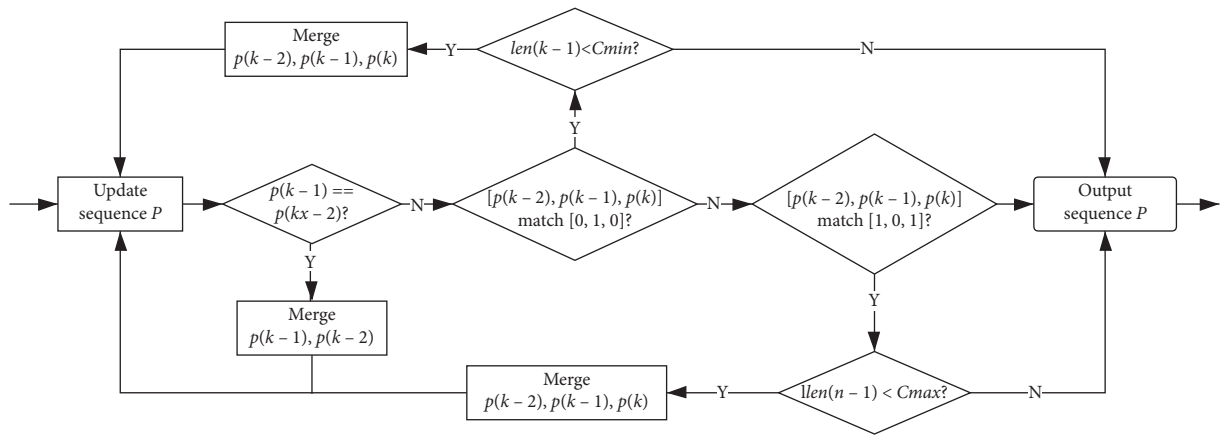


FIGURE 6: The merge algorithm processing flow.

Figure 7 illustrates the execution of the merge algorithm for an independent double-knock signal. In Figure 7(a), when the 3rd segment is cut out, the characteristic sequence  $P$  is  $[1,1,0]$  at this time. As  $p_1 = p_2$ , the first and the second segments are merged into a new segment with a fluctuation characteristic of 1. The characteristic sequence  $P$  is updated to  $[1,0]$ . Then, the 4th segment is cut out with a fluctuation characteristic of 0; thus,  $P$  is updated to  $[1,0,0]$  as in Figure 7(b). This does not fall into the three aforementioned merge cases. Next, the 5th segment is cut out with a fluctuation characteristic of 1. At now, the sequence  $P$  is changed to  $[1,0,0,1]$ . The fluctuation characteristics of the last three segments are checked. As  $p_2 = p_3 = 0$ , the two segments are merged into a new segment with a fluctuation characteristic of 0. After that, the sequence  $P$  is changed to  $[1,0,1]$ , as in Figure 7(c). Suppose the size of the new segment is less than  $C_{\max}$ , that meets the merge Case 2. The three segments are merged into a big segment with a fluctuation characteristic of 1 as shown in Figure 7(d). Through continuous online cutting and merge processing, the complete segment of a knock gesture can be extracted. The pseudo-code of the merge algorithm is shown in Algorithm 2.

**3.3.3. Bit Cutting Experiments.** Two experimental scenarios are designed. In Scenario 1, the smartphone is placed on the

desktop; in Scenario 2, the smartphone is held on user's hand. A total of 8 volunteers participated in the experiments. Each volunteer is required to perform 4 knock gestures during an interaction. A round of experiments contains 16 interactions, corresponding to the bit sequences "0000"- "1111." Ten rounds of experiments were performed, and 2,560 gesture samples for each scene are obtained.

A metric of cut-out rate is used to evaluate the effect of the bit cutting process. The cut-out rate is defined as follows:

$$\text{cut of rate} = \frac{\text{number of bits actually cut out}}{\text{number of bits theoretically cut out}} \quad (6)$$

The setting of parameters is shown in Table 2.

The experiments mainly analyse the cut-out rate of the binary gestures under different cumulative error thresholds  $E_{\max}$ . The threshold  $E_{\max}$  is set as follows [19]:

$$E_{\max} = E * 2^m. \quad (7)$$

Here,  $E$  is 0.01, and  $m$  varies from 0 to 12. The experimental results are shown in Figure 7.

As illustrated in Figure 8, the cut-out rate decreases as  $E_{\max}$  increases overall. When  $E_{\max}$  is large, some gestures with low knock strength will be recognized as quiet periods incorrectly. That resulted in a situation of less cut-out, and the cut-out rate is less than 1.

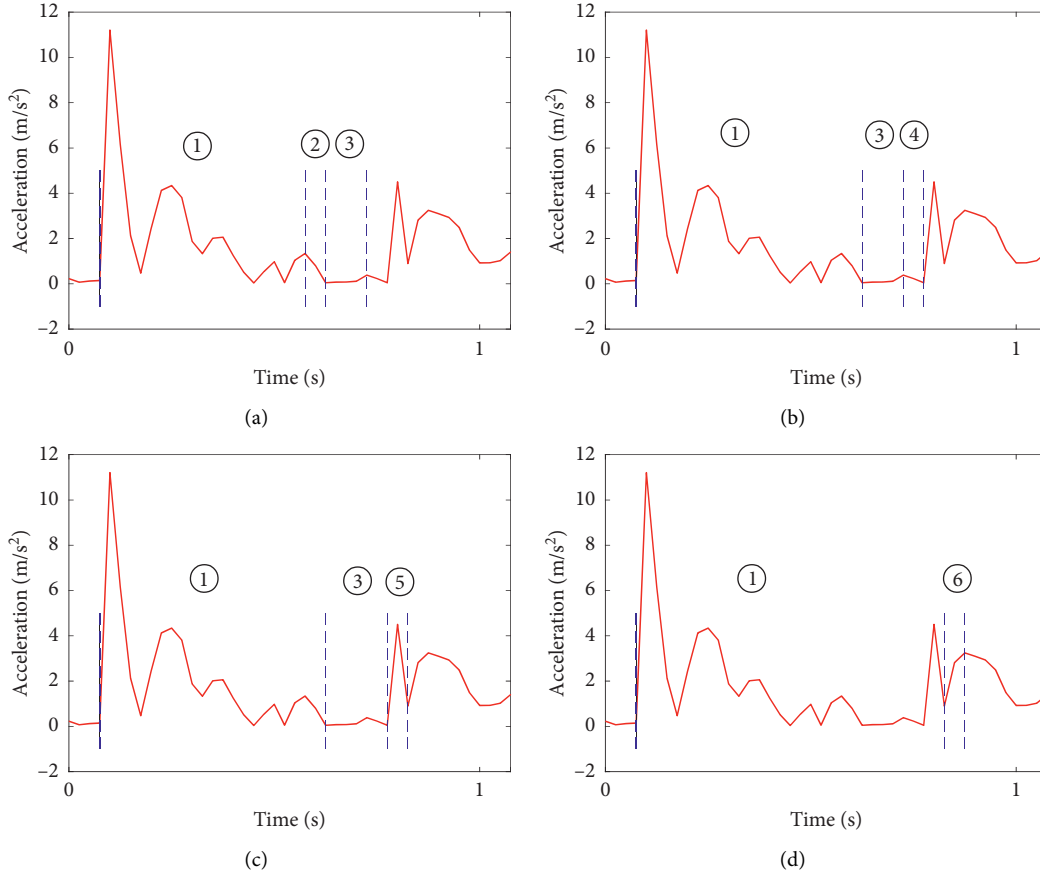


FIGURE 7: Schematic diagram of the merge algorithm. (a)  $P = [1, 1, 0]$ . (b)  $P = [1, 0, 0]$ . (c)  $P = [1, 0, 1]$ . (d)  $P = [1, 1]$ .

In the scenario of handheld, we can see the reasonable range of  $m$  is 0–7. However, the reasonable range of  $m$  is 0–4 in the scenario of the desktop. When a volunteer held the phone, a small shaking of the hand will cause continuous small fluctuations in the accelerometer signal. To discriminate between fluctuations caused by a handshake and those caused by knock gestures,  $E_{\max}$  needs to be larger. In order to adapt to different scenarios, the setting of  $E_{\max}$  is studied in the next subsection.

**3.3.4. Setting of  $E_{\max}$ .** If an initial quiet period is detected,  $E_{\max}$  is set as in the following equation:

$$E_{\max} = k \cdot N \cdot E_{\text{avgb}}. \quad (8)$$

Here,  $k$  is the linear adjustment coefficient,  $N$  is the current window size of SW/SWAB, and  $E_{\text{avgb}}$  is the average error of the initial quiet period. In this way, the setting of  $E_{\max}$  can be dynamically adjusted according to  $E_{\text{avgb}}$  and the current window size. This achieves the purpose of scene adaptation.

The influence of  $k$  value on the bit cutting is analysed.  $k$  varies in [0.001, 0.01, 0.1, 0.5, 1, 3, 5, 10]. The experimental results are shown in Figure 9.

As shown in Figure 9, a reasonable range of  $k$  tends to be the same in both scenarios. Scene adaptation is achieved to a certain degree by adaptively adjusting  $E_{\max}$ .

For different scenarios, only parameter  $k$  needs to be determined. When  $k$  is small, it has little effect on the cut-out rate. When  $k$  exceeds a certain threshold, the cut-out rate decreases rapidly. A smaller  $k$  means a small cumulative error threshold. This results in more segments being cut out, but a good bit cut-out rate can also be obtained by the merge algorithm. In contrast, a larger  $k$  means a large cumulative error threshold. This leads to less cut-out, and the cut-out rate is less than 1. From Figure 9, we can know that the cut-out rate is better when  $k$  is less than or equal to 1.

**3.3.5. Effectiveness of Bit Cutting.** In this section, we evaluated the effectiveness of the proposed bit cutting algorithm. The online bit cutting process is compared to an offline process using Support Vector Machine (SVM) [20]. The offline process is as follows.

A heuristic algorithm is used to cut the gesture signal sequence into multiple quiet and fluctuating segments. Then, the signal segments that are correctly cut out will be used to train an SVM model. Two features are extracted for each sampling point in a signal segment, namely, the 3-axis synthetic acceleration and the synthetic acceleration difference between the current and previous sampling point. The label of the sample point is the category of its segment, which is a quiet segment or a fluctuating segment. After the

```

Input: Segment, the signal segments produced by SW or SWAB
P, the corresponding fluctuation characteristic of segments in Segment
k, the count of segment in Segment
Output: Segment, P, k after merging process
/* the merge algorithm runs only there are more than 3 segments. */;
if  $k > 3$  then
  if  $P[k-2] == P[k-1]$  then
    /*  $[0, 0, 0] \rightarrow [0, 0]$ ,  $[0, 0, 1] \rightarrow [0, 1]$  */;
    /*  $[1, 1, 0] \rightarrow [1, 0]$ ,  $[1, 1, 1] \rightarrow [1, 1]$  */;
    Segment[ $k-2$ ]  $\leftarrow$  Segment[ $k-1$ ];
    remove ( $k-1$ )th item in Segment;
     $k \leftarrow k-1$ ;
  else if [ $P[k-2], P[k-1], P[k]$ ] ==  $[0, 1, 0]$  then
    /*  $[0, 1, 0] \rightarrow [0]$  */;
    if count of ( $k-1$ )th segment in Segment <  $C_{min}$  then
      Segment[ $k-2$ ]  $\leftarrow$  Segment[ $k$ ];
      remove ( $k-1$ )th item in Segment and P;
       $k \leftarrow k-2$ ;
    end
  else if [ $P[k-2], P[k-1], P[k]$ ] ==  $[1, 0, 1]$  then
    /*  $[1, 0, 1] \rightarrow [1]$  */;
    if count of ( $k-1$ )th segment in Segment >  $C_{max}$  then
      Segment[ $k-2$ ]  $\leftarrow$  Segment[ $k$ ];
      remove ( $k-1$ )th item in Segment and P;
       $k \leftarrow k-2$ ;
    end
  else
    doing nothing;
  end
end
Return Segment, P, k

```

ALGORITHM 2: The pseudocode of the merge algorithm.

TABLE 2: Parameters setting.

Parameters	Value
Smartphone	Huawei honor 8x
$\alpha$	0.8
$\beta$	5
$C_{max}$	15
$C_{min}$	3

above processing, we can get an SVM model to predict the category of each sampling point. Finally, the sample points are merged into segments according to their category labels. A similar merge process as shown in Figure 6 is utilized in the offline process.

The SVM algorithm has a global view, which simplifies the classification problem. All data samples are labelled and the 10-fold cross-validation is used to obtain the average bit cut-out rate of the SVM algorithm. For the bit cutting process,  $E_{max}$  is set based on equation (7), and  $k$  is 0.5. The experimental results are shown in Figure 10. The online bit cutting process designed in this paper achieves a cut-out rate comparable to the offline SVM algorithm. This shows that the proposed bit cutting process is suitable for online cutting of binary gesture signals.

**3.3.6. Comparison of Different Gestures.** In this section, the proposed bit cutting algorithm is applied to knock, pitch, and flip gestures sequence. The cut-out rate and the bit completion time are compared for these three gestures. Except for  $\beta$ , the parameters setting is the same as that in Table 2. As discussed in Section 3.3.2, the coefficient  $\beta$  should be greater than 1. Here,  $\beta$  varies from 1 to 10. As shown in Figure 11, the bit cutting algorithm is effective for all three gesture sequences. When  $\beta$  is set to 3, 4, and 5, the cut-out rate of the three gesture sequences is close to 1. That means all signal segment is cut-out correctly. As  $\beta$  increases, some segments in a fluctuation period were marked as segments belonging to a quiet period incorrectly. That causes the cut-out rate of the flip gesture sequence to increase to about 1.2.

Next, we counted the length of all correctly cut-out signal segments and obtained the average completion time to express bit “0” and “1.” As illustrated in Figure 12, the bit completion time of pitch and flip gestures are longer than the knock gesture. The single-knock action takes about 0.3 seconds on average to express the bit “0,” while the pitch and flip actions take more than 0.5 seconds. The double-knock action takes about 0.65 seconds on average to express the bit “1,” while the pitch and flip actions take more than 1.0 seconds. To issue the same command to a phone, the time spent using the knock gesture is only about half of the pitch

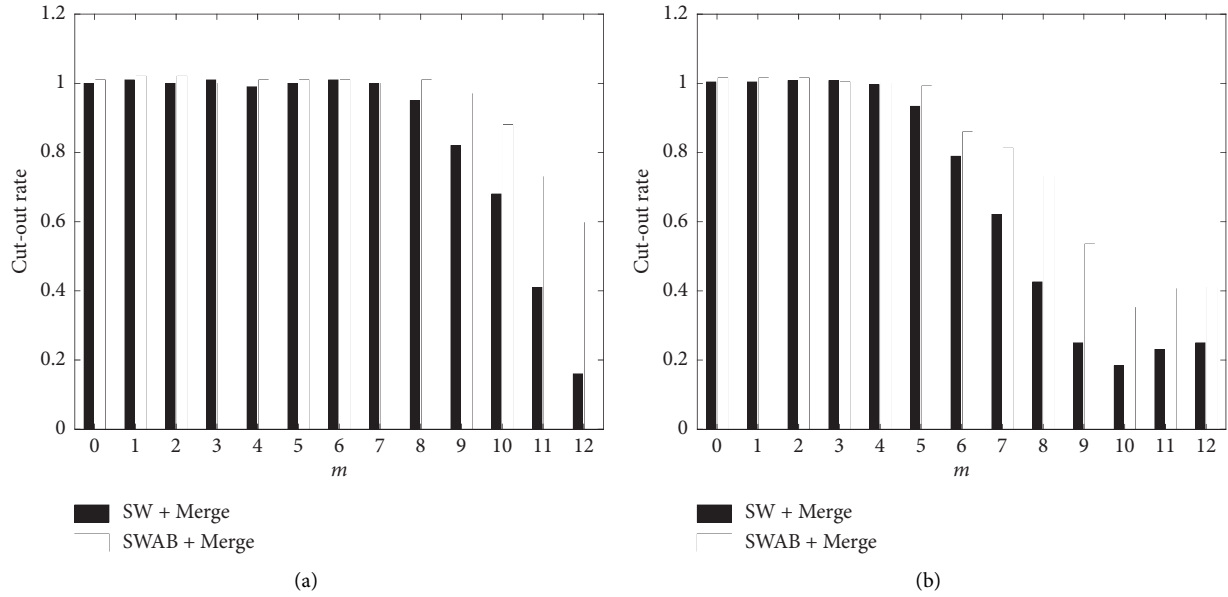


FIGURE 8: Comparison of cut-out rates in different scenarios. (a) Handheld. (b) Desktop.

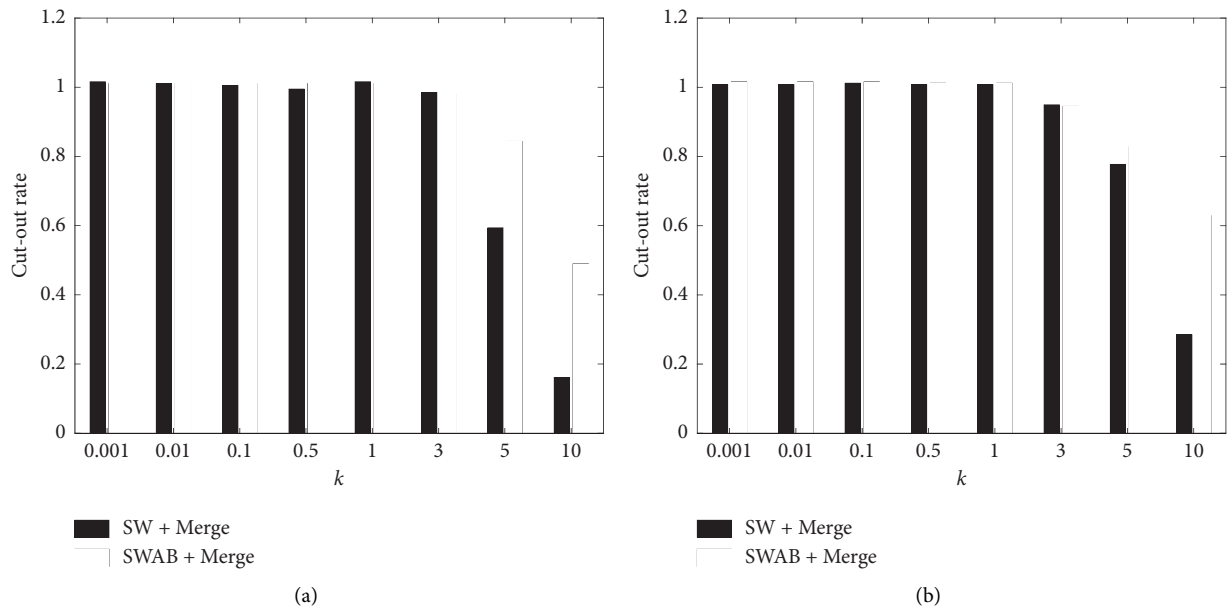


FIGURE 9: Comparison of cut-out rates for different  $k$ . (a) Handheld. (b) Desktop.

and flip gestures. Thus, the interaction efficiency of knock gesture outperforms the other two.

Since the proposed algorithm is better at cutting knock and pitch gesture sequences, how to recognize the cut-out signal segments of these two gestures to their binary meaning is studied in Section 4.

#### 4. Binary Gesture Recognition

After bit cutting, a complete signal segment of a gesture sequence is obtained. To distinguish between single and double gesture action, the DTW, traditional machine learning, and BLSTM methods are exploited in this section.

**4.1. DTW Method.** Dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences, which may vary in length [21]. The temporal sequences of signature will be denoted as matrices like  $S_{P \times Z}$ , where  $P$  is the number of points in the cut-out signal segment and  $Z$  is the number of features extracted from each point. Here, the 3-axis raw acceleration data is used. As a result, the  $i^{\text{th}}$  point in the sequence is a 3-dimensional vector. In order to verify whether a sample ( $S_{P \times Z}$ ) matches its corresponding template ( $T_{Q \times Z}$ ), a dissimilarity score  $dis$  is computed between them based on the DTW algorithm.  $dis$  is a cumulative distance of the two gesture signal segments.

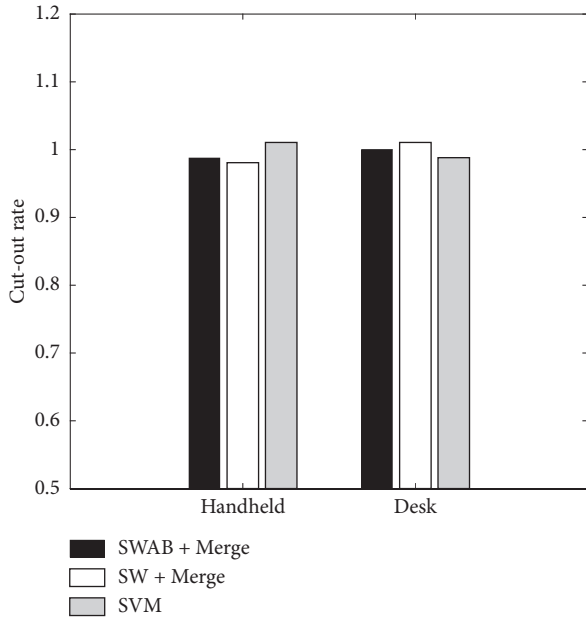


FIGURE 10: Comparison of cut-out rates of different algorithms.

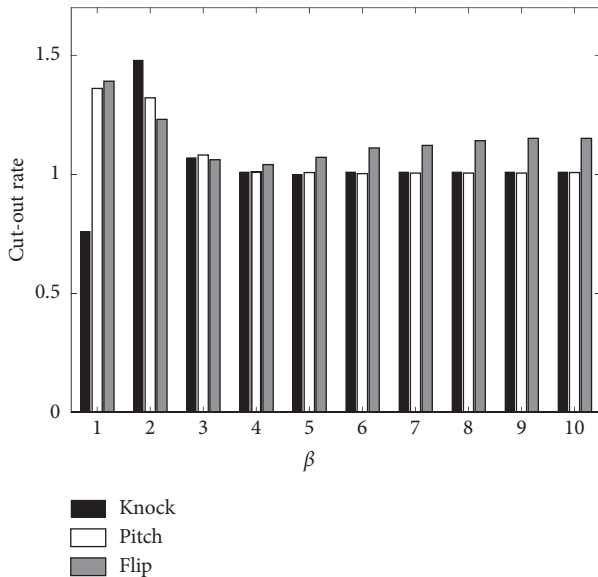


FIGURE 11: Cut-out rates for different  $\beta$ .

As seen in Table 1, a single-action and a double-action are defined in each gesture category. Therefore, a single-action signal segment and a double-action signal segment are manually selected for each volunteer as reference templates. When a signal segment is cut out, two dissimilarity scores are calculated between the segment and the two reference templates. The segment is classified as consistent with the template of a smaller dissimilarity score.

**4.2. SVM Methods.** Support Vector Machines (SVMs) are widely used for classification and regression tasks. Here, gesture recognition is treated as a binary classification

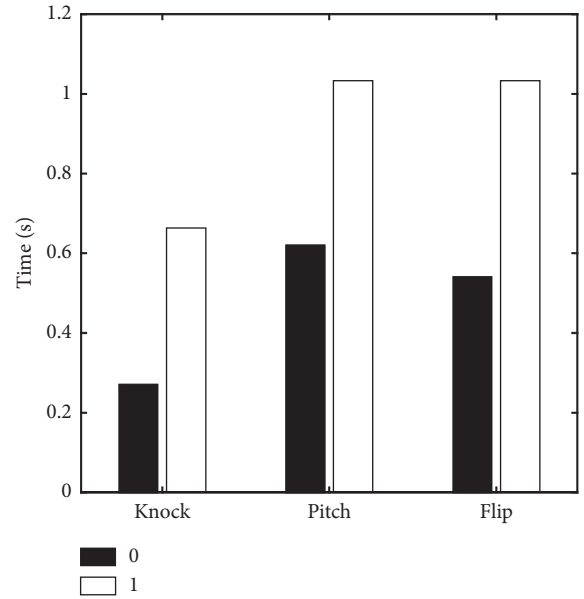


FIGURE 12: The average completion time for different binary gestures.

problem. SVM constructs a hyperplane in a high-dimensional space to separate two class of gesture, single-action, and double-action gestures. We use LIBSVM as a classification algorithm and use the RBF kernel as a kernel function. Three features are extracted to construct a 3-dimensional feature vector for each gesture signal segment. They are the gesture size, gesture energy, and the first-order components of the signal after Discrete Cosine Transforms (DCT).

**4.2.1. Gesture Size.** The size of a gesture refers to the duration of the gesture. It is defined as the number of sampling points in a cut-out gesture segment. Obviously, a double-action gesture usually takes longer than a single-action gesture.

**4.2.2. Gesture Energy.** The energy consumption of an object’s movement is closely related to its speed and acceleration. Bouten’s research in recent years has proved that the absolute integral of the acceleration and angular velocity of an object’s movement have a linear relationship with energy consumption [23]. This provides a theoretical basis for evaluating gestures’ movements with an acceleration sensor. When the output signal is a digital signal, the following formula can be used to calculate the energy of a gesture:

$$E_n = \sum_{i=1}^n (|A_x| + |A_y| + |A_z|). \tag{9}$$

Among them,  $A_x$ ,  $A_y$ , and  $A_z$  are the 3-axis values of the acceleration sensor. Since we have performed vector synthesis on the 3-axis data based on equation (1), the knock energy is defined as follows:

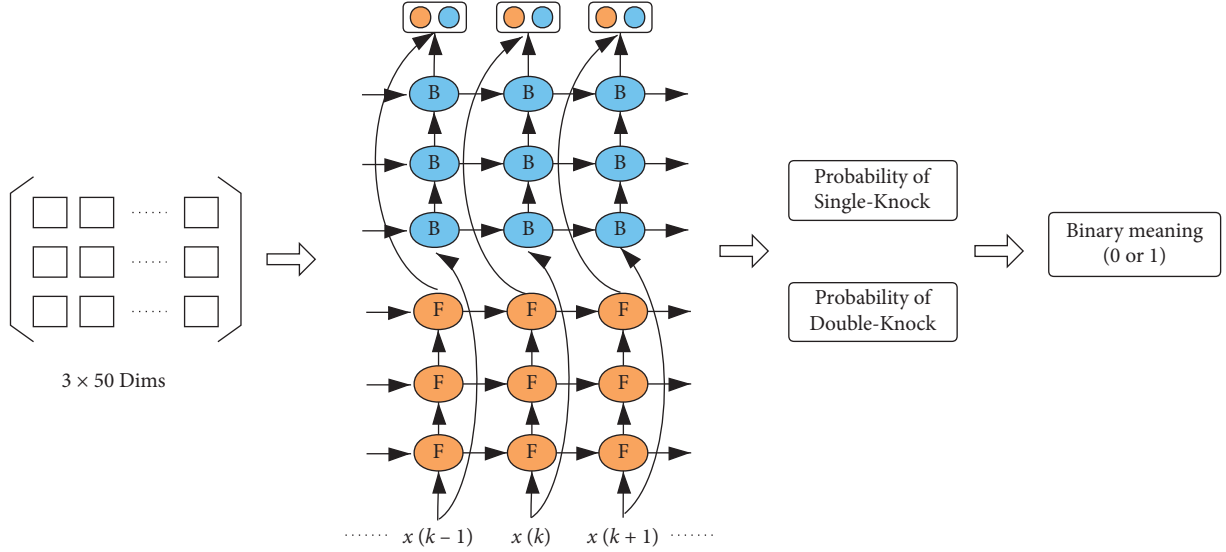


FIGURE 13: Knock gesture recognition using a 3-layer BLSTM model.

$$E_n = \sum_{i=1}^n (|A_i|). \quad (10)$$

4.2.3. *DCT*. A one-dimensional DCT is performed on a knock gesture signal segment. DCT converts the gesture signal into a set of frequencies. The first frequency in the set is the most meaningful. Therefore, the first-order components of the signal after DCT is selected as one of the features.

Two other machine learning methods, including Naive Bayes and Decision Tree, are also used to recognize binary gestures for comparison purposes [22]. These algorithms use the same feature vector as SVM for classification.

4.3. *BLSTM Method*. BLSTM is an extension of traditional LSTM that can improve model performance on sequence classification problems [24]. A 3-layer BLSTM architecture is used to model the gesture data in this paper. The process of knock gesture is illustrated in Figure 13.

Since the maximum duration of a knock gesture does not exceed 1 second, and the sample frequency is set to 50 Hz, up to 50 samples are captured for a cut-out gesture segment. Instead of using the synthesis and filtered values, the 3-axis raw acceleration data is used. Thus, a matrix of 3 by 50 is fed into the BLSTM model. The forward and the backward output are concentrated together to generate the probability for two knock gestures. The gesture of higher probability is selected as the predicted result, i.e., 0 for a single-knock and 1 for a double-knock.

The parameters of the BLSTM model are shown in Table 3. The same model is also applied to recognize the pitch gesture. Because the bit completion time of pitch gestures is longer than knock gestures, a matrix of 3 by 100 is used as input to the model.

TABLE 3: Parameters of the BLSTM model.

Parameters	Value
Learning rate	0.001
Batch size	32
Optimizer	Adam
Loss	Categorical cross entropy
Epochs	100
Dropout	0.5

4.4. *Experimental Results*. A metric defined in equation (11) is used to evaluate the recognition accuracy.

$$M_{\text{acc}} = \frac{(TP + TN)}{(P + N)}. \quad (11)$$

Here,  $P$  is the number of segments that belong to a single-action gesture.  $N$  is the number of segments that belong to a double-action gesture.  $TP$  is the number that is predicted to be a single-action gesture. And  $TN$  is the number that is predicted to be a double-action gesture.

A metric defined in the following equation is used to evaluate the imbalance of recognition for the two action gestures.

$$M_{\text{bal}} = \frac{M_{\text{acc}}(0)}{M_{\text{acc}}(1)}. \quad (12)$$

Here,  $M_{\text{acc}}(0)$  represents the recognition accuracy of single-action gesture, and  $M_{\text{acc}}(1)$  represents the recognition accuracy of double-action gesture.  $M_{\text{bal}}$  is expected to be around 1, which means the recognition accuracy for the two binary actions is similar. Moreover, the metrics of micro F1 and recall are also evaluated.

The experimental results are shown in Figure 14. All gesture recognition methods have achieved recognition accuracy of more than 90%. The BLSTM method outperformed the other algorithms and achieved the highest recognition accuracy of 98%. The metric of micro F1 also

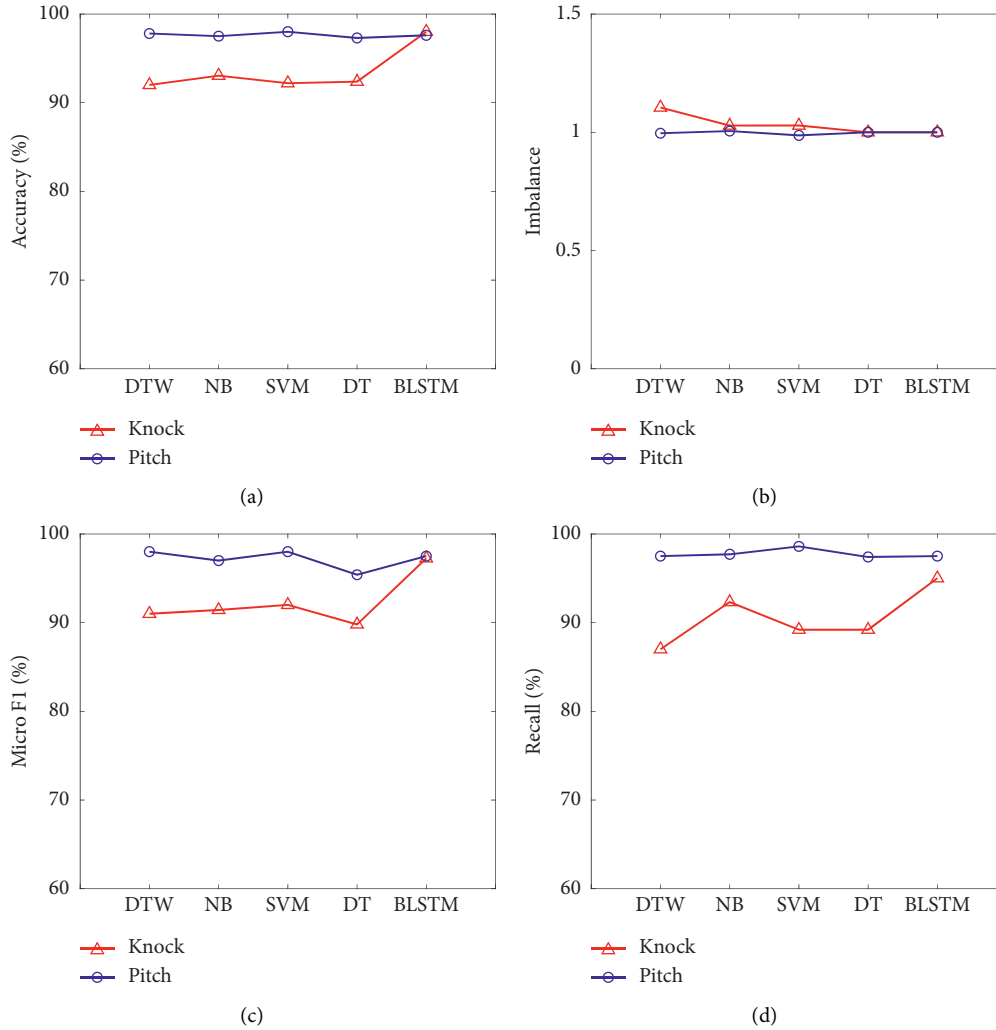


FIGURE 14: Gesture recognition results. (a) Recognition accuracy. (b) Recognition imbalance. (c) Micro F1. (d) Recall.

indicated that DTW, NB, SVM, and BLSTM methods can recognize the cut-out signal segments into its binary meaning in high accuracy.

From the perspective of recognition imbalance, the recognition accuracy of DTW for single-knock gesture is higher than a double-knock gesture, making its  $M_{bal}$  greater than 1. However, the recognition accuracy of DTW for single-pitch and double-pitch gesture are close. The imbalance of other recognition methods is good, and the experimental results are close to 1, among which the BLSTM method is the best.

As seen from Figure 14, these methods achieve superior performance in recognizing pitch gestures than knock gestures. An important reason is that the completion time of the pitch gesture is longer than that of the knock gesture. Compared with knock gestures, the difference in the feature of gesture duration between a single-pitch and a double-pitch is more significant; The same is true with regard to the energy difference between the two gesture types.

The experimental results show the effectiveness of using knock and pitch gestures for interaction. Only two simple gestures are required. High recognition accuracy can be

achieved for both gestures and avoid imbalance problem at the same time.

## 5. Prototype Application and Discussion

In general, the knock gesture is simple and clear, which is convenient for users to operate. The bit completion time of knock gesture is shorter than that of pitch gesture. In addition, the knock gesture signal segment can be recognized in high recognition accuracy. Therefore, the knock gesture is selected to implement interaction between human and mobile applications.

In this prototype application, users utilize the single-knock and double-knock gestures to command the applications in an Android smartphone to send SMS messages. The prototype application is useful in some scenarios where private interaction is required; the user cannot speak or cannot light up the screen, which may attract others' attention. The binary knock gestures are inconspicuous and can be used to send text messages covertly.

As the BLSTM model performs best both in recognition accuracy and recognition imbalance, it is selected to



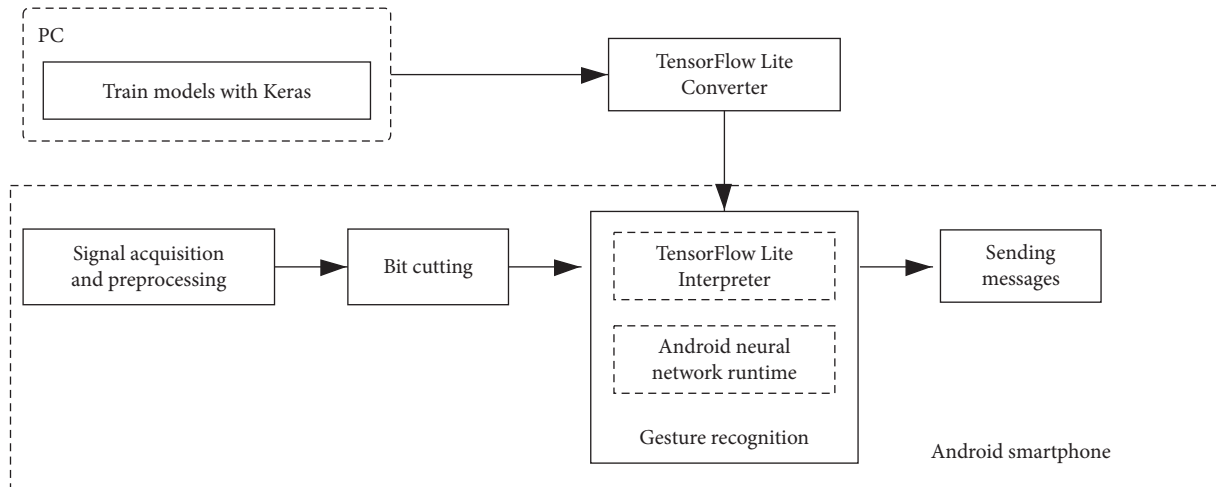


FIGURE 15: The development process of the prototype application.

implement in our prototype application. The framework of TensorFlow Lite [25] is used to integrate BLSTM into smartphones. The development process of the prototype application is shown in Figure 15.

A 3-layer BLSTM model is trained with Keras [26] in PC. Then it is converted into a TensorFlow Lite model using the TensorFlow Lite converter. The TensorFlow Lite interpreter executes the model on smartphones to make predictions based on input accelerometer data. If the predicted binary sequence is matched with the preset command, the application automatically sends a short message to the corresponding phone number.

The prototype application is tested with four different scenarios on how people interact with a smartphone. In the last three scenarios, users interacted in an eyes-free manner. (1) Normal: a person is sitting on a chair and holding a mobile phone on a desk. (2) Eyes-free: a person is sitting on a chair and holding a mobile phone beneath a desk. (3) Covert: a person is standing still with the phone in his pants pocket. (4) Walking: a person is walking at a constant speed with the phone in his pants pocket.

The metrics of the cut-out rate and accuracy are evaluated. Figure 16 illustrated the experimental results. In the scenarios of normal, eyes-free, and covert, they all achieved a cut-out rate close to 1. Most of the bit signal segments are split out from the gesture signal sequences successfully. Meanwhile, these bits are recognized with high accuracy. However, when people are moving, it greatly affects the cut-out effect. The proposed scheme is more suitable for interacting with smartphones when people are in a stationary state.

In addition to the above interaction cases, binary gestures can be used as a supplementary input modality for many scenarios. For example, it can be used as an interaction method of a blind assistive system. In [13], a blind person can establish a voice call to a predefined number using voice command. However, they got some error as a sound wave is affected much for noise and humidity. In such an environment, the blind person can use binary gestures instead of voice. In [28], a set of hand

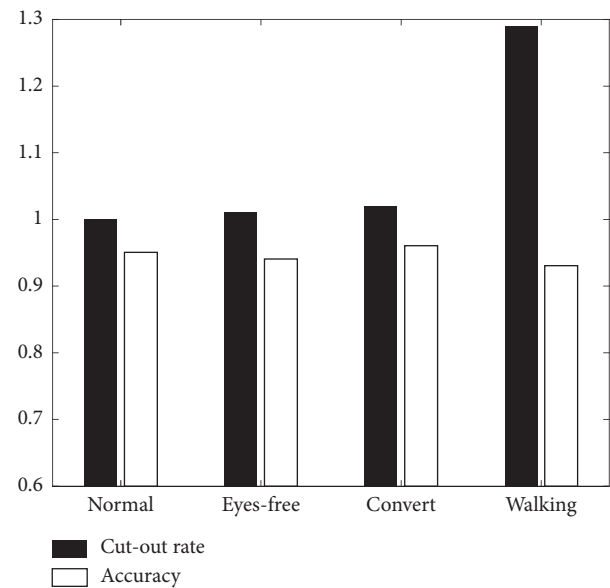


FIGURE 16: The development process of the prototype application.

gestures is proposed to control the smart lighting system. These vision-based hand gestures are more complex and difficult in terms of their recognition. Under such circumstances, smartphones can be adapted as a user interaction interface. By encoding these tasks into a binary command set, users can control the lighting system by binary motion-based gestures.

## 6. Conclusion

A novel user-smartphone interaction scheme using binary gestures is proposed in this paper. Firstly, four kinds of binary gestures are evaluated. The gestures of flip, pitch, and knock are selected as candidate interaction gestures. Then, the gesture extraction process is investigated in detail. The accelerometer signal is captured and pre-processed. An online signal cutting and merging

algorithm is designed to extract the independent gesture signal segment from the binary gesture sequence. Experiments show that the proposed method outperforms its counterparts in cutting knock and pitch gesture sequences. Next, five algorithms, including DTW, Naive Bayes, Decision Tree, Support Vector Machine, and BLSTM, are exploited to recognize the flip and knock gesture. Finally, an Android application is developed based on the binary command channel using knock gestures.

The proposed scheme only requires two meta gestures. And rich information can be expressed through the permutation and combination of the two gestures. As the binary gestures are much simpler than traditional gestures, our method achieves high recognition accuracy and avoids the imbalance problem.

The proposed scheme provides an alternative for eyes-free interaction scenarios. It is applicable to visually disabled user-smartphone interactions, distracted interaction, and covert operations.

As future work, we will enhance the ability to express more complex human-smartphone interaction commands.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

This work was supported by the following foundations: the National Natural Science Foundation of China (Grant no. 31872847), the Natural Science Foundation of Shaanxi Province, China (Grant nos. 2019JM-244), the Industry-University Collaborative Education Program granted by the Ministry of Education of China (201902323022 and 201802217002), Weifang Science and Technology Development Plan (Grant no. 2017GX021), and Shandong University Scientific Research Development Plan (Grant no. J17KB183).

## References

- [1] K. Katsuragawa, A. Kamal, and Q. F. Liu, "Bi-Level thresholding: analyzing the effect of repeated errors in gesture input," *ACM Transactions on Interactive Intelligent Systems*, vol. 9, no. 2-3, pp. 1–30, 2019.
- [2] B. Chaudhuri, L. Perlmutter, and J. Petelka, "GestureCalc: an eyes-free calculator for touch screens," in *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 112–123, Pittsburgh PA USA, October 2019.
- [3] S. Azenkot, J. O. Wobbrock, S. Prasain, and R. E. Ladner, "Input finger detection for nonvisual touch screen text entry in Perkinput," in *Proceedings of Graphics Interface 2012*, pp. 121–129, Toronto, Canada, May 2012.
- [4] A. Vtyurina, A. Fourney, and M. R. Morris, "Bridging screen readers and voice assistants for enhanced eyes-free web search," in *Proceedings of the World Wide Web Conference*, pp. 3590–3594, San Francisco, CA, USA, May 2019.
- [5] J. Ruiz, Y. Li, and E. Lank, "User-defined motion gestures for mobile interaction," in *Proceedings of the International Conference on Human Factors in Computing Systems*, pp. 197–206, Vancouver, Canada, May 2011.
- [6] M. Negulescu, J. Ruiz, Y. Li, and E. Lank, "Tap, swipe, or move: attentional demands for distracted smartphone input," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 173–180, Rome, Italy, May 2012.
- [7] M. S. R. Tanveer, M. M. A. Hashem, and M. K. Hossain, "Android assistant EyeMate for blind and blind tracker," in *Proceedings of 2015 18th International Conference on Computer and Information Technology*, pp. 266–271, Istanbul, Turkey, September 2015.
- [8] S. J. Castellucci, I. S. MacKenzie, M. Misra, L. Pandey, and A. S. Arif, "TiltWriter: design and evaluation of a no-touch tilt-based text entry method for handheld devices," in *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pp. 1–8, Pisa, Italy, Italy 2019.
- [9] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Grealy, "Systematic literature review of hand gestures used in human computer interaction interfaces," *International Journal of Human-Computer Studies*, vol. 129, no. 9, pp. 74–94, 2019.
- [10] F. Hong, M. Wei, S. You, Y. Feng, and Z. Guo, "Waving authentication: your smartphone authenticate you on motion gesture," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 263–266, Seoul, South Korea, April 2015.
- [11] Y. Jhang, Y. Chu, and T. Tai, "Sensor based dynamic hand gesture recognition by PairNet," in *International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, IEEE, Chengdu, China, pp. 994–1001, December 2019.
- [12] A. Kamal, Y. Li, and E. Lank, "Teaching motion gestures via recognizer feedback," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 73–82, Los Angeles, CA, USA, March 2014.
- [13] J. Choi, K. Song, and S. Lee, "Enabling a gesture-based numeric input on mobile phones," in *Proceedings of 2011 IEEE International Conference on Consumer Electronics*, pp. 151–152, Las Vegas, NV USA, January 2011.
- [14] S. S. A. Shimon, S. Morrison-Smith, N. John, G. Fahimi, and J. Ruiz, "Exploring user-defined back-of-device gestures for mobile devices," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 227–232, Copenhagen Denmark, August 2015.
- [15] P. Mittal and N. Singh, "Speech based command and control system for mobile phones: issues and challenges," in *2016 Second International Conference on Computational Intelligence & Communication Technology*, pp. 729–732, Ghaziabad, India, February 2016.
- [16] Motion sensors, [https://developer.android.com/guide/topics/sensors/sensors\\_motion](https://developer.android.com/guide/topics/sensors/sensors_motion), 2020.
- [17] Sensors overview, [http://developer.android.com/guide/topics/sensors/sensors\\_overview.html](http://developer.android.com/guide/topics/sensors/sensors_overview.html), 2020.
- [18] A. Hatori and H. Kobayashi, "A preliminary study of iot-device control using gestures recognition," in *Proceedings of*

- 56th Annual Conference of the Society of Instrument and Control Engineers of Japan, pp. 976–979, Kanazawa, Japan, September 2017.
- [19] E. Keogh, S. Chu, D. Hart, and M. Pazzani, “An online algorithm for segmenting time series,” in *Proceedings of the 2001 IEEE International Conference on Data Mining*, IEEE, San Jose, CA, USA, pp. 289–296, December 2001.
  - [20] R. Kumar and P. Singhal, “Review on offline Signature verification by SVM,” *International Research Journal. Engineering and Technology*, vol. 4, no. 6, pp. 1771–1773, 2017.
  - [21] D. Kajiwaro and K. Murao, “Gesture recognition method with acceleration data weighted by sEMG,” in *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2019 ACM International Symposium on Wearable Computers*, pp. 741–745, New York, NY, USA, September 2019.
  - [22] F. Pedregosa, G. Varoquaux, and A. Gramfort, “Scikit-learn: machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 10, pp. 2825–2830, 2011.
  - [23] C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, “A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity,” *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 3, pp. 136–147, 1997.
  - [24] M.-C. Lee and S.-B. Cho, “A recurrent neural network with non-gesture rejection model for recognizing gestures with smartphone sensors,” *Lecture Notes in Computer Science*, vol. 8251, pp. 40–46, 2013.
  - [25] A. Campoverde and G. Barros, “Detection and classification of urban actors through TensorFlow with an android device,” *Advances in Intelligent Systems and Computing*, vol. 1099, pp. 167–181, 2019.
  - [26] N. Ketkar, “Introduction to keras,” in *Deep Learning with Python*, Apress, Berkeley, CA, USA, 2017.
  - [27] Motion and position sensors, <https://google-developer-training.github.io/android-developer-advanced-course-concepts/unit-1-expand-the-user-experience/lesson-3-sensors/3-2-c-motion-and-position-sensors/3-2-c-motion-and-position-sensors.html>, 2020.
  - [28] D. Park, Y. S. Lee, and S. Song, “User centered gesture development for smart lighting,” in *Proceedings of HCI Korea*, pp. 146–150, Seoul, South Korea, December 2016.

## Research Article

# Analysis and Evaluation of Braille to Text Conversion Methods

**Sana Shokat,<sup>1</sup> Rabia Riaz,<sup>1</sup> Sanam Shahla Rizvi,<sup>2</sup> Khalil Khan,<sup>1</sup> Farina Riaz,<sup>3</sup> and Se Jin Kwon<sup>4</sup>**

<sup>1</sup>The University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

<sup>2</sup>Raptor Interactive (Pty) Ltd., Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa

<sup>3</sup>University of Southern Queensland, Toowoomba, Australia

<sup>4</sup>Department of Computer Engineering, Kangwon National University, Samcheok 25806, Republic of Korea

Correspondence should be addressed to Se Jin Kwon; [sjkwon@kangwon.ac.kr](mailto:sjkwon@kangwon.ac.kr)

Received 14 February 2020; Revised 5 May 2020; Accepted 4 June 2020; Published 26 July 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Sana Shokat 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.

Technology is advancing rapidly in present times. To serve as a useful and connected part of the community, everyone is required to learn and update themselves on innovations. Visually impaired people fall behind in this regard because of their inherent limitations. To involve these people as active participants within communities, technology must be modified for their facilitation. This paper provides a comprehensive survey of various user input schemes designed for the visually impaired for Braille to natural language conversion. These techniques are analyzed in detail with a focus on their accessibility and usability. Currently, considerable effort has been made to design a touch-screen input mechanism for visually impaired people, such as Braille Touch, Braille Enter, and Edge Braille. All of these schemes use location-specific input and challenge visually impaired persons to locate specified places on the touch screen. Most of the schemes require special actions to switch between upper and lowercase and between numbers and special characters, which affects system usability. The key features used for accessing the performance of these techniques are efficiency, accuracy, and usability issues found in the applications. In the end, a comparison of all these techniques is performed. Outcomes of this analysis show that there is a strong need for application that put the least burden on the visually impaired users. Based on this survey, a guideline has been designed for future research in this area.

## 1. Introduction

Visually impaired people are an important part of every community [1]. They are also concerned in learning the details of everything they encounter in their daily life [2]. The total number of visually impaired people is 2.2 billion; among them, 36 million are completely blind, and rest of the 1 billion have moderate to severe vision impairment [3].

Approximately thirty-seven million of the six billion populations worldwide are suffering from blindness. Unfortunately, 80% of blind people live in developing countries with restricted facilities for them [4].

Smartphones have become an integral part of everyday life. An expected increase of smartphone users will increase up to nine billion by 2022 [5]. The widespread use of smartphones has brought significant changes in how people learn. Research indicates that about one-third of smartphone

usage consists of educational activities. Although smartphone usage has increased exponentially, it has low prevalence among people with visual disabilities. There are many complex accessibility issues that must be resolved in order to enable the full inclusion of this community [6]. Accessibility issues have been an important research domain over the last few years promoting the development of thousands of smartphone applications to help people with a visual disability, e.g., voiceOver services, talkback services, screen readers, and navigators.

These researches resulted in a dramatic increase in mobile-screen reader usage for the visually impaired, from 12% in 2009 to 88% in 2019 [7]. Despite the benefits that smart devices can offer, if the learning applications are not properly designed, their touch-screen interfaces may place an extra burden on blind learners. There are features such as VoiceOver for iPhone that help blind users interact with

their device and browse content. However, educational applications often fail to consider the interaction patterns of blind learners with smart devices.

The language visually impaired people use for reading and writing is known as Braille, which was designed by Louis Braille. It is composed of six raised dots that can be easily written by visually impaired [8]. His design is illustrated in Figure 1. Each Braille character is represented using a combination of six dots arranged in a 3 by 2 matrix [9].

The Braille code system has been widely adopted in several communities because of its simplicity and comfort. Braille has been supported by different languages such as English, Arabic, and Hindi, among others [10]. However, few studies have been conducted on Braille for smartphones.

Research on Braille to text conversion has been carried out in the USA, Canada, India, Pakistan, and France. The literature shows that majority of the conducted research is limited to the USA and Canada. This indicates that there is a considerable demand for such a study in the rest of the world [6].

With the advancement of technology, Braille scripting mechanisms became an important research domain. Within this category, an initial device called Perkins Braille was introduced to facilitate Braille writing. Space, backspace, and line space keys were designed in Perkins Braille, as well as keys corresponding with each of the six dots in the Braille code [11]. In 2008, a lighter and quieter version was developed and launched that included an erase key and integrated carrying handle, which was not available in Perkins Braille.

Another adaptation of the Perkins Braille, the SMART Braille, was created by David S. Morgan and released in 2011 [12]. Along with the existing features, Smart Braille also included text-to-speech functionality in several languages. With the advent of computers, many users created Braille output by connecting a computer and Braille embosser. Visually impaired users were able to read the computer screen using screen reader computer software and/or Braille displays. Another similar Braille recognition system was designed by [13]. In this scheme, images were distributed into three threshold values, and Braille characters were subsequently recognized. Effectively, this interpretation was used to create a suitable dictionary. Recent research has focused on eliminating the need for separate hardware in Braille scripting. Application-level software has been designed to facilitate Braille users.

This survey focuses on gathering the difficulties faced by visually impaired while using a computing-based Braille input mechanism. Many technology-oriented applications for the visually impaired are available. Only those applications are considered for these surveys that are part of research taking place in different countries. Studied applications were analyzed and compared based on matrices related to usability issues for touch-screen-based Braille input methods. These evaluations bring forward the strengths and weaknesses of current schemes with a special focus on usability. No such study exists in the literature, so this paper can provide guidelines to the researchers for designing future applications that are highly usable for visually impaired people.

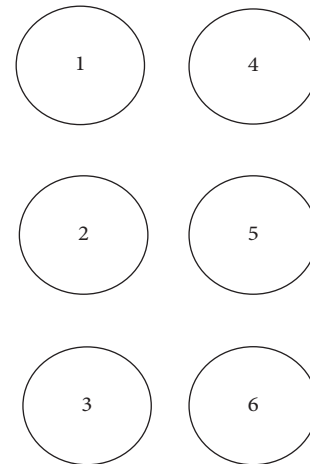


FIGURE 1: Braille Dots.

This paper comprises the following sections. Section 2 gives a detail insight into the previous studies that have been designed for entering Braille data. Section 3 describes the methodology in detail. Input methods are compared and evaluated in Section 4. Usability issues gathered from different studies are also included in this section. Section 5 concludes the paper and gives future recommendations.

## 2. Previous Work

This survey paper provides a review of the current state-of-the-art Braille input methods. In this section, we provide a detailed insight into the problems with these schemes when these are used by visually impaired people. Based on the current survey, we have identified new directions for future research. In recent years, many studies have been conducted to make Braille more technology-assisted. To analyze these studies, we have broadly divided the Braille input mechanism into two main categories:

- (i) *Scanned Input*
- (ii) *Touch-Screen-based Input*

**2.1. Scanned Input.** In the scanned Braille input, Braille Dots are extracted from Braille sheets using a scanner and, then, converted into text using optical character recognition, as shown in Figure 2. In this mechanism, visually impaired users give input on sheets without any interaction with a computing device.

**2.1.1. Arabic Optical Braille Recognition System.** A study was conducted that takes input from a flatbed scanner, as it clearly displayed the Braille Dots. The scanned image was converted into grayscale, the image frame was cropped, and the resulting image was stored in a 2D array. To remove the skewness in the framed image, an algorithm was designed. Finally, the Braille cells were recognized. They achieved approximately 99% accuracy for Braille written in Arabic with single- and double-sided scanned documents. They did not evaluate their system against any other application [14].

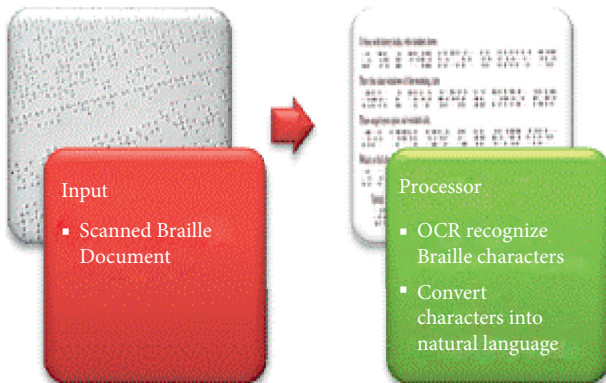


FIGURE 2: Scanned Braille input.

**2.1.2. Deterministic Turing Machine for Context-Sensitive Translation.** To reduce the written communication gap between visually impaired and sighted people, a mechanism was developed to convert Braille codes into Urdu text [15]. They used a scanned Braille document as the input, which was, then, converted into a binary grayscale image. Multiple experiments were conducted for adjusting the threshold value of the scanned document. The primary focus of this study was to evaluate the context-sensitivity of Urdu-Braille. The main issue in the development of Urdu-Braille is that Braille is written left-to-right, whereas Urdu is written right-to-left. This issue was resolved by inventing a deterministic Turing machine that translates Urdu into Unicode and, then, reads the Unicode from left-to-right.

**2.1.3. Text Translation of a Scanned Hindi Document into Braille via Image Processing.** Hindi to Braille conversion was performed so that visually impaired people could have access to a wider range of Hindi literature [16]. They designed a database for consonant and matras generation using an image segmentation technique. Segmented letters were matched with the generated Hindi database using Principal Component Analysis before conversion into equivalent Braille codes. Braille cells for the Hindi text were generated with the help of mapping tables. Every image in the database was sized to  $187 \times 128$  pixels to minimize the memory requirement. This scheme also requires less programming time as it utilizes a letter position in the Hindi database for letter matching.

**2.1.4. Braille Instant Translator.** The Braille Instant Translator converts Braille documents into the English language by identifying Braille Dots embossed on a page [17]. The scanned Braille document uses a camera-generated image as input and identifies the Braille Dots to emboss on the page in a grayscale image format. The extracted image was enhanced through standard deviation evaluation. A histogram was generated from the pixel values for image binarization. The brighter regions were assigned a value of 1, and the remaining regions a value of 0. MATLAB image-processing techniques were used to identify dots that were, then, matched with the English alphabet. Effectually, results

indicated that the study achieved 80% accuracy, but implications are limited to only Grade 1 Braille.

**2.1.5. Web-Based Urdu-Braille Translator.** A web-based Urdu-Braille translator was developed by Iqbal et al. [18] for parents of the visually impaired. The translated Urdu script can be exported to PDF or directly sent to an embosser. Four modules were developed to enable an explicit learning process: a multiple-choice interface, an Urdu-Braille Reading module, a Braille-to-Urdu word mapping module, and a fill-in-the-blank module for Braille-to-Urdu word translation. To prove the effectiveness of their Urdu-Braille translator, a usability study was conducted on the parents of 15 blind people. For these tests, the participants used the Urdu script from BBC Urdu and translated it with the Urdu-Braille translator. Results indicated that, with each new test, the number of correct answers increased, verifying the effectiveness of the translator when applied to Braille learning. Given the educational impacts of the translated Urdu script, parents were able to use the online translator's Braille-learning application for more involvement with their visually impaired kids.

**2.1.6. Feature Extraction Using SDAE.** Braille pattern feature extraction is a cumbersome task. To reduce complexity, Stacked Denoising Auto Encoder (SDAE) was applied for Braille recognition [19]. SDAE is a deep learning technique used for automatic feature extraction and dimension reduction. To create the initial dataset, the authors produced a sample dataset of Braille images by segmenting the original photos taken with a digital camera from a Braille book. This scheme used SDAE for Braille feature extraction in order to obtain initial weights. The output of SDAE was used as input in SoftMax to build the classifier for training, which employed a traditional supervised learning algorithm with an initial dataset. Consequently, a deep network was initialized, and then, weights were fine-tuned using a back-propagation algorithm. A comparison of three different networks was performed, including multilayer perceptron (MLP), radial basis function (RBF), and SoftMax. Results indicate that, in Braille recognition, SDAE outperforms the traditional feature extraction algorithms and SoftMax outperforms MLP and RBF in conjunction with SDAE. Results demonstrate that, when compared with traditional methods, deep learning techniques facilitate recognition of Braille patterns due to effective automatic feature extraction and a reduced preprocessing time.

**2.1.7. Braille Translator: Braille to Speech Converter.** An application was designed by Falcon et al. [20] that converts the scanned Braille document into text and, then, speaks the translated text. Dynamic thresholding technique is applied in this scheme to extract the important information from the scanned document. Scan Braille Dots are then recognized using pattern recognition and recovered using the Braille grid. The final generated image is in the form of Braille Dots in proper lines. For character translation, each dot is read

before being converted into the binary and speech format. The authors were able to achieve an efficiency of 99%. It took only 26 seconds to translate a two-sided document into text and speech.

*2.1.8. ODIA Braille.* Another application named ODIA (a language used in East India) Braille was designed by the authors [21]. In this scheme, a scanned Braille document was converted into text and vice versa using the image-processing technique. Initially, a greyscale image is acquired from the scanned document. This image is enhanced using feature extraction and Braille cell segmentation technique. After this, recognized patterns are stored in the database. This database is verified by using different Braille samples for conversion.

*2.1.9. A Braille Recognition System by Using the Mobile Phone with an Embedded Camera.* A mobile camera-based application is designed in [22] which captures a Braille image. Using this application, a visually impaired person can convert a Braille written text anywhere easily. Noise removal is performed to extract the important features from the image using segmentation and fast dynamic thresholding techniques. A grid, constituting the location of the dots, is converted into vector form, from which Braille characters are recognized and translated into English characters. In this scheme, the time required to convert an image depends upon the quality of the image. Images with more noise will take a longer time to process. This application guarantees a 100% noise reduction. After noise removal, it only takes 2 seconds for the conversion process.

*2.1.10. Recognition of Ethiopic Braille Characters.* A Mechanism to recognize Ethiopic Braille characters was designed in [23]. The authors designed a new skewness correction technique. In the first step, noise is removed, and then, segmentation is performed using the direction field to detect the exact regions of the Braille Dots. In the next step, skewness correction is performed. After this, important Braille cell values such as the height of the cell, the width of the cell distance between different cells, and Braille character lines are identified. Initially, half characters are detected and recognized. Then complete Braille cell formulation and translation is performed. They tested their prototype in MATLAB. An average accuracy of 96.5%–98.5% is achieved for poor to medium quality images, while 99.9% accuracy is achieved for good quality images. Currently, this scheme can only convert one-sided Braille documents.

Table 1 provides a comparison of these schemes for which the Braille input is extracted from scanned documents or camera images.

*2.2. Touch Screen.* In this method, Braille data is input using a touch screen. As outlined in Figure 3, these images are compared with the Braille dataset.

The matched Braille character is then conveyed to the user using various output techniques.

*2.2.1. NavTap and Braille Tap.* The NavTouch keyboard for mobile devices was developed by Guerreiro et al. [24]. This layout of the keyboard divides the alphabet into five rows, starting with a vowel each row. By performing navigation gestures in four different directions, users can navigate through these rows: up, down, left, and right. Both vertical and horizontal navigations are cyclical. Audio feedback is also provided in order to locate the desired letter. Three groups of five users, who had no previous experience in mobile text entry, evaluated designs. Users were first trained on the text entry method. Results indicated that NavTap outperformed other layouts. However, these keyboards require both hands for operation, which is difficult for visually impaired people.

*2.2.2. No-Look Notes.* Bonner designed No-Look Notes an eyes-free, gesture-based text entry system for multitouch devices. In this layout, characters are arranged on the screen in eight pie-shaped menus, in which each segment includes three to four alphabets that correspond to the international standard mapping a phone keypad to letters. A user can move his or her finger on the pie menu, and the voice feedback will respond to assist them in selecting the appropriate character. Gestures are used to enter spaces or undo any action. This layout was tested on a group of 10 visually impaired users and evaluated based on matrices, such as words-per-minute (wpm), as well as the relationship between speed, accuracy, and errors. Results indicated a 100% increase in wpm for No-Look Note as compared with VoiceOver, but both exhibited approximately the same error rate. Not only is this scheme specific to English text entry but also to users complaining about fatigue during the entry procedure [25].

*2.2.3. V-Braille.* V-Braille is a method that conveys Braille characters in mobile devices using vibrations [26]. In V-Braille, the screen is divided into six portions of 3 rows and 2 columns. Each area represents a single Braille Dot. The phone vibrates when the user touches a region that corresponds to the raised dots in the current character. For example, touching the area for dot 2 and 5 provides a strong vibration to help users identify screen dots that are vertically adjacent. V-Braille is a useful output method for deaf-blind mobile device users, since V-Braille only provides haptic feedback. To evaluate this design, a user study was conducted. There were 6 male and 3 female participants in the study. The first task assigned to the users was reading 10 random characters, and the second task was reading short sentences. The average time calculated for reading a single character was between 4.2 and 26.6 seconds. More than 50% of the users were able to read the characters in less than 10 seconds. The time calculated for reading a short sentence was 130–781 seconds. More than 70% of participants clearly understood the meaning of the sentence. Subsequently, a semistructured interview was also conducted to elaborate on the overall environment of the application. Participants reported feeling happy and relaxed using this application. Designers of Braille Play games [27] built upon V-Braille,

TABLE 1: Braille to natural language conversion using scanned images.

References	Tool learning complexity	Language supported	Efficiency	Technique used
[14]	Not applicable	Arabic	99%	Arabic OBR system
[15]	Not applicable	Urdu	Not applicable	Turing machine for context-sensitive translation of Urdu-Braille
[16]	Image segmentation technique	Hindi	Not applicable	Principal component analysis
[17]	Easy to learn, and the Output can be obtained in a pdf form or directly to the embosser	Urdu	Not applicable	Web-based Urdu-Braille translator
[18]	Easy to learn as only grade 1 Braille was used	English	80%	Image-processing techniques
[19]	Easy to learn	English	92%	SDAE using SoftMax
[20]	Easy to use. Text-to-speech facility is also available.	English	99%	Dynamic thresholding technique
[21]	Easy to learn	Odia	Successful one-to-one mapping from ODIA to Braille	Feature extraction and Braille pattern recognition
[22]	Easy to learn, and the mobile can be taken anywhere.	English	100% noise reduction	Mobile camera-based application
[23]	Not applicable	Ethiopic	98.5%	Direction filed tensors are used

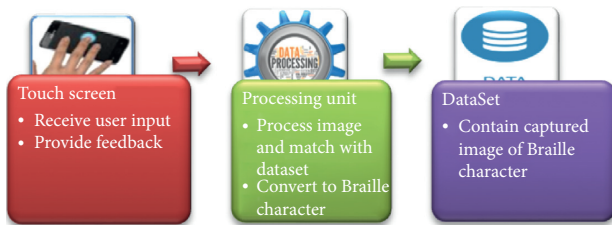


FIGURE 3: Braille input from a touch screen.

adding a speech output that tells the users which dot they are touching. They also created a V-Braille interface for entering characters, where the user double taps regions of the screen to raise the corresponding dots.

**2.2.4. TypeIn Braille.** TypeIn Braille, developed by Mascetti et al., tried to overcome the limitations of the existing keyboards. The TypeIn Braille keyboard uses gestures to enter Braille characters. Three different gestures of single- and multitouch were used for activating and deactivating Braille Dots. Audio or vibration feedback confirms the user input. Flick gestures were defined for input and editing options, such as the end of character, space insertion, text deletion, capitalization, text selection, and cursor relocation. Gesture memorization, time consumption for moving between writing, exploration, and selection mode are the basic drawbacks of the system [28].

**2.2.5. Braille Touch.** A study conducted on Braille keyboard design used six specific screen locations on touch-screen devices for entry of the Braille code [29]. Data entry required both of the user's hands, specifically three fingers from each hand for entry of a single character. Users faced problems simultaneously operating multiple fingers and both hands. To operate this keyboard, blind users must distance the phone from their faces, which can be uncomfortable and

lead to a breach of privacy. Users were forced to write on the fixed locations mentioned on the screen, which burdens visually impaired people who struggle to keep track of the location. Several usability experts evaluated Braille Touch and two visually impaired users. The most common feedback was to provide a position-free text entry that did not require distance between users and their phones.

**2.2.6. Braille Type.** Oliveria et al. proposed a Braille Type. This keyboard was designed to avoid the multitouch technique [30]. Users must perform a single touch to enter a Braille Dot, and after a period, the user receives an audio confirmation. The timer for audio feedback can be adjusted per user's expertise. The user must double-tap the screen anywhere to accept the entered Braille character. The left swipe gesture is used to clear marked Braille cells. They evaluated Braille Type on 15 blind users. Results indicated a text entry speed of 1.45 wpm, a significant improvement from the 2.11 wpm achieved by VoiceOver. The error rate of Braille Type was comparatively low at 8.91% as opposed to VoiceOver's 14.12%. Overall results indicated that Braille Type, when compared with VoiceOver, was easy to learn and exhibited a low error rate, but performed slower. Due to less screen space, the integration of this keyboard with other applications is not possible.

**2.2.7. Mobile Brailier.** A new keyboard layout was designed based on the similar design patterns to those in TypeIn Braille [31]. Along with the tapping function, this keyboard also used swiping gestures for inserting and deleting characters. To enter two dots in a consecutive row, the user must tap with two fingers at the same time. To deactivate the left and right dots, a user must swipe to the right, which removes the last entered character. Mobile Brailier prototyped and compared five different input methods: One-Finger, Split-Tap, Two-Finger, Thumb-Typing, and Nine-Digit. Along



with the tapping function, the right and left swipes are used for adding spaces and backspaces, respectively. The study was conducted on 15 visually impaired people from the greater New York City area, and the prototypes were tested on the android environment. Users answered qualitative and open-ended questions in each application. Results indicated that the One-Finger method was preferred by the visually impaired because of its simplicity. Gesture memorization was among the primary drawbacks. The typing speed of other techniques was faster than the single-finger method, but those techniques were not preferred due to gesture difficulty. Recommendations for these input methods were provided based on the user experience in order to improve prototype deficiencies.

*2.2.8. Braille Key.* Braille Key was designed with four large keys in the screen corners. Two buttons on the upper side are used to enter two columns of Braille code. A user must perform a single touch for entering the first dot, a double-tap to enter the second dot, and a long tap to enter the third dot. The lower two buttons are used for editing the text. The developed prototype was compared with Apple's well-established VoiceOver. Five visually impaired people participated in the study. Users were given a brief 10–15 minute awareness session on how to use both of the applications. Two sentences were entered without correcting any mistakes. Text entry speed and typing accuracy were measured for both applications. Effectively, Braille Key text entry speed and accuracy outperformed the iPhone's VoiceOver. The primary limitation of Braille Key is the identification of the button position on the screen [32].

*2.2.9. Perkinput.* Perkinput is a novel technique designed by Azenkot that uses the input finger detection (IFD) method. The input is entered into the device through multipoint touch. Tracking algorithms are used to detect the input finger reference point [33]. This method uses a 6-bit Braille code with audio feedback and provides single- and double-hand options. A study was conducted on eight users to evaluate this input method. Results found that Perkinput outperformed the iPhone's VoiceOver in speed and accuracy. A case study was conducted for performance evaluation, confirming that Perkinput improved writing proficiency without errors.

The Perkinput keyboard is particularly advantageous because it eliminates the fixed-key concept, successfully resolving navigational problems. However, creating a reference point requires users to single, double, and triple tap, the latter of which is time-consuming and often results in users forgetting the location of the reference point.

*2.2.10. Braille Calculator.* Learning mathematics is as important for the blind as it is for the sighted people. The Braille Calculator is designed so that visually impaired people can easily learn mathematics [34]. The input is taken from a touch-screen device. The Braille Calculator uses a 4-wire

resistive 2.8-inch touch-screen. The screen is divided into six portions for text entry.

The touch-screen is interfaced with the Atmega 328 microcontrollers. An analog to digital converter (ADC) gives the user's impression of the location coordinates. The user is provided with an audio feedback in response to the dot entered. A step-by-step input is received for solving complex equations. Finally, the user is provided audio feedback regarding the solution. The Braille Calculator uses the Atmega 328 microcontrollers and a Secure Digital Card (SD) interface.

*2.2.11. VB Ghost.* VB Ghost, based on the Ghost word game, is an educational smartphone game for people with low or no vision [35]. The V-Braille interface was used as the basis of this game. The game was specially designed for educational and recreational purposes. This application was developed to reduce the accessibility issues that arise while using Braille on touch screens.

Both the hepatic and audio feedback were provided. Vibration occurs at the place where a raised dot occurs. In this game, a word fragment is presented and the player has to complete it. When satisfied with their letters, a player can submit the word by pressing the enter key or swiping with two fingers. The primary purpose of this game is to demonstrate the potential of developing fun, accessible, and educational games for visually impaired users.

*2.2.12. Braille Play.* VB Ghost was further improved and developed in the form of a complete suite named Braille Play. The suite consists of four different games: VBReader, VBWriter, VBHangman, and VBGhost. A longitudinal study conducted resulted in only one child capable of playing the game independently. However, some children were able to acquire the basic Braille-learning skills [27].

To analyze accessibility issues for blind people, four different smartphone applications were evaluated, namely, Blind Navigator, Easy Phone for the Blind, Blind Launcher, and Call Dialer [36]. To perform the study, ten visually impaired people were selected from various educational institutions. Ease of use, learning ability, absence of errors, efficiency, and voice understanding were the primary matrices for comparison. Survey forms were used for collecting feedback. Results indicated that, currently, most visually impaired people use Symbian phones, but blind people who already use smartphones are not ready to use any other device. Furthermore, 70% of the blind people easily understood the message in their Pakistani native language.

*2.2.13. Eye Droid Keyboard.* Another keyboard, Eye droid, was designed for entering Braille patterns using different gestures. To calculate the minimum swipe distance, input coordinates of the Braille Dots entered, that is, X1, Y1 and X2, Y2, are extracted, and the swipe threshold velocity is calculated using Velocity-X and Velocity-Y. The different gestures employed were left-to-right, right-to-left, and bottom-to-top swiping, as well as screen tapping [37]. These

gestures, correspond with functions including activation and deactivation of left and right dots, activation of both dots, and activation of a single dot. The subsequent Eye droid-B scheme was compared with the earlier Eye droid-A design. Survey participants found Eye droid-B to be faster and easier to use when compared with Eye droid-A. This keyboard resolved navigational problems by eliminating the location-specific buttons that troubled users. Alternatively, users only need to memorize the defined gestures.

*2.2.14. Edge Braille.* Edge Braille introduces another text entry method that designed buttons in the touch-screen corners. This method allows a user to draw a continuous line by swiping along the screen edges for the entry of a specific character. When the user slides his or her finger on the screen, a vibrio tactile and voice feedback is returned to inform the user of Braille Dot activation or deactivation [38]. The input speed of Edge Braille was compared with TypeIn Braille and Perkinput. Results showed that the input speed of Edge Braille was faster than that of Braille Type and slower than that of Braille Touch. Furthermore, users found the Edge Braille interface easy to operate. However, numerous problems still exist with this design, since it is impossible to draw lines for all characters based on dot position. This method limits users to enter only alphabets and numbers and does not allow for “editing.” Additionally, activation and deactivation of different dots reduces application speed.

*2.2.15. Braille Easy.* Braille Easy was developed for the entry of Arabic and English Braille codes within a mobile application. This system also used gestures, but for activation of the first and second column, users are required to tap once, twice, and three times [39]. This keyboard was significantly difficult for users to operate because it requires the memorization of different reference points. The keyboard speed was evaluated at 7 wpm, but the error rate was not specified. Furthermore, the keyboard only supports Grade 1 Braille.

*2.2.16. Braille Ecran.* A tactile interface cover for touch-screen phones was designed by [40]. This cover, called Braille Ecran, which provides an interface that consists of six Braille Dots. Each dot was given a tangible button that users can find and press. The significant advantages of this design include the “editing” capability, as well as vibration and audio feedback. However, users encountered considerable problems with the system, including confusion when operating buttons due to a single key’s multifunctionality. Additionally, data entry was associated with a high probability of error due to the close proximity of buttons.

*2.2.17. Single-Tap Braille.* Single-Tap Braille is a position-free text entry method. A user can enter text, numbers, and punctuation by tapping anywhere on the screen. An algorithm runs in the background, interpreting the user’s finger tapping for the identification of the specific corresponding character. The significant limitation of this model is the location-specific memory required; blinds often struggle to

remember where they tapped once they have picked up their finger, eliminating their ability to complete subsequent actions correctly. No audio or vibrotactile feedback was provided in this scheme [41].

*2.2.18. Braille Enter.* Considering the problems highlighted in Single-Tap Braille, Braille Enter was designed to improve upon the method, consequently reducing on-screen navigation problems. In this method, the text is entered by activating and deactivating the Braille six-dot pattern, as shown in Figure 4. Additionally, the model allows for single-hand use and supports the entry of upper and lower-case letters, numbers, and special characters. Special functions are also available, such as adding and removing spaces. Furthermore, audio feedback was also provided to the users. The input received was used for activation and deactivation of the Braille Dots [42]. The press gesture entered the active dots, and a long tap deactivated the dots. The swipe function changed the character mode. The primary problem with Braille Enter is that users must enter all six dots even if only one dot is necessary, resulting in an excessively time-consuming process.

*2.2.19. Braille Sketch.* Braille Sketch, a gesture-based input method, was designed for visually impaired users on touch-screen devices [43]. A user simply draws a gesture for text entry. Audio feedback is provided when words are completed as opposed to letters to reduce time consumption. For error correction typing, an auto typing algorithm was used. A study was conducted on ten participants with visual impairments to evaluate the method. Each participant completed five typing sessions, and results demonstrated that Braille Sketch supports a text entry speed of 14.53 wpm with a 10.6% error rate.

In summary, in contrast to the immediate letter-level, Braille Sketch provides audio feedback to encourage users to type more quickly. To correct typing errors, an auto-correction algorithm is used.

### 3. Methodology

The research problem that motivated the conduction of this study was to highlight the usability issues faced by visually impaired people while using the latest technologies for Braille writing. This study also highlights a deeper understanding of which methods are in use for converting Braille into natural languages to enhance the scope of work performed in Braille language.

This is a survey-based research in which several searches were made to collect relevant research articles. These articles were collected from authentic resources such as Web of Science, IEEE Xplore, and Springer. Different queries were placed for searching such as the “Braille input method,” “touch-screen-based Braille input method,” and “touch-screen-based input method for visually impaired people.” After skimming the results, only those papers were selected that take input from the visually impaired in Braille language, and then, they were processed into natural language. These research papers were

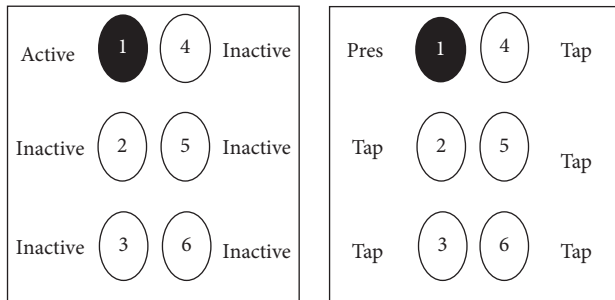


FIGURE 4: Braille Enter input screen.

broadly categorized as scanned-based and touch-screen-based on their input mechanism. In scanned-based, Braille sheets are scanned and input is given to the computer for further processing, while touch-screen-based directly took input from the visually impaired on touch-screen devices and, then, converted that input into its equivalent character.

After categorizing, a visit was made to the National Special Education Center (NSEC) “Mannak Payyan” which is the only local school for people with impairments. Both methods were discussed with the students of the National Special Education Center. For scanned input, since the Braille is written on paper, the students felt no difference in using computing technologies. Touch-screen-based input provided more motivation for the use of the latest technology. Thus, questions related to general user experience on entering Braille on touch screens were asked. A theoretical output was made after analyzing this raw data from which concern problems experienced by the visually impaired people were extracted and used as matrices for evaluating the usability issues. Therefore, we can design a system that helps the visually impaired community by fulfilling all their needs.

#### 4. Comparative Analysis

A performance analysis was performed against VoiceOver based on the input efficiency of different techniques. On average, users require sixteen seconds to enter one word (approximately five Braille characters) using VoiceOver, which was designated as the standard [30].

In sixteen seconds, each method achieved the following word entries: Braille Type with 0.687 words, TypeIn with 1.211 words, Perkinput with 1.516 words, Braille Key with 0.524 words, and Edge Braille with 1.14 words. Evidently, TypeIn Braille and Perkinput outperform the other schemes with respect to data entry speed, as illustrated in Figure 5.

Table 2 summaries the efficiency and accuracy of Braille input schemes currently in use.

Our analysis found that Edge Braille exhibits not only the highest input efficiency at 7.17 wpm but also the lowest accuracy. Alternatively, V-Braille exhibits a low input efficiency (1.32 wpm) but achieves 90% accuracy. Braille Enter exhibits a similar pattern, achieving a text entry speed of 2.45 wpm with 85.88% accuracy.

These results indicate that, in order to achieve high accuracy, input efficiency is compromised. Research is

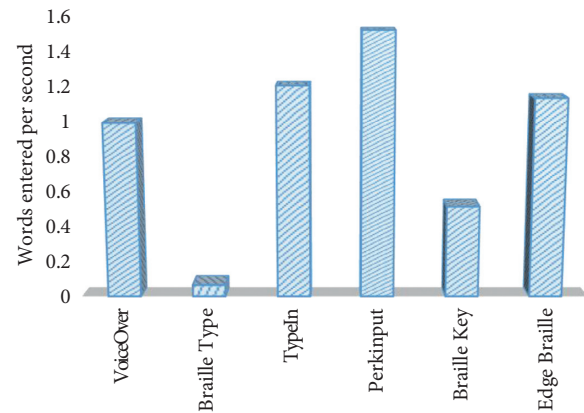


FIGURE 5: Braille input efficiency of different techniques.

needed to design a Braille input technique that delivers satisfactory efficiency along with high accuracy.

We also analyzed how age affects learning Braille on touch screens. Most of the devices used for the assessment ran an android operating system. Figure 6 demonstrates that the age of the participant does not affect the learning time.

**4.1. Usability Analysis.** The usability analysis of touch-screen-based Braille input mechanisms is summarized in Table 3. This analysis was performed on different input schemes based on factors such as tool learning complexity, ease of use, feedback, language support, screen location dependency, and gestures used. Furthermore, two types of feedback, audio, and tactile were explored.

English is the most commonly used language in the current schemes. Visually impaired users found that applications that did not force them to touch specific locations for Braille Dot entry were easy to use. Alternatively, schemes requiring multitouch or memorization of a large number of gestures were found difficult to use.

All of these schemes were designed to enable visually impaired members of the society active participation in the advancing technology. These input methods focus on facilitating touch-screen Braille input for blind users. The primary disadvantages of the current schemes are discussed in the following sections.

**4.1.1. Screen Location Identification.** Keeping track of a specific location on a touch-screen device for entry of Braille Dots is a tiresome task for visually impaired users. Among these schemes, Edge Braille was considered the easiest to use, as the visually impaired were better able to identify the edges of the device.

**4.1.2. Screen Location Identification.** Keeping track of a specific location on a touch-screen device for entry of Braille Dots is a tiresome task for visually impaired users. Among these schemes, Edge Braille was considered the easiest to use, as the visually impaired were better able to identify the edges of the device.

TABLE 2: Performance analysis of touch-screen-based Braille input techniques.

Application name	Efficiency (wpm)	Accuracy/standard deviation	References
Single-Tap Braille	4.71	11.23	[6]
Braille Tap	3.35	3 MSD	[24]
Braille Key	1.8	5 MSD	[23]
Multitap	0.78	15.6 MSD	[24]
NavTap	1.25	9.99 MSD	[24]
VBraille	1.32	90%	[26]
Braille Touch	6.3	—	[29]
Braille Type	1.45	46.15%	[30]
VoiceOver	2.11	29.40%	[30]
Mobile Brailler	2.1	—	[31]
Braille Key	1.8	5 MSD	[32]
Edge braille	7.17	15%	[38]
QWERTY	3.72	20.54	[39]
Braille Enter	2.45	85.88%	[42]

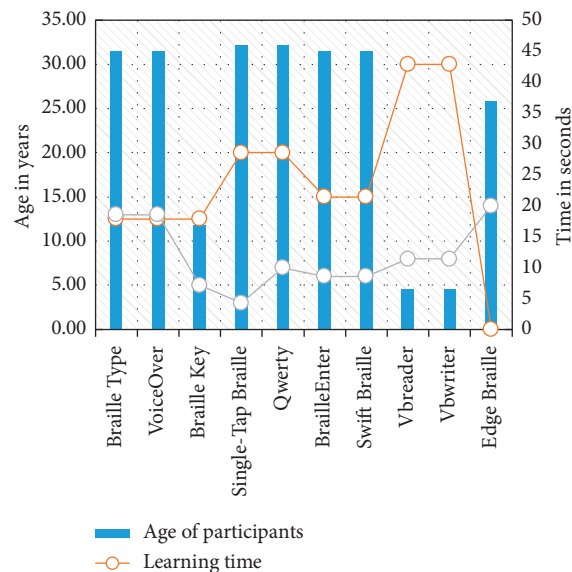


FIGURE 6: Relationship between the age of participants and learning time.

**4.1.3. Use of Both Hands.** Around 30% of the touch-screen-based schemes studied in this survey require that the visually impaired use both hands for entering Braille characters [24, 29, 31, 39]. This is a cumbersome task, especially if a blind user is walking or multitasking.

**4.1.4. Multitaps.** Long press, double-tap, and triple tap each require advanced usability of touch-screen devices. The following schemes require the use of multitaps for entering Braille text [28, 32, 38, 39, 42]. These mechanisms were found confusing to blind users, as mentioned by the examined literature.

**4.1.5. Use of Gestures.** The utilization of too many gestures significantly increases the Braille keyboard learning time, which was especially evident in [27, 28, 37, 42]. Specifically, mapping tasks with gestures, along with entering Braille

Dots, caused a significantly higher error rate when tested on visually impaired users.

The provision of an eyes-free, comfortable text entry method is a vital research area. In current schemes, all the burden of correct Braille entry is placed on the visually impaired user. There is a substantial need for a mechanism that moves the burden from the visually impaired users to the technology, enabling greater accessibility and usability for this specific community.

On the basis of usability analysis, we have designed four new categories that can help researchers to design a better application for the visually impaired.

- (i) Interactive display: for a visually impaired user, an interactive display would be one that is not screen-specific and also provides feedback. It could be an auditory feedback or tactile feedback
- (ii) Efficiency of use: people with visual impairments find those applications more attractive that are less

TABLE 3: Comparison of touch-screen-based Braille input methods.

Sr #	Name	Sample size	Tool learning complexity	Usable	Single/double-handed usage	Tactile/audio feedback	Language supported	Allows editing	Screen location dependency	Gestures used	Technique used	References
1	Single-Tap Braille	7	Low	Low	Single	None	English	NA	No	Tapping	Algorithm for recognizing the taps	[6]
2	NavTap and Braille Tap	15	High	Low	Double	Audio	English	NA	Yes	(i) Up (ii) Down (iii) Left (iv) Right	Minimum string distance rate	[24]
3	No-Look Notes	10	N/A	Medium	Double	Audio	English	Yes	Yes	(i) Split-tapping (ii) Slide rule	T-test/Robust entry techniques	[25]
4	V-Braille	9	N/A	Medium	Single	Tactile	English	N/A	Yes	N/A	Semi-structured interviews	[26]
5	Braille Play	8	High	Low	Single	Audio	English	N/A	Yes	Moving between different options	Gesture recognition back-tracking algorithm	[27]
6	TypeIn Braille	7	High	Low	Both	Audio	English	N/A	N/A	(i) Single-Tap left (ii) Single-Tap right (iii) Double-Tap (iv) Triple-Tap	Usability study	[28]
7	Braille Touch	6	Low	Low	Double	N/A	English	N/A	Yes	N/A	Usability study	[29]
8	Braille Type	15	Low	Medium	Single finger	Audio	English	Yes	Yes	(i) Tap (ii) Left swipe	One-way ANOVA	[30]
9	Mobile Braille	0	High	Medium	Both	N/A	English	Yes	Yes	Swiping	Euclidian distance for calculating touch points	[31]
10	Single-finger prototype	15	Low	High	Single	Audio	English	N/A	Yes	Tapping	Usability study	[31]
11	Braille Key	5	Low	High	Single	Audio	English	N/A	Yes	(i) Single-Tap (ii) Double-Tap (iii) Long press	One-way ANOVA	[32]
12	Braille Calculator	0	Low	Medium	Single	Audio	English	N/A	Yes	Tapping	Adriano UN board	[34]
13	VB Ghost	0	Low	Medium	Single	Tactile and audio	English	N/A	Yes	Tapping	Longitudinal study	[35]
14	Eye droid	12	Low	Medium	Single	Audio	English	No	No	(i) Swipe left (ii) Swipe right (iii) Swipe from bottom-to-top (iv) Swipe from top-to-bottom	Gesture recognition algorithm, experimental study.	[37]

TABLE 3: Continued.

Sr #	Name	Sample size	Tool learning complexity	Usable	Single/double-handed usage	Tactile/audio feedback	Language supported	Allows editing	Screen location dependency	Gestures used	Technique used	References
15	Edge Braille	14	Medium	Medium	Single	Tactile	English	N/A	Yes	Moving between different dots to draw continuous patterns	Lab-based study, ANOVA	[38]
16	Braille Easy	6	Medium	Low	Single	NA	English	N/A	NA	(i) Single-Tap (ii) Double-Tap (iii) Triple-Tap	N/A	[39]
17	Braille Enter	6	Low	Medium	Single	Audio	English	Yes	No	(i) Press (ii) Long tap (iii) Swipe	Braille enter algorithm	[40]
18	Braille Sketch	10	Low	High	Single	Voice	English	N/A	No	N/A	Auto correction algorithm	[42]
19	Braille Ecran	1	High	Medium	N/A	Tactile and audio	English	Yes	Yes	N/A	Predictive and experimental evaluation	[43]

TABLE 4: Usability-based categorization of current schemes.

Touch-screen-based input methods		New category
Single-Tap Braille	[6]	(1) Interactive display
Eye droid	[37]	
Braille Enter	[42]	
Braille Sketch	[43]	
Single-finger prototype	[31]	(2) Efficiency of use
Braille Key	[32]	
Braille Calculator	[34]	
VB Ghost	[35]	
Eye droid	[37]	
Edge Braille	[38]	
Braille Enter	[40]	
Braille Sketch	[42]	
Single-Tap Braille	[6]	(3) Easy memorization
Braille Touch	[29]	
Braille Type	[30]	
Mobile Brailler	[31]	
Single-finger prototype	[31]	
Braille Calculator	[34]	
VB Ghost	[35]	
Braille Sketch	[42]	
Braille Ecran	[40]	
No-Look Notes	[25]	(4) Recovery from errors
Braille Type	[30]	
Mobile Brailler	[31]	
Braille Enter	[40]	
Braille Ecran	[43]	

complex, can be handled using a single hand, and are simple to use.

- (iii) Memorability: location-free specific applications or application that needs few gestures to remember are more appreciable by the visually impaired people.
- (iv) Recovery from errors: applications that allow editing or reentering text helps in recovering from errors.

Table 4 presents the current schemes as per new categorization.

## 5. Conclusions

This survey paper focused on technological assistance available for visually impaired people. Braille input mechanisms can be categorically divided as scanned- and touch-screen-based input methods. In the scanned input, hand-written Braille sheets are scanned using scanners. Various studies have applied machine learning techniques such as optical Braille recognition, deterministic Turing machine for context-sensitive translation, feature extraction, and image processing to extract Braille Dots and convert them into a specific language. In the touch-screen-based input method, Braille Dots are entered using a touch screen on handheld devices such as mobile phones or tablets. Braille consists of six dots, and the basic mechanism of entering Braille using a touch screen requires entering Braille Dots by activating and deactivating pixels on the screen. Once the input is acquired, various algorithms process the extracted Braille Dots and

convert them into their equivalent natural language characters or words. These touch-screen methods use haptic, audio, and tactile feedback to assist visually impaired users. This study compared different input methods on the basis of the entry speed and accuracies achieved, techniques used in the input methods, the number of participants on which the study has been conducted, gestures used, usability level, language used, feedback provided, and screen location independency. These comparisons enabled us to understand the strengths and weaknesses of the current applications. Based on this literature review, we plan to design an application that provides high usability to the visually impaired students. In the future, a newly designed application can be compared with the previous techniques to improve its performance. Machine learning techniques can be applied for acquiring better accuracy for Braille to text conversion.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

This work was supported by Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF2017R1D1A3B04031440).

## References

- [1] R. P. Mills, D. L. Budenz, P. P. Lee et al., "Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease," *American Journal of Ophthalmology*, vol. 141, no. 1, pp. 24–30, 2006.
- [2] P. M. Leonardi, "Innovation blindness: culture, frames, and cross-boundary problem construction in the development of new technology concepts," *Organization Science*, vol. 22, no. 2, pp. 347–369, 2011.
- [3] WHO, *Blindness and Vision Impairment*, WHO, Geneva, Switzerland, 2018.
- [4] Pakistan Today, 2.5 Percentage of Pakistanis Suffer from Blindness| Pakistan Today, Mindblaze Technologies, <https://www.pakistantoday.com.pk/2011/10/14/2-5-percent-of-pakistanis-suffer-from-blindness/>.
- [5] Ericsson, "The ericsson mobility report," 2017, <https://www.ericsson.com/en/mobility-report>.
- [6] M. Alnfai and S. Sampalli, "An evaluation of SingleTap Braille keyboard: a text entry method that utilizes braille patterns on touchscreen devices," in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, Galway, Ireland, October 2016.
- [7] WebAIM, "WebAIMaccessibility in mind," 2019, <https://webaim.org/projects/screenreadersurvey7/>.
- [8] L. Braille, *Method of Writing Words, Music, and Plain Songs by Means of Dots, for Use by the Blind and Arranged for them*, Institution Royale des JeunesAveugles, Paris, France, 1829.

- [9] E. C. Jibril and M. Meshesha, "Recognition of Amharic braille documents," in *Proceedings of the 5th International Conference on the Advancement of Science and Technology*, Chennai, India, 2017.
- [10] D. Saad, Al-Shamma, and S. Fathi, "Arabic braille recognition and transcription into text and voice," in *Proceedings of the 5th Cairo International Biomedical Engineering Conference*, pp. 227–231, IEEE, Cairo, Egypt, December 2010.
- [11] J. O. Bickford and R. A. Falco, "Technology for early braille literacy: comparison of traditional braille instruction and instruction with an electronic notetaker," *Journal of Visual Impairment & Blindness*, vol. 106, no. 10, pp. 679–693, 2012.
- [12] L. M. Michelson, J. Kathleen, and D. Morgan, "Using a new electronic Braille to improve Braille learning at the Florida school for the deaf and blind," *Journal of Visual Impairment & Blindness*, vol. 109, no. 3, pp. 226–231, 2015.
- [13] A. Antonacopoulos and D. Bridson, "A robust braille recognition system," in *Document Analysis Systems VI*, Springer, Berlin, Germany, 2004.
- [14] A. Al-Salma, Y. Al Ohali, M. Al Kanhal, and A. AlRajih, "An Arabic optical braille recognition system," in *Proceedings of the 1st International Conference in Information and Communication Technology and Accessibility, ICTA*, Hammamet, Tunisia, January 2007.
- [15] M. A. Fahiem, "A deterministic turing machine for context sensitive translation of Braille codes to Urdu text," in *Combinatorial Image Analysis*, Springer, Berlin, Germany, 2008.
- [16] U. Beg, K. Parvathi, and V. Jha, "Text translation of scanned Hindi document to braille via image processing," *Indian Journal of Science and Technology*, vol. 10, p. 33, 2017.
- [17] M. Z. Iqbal, S. Shahid, and M. Naseem, "Interactive Urdu braille learning system for parents of visually impaired students," in *Proceedings of the 19th International ACM SIG ACCESS Conference on Computers and Accessibility*, Baltimore, MD, USA, 2017.
- [18] S. Iqbal, A. Ali, M. Younus, M. Huzaifa, and Z. Abbas, *Braille Instant Translator*, pp. 327–328, National University of Computer and Emerging Sciences, Islamabad, Pakistan, 2017.
- [19] L. Ting, X. Zeng, and S. Xu, "A deep learning method for Braille recognition," in *Proceedings of the 6th International Conference on Computational Intelligence and Communication Networks*, November 2014.
- [20] N. Falcon, C. M. Travieso, J. B. Alonso, and M. A. Ferrer, "Image processing techniques for braille writing recognition," in *Computer Aided Systems Theory*, Springer, Berlin, Germany, 2005.
- [21] K. Parvathi, B. M. Samal, and J. K. Das, "ODIA Braille: text transcription via image processing," in *Proceedings of the 1st International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, February 2015.
- [22] S. Zhang and Y. Kazuyoshi, "A braille recognition system by the mobile phone with embedded camera," in *Proceedings of the Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, September 2007.
- [23] M. Hassen and Y. Assabie, "Recognition of ethiopic braille characters," in *Proceedings of the International Conference on Management of Emergent Digital Eco Systems*, ACM, Lyon, France, 2012.
- [24] T. Guerreiro, P. Lagoa, P. Santana, D. Gonçalves, and J. Jorge, "NavTap and BrailleTap: non-visual texting interfaces," in *Proceedings of the Rehabilitation Engineering and Assistive Technology Society of North America Conference*, Resna, Toronto, Canada, 2008.
- [25] M. N. Bonner, J. T. Brudvik, G. D. Abowd, and W. K. Edwards, "No-look notes: accessible eyes-free multi-touch text entry," in *Pervasive Computing*, Springer, Berlin, Germany, 2010.
- [26] C. Jayant, C. Acuario, W. Johnson, J. Hollier, and R. E. Ladner, "V-braille: haptic braille perception using a touchscreen and vibration on mobile phones," in *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2010*, Orlando, FL, USA, October 2010.
- [27] L. R. Milne, C. L. Bennett, A. Shiri, and R. E. Ladner, "Braille Play: educational smartphone games for blind children," in *Proceedings of the 16th International ACM SIG ACCESS Conference on Computers & Accessibility*, Rochester, NY, USA, 2014.
- [28] S. Mascetti, C. Bernareggi, and M. Belotti, "TypeIn Braille: a Braille-based typing application for touchscreen devices," in *Proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility-ASSETS'11*, Dundee, Scotland, October 2011.
- [29] B. Frey, C. Southern, and M. Romero, "Braille touch: mobile texting for the visually impaired," in *Universal Access in Human-Computer Interaction*, Springer, Berlin, Germany, 2011.
- [30] J. Oliveira, T. Guerreiro, H. Nicolau, J. Jorge, and D. Gonçalves, "Braille Type: unleashing braille over touch screen mobile phones," in *Universal Access in Human-Computer Interaction*, Springer, Berlin, Germany, 2011.
- [31] N. Paisios, A. Rubinsteyn, and S. Lakshmi narayanan, "Mobile Braille: making touch-screen typing accessible to visually impaired users," in *Accessibility for Pervasive Computing*, Springer, Newcastle, UK, 2012.
- [32] N. S. Subash, S. Nambiar, and V. Kumar, "Braille Key: an alternative Braille text input system: Comparative study of an innovative simplified text input system for the visually impaired," in *Proceedings of the 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, IEEE, Kharagpur, India, December 2012.
- [33] S. Azenkot, "Eyes-Free input on mobile devices," ProQuest Dissertations and Theses, Washington, DC, USA, 2014.
- [34] Y. V. Gidh, M. S. Latey, A. Roy, K. Shah, and S. Ingle, "Braille calculator," *International Journal of Engineering and Computer Science*, vol. 2, no. 2, pp. 382–481, 2013.
- [35] L. R. Milne, C. L. Bennett, and R. E. Ladner, "VBGhost: a braille-based educational smartphone game for children," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, Bellevue, WA, USA, 2013.
- [36] N. Sultan, K. Siddiq, T. Rashid, and M. Farooque, "Evaluation of smart phone applications accessibility for blind users," *International Journal of Computer Applications*, vol. 127, no. 3, pp. 9–16, 2015.
- [37] M. Shabnam and S. Govindarajan, "Braille-coded gesture patterns for touch- screens: a character input method for differently enabled persons using mobile devices," in *Proceedings of the International Conference on Communication, Computing and Information Technology*, Chennai, India, 2014.
- [38] E. Mattheiss, G. Regal, J. Schrammel et al., "EdgeBraille: braille-based text input for touch devices," *Journal of Assistive Technologies*, vol. 9, no. 3, pp. 147–158, 2015.
- [39] B. Sepic, A. Ghanem, and S. Vogel, *Braille Easy: One-Handed Braille Keyboard for Smartphones*, Qatar Computing Research Institute, Health Technology and Informatics, Qatar, UAE, 2015.



- [40] J. Siqueira, F. A. Soares, and C. R. Silva, "Braille Ecran: a braille approach to text entry on smart phones," in *Proceedings of the IEEE 40th Annual Computer Software and Applications Conference*, IEEE, Atlanta, GA, USA, 2016.
- [41] M. Alnfai and S. Sampalli, "SingleTap Braille: developing a text entry method based on braille Patterns using a single tap," in *Proceedings of the 11th International Conference on Future Networks and Communications*, Quebec, Canada, 2016.
- [42] M. Alnfai and S. Sampalli, "BrailleEnter: a touch screen braille text entry method for the blind," *Procedia Computer Science*, vol. 109, pp. 257–264, 2017.
- [43] M. Li, M. Fan, and K. N. Truong, "Braille Sketch: a gesture-based text input method for people with visual impairments," in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, Baltimore, MD, USA, 2017.

## Research Article

# Certificateless Proxy Reencryption Scheme (CPRES) Based on Hyperelliptic Curve for Access Control in Content-Centric Network (CCN)

Zahid Ullah,<sup>1</sup> Asim Zeb ,<sup>2</sup> Insaf Ullah,<sup>3</sup> Khalid Mahmood Awan,<sup>4</sup> Yousaf Saeed,<sup>5</sup> M. Irfan Uddin,<sup>6</sup> Mahmoud Ahmad Al-Khasawneh ,<sup>7</sup> Marwan Mahmoud,<sup>8</sup> and Mahdi Zareei<sup>9</sup>

<sup>1</sup>Department of Physical and Numerical Sciences, Qurtuba University of Science and Information Technology, Peshawar Campus, 25000 KP, Pakistan

<sup>2</sup>Department of Computer Science, Abbottabad University of Science and Technology, 22500 Havelian, KP, Pakistan

<sup>3</sup>HIET, Hamdard University Karachi, Islamabad Campus, 44000 Islamabad, Pakistan

<sup>4</sup>Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock, Pakistan

<sup>5</sup>Department of Information Technology, University of Haripur, 22620 Haripur, Pakistan

<sup>6</sup>Institute of Computing, Kohat University of Science and Technology, 26000 Kohat, KP, Pakistan

<sup>7</sup>Faculty of Computer & Information Technology, Al-Madinah International University, Kuala Lumpur, Malaysia

<sup>8</sup>King Abdulaziz University, Jeddah, Saudi Arabia

<sup>9</sup>Tecnologico de Monterrey, School of Engineering and Sciences, Zapopan 45201, Mexico

Correspondence should be addressed to Asim Zeb; [asimzeb1@gmail.com](mailto:asimzeb1@gmail.com)

Received 26 January 2020; Revised 23 May 2020; Accepted 10 June 2020; Published 25 July 2020

Academic Editor: Sungchang Lee

Copyright © 2020 Zahid Ullah 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.

Information-centric networking is the developing model envisioned by an increasing body of the data communication research community, which shifts the current network paradigm from host-centric to data-centric, well-known to information-centric networking (ICN). Further, the ICN adopts different types of architectures to extend the growth of the Internet infrastructure, e.g., name-based routing and in-network caching. As a result, the data can be easily routed and accessed within the network. However, when the producer generates contents for authentic consumers, then it is necessary for him/her to have a technique for content confidentiality, privacy, and access control. To provide the previously mentioned services, this paper presents a certificateless proxy reencryption scheme (CPRES) based on the hyperelliptic curve for access control in the content-centric network (CCN). Using certificateless PRE, the power of the key generation center (KGC) is limited to only the generation of partial keys to secure the access to the content. With the help of these partial keys, the producer further calculates keys for encryption and reencryption process. The simulation results show that the proposed scheme provides secure access to content during end-to-end communication. Moreover, the proposed CPRES scheme outperforms in terms of low computational energy and efficient utilization of communication bandwidth.

## 1. Introduction

Information-centric networking (ICN) is an approach to develop the Internet infrastructure to directly support the unique named data [1]. The ICN attracts much attention in the continuing search for a future communication model of the Internet [2]. It shifts the networking model from the current host-centric model, where all requests for content

are made to a host identified by its Internet protocol (IP) address(es), to the data-centric model [3]. Table 1 depicts the differences among both the networks, i.e., host-centric and ICN [4]. An ICN named content can be stored anywhere in the network, and each content object can be uniquely addressed and requested.

Content-centric networking (CCN) is the most encouraging architecture of ICN paradigms, which performs

TABLE 1: Host centric vs ICN.

	Host centric	ICN
Routing	Using IP addresses	Name-based routing
Caching	Specific caching points	Each node can cache the content
Security	Communication channels	Secure the content
API	Data send to a specific address	Publish and subscribe contents

communication by using two specialized kinds of packets, i.e., interest packet and data packet, which carry a name to uniquely identify the requested content [5]. The interest packet is used to advertise a user's request to obtain the interested data, as shown in Figure 1, while the data packet is used to return the corresponded content to the user [6]. Compared with the host-based conversation model of current IP architectures, the content delivery in ICN follows a receiver pushed back method. Once the requested content is matched in ICN, the data are transferred to the receivers with the reverse method.

Therefore, the objective of ICN is to find, publish, and distribute network contents rather than the reachability of end hosts and keep host-to-host discussions between them [6]. For more clarifications, the system model of ICN is shown in Figure 2, where it includes four basic parties [3, 7], namely, content producers; secondly, routers; thirdly, edge service router; and lastly, content consumers. Here, the content producer is responsible for generating the content, converting data to named data objects with desired security bindings and protections, and publishing it in the network.

The routers are responsible to forward requests for data objects and also provide a platform for communication between the consumers and the producer. Routers are composed of three primary elements: (i) forwarding information base (FIB), (ii) pending interest table (PIT), and (iii) content store (CS) [3]. The FIB is used to route incoming interests to the appropriate output port towards the desired content producer. Much like traditional IP routing tables, the FIB is populated using standard routing protocols or static routes and matches content names in interest packets to FIB entries using the longest prefix match. The PIT serves as a cache of the interest state such that content objects that satisfy interests may follow the reverse interest path back to the requester. This preserves upstream and downstream network flow. Finally, the CS is an optional cache for content objects that, if present, is first searched prior to forwarding an interest upstream. These caches serve to reduce content object retrieval latency and bandwidth consumption in the network.

The edge service routers placed at the edge of the ICN network domain have the additional features that allow publishers to deploy certain services such as processing data, forwarding encrypted data to the proper destination, and also storing the content [7]. Lastly, the content consumer downloads the encrypted content from the edge service router through their interest and decrypts with the help of the desired decryption key.

As the Internet shifts from IP-based communication to a content name-based approach, this model will face some critical challenges, for example, mobility, security, access control, routing, naming, and caching [8].

By keeping in view the above observations, access control is one of the most significant techniques for authentication and

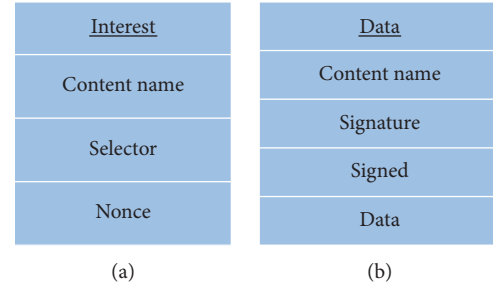


FIGURE 1: (a) Interest packet and (b) data packet.

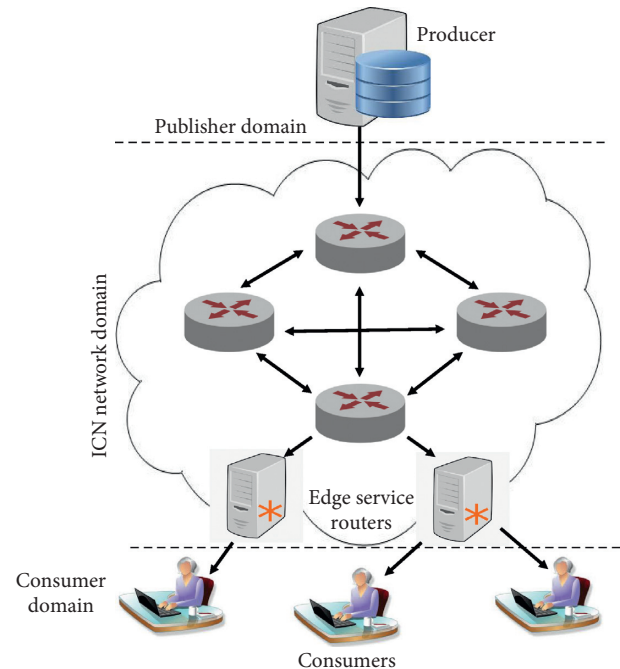


FIGURE 2: ICN system model.

accessing of contents of the ICN architecture. As contents are retrieved from distributed in-network caches, there should be a security mechanism, which ensures the contents' protection and users' authorizations [9]. Since a number of proposals are available in the literature, which can be fruitful for access control, but to the best of our investigation, the certificateless proxy reencryption is the most prominent and securable scheme. So, a certificateless proxy reencryption scheme is the best choice for improving the efficiency and security level because it generates the partial secret key to reduce the extra efforts of the key generation center (KGC) and control the misuse of the secret key.

Motivated by the above insight, the certificateless proxy reencryption scheme based on the hyperelliptic curve for access control is a newly recommended scheme for CCN in this paper. The certificateless proxy reencryption eliminates the key escrow problem that is found in the identity-based proxy reencryption scheme (IB-PRE) [10]. According to our investigatory study, the security hardness and efficiency of existing IB-PRE and certificateless proxy reencryption are based on the standard cryptosystems like Rivest, Shamir, and Adleman (RSA), elliptic curve (EC), and bilinear pairing (BP). The RSA uses a 1024-bit key and public and private parameter sizes, while EC uses 160 bits, where BP is 13.65 ms worse than RSA and 13.93 ms worse than the elliptic curve according to the experimental results in [11], and also 14.42 ms worse than the hyperelliptic curve from the assumption in [12]. The proposed scheme hyperelliptic curve uses 80 bits for the parameter size providing the same level of security along with low computational and communication cost.

*1.1. Motivations and Contributions.* To provide a better and secure networking structure to the information-centric network, the researchers are interested to put more efforts in this field to push the research forward. In this sequence, recently, Wood [10] proposed an identity-based proxy reencryption (IB-PRE) scheme based on elliptic curve cryptography for CCN. But, any proper mechanism for security analysis and algorithm was not specified. Also, the key escrow problem ambiguity was indicated in the IB-PRE scheme. Furthermore, in a recent research in 2019, Wang et al. [13] proposed another PRE scheme using BP cryptography based on the random oracle model. So, the current trends among the cryptographic researchers are that they believe on practical analysis instead of theoretical, e.g., the random oracle model. Furthermore, besides from these two schemes, which are specific to the ICN, a number of public key infrastructure (PKI), identity based, and certificateless signature methods are available in the literature for providing applications to different communication systems [13–17]. The computational and communication cost of this crypto system is so much higher because of using the known cryptographic protocol parameters and key sizes, i.e., RSA uses 1024 bits, where BP is almost 13.65 times worse than RSA, 13.93 times than EC, and 14.42 than the hyperelliptic curve [12], respectively. So, to continue the same debate, by using the results in [12], the EC is 0.28 times faster than RSA and the hyperelliptic curve is 0.48 times faster than EC and 0.77 times quicker than RSA.

As concluded from the above discussion, we found that there is no such scheme, which has formal security analysis and is not suffering from extra computational and communication cost. So, the motivation of our research is to propose a unique CPRES scheme to solve the above-mentioned problems in the form of the certificateless proxy reencryption scheme based on the hyperelliptic curve for access control in content-centric networking. Our contribution is listed in the following steps.

- (i) We proposed a certificateless proxy reencryption scheme based on the hyperelliptic curve for access control in content-centric networking.

- (ii) Our scheme utilizes an 80-bit key instead of the bilinear pairing and the elliptic curve which use 1024-bit key and 160-bit key, respectively.
- (iii) Our scheme removes the key escrow problem of identity-based PRE by using CL-PRE.
- (iv) In terms of computational and communication cost, our scheme is more efficient as compared to the models proposed in [7, 8, 18, 19] and other existing schemes [13, 17, 20–25].
- (v) We provide our security analysis through a recognized security validation tool known as AVISPA.

## 2. Related Work

*2.1. Access Control.* Access control (AC) is the main selected area of the proposed scheme. A number of schemes are proposed for AC in CCN to provide accessibility to only authorized users. The researchers divide an access control method into two ways: namely, encryption-based access control and encryption independent [26]. The encryption-based access control mechanism is further categorized into four ways, i.e., broad encryption, PKI-based encryption, attribute-based encryption, and identity-based encryption. Furthermore, the PKI-based encryption is implemented in three ways, i.e., session based, proxy reencryption, and probabilistic model. This article relates to the proxy reencryption mechanism; so, here, we focus on proxy reencryption access control mechanisms.

The reencryption process is performed by an intermediate proxy node for each consumer; Wood et al. [18] proposed a flexible scheme using the combination of identity-based encryption and proxy reencryption for secure communication. Before the content distribution, the producer encrypts the content with a symmetric key. The consumer can retrieve content from either the producer or the cache node. After receiving the encrypted content by the consumer, it requests a symmetric key from the producer, and the producer verifies the consumer validity and access level and then sends the encrypted symmetric key using the consumer identity to a verifier consumer. The consumer uses this key for decryption of the content.

Another context for AC is proposed by Mangili et al. [19]. In this context, the content is divided into partitions and then fragments. Further, the producer performed two-level encryptions: firstly, the fragments are encrypted using a symmetric key into a chunk, and this chunk is stored in an encrypted form; secondly, the encryption is performed for collusion elimination and confidentiality which uses the “key regression” method for generation of the key chain based on the key derivation algorithm [27]. Using a secure encrypted access obtained from the producer, the authorized consumer regenerates the second-level encryption key. The producer reencrypts the encrypted chunks only for the authorized consumer to protect the collusion.

A unique AC framework was proposed by Zheng et al. [7] for ICN. In this framework, the encryption process is performed by the edge routers. Firstly, the publisher encrypts the content with the public key and  $k_1$  as a random key. When the

consumer sends a request for content access, the edge router selects  $k_2$  as a random key and performs the reencryption on encrypted content. The edge router uses the publisher's public key to encrypt the random key  $k_2$ , attaches it with the content, and then sends it to the consumer. Before the decryption, the consumer sends their identity, content, name, and  $k_2$  to the publisher for verification. The publisher generates another key  $k$ , after the verification of the consumer access level and identity using the private key, along with  $k_1$  and  $k_2$  for the consumer. The consumer decrypts the content using key  $k$ . The decryption key  $k$  is different for every consumer due to the generation of key  $k_2$  randomness of each request.

*2.2. Certificateless Proxy Reencryption (CL-PRE).* For the first time, Blaze et al. [28] presented the concept of PRE in 1998. It was, however, bidirectional and colluding insecure. Following Blaze et al.'s PRE scheme, Ateniese et al. [29] improved it in the form of a unidirectional PRE scheme based on paillier encryption. Later, they proposed two more schemes: chosen plaintext attack (CPA) secure schemes based on the bulletin board system with pairing and two-level encryption schemes. The first chosen ciphertext attack (CCA) was improved by Canneti and Hohenberger [30] in the form of the secure bidirectional multihop PRE scheme. Further, this work was extended by Libert and Vergnaud [31] to make it the chosen ciphertext attack (CCA2) scheme in order to make it more secure and to make reencrypted ciphertext publicly verifiable. First, the CCA2 secure pairing-free bidirectional PRE scheme based on ElGamal encryption and Schnorr's signature was proposed by Deng et al. in [32]. They made it efficient than previous paradigms and left the possibility for the construction of a CCA2 secure PRE scheme in a standard model. It was ultimately solved by Wang et al. in [33] using Cramer-Shoup encryption [34]. They compared their efficiency with the work of Canneti and Hohenberger [30].

To solve the certification management problem in PRE, Green and Ateniese [14] proposed employed conventional PRE in an identity-based (IB) setup, for the first time in 2007. Many other unidirectional IB-PRE schemes have been proposed [35, 36] in the same year. However, the schemes in [35, 37] are insecure against the collusion attack in which a private key of the delegator can be extracted by proxy. Later, Wang et al. proposed in [15] another IB-PRE scheme based on the random oracle model, and Mizuno and Doi [38] designed one more IB-PRE algorithm based on the chosen plaintext attack security using a standard model. Using the standard model, another CCA-secure IB-PRE scheme was proposed by Shao and Cao in [39]. The first CCA-secure single-hop IB-PRE based on the standard model to maintain conditional reencryption was introduced by Liang et al. in [40]. Further, in 2014, Liang et al. continued their work and designed a cloud-based revocable IB-PRE scheme in which ciphertexts are reencrypted by proxy under an identity and time period in [41]. However, Wang et al. proved in [36] that Liang's scheme in [40] is weak against collusion and reencryption key dummy attack although the withdrawal users decrypt the encrypted data after time expires which was allowed by it. They further proposed the improved version using the standard model based on expensive pairing operations.

Another ambiguity is exposed in identity-based encryption in the form of the key escrow problem. It provides growth, for instance, to certificateless PRE (CL-PRE). CL-PRE developed with pairing for the first time was presented by Sur et al. in [42], and since then, this development has attracted more attention from academia and research community. They claimed their scheme to be CCA-secure, but Zheng et al. proved in [43] that the concrete attack is possible in their scheme. CL-PRE scheme for data distributing with the public cloud using encryption-based access control and key management was designed by Xu et al. [20] in 2012. They claimed its security against a chosen plaintext attack. To increase the security and efficiency level, they further designed the multiproxy and randomized CL-PRE scheme. In 2013, replayable CCA-secure PRE scheme based on the random oracle model was proposed by Guo et al. [23] to verify that Xu et al.'s scheme in [20] is weak against type I adversary. The above schemes [20,23,42] were based on expensive bilinear pairing operations. To conclude the PRE literature, only few pairing-free CL-PRE schemes exist. The first pairing-free CL-PRE scheme was proposed by Lee and Han [24] in 2014. Also, they compared their work with Xu et al.'s [20] and Sur et al.'s [42] schemes and proved that their scheme is better in terms of confidentiality and computation time. In 2014, to improve the security models in [24], a CCA-secure bidirectional CL-PRE scheme was proposed by Wang et al. [16]. However, for reencryption process, proxy has required secret keys of both the sender and the receiver.

Qin et al. [25] proposed another CL-PRE scheme in 2015 for data distributing in cloud and compared its security with CCA based on the strong security model. However, any formal security analysis was not provided by them. The simulation results proved that their scheme performance is better than Xu et al.'s scheme [20], Sur et al.'s scheme [42], and Lee and Han's scheme [24] in terms of storage and communication overhead.

Another CCA-secure unidirectional and single-hop CL-PRE scheme was proposed by Srinivasan and Rangan [22]. They broke the confidentiality of the scheme in [24] and proved that it is insecure. They also compared their work in terms of efficiency with Guo et al.'s scheme [23]. The proposed scheme of Srinivasan and Rangan [22] required several precalculations to perform the key generation process. It could also be stored locally. As a result, it increased the storage capacity, which was not suitable for constrained resource devices.

Recently, in 2018, Bhatia et al. [17] proposed another CL-PRE scheme for health care environment based on elliptic curve cryptography which uses a 160-bit key size. They compared their scheme efficiency with the schemes in [20, 22–25, 42] in terms of computational and communication cost. Furthermore, in a recent research in 2019, the PRE scheme for access control in ICN was proposed by Qiang Wang et al. [13] which is based on the random oracle model using bilinear pairing cryptography.

### 3. Materials and Methods

*3.1. Preliminaries.* First time in 1988, Koblitz designed the EC simplification form to uphold class of the curve, known as hyperelliptic curve (HEC). The HEC performance is more remarkable when compared to that of the elliptic

curve (EC), and it uses a smaller key with the same security level [44]. To break the HEC security is more difficult due to the solution of the hyperelliptic curve discrete logarithm problem (HECDLP) [45]. Also, HEC provides more suitable environment for resource-constrained devices.

Let us suppose  $\mathcal{C}\mathcal{R}\mathcal{V}$  is the curve on the field  $\mathbb{F}_n$  and  $\mathbb{F}_n$  is the finite set on this field in order  $n$ . The length of the type one curve on the field  $\mathbb{F}_n$  is as long as " $n$ "  $\log_2 n \approx 2^{160}$ . Also, the length of the type two curve on the field  $\mathbb{F}_n$  with  $|\mathbb{F}_n| \approx 2^{80}$  is 80 bits [44, 45].

Let the finite field of HEC be  $\mathbb{F}$ , the algebraic closure be  $\overline{\mathbb{F}}$  over the field  $\mathbb{F}$ , and  $\mathcal{C}\mathcal{R}\mathcal{V} > 1$  be the type of curve of HEC on  $\mathbb{F}$ . The solution set is described as  $(\mathcal{J}, j) \in \mathbb{F} * \mathbb{F}$ . Equation (1) represents the HEC which is as follows:

$$\mathcal{C}\mathcal{R}\mathcal{V}: j^2 + h(\mathcal{J})j = f(\mathcal{J}). \quad (1)$$

So,  $h(\mathcal{J}) \in \mathbb{F}[\mathcal{J}]$  and  $f(\mathcal{J}) \in \mathbb{F}[\mathcal{J}]$  are polynomial of degree  $\mathcal{G}$  and monic polynomial of degree  $2\mathcal{G} + 1$ , respectively. To calculate equation (1), there is no solution set of  $(\mathcal{J}) \in \mathbb{F} * \mathbb{F}$ . Hyperelliptic curve at  $\mathcal{G} = 1$  is the specific case of the elliptic curve [44].

Furthermore, the hyperelliptic curve discrete logarithm problem (HECDLP) is populated by its own in the field of cryptography because of providing the hard security level. It is used in different cryptographic approaches, e.g., ElGamal [46], based on the discrete logarithm problem.

The HECDLP is defined as suppose  $D$  is the divisor from  $\mathcal{C}\mathcal{R}\mathcal{V}$  and  $\ell$  is the integer which belongs to  $\mathbb{F}_n$ , so finding  $\ell$  from  $y = \ell.D$  is said to be HECDLP.

**3.2. Architecture of Proposed Model.** The proposed certificateless proxy reencryption scheme for AC in CCN is described in Figure 3, which contains four basic parties, i.e., key generation center (KGC), producer, edge service router, and consumer, respectively. Firstly, the producer and the consumer send their identity ( $ID_p$  and  $ID_c$ ) to the KGC. The KGC calculates the master public key  $\mathcal{L} = \delta.L$  and publishes the parameters  $\psi = \{HEC, \mathbb{F}_n, n, n \leq 280, \mathcal{L}, L, h\}$ . Further, the KGC delivers the partial private key  $\mathcal{E}_p = (\alpha_p, \beta_p)$  using the secure network and the partial public key  $\mathcal{Q}_p = (\mathcal{X}_p, \mathcal{Y}_p, \mathcal{Z}_p, \gamma_p)$  using the insecure network to each participant with their identity  $ID_p$ , and then each participant, using their identity  $ID_p$ , sets a secret value  $\mathcal{U}_p = (\mathcal{J}_p, \mathcal{H}_p)$  and generates private and public keys  $\mathcal{P}_p = (\alpha_p, \beta_p, \mathcal{J}_p, \mathcal{H}_p)$  and  $\mathcal{P}\mathcal{B}_p = (\mathcal{X}_p, \mathcal{Y}_p, \mathcal{Z}_p, \gamma_p, \mathcal{B}_p, \mathcal{J}_p)$ . Also, the producer generates a reencryption key  $\Omega$  for level-2 encryption. In this process, it takes the input, identity  $ID_p$ , public and private keys ( $\mathcal{P}_p$  and  $\mathcal{P}\mathcal{B}_p$ ) of the producer, public key of the consumer  $\mathcal{P}\mathcal{B}_c$ , and the identity of the consumer  $ID_c$ . Now, the level-1 encryption is performed by the producer on the content (CNT) by taking input the public key  $\mathcal{P}\mathcal{B}_p$  of the producer and public parameters  $\psi$  and this encrypted content is sent along with the reencryption (level-2) key  $\Omega$  to the concerned edge service router using a secure channel. Further, the edge service router performed reencryption (level-2) process using the reencryption key  $\Omega$  and public parameters  $\psi$ , and also computes  $\mathcal{C}1^* = \mathcal{C}1 \oplus \Omega$  and  $\mathcal{C}2^* = \mathcal{C}2$  and

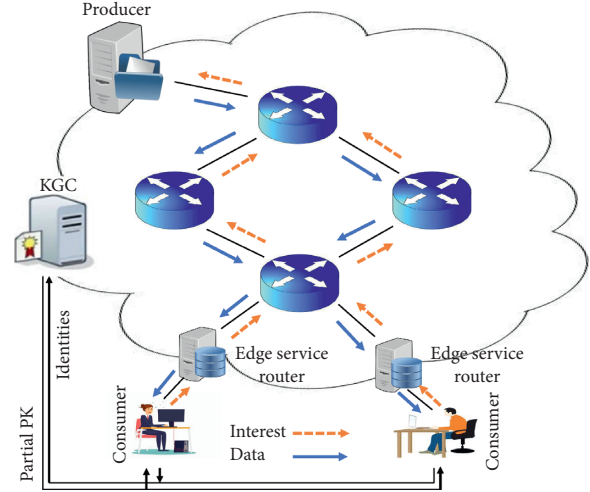


FIGURE 3: Architecture of the proposed CL-PRE scheme for access control in CCN.

sends the pair  $\Phi = (\mathcal{C}1^*, \mathcal{C}2^*)$  to the consumer. Finally, the consumer takes input  $\Phi = (\mathcal{C}1^*, \mathcal{C}2^*)$  and  $(\mathcal{E}_c, \mathcal{J}_c, \mathcal{H}_c)$  to decrypt the content.

**3.3. Basic Notation.** Table 2 represents the basic notations that are used in the proposed algorithm.

## 4. Construction of Proposed Algorithm

The proposed certificateless proxy reencryption scheme CPRES algorithm includes the following nine phases:

**Setup.** In this phase, the KGC selects a security  $\Upsilon$  and hyperelliptic curve (HEC) over the field  $\mathbb{F}_n$  of order  $n \leq 280$ , suppose  $\mathcal{L}$  is the divisor on HEC of order  $n$ . Further, KGC picks a secret key  $\delta \in \{1, 2, \dots, n-1\}$  and calculates a master public key as  $L = \delta.L$ . Finally, the parameters  $\psi = \{HEC, \mathbb{F}_n, n, n \leq 280, \mathcal{L}, L, h\}$  are published.

**Partial Private Key Extract (PPKE).** In this PPKE phase, the KGC first randomly selects three numbers  $x, y, z \in \{1, 2, \dots, n-1\}$  and calculates  $\mathcal{X}_p = x.L$ ,  $\mathcal{Y}_p = y.L$ , and  $\mathcal{Z}_p = z.L$ . It further computes  $\alpha_p = x + \delta.(ID_p, \mathcal{X}_p)$ ,  $\beta_p = y + \delta.(ID_p, \mathcal{Y}_p)$ , and  $\gamma_p = z + \delta.(ID_p, \mathcal{Z}_p, \mathcal{X}_p, \mathcal{Y}_p)$ . Then, KGC delivers a partial private key  $\mathcal{E}_p = (\alpha_p, \beta_p)$  utilizing the secure network and the partial public key  $\mathcal{Q}_p = (\mathcal{X}_p, \mathcal{Y}_p, \mathcal{Z}_p, \gamma_p)$  utilizing the insecure network, to each participant with identity  $ID_p$ .

**Set Secret Value (SSV).** In SSV, each participant with identity  $ID_p$  selects two random numbers  $\mathcal{J}_p$  and  $\mathcal{H}_p \in \{1, 2, \dots, n-1\}$ , as a secret value  $\mathcal{U}_p = (\mathcal{J}_p, \mathcal{H}_p)$ .

**Generate Private Key (GPK).** In GPK, each participant with identity  $ID_p$  generates the private key  $\mathcal{P}_p = (\alpha_p, \beta_p, \mathcal{J}_p, \mathcal{H}_p)$ . In this process, it takes input the partial private key  $\mathcal{E}_p$  and secret value  $\mathcal{U}_p$ .

**Generate Public Key (GPBK).** In GPBK, each participant with identity  $ID_p$  first computes  $\mathcal{B}_p = \mathcal{J}_p.L$  and  $\mathcal{J}_p = \mathcal{H}_p.L$  and generates the public key  $\mathcal{P}\mathcal{B}_p = (\mathcal{X}_p, \mathcal{Y}_p,$

TABLE 2: Notations of the proposed scheme.

S.no	Notation	Description
1	HEC	Hyper elliptic curve from the finite field $F_n$ having order $n$
2	$n$	It is a large prime number, and the order is $n \leq 280$
3	$\psi$	Public parameter set
4	$Y$	Symbolizes the input security parameter from the hyperelliptic curve
5	$\delta$	Master secret key of KGC
6	$\mathcal{L}$	Master public key of KGC
7	$h$	Irreversible or one-way hash function
8	$\mathcal{L}$	Divisor on HEC
9	$\mathcal{E}_{pr}, \mathcal{E}_{cr}$	Partial private keys for producer and consumer
10	$\mathcal{Q}_{pr}, \mathcal{Q}_{cr}$	Partial public keys for producer and consumer
11	$\mathcal{I}_{pr}, \mathcal{H}_{pr}$	Secret values of producer
12	$\mathcal{I}_{cr}, \mathcal{H}_{cr}$	Secret values of consumer
13	$\mathcal{P}_{pr}, \mathcal{P}_{cr}$	Full private keys of producer and consumer
14	$\mathcal{PB}_{pr}, \mathcal{PB}_{cr}$	Public keys of producer and consumer
15	IDpr, IDcr	Identities of producer and consumer
16	$\Omega$	Reencryption key
17	CNT	It means the contents (plain text)
18	$Lfk$	Level-1 encryption key
19	$Lfk^*$	Level-2 decryption key
20	Npr	Fresh nonce
21	$\mathcal{E}_{pr}$	Level-1 encryption
22	$\Phi$	Level-2 encryption
23	$\oplus$	Used for encryption and decryption

$\mathcal{E}_p, \gamma_p, \mathcal{B}_p, \mathcal{I}_p$ ). In this process, it takes input the partial public key  $\mathcal{Q}_p$  and secret value  $\mathcal{U}_p$ .

*Generate Reencrypt Key (GREK).* In GREK, the producer generates a proxy reencryption key  $\Omega$  for level-2 encryption. In this process, it takes input the identity of the producer IDpr, the public and private keys ( $\mathcal{P}_{pr}$  and  $\mathcal{PB}_{pr}$ ), the public key of the consumer  $\mathcal{PB}_{cr}$ , and the identity of the consumer IDcr. The following steps more clearly explain the generation of the proxy reencryption key:

Compute  $Q_{pr} = \mathcal{X}_{cr} + L(\text{IDcr}, \mathcal{X}_{cr})$   
 Compute  $Q_{pr} = (\mathcal{I}_{pr}, Q_{pr}, \alpha_{pr}, \mathcal{I}_{cr}, \text{IDpr}, \text{IDcr}, \mathcal{PB}_{pr}, \mathcal{PB}_{cr})$   
 Compute  $\Omega = ((\alpha_{pr} + \mathcal{I}_{pr})(\mathcal{X}_{pr}, \mathcal{Y}_{pr}, \mathcal{B}_{pr}, \mathcal{I}_{pr}) + \alpha_{pr} + \mathcal{H}_{pr}) Q_{pr}$

*Level-1 Encrypt.* In this L-1 phase, the producer generates the level-1 encryption on content (CNT), by taking input the public key  $\mathcal{PB}_{pr}$  of the producer and public parameters  $\psi$ . The following are the steps:

Choose nonce Npr  
 Choose  $\mathcal{O} \in \{1, 2, \dots, n-1\}$   
 Compute  $\mathcal{R} = h(\text{CNT}, \text{Npr}, \mathcal{B}_{pr}, \text{IDpr}, \mathcal{I}_{pr})$   
 Compute  $\mathcal{E}_1 = \mathcal{R} \cdot \mathcal{L}$ , compute  $\mathcal{E}_3 = \mathcal{O} \cdot \mathcal{L}$   
 Compute level-1 encryption key  $Lfk = (\mathcal{R}((\mathcal{X}_{pr} + (\text{IDpr}, \mathcal{X}_{pr}) + \mathcal{B}_{pr})(\mathcal{X}_{pr}, \mathcal{Y}_{pr}, \mathcal{B}_{pr}, \mathcal{I}_{pr}) + \mathcal{Y}_{pr} + \mathcal{L}(\text{IDpr}, \mathcal{Y}_{pr}) + \mathcal{I}_{pr}))$   
 Compute  $\mathcal{E}_2 = (\text{CNT}, \text{Npr}) \oplus Lfk$   
 Compute  $\mathcal{E}_4 = \mathcal{O} + (\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3)$  and return  $\mathcal{E}_{pr} = (\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4)$  for proxy

*Level-2 (Reencrypt).* In this L-2 phase, the edge server router generates the level-2 encryption on level-1 cipher text, by taking input the reencryption key  $\Omega$  and public parameters  $\psi$ . The edge service router first computes  $\mathcal{E}_1^* = \mathcal{E}_1 \oplus \Omega$  and  $\mathcal{E}_2^* = \mathcal{E}_2$  and sends the pair  $\Phi = (\mathcal{E}_1^*, \mathcal{E}_2^*)$  to the consumer.

*Decryption.* This process takes input  $\Phi = (\mathcal{E}_1^*, \mathcal{E}_2^*)$  and  $(\mathcal{E}_{cr}, cr, \mathcal{H}_{cr})$  and produces the plaintext. The consumer performs the following steps:

Compute  $Q_{cr} = \mathcal{X}_{pr} + L(\text{IDpr}, \mathcal{X}_{pr})$   
 Compute  $Q_{cr} = (\mathcal{B}_{pr}, \alpha_{cr}, Q_{cr}, \mathcal{H}_{cr}, \text{IDpr}, \text{IDcr}, \mathcal{PB}_{pr}, \mathcal{PB}_{cr})$   
 Compute  $Lfk^* = (\mathcal{E}_1^*) / Q_{cr}$   
 Decrypt  $(\text{CNT}, \text{Npr}) = \mathcal{E}_2^* \oplus Lfk^*$

## 5. Security Analysis

Detailed analysis of the proposed scheme with respect to showing the resistance against the intruders included confidentiality (level-1 and level-2) and replay attack which are given below.

*5.1. Confidentiality of Level-1 Encryption.* Confidentiality is a rule to block the access of an unauthorized user to the secure and protected data. So, in this proposed scheme, when the intruders want to get the actual content, they must have a level-1 encryption secret key, that is,  $Lfk$ , and  $Lfk = (\mathcal{R}((\mathcal{X}_{pr} + (\text{IDpr}, \mathcal{X}_{pr}) + \mathcal{B}_{pr})(\mathcal{X}_{pr}, \mathcal{Y}_{pr}, \mathcal{B}_{pr}, \mathcal{I}_{pr}) + \mathcal{Y}_{pr} + \mathcal{L}(\text{IDpr}, \mathcal{Y}_{pr}) + \mathcal{I}_{pr}))$ . It is very hard for intruders to find  $Lfk$  because in  $Lfk$ , the producer concatenates

TABLE 3: Computational cost comparisons on the basis of major operations.

Schemes	Involves participants				
	Reencryption	Encryption	Proxy reencryption	Decryption	Total
Xu et al.'s [20]	2 EXPO	3 EXPO + 1 BPR	3 EXPO + 1 BPR	2 BPR	8 EXPO + 4 BPR
Guo et al.'s [23]	5 EXPO	3 EXPO + 2 BPR	5 EXPO + 1 BPR	8 EXPO + 2 BPR	21 EXPO + 5 BPR
Lee and Han's 1 [24]	3 EXPO	4 EXPO	1 EXPO	4 EXPO	12 EXPO
Lee and Han's 2 [24]	3 EXPO	5 EXPO	3 EXPO	10 EXPO	21 EXPO
Wang et al.'s [25]	3 EXPO	4 EXPO	1 EXPO	7 EXPO	15 EXPO
Srinivasan and Rangan's [22]	5 EXPO	8 EXPO	1 EXPO	6 EXPO	20 EXPO
Bhatia et al.'s [17]	4 PM	5 PM	1 PM	7 PM	17PM
Wang et al.'s [13]	1 EXPO	2 SM + 2 BPR	1 BPR	1 BPR + 1 SM	1 EXPO + 3 SM + 4 BPR
Proposed	4 HDM	5 HDM	1 HDM	3 HDM	13 HDM

his/her own private key, i.e.,  $\mathcal{I}_p$ , with other parameters. Further, the intruder calculates  $\mathcal{I}_p$  from  $\mathcal{I}_p = \mathcal{H}_p$  which is harder due to the hyperelliptic curve discrete logarithm problem (HECDLP).

**5.2. Confidentiality of Level-2 Encryption.** In this phase, the confidentiality of the proposed scheme is analyzed for both cases for intruders and also for the key generation center (KGC), i.e., the part of the network.

*Case 1.* Again, when the intruders want to get the content, they must have a level-2 encryption (reencryption) secret key, that is,  $\Omega$ , and  $\Omega = ((\alpha_{pr} + \mathcal{I}_{pr}) (\mathcal{X}_{pr}, \mathcal{Y}_{pr}, \mathcal{B}_{pr}, \mathcal{I}_{pr}) + \alpha_{pr} + \mathcal{H}_{pr}) Q_{pc}$ . Due to the use of the producer partial private key  $\alpha_{pr}$  and  $\alpha_p = x + \delta$  (IDp,  $\mathcal{X}_p$ ) it is very hard for intruders to calculate the level-2 encryption secret key.

*Case 2.* Also, for KGC they must need  $\mathcal{B}_{pr}$  and  $\mathcal{B}_{pr} = \mathcal{I}_p \cdot \mathcal{L}$ . To find  $\mathcal{B}_{pr}$  again, they must calculate hyperelliptic curve discrete logarithm problem (HECDLP) that is infeasible for KGC.

**5.3. Replay Attack.** In our proposed algorithm, the producer generates and associates a nonce (Npr) value with every content like (CNT, Npr). This nonce value is the identity of every content. If any active intruder tries to send messages regularly for disturbance or breaking the communication, the producer can easily identify due to this nonce identity value. So, our proposed scheme is fully safe from replay attack.

## 6. Performance Evaluation

We evaluate our proposed approach in terms of different properties, e.g., computational and communication overhead, in Tables 3 and 4 and Figures 4 and 5, respectively.

**6.1. Computational Cost.** The comparison of the proposed scheme in terms of the computational cost with the latest contribution to the certificateless proxy reencryption scheme, i.e., Xu et al. [20], Guo et al. [23], Lee and Han [24], Wang et al. [25], Srinivasan and Rangan [22], Bhatia et al.

[17], and Wang et al. [13], is illustrated. To show this, we select the major operations, for example, bilinear pairing operation (BPR), modular exponential (EXPO), elliptic curve point multiplication (PM), and hyperelliptic curve divisor multiplication (HDM), in the proposed scheme and those by Xu et al. [20], Guo et al. [23], Lee and Han [24], Wang et al. [25], Srinivasan and Rangan [22], Bhatia et al. [17], and Wang et al. [13] for computational cost comparisons. Further, the cost of the abovementioned major operations is shown in Table 3, with respect to proposed and the existing schemes. Also, the computational cost comparison is calculated with respect to milliseconds (ms), illustrated in Table 4. To demonstrate the computational time in milliseconds of different cryptographic operations, we use the theoretical results of schemes [12, 47] such as a single BPR consumes 14.90 ms, EXPO consumes 1.25 ms, scalar multiplication on  $G$  takes 4.31 ms, PM consumes 0.97 ms, and HDM consumes 0.48 ms, respectively. As a result, the proposed scheme reduces the computational cost up to 91.26% from the recent research scheme [13], and the differentiation from other schemes is shown in Figure 4.

Further, a recognized formula ((existing framework – proposed method) divided by (existing framework)) to calculate the reduction of the computational cost in millisecond is used, see [12]. Now, the difference of the proposed scheme's computational cost from other schemes is as follows: difference from Xu et al.'s scheme [20] is  $(8 \text{ EXPO} + 4 \text{ BPR} - 13 \text{ HDM}) / (8 \text{ EXPO} + 4 \text{ BPR}) = (69.6 - 6.24) / 69.6 * 100 = 91.03\%$ , from Guo et al.'s scheme [23] is  $(21 \text{ EXPO} + 5 \text{ BPR} - 13 \text{ HDM}) / (21 \text{ EXPO} + 5 \text{ BPR}) = (100.75 - 6.24) / 100.75 * 100 = 93.806 \text{ vvv}\%$ , from Lee and Han's scheme 1 [24] is  $(12 \text{ EXPO} - 13 \text{ HDM}) / (12 \text{ EXPO}) = (15 - 6.24) / 15 * 100 = 58.4\%$ , from Lee and Han's scheme 2 [24] is  $(21 \text{ EXPO} - 13 \text{ HDM}) / (21 \text{ EXPO}) = (26.25 - 6.24) / 26.25 * 100 = 76.22\%$ , from Wang et al.'s scheme [25] is  $(15 \text{ EXPO} - 13 \text{ HDM}) / (15 \text{ EXPO}) = (18.75 - 6.24) / 18.75 * 100 = 66.72\%$ , from Srinivasan and Rangan's scheme [22] is  $(20 \text{ EXPO} - 13 \text{ HDM}) / (20 \text{ EXPO}) = (25 - 6.24) / 25 * 100 = 75.04\%$ , from Bhatia et al.'s scheme [17] is  $(17 \text{ PM} - 13 \text{ HDM}) / (17 \text{ PM}) = (16.49 - 6.24) / 16.49 * 100 = 62.15\%$ , and from Wang et al.'s scheme [13] is  $(1 \text{ EXPO} + 3 \text{ SM} + 4 \text{ BPR} - 13 \text{ HDM}) / (1 \text{ EXPO} + 3 \text{ SM} + 4 \text{ BPR}) = (73.78 - 6.24) / 73.78 * 100 = 91.54\%$ , respectively. In Figure 5, we illustrate the difference of computational cost of the proposed scheme from that of Xu et al.'s [20], Guo et al.'s [23], Lee and Han's 1 [24],



TABLE 4: Computational cost comparisons on the basis of millisecond.

Schemes	Involves participants					Total (ms)
	Reencryption (ms)	Encryption (ms)	Proxy reencryption (ms)	Decryption (ms)		
Xu et al.'s [20]	2.5	18.65	18.65	29.08	69.6	
Guo et al.'s [23]	6.25	33.55	21.15	39.8	100.75	
Lee and Han's 1 [24]	3.75	5	1.25	5	15	
Lee and Han's 2 [24]	3.75	6.25	3.75	12.5	26.25	
Wang et al.'s [25]	3.75	5	1.25	8.75	18.75	
Srinivasan and Rangan's [22]	6.25	10	1.25	7.5	25	
Bhatia et al.'s [17]	3.88	4.85	0.97	6.79	16.49	
Wang et al.'s [13]	1.25	38.42	14.90	19.21	73.78	
Proposed	1.92	2.4	0.48	1.44	6.24	

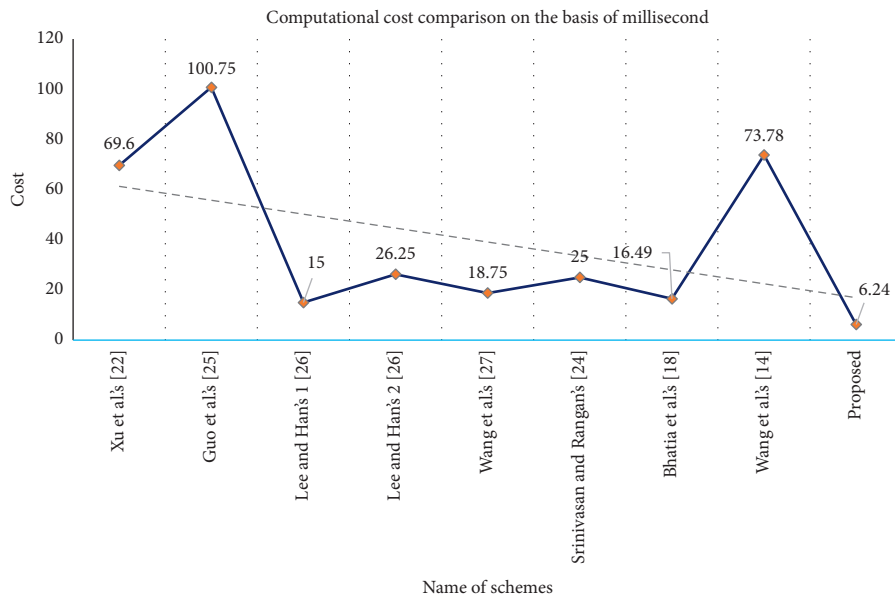


FIGURE 4: Computational cost comparison in millisecond.

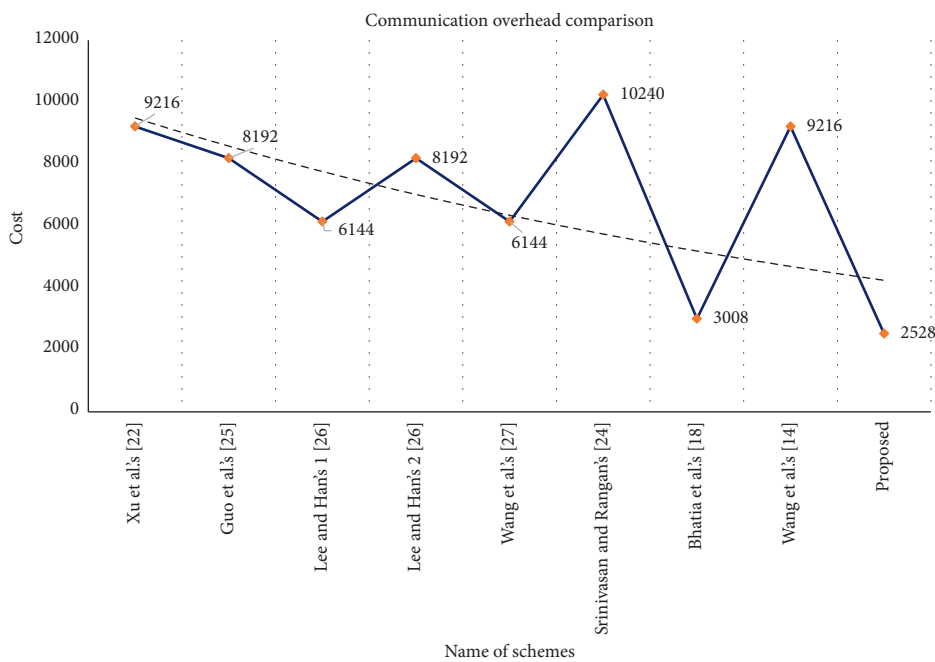


FIGURE 5: Communication overhead comparison.

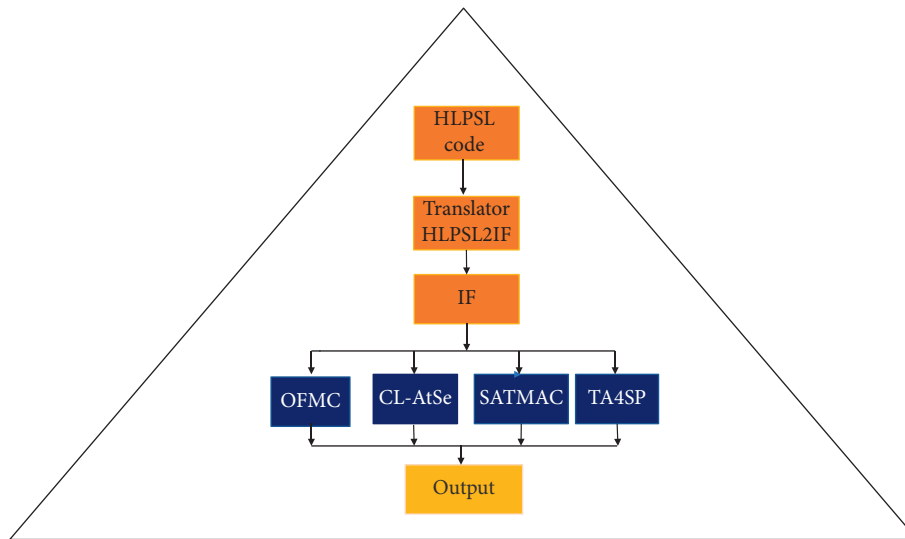


FIGURE 6: Basic architecture of the AVISPA tool.

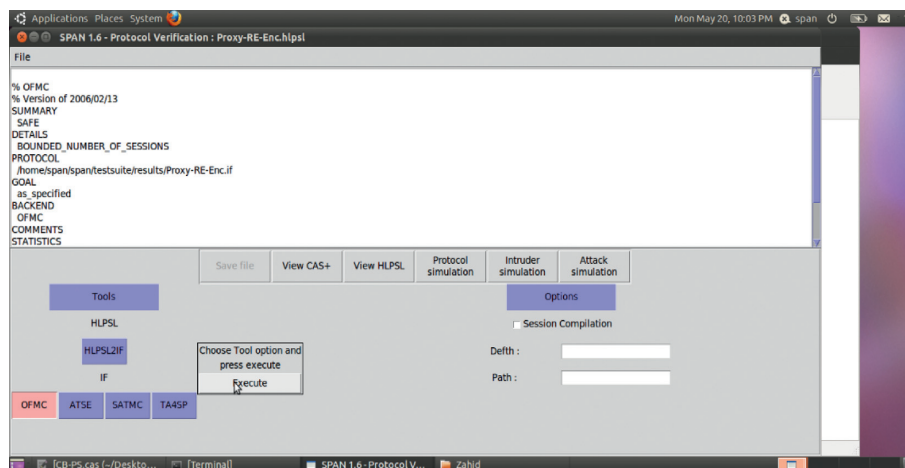


FIGURE 7: OFMC results.

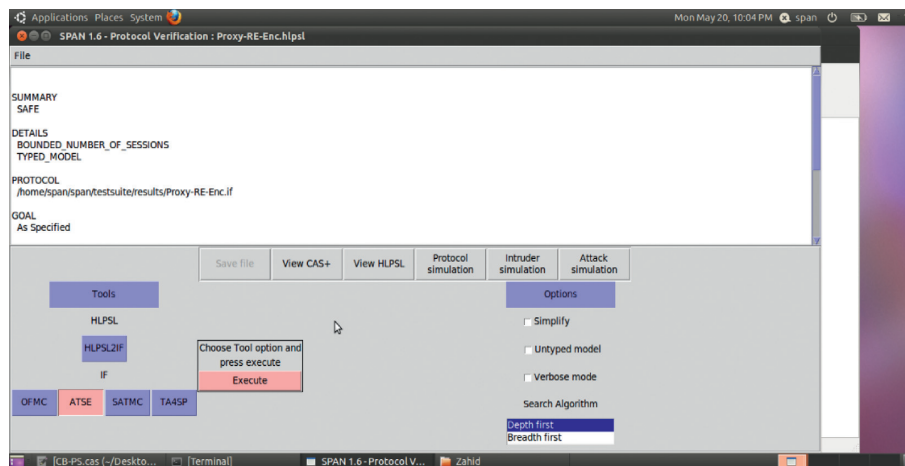


FIGURE 8: ATSE results.

TABLE 5: HLPSSL code for the producer.

---

Role role_Producer(Edgeservicerouter:agent, Producer:agent, Consumer:agent, Pbpr:public_key,Pbcr:public_key,SND,RCV:channel(dy)) played_by Producer def = local State:nat,Lfk:symmetric_key,Encrypt:hash_func,Npr:text,Cnt:text init transition 1. State = 0 $\wedge$ RCV(start) = $ >$ State' := 1 $\wedge$ SND(Producer.Consumer) 2. State = 1 $\wedge$ RCV(Consumer.{Npr'}_Pbpr) = $ >$ State' := 2 $\wedge$ Lfk' := new() $\wedge$ Cnt' := new() $\wedge$ secret(Cnt',sec_2,{Producer}) $\wedge$ witness(Producer, Edgeservicerouter, auth_1,Cnt') $\wedge$ SND(Producer.{Encrypt(Npr'.Cnt')}_Lfk') end role	State:= 0
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------

---

TABLE 6: HLPSSL code for the edge service router.

---

Role role_Edgeservicerouter(Secondlastnode:agent, Producer:agent, Consumer:agent,Pbp r:public_key,Pbcr:public_key,SND,RCV:channel(dy)) played_by Secondlastnode def = local State:nat,Lfk:symmetric_key,Cnt:text, Omega:symmetric_key,Encrypt:hash_func,C1:text,Npr:text init State: = 0 transition 3. State = 0 $\wedge$ RCV(Producer.{Encrypt(Npr'.Cnt')}_Lfk') = $ >$ State' := 1 $\wedge$ request(Edgeservicerouter,Producer,auth_1,Cnt') $\wedge$ secret(Cnt',sec_2,{Producer}) $\wedge$ Omega' := new() $\wedge$ C1' := new() $\wedge$ secret(Cpr',sec_4,{Consumer}) / witness(Edgeservicerouter, Consumer,auth_3,C1') $\wedge$ SND(Secondlastnode.{Encrypt(C1'.Npr')}_Omega') end role	
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

---

TABLE 7: HLPSSL code for the consumer.

---

Role role_Consumer(Edgeservicerouter:agent, Producer:agent, Consumer:agent, Pbpr:pub lic_key,Pbcr:public_key,SND,RCV:channel(dy)) played_by Consumer def = local State:nat,Omega:symmetric_key,Encrypt:hash_func,Cpr:text,Npr:text init State:= 0 transition 1. State = 0 $\wedge$ RCV(Producer.Consumer) = $ >$ State' := 1 / Npr' := new() $\wedge$ SND(Consumer.{Npr'}_Pbpr) 7. State = 1 $\wedge$ RCV(Edgeservicerouter.{Encrypt(Cpr'.Npr')}_Omega') = $ >$ State' := 2 $\wedge$ secret(Cpr',sec_4,{Consumer}) end role	
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

---

TABLE 8: HLPSSL code for the session.

---

Role session1(Edgerouternode:agent, Producer:agent,Consumer:agent, Pbpr:public_key,Pbcr:public_key) def = local SND3,RCV3,SND2,RCV2,SND1,RCV1:channel(dy) composition role_Edgerouternode(Edgerouternode, Producer,Consumer, Pbpr,Pbcr,SND3, RCV3) $\wedge$ role_Consumer(Edgerouternode, Producer,Consumer, Pbpr,Pbcr,SND2,RCV2) $\wedge$ role_Producer(Edgerouternode, Producer,Consumer, Pbpr,Pbcr,SND1,RCV1) end role session2 Edgerouternode:agent, Producer:agent, Consumer:agent, Pbpr:public_key,Pbcr:public_key) def = local SND1,RCV1:channel(dy) composition role_Producer(Edgerouternode, Producer,Consumer, Pbpr,Pbcr,SND1,RCV1) end role	
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

---

Lee and Han's 2 [24], Wang et al.'s [25], Srinivasan and Rangan's [22], Bhatia et al.'s [17], and Wang et al.'s [13] existing schemes.

**6.2. Communication Overhead.** The term communication overhead in the computer network refers to how much time the communication channel spends to send a single message. It is directly proportional to how long is your message. It means that how much extra bits will be sent along with the actual message. Further, it depends on the scheme that is implemented for desired network communication. Now here, we compare our proposed scheme with the existing schemes,

i.e., Xu et al.'s [20], Guo et al.'s [23], Lee and Han's [24], Wang et al.'s [25], Srinivasan and Rangan's [22], Bhatia et al.'s [17], and Wang et al.'s [13], with respect to communication overheads and illustrate that how much communication overhead is reduced by the proposed scheme. We accept that  $|G2| \cong |G1| \cong |G| \cong 1024$  bits,  $|P| \cong 1024$  bits,  $|q| \cong 160$  bits,  $|n| \cong 80$  bits, and  $|\mathcal{M}| = 1024$  bits, respectively. The required communication overhead by Xu et al.'s scheme [20] is  $2|\mathcal{M}| + 7|G| = 9216$ , by Guo et al.'s scheme [23] is  $2|\mathcal{M}| + 6|G| = 8192$ , by Lee and Han's scheme 1 [24] is  $2|\mathcal{M}| + 4|P| = 6144$ , by Lee and Han's scheme 2 [24] is  $2|\mathcal{M}| + 6|P| = 8192$ , by Wang et al.'s scheme [25] is  $2|\mathcal{M}| + 4|P| = 6144$ , by Srinivasan and Rangan's scheme [22] is  $2|\mathcal{M}| + 8|P| = 10240$ , by Bhatia

TABLE 9: HLPSSL code for the environment.

---

```

Role environment() def =
  const
  hash_0:hash_func,pbr:public_key,alice:agent, producer:agent,bob:age
nt,pbcr:public_key,const_1:agent, const_2:agent, const_3:public_key,const_
4:public_key,auth_1:protocol_id,sec_2:protocol_id,auth_3:protocol_id,sec_
_4:protocol_id
  intruder_knowledge = {alice,bob,producer}    composition
  session2(const_1,i,const_2,const_3,const_4)
session1(producer, alice,bob,pbr,pbcr) end role goal
  authentication_on auth_1 secrecy_of sec_2 authentication_on auth_3
secrecy_of sec_4 end goal environment()

```

---

et al.'s scheme [17] is  $2|\mathcal{M}| + 6|q| = 3008$ , by Wang et al.'s scheme [13] is  $2|\mathcal{M}| + 7|G| = 9216$ , and for the proposed scheme is  $2|\mathcal{M}| + 6|n| = 2528$ , respectively. Moreover, we achieve that the proposed scheme is  $9216-2528/9216 * 100 = 72.569\%$  faster than that in [20],  $8192-2528/8192 * 100 = 69.140\%$  faster than that in [23],  $6144-2528/6144 * 100 = 58.854\%$  faster than that in [24] (for 1),  $8192-2528/8192 * 100 = 69.140\%$  faster than that in [24] (for 2),  $6144-2528/6144 * 100 = 58.854\%$  faster than that in [25],  $10240-2528/10240 * 100 = 75.312\%$  faster than that in [22],  $3008-2528/3008 * 100 = 15.957\%$  faster than that in [17], and  $9216-2528/9216 * 100 = 72.569\%$  faster than that in [13], respectively. As a result, from the abovementioned findings, our proposed scheme is faster than the recent research scheme [13] up to 72.569%; Figure 5 illustrates the differentiation.

## 7. Conclusion

The access control management faces high security issues in CCN at the time, when the content provider distributes the contents within the network. For this purpose, we address a secure content architecture for access control in CCN known as CPRES. The proposed CPRES believes on four basic parties on the network, i.e., producer, KGC, edge service router, and consumer. When the consumer (one of the basic element) retrieves encrypted content from the edge service router, he/she just contacts with KGC instead of the producer to authenticate themselves and fetch keys for content decryption. Our scheme accurately fulfils the security requirements, i.e., confidentiality L-1 and L-2 encryption, and replay attacks. Also, the CL-PRE plays a unique role to generate partial keys for improving the security of content accessing, showing that the proposed scheme reduced the computational and communication cost as compared to the existing schemes up to 58.4% to 93.80% and 15% to 72.569%, respectively. So, the proposed CPRES is more attractive to use in the resource-constrained mobile devices.

## Appendix

### Implementation and Validation Using AVISPA Tool

AVISPA is a security claim verification tool, which ensures the scheme protection, concerning two well-known attacks, called man-in-the-middle and replay. The simulation code is generally executed in  $\mathcal{HLPSSL}$ , identified as the high-

level protocol specification language. Typically, the basic architecture of the AVISPA tool is given in Figure 6. Each and every participant is usually free and contains some information in kind of guidelines for communication among other additional participants using channels. According to the architecture, the AVISPA tool first composes the code in  $\mathcal{HLPSSL}$  and translates it directly into an intermediate format ( $\mathcal{JF}$ ) simply by the help of the  $\mathcal{HLPSSL} \rightarrow \mathcal{JF}$  translator.  $\mathcal{JF}$  is a further lower-level language as compared to  $\mathcal{HLPSSL}$  and directly read by AVISPA's backends. AVISPA is executed in four backends: (1) OFMC (on-the fly model checker), (2) CL-AtSe (constraint logic-based attack searcher), (3) SATMC (SAT-based model checker), and (4) TA4SP (tree automata based on automatic approximations for the analysis of security protocols). On the basis of these backends, the output format is created in addition to describing the result and then confirms whether or not the scheme is secure from attacks [48].

Further, this section summarizes our proposed certificateless proxy reencryption scheme based on the hyper-elliptic curve for access control in CCN roles in a recognized security simulation tool known as AVISPA. The proposed scheme algorithm is written in the  $\mathcal{HLPSSL}$  language for checking the validation of security attacks through two backends of the AVISPA tool, i.e., OFMC and ATSE. The simulation results are fully safe against these two backends from the intruder's attack that are shown in Figures 7 and 8. The  $\mathcal{HLPSSL}$  code has five roles in our proposed algorithm. To understand these roles in the  $\mathcal{HLPSSL}$  code it is undermined that the symbols used in the proposed algorithm are shown after the arrow symbol ( $\leftrightarrow$ ) and the  $\mathcal{HLPSSL}$  code symbols are shown before the arrow symbol. So, in Table 5, in the producer role,  $Lfk \leftrightarrow Lfk, \text{Encrypt} \leftrightarrow \oplus, Npr \leftrightarrow Npr, Cnt \leftrightarrow CNT, \{\text{Encrypt}(Npr'.Cnt')\}_Lfk' \leftrightarrow (CNT, Npr) \oplus Lfk, Pbr \leftrightarrow \mathcal{PBpr}$ , and  $Pbcr \leftrightarrow \mathcal{PBcr}$ ; in Table 6, in the edge service router role,  $\Omega \leftrightarrow \Omega, \mathcal{C}1 \leftrightarrow \mathcal{C}1 = \mathcal{R}.L$  and  $\{\text{Encrypt}(C1'.Npr')\}_\Omega \leftrightarrow \mathcal{C}1^* = \mathcal{C}1 \oplus \Omega$ . Similarly, Tables 7-9 provide the  $\mathcal{HLPSSL}$  code for the consumer role, session role, and environment role, respectively. The symbols of Tables 7-9 are already explained above. Further, the consumer role handles the decryption operations. The session role determines how many sessions are made among the nodes. The environment's role is generally related to security of the desired algorithm. Finally, in Figures 7 and 8, the simulation results for the proposed scheme illustrate that our scheme gives fully safe results

against the two backends, OFMC and ATSE, of the AVISPA tool.

## Data Availability

The data used to support the findings of this study are uploaded to the GitHub repository (xx).

## Conflicts of Interest

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

## Acknowledgments

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant no. (DF-459-156-1441). The authors, therefore, gratefully acknowledge the DSR technical and financial support.

## References

- [1] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile information-centric networking: research issues and challenges," vol. 20, no. 3, pp. 2353–2371, 2018.
- [2] K. Xue, X. Zhang, Q. Xia, D. S. L. Wei, H. Yue, and F. Wu, "SEAF: a secure, efficient and accountable Access control framework for information centric networking," in *Proceedings of the IEEE Computer and Communications*, pp. 2213–2221, Honolulu, HI, USA, April 2018.
- [3] J. Kuriharay, E. Uzun, and C. A. Wood, "An encryption-based access control framework for content-centric networking," in *Proceedings of the 2015 IFIP Networking Conference (IFIP Networking)*, Toulouse, France, May 2015.
- [4] E. G. Abdallah, H. S. Hassanein, and M. Zulkernine, "A survey of security attacks in information-centric networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1441–1454, 2015.
- [5] S. Siddiqui, A. Waqas, A. Khan, F. Zareen, and M. N. Iqbal, "Congestion controlling mechanisms in content centric networking and named data networking-a survey," in *Proceedings of the 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, Pakistan, January 2019.
- [6] A. Boukerche and R. W. L. Coutinho, "LoICen: A novel location-based and information-centric architecture for content distribution in vehicular networks," *Ad Hoc Networks*, vol. 93, Article ID 101899, 2019.
- [7] Q. Zheng, G. Wang, R. Ravindran, and A. Azgin, "Achieving secure and scalable data access control in information-centric networking," in *Proceedings of the 2015 IEEE International Conference on Communications (ICC)*, pp. 5367–5373, London, UK, June 2015.
- [8] I. U. Din, B. S. Kim, S. Hassan, M. Guizani, M. Atiquzzaman, and J. Rodrigues, "Information-centric network-based vehicular communications: overview and research opportunities," *Sensors*, vol. 18, no. 1, p. 3857, 2018.
- [9] B. Ahlgren, C. Dannewitz, C. Imbrenda, and D. Kutscher, "A survey of information-centric networking," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 26–36, 2012.
- [10] C. Cavanagh and U. C. Irvine, *UC Irvine Electronic Theses and Dissertations*, vol. 228, 2016.
- [11] C. Zhou, Z. Zhao, W. Zhou, and Y. Mei, "Certificateless key-insulated generalized signcryption scheme without bilinear pairings," *Security and Communication Networks*, vol. 2017, Article ID 8405879, 17 pages, 2017.
- [12] A. Rahman, I. Ullah, M. Naeem et al., "A lightweight multi-message and multi-receiver heterogeneous hybrid signcryption scheme based on hyper elliptic curve," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 160–167, 2018.
- [13] Q. Wang, W. Li, and Z. Qin, "Proxy Re-encryption in access control framework of information-centric networks," *IEEE Access*, vol. 7, pp. 48417–48429, 2019.
- [14] M. Green and G. Ateniese, "Identity-based proxy re-encryption," *Applied Cryptography and Network Security*, Springer, Berlin, Heidelberg, Germany, 2007.
- [15] L. Wang, L. Wang, M. Mambo, and E. Okamoto, "New identity-based proxy Re-encryption schemes to prevent collusion attacks," in *Proceedings of the International Conference on Pairing-Based Cryptography*, Beijing, China, 2010.
- [16] L. L. Wang, K. F. Chen, X. P. Mao, and Y. T. Wang, "Efficient and provably-secure certificateless proxy re-encryption scheme for secure cloud data sharing," *Journal of Shanghai Jiaotong University*, vol. 19, no. 4, pp. 398–405, 2014.
- [17] T. Bhatia, A. K. Verma, and G. Sharma, "Secure sharing of mobile personal healthcare records using certificateless proxy re-encryption in cloud," *Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 6, Article ID e3309, 2018.
- [18] C. A. Wood and E. Uzun, "Flexible end-to-end content security in CCN," in *Proceedings of the 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, pp. 858–865, Las Vegas, NV, USA, January 2014.
- [19] M. Mangili, F. Martignon, and S. Paraboschi, "A cache-aware mechanism to enforce confidentiality, trackability and access policy evolution in content-centric networks," *Computer Networks*, vol. 76, pp. 126–145, 2015.
- [20] L. Xu, X. Wu, and X. Zhang, "CI-PRE," in *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, New York, NY, USA, May 2012.
- [21] M. Ion, J. Zhang, and E. M. Schooler, "Toward content-centric privacy in ICN: attribute-based encryption and routing," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, p. 513, 2013.
- [22] A. Srinivasan and C. P. Rangan, "Certificateless proxy re-encryption without pairing," in *Proceedings of the 3rd International Workshop on Security in Cloud Computing*, pp. 41–52, Dubai, 2015.
- [23] H. Guo, Z. Zhang, J. Zhang, and C. Chen, "Towards a secure certificateless proxy re-encryption scheme," in *Proceedings of the International Conference on Provable Security*, pp. 330–346, Melaka, Malaysia, October 2013.
- [24] H. S. Lee and D. G. Han, "Information security and cryptology-ICISC 2013," in *Proceedings of the International Conference on Information Security and Cryptology*, pp. 67–88, Seoul, Korea, December 2014.
- [25] Y. Wang, H. Xiong, S. Argamon, X. Y. Li, and J. Z. Li, "Big data computing and communications," in *Proceedings of the First International Conference, BigCom 2015*, pp. 205–206, Taiyuan, China, August 2015.
- [26] R. Tourani, S. Misra, T. Mick, and G. Panwar, "Security, privacy, and access control in information-centric networking: a survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 556–600, 2018.

- [27] K. Fu, S. Kamara, and T. Kohno, "Key regression: enabling efficient key distribution for secure distributed storage," vol. 149, 2006 Comput. Sci. Dep. Fac. Publ. Ser.
- [28] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," in *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 127–144, Konstanz, Germany, May 1998.
- [29] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," *ACM Transactions on Information and System Security*, vol. 9, no. 1, pp. 1–30, 2006.
- [30] R. Canetti and S. Hohenberger, "Chosen-ciphertext secure proxy re-encryption," in *Proceedings of the 14th ACM conference on Computer and Communications Security*, Alexandria, VA, USA, 2007.
- [31] B. Libert and D. Vergnaud, "Unidirectional chosen-ciphertext secure proxy re-encryption," in *Proceedings of the IACR International Conference on Public-Key Cryptography*, pp. 360–379, Barcelona, Spain, March 2008.
- [32] R. H. Deng, J. Weng, S. Liu, and K. Chen, "Chosen-ciphertext secure proxy re-encryption without pairings," in *Proceedings of the International Conference on Cryptology and Network Security*, pp. 1–17, Hong Kong, China, December 2008.
- [33] X. A. Wang, J. Ma, and X. Yang, "A new proxy re-encryption scheme for protecting critical information systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, no. 6, pp. 699–711, 2015.
- [34] H. Shacham, "A cramer-shoup encryption scheme from the linear assumption and from progressively weaker linear variants," 2007.
- [35] Y. Ren, D. Gu, S. Wang, and X. Zhang, "Hierarchical identity-based proxy re-encryption without random oracles," *International Journal of Foundations of Computer Science*, vol. 21, no. 6, pp. 1049–1063, 2010.
- [36] L. Batten, G. Li, W. Niu, and M. Warren, "Applications and techniques in information security," in *Proceedings of the International Conference on Applications and Techniques in Information Security*, Melbourne, VIC, Australia, November 2014.
- [37] Q. Tang, P. Hartel, and W. Jonker, "Inter-domain identity-based proxy re-encryption," in *Proceedings of the International Conference on Information Security and Cryptology*, pp. 332–347, Beijing, China, December 2009.
- [38] T. Mizuno and H. Doi, "Secure and efficient IBE-PKE proxy re-encryption," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E94-A, no. 1, pp. 36–44, 2011.
- [39] J. Shao and Z. Cao, "Multi-use unidirectional identity-based proxy re-encryption from hierarchical identity-based encryption," *Information Sciences*, vol. 206, pp. 83–95, 2012.
- [40] K. Liang, Z. Liu, X. Tan, D. S. Wong, and C. Tang, "A CCA-secure identity-based conditional proxy re-encryption without random oracles," in *Proceedings of the International Conference on Information Security and Cryptology*, pp. 231–246, Seoul, Korea, November 2013.
- [41] K. Liang, J. K. Liu, D. S. Wong, and W. Susilo, "An efficient cloud-based revocable identity-based proxy re-encryption scheme for public clouds data sharing," in *Proceedings of the European Symposium on Research in Computer Security*, pp. 257–272, Luxembourg, Luxembourg, September 2014.
- [42] S. Saxby, "Communications and multimedia security," *Computer Law & Security Review*, vol. 22, no. 4, p. 338, 2006.
- [43] Y. Zheng, S. Tang, C. Guan, and M. R. Chen, "Cryptanalysis of a certificateless proxy re-encryption scheme," in *Proceedings of the 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, pp. 307–312, Xi'an, China, September 2013.
- [44] S. A. Ullah, "Review of signcryption schemes based on hyper elliptic curve," in *Proceedings of the 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM)*, Chengdu, China, August 2017.
- [45] Nizamuddin, C. Shehzad Ashraf, and N. Amin, "Signcryption schemes with forward secrecy based on hyperelliptic curve cryptosystem," in *Proceedings of the 8th International Conference on High-capacity Optical Networks and Emerging Technologies*, pp. 244–247, Riyadh, Saudi Arabia, December 2011.
- [46] A. J. Ordonez, R. P. Medina, and B. D. Gerardo, "Modified El gamal algorithm for multiple senders and single receiver encryption," in *Proceedings of the 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia, April 2018.
- [47] I. Ullah, N. Amin, J. Khan et al., "A novel provable secured signcryption scheme PSSS: a hyper-elliptic curve-based approach," *Mathematics*, vol. 7, no. 8, p. 686, 2019.
- [48] R. Ali and A. K. Pal, "Three-factor-based confidentiality-preserving remote user authentication scheme in multi-server environment," *Arabian Journal for Science and Engineering*, vol. 42, no. 8, pp. 3655–3672, 2017.

## Research Article

# Robust Spectrum Sensing via Double-Sided Neighbor Distance Based on Genetic Algorithm in Cognitive Radio Networks

Noor Gul,<sup>1</sup> Muhammad Sajjad Khan,<sup>1,2</sup> Junsu Kim,<sup>2</sup> and Su Min Kim <sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, Faculty of Engineering and Technology, International Islamic University, Islamabad 44000, Pakistan

<sup>2</sup>Department of Electronics Engineering, Korea Polytechnic University, 237 Sangidaehak-ro, Siheung-si, Gyeonggi-do 15073, Republic of Korea

Correspondence should be addressed to Su Min Kim; [suminkim@kpu.ac.kr](mailto:suminkim@kpu.ac.kr)

Received 17 March 2020; Revised 30 June 2020; Accepted 8 July 2020; Published 23 July 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Noor Gul 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 cognitive radio networks (CRNs), secondary users (SUs) can access vacant spectrum licensed to a primary user (PU). Therefore, accurate and timely spectrum sensing is vital for efficient utilization of available spectrum. The sensing result at each SU is unauthentic due to fading, shadowing, and receiver uncertainty problems. Cooperative spectrum sensing (CSS) provides a solution to these problems. In CSS, false sensing reports at the fusion center (FC) received from malicious users (MUs) drastically degrade the performance of cooperation in PU detection. In this paper, we propose a robust spectrum sensing scheme to minimize the effects of false sensing reports by MUs. The proposed scheme focuses on double-sided neighbor distance (DSND) based on genetic algorithm (GA) in order to filter out the MU sensing reports in CSS. The simulation results show that the sensing results are more accurate and reliable for the proposed GA majority-voting hard decision fusion (GAMV-HDF) and GA weighted soft decision fusion (GAW-SDF) compared to conventional equal gain combination soft decision fusion (EGC-SDF), maximum gain combination soft decision fusion (MGC-SDF), and majority-voting hard decision fusion (MV-HDF) schemes in the presence of MUs.

## 1. Introduction

Rapid developments in wireless communication system demand new wireless services in both used and unused parts of electromagnetic spectrum [1]. The underutilization of the spectrum fallout in spectrum holes representing the frequency band assigned to a legitimate primary user (PU), but it is not utilized by the PU at certain time and specific geographical locations. The motivation to introduce cognitive radio technology is increasing demands for higher data rates under underutilized spectral scarcity issues [2–4]. To solve the spectrum scarcity issues, federal communications commission (FCC) permits secondary users (SUs) to dynamically utilize the spectrum in different services or even to lease the spectrum to a third party [5, 6]. The cognitive radio network (CRN) consists of an intelligent wireless communication system embedded with key functionalities

to provide seamless communications at all times and all geographical places based on the needs with proficient utilization of the spectrum resources [7].

One of major issues in CRN is to properly detect the status of PU channel. Proper detection of the status of PU channel is critical at the SU for minimizing interference to the PU. In CRNs, SUs collect information on the PU existence based on various detection techniques such as feature detector, matched filter detector, and energy detector [3, 4]. The energy detector can be the best choice to differentiate the PU signal from the noise, thanks to its simplicity and minimal computation, but it poorly performs in weak signal to noise ratio (SNR) environments.

Cooperative spectrum sensing (CSS) performs well in fading and shadowing environments, where multiple radios provide an independent realization of related random variable in the course of distributed transmission [8–10]. The

probability that all SUs are in deep fades is incredibly low, which enables CSS to employ fewer sensitive detectors with cheap hardware, hence reducing the overall cost and complexity of the system. The artificial bee colony clustering (ABCC) algorithm in [11] is competent to reduce and stabilize the energy expenditure of the cooperative users. In CSS, SUs make their own local decisions about the PU existence and forward it to the fusion center (FC) for further analysis [12, 13].

CSS is exposed to the false sensing reports of malicious users (MUs), therefore identification and exclusion of MU reports in the cooperative scheme is essential for minimizing their adverse effects. An abnormality detection approach of data mining is discussed in [14–16]. In [17], MUs with the primary user emulation attack (PUEA) to imitate the PU behavior is discussed. A robust CSS scheme under the attack of MUs sending an all-time busy status of the PU to the FC is discussed in [18]. In [19], the Kullback–Leibler (KL) divergence method is used against MUs with always busy and always free signaling of the PU channel without SNR requirements. SUs in the soft combination schemes report their energy statistics to the FC without local processing at individual SU [20–22]. In [23], the agents are allowed to cooperate in completing individual tasks to solve multiagent tasks with improved efficiency and reduced communication cost. A hard decision scheme in [24] maintains low communication overhead compared to other soft combination schemes. The population-based search algorithm with inherited ability of gripping several optimization jobs at once is proposed in [25].

The genetic algorithm (GA) is used to determine optimized solutions using biologically stimulated techniques, such as natural selection, genetic inheritance, recombination, and crossover [26, 27]. The remarkable generality and versatility of the GA make it useful in a variety of settings in the wireless communication to reduce the error probability of the CSS [28, 29].

In this paper, the CSS sensing performance is optimized in the presence of MUs reporting false information to the FC, by reducing miss detection and false alarm probabilities, resulting in overall reduction in error probability. In our previous study [30], SUs perform their local sensing and report soft energies to the FC and also store the information in their local database. After then, the FC determines the KL divergence score against each SU and also acknowledges this same information to the SU. A normally declared user based on the KL divergence score tries to send mean of the previous energy reports to the FC based on its current observation. Similarly, in our proposed GA-based scheme [31], no additional steps are taken for MU identification and FC makes a global decision based on the best selection results given by the GA to optimize detection and false alarm probabilities. Our previous work in [26] is based on the combination of double-sided neighbor distance (DSND) algorithm with GA first identify MUs using DSND and then the GA is used in selection of best spectrum sensing results at the end of the given number of iterations. The best selection results of the GA are followed by the majority-voting hard decision fusion (MV-HDF) to make a global decision. This paper is an extension of the previous work, where MUs and normal SUs report their local sensing results to the FC.

When enough statistics are collected against SUs, abnormalities are first identified by the GA with the DSND algorithm, as in [14, 26]; GA then randomly mutates the sensing data of the detected abnormalities along with crossover operation to search more suitable sensing information against the reporting SUs. The GA selection is further used to find best sensing data based on the Hamming distances of all SUs from its neighbors during each history interval, and the minimum Hamming distance report is decided as the best sensing results on behalf of all SUs for majority-voting hard decision fusion (MV-HDF). The best selection results are further used for assigning weights to SU reports in the soft decision fusion (SDF)-based global decision at the FC. Unlike our previous work where the MV-HDF performance was compared with SDF and HDF schemes, in the extended work, the proposed scheme effectiveness has been further confirmed under different number of cooperative SUs and various SNRs. Simulation results at different levels of cooperative SUs and SNR confirmed that, in the presence of MUs, the proposed DSND-based GA system is able to produce more precise detection outcomes for the SDF and HDF schemes. The proposed GA weighted SDF (GAW-SDF) and GA majority-voting HDF (GAMV-HDF) are able to beat simple equal gain combination soft decision fusion (EGC-SDF), maximum gain combination soft decision fusion (MGC-SDF), and simple majority-voting hard decision fusion (MV-HDF) schemes during PU channel recognition by keeping the probability of error results optimum with high detection and low false alarm results at different levels of SNR and cooperative users.

The main contributions of this paper are summarized as follows:

- (i) A novel weighted soft decision scheme is proposed to combine the sensing results reported from both normal SUs and MUs
- (ii) The proposed weighted soft decision scheme utilizes both soft and hard combinations to achieve better performance, whereas the previous work is only suitable for hard combination scheme
- (iii) Through extensive simulations, the effectiveness of the proposed scheme is evaluated in terms of detection, false alarm, and error probabilities in different ranges of SNRs and number of users compared with the existing schemes

The rest of the paper is organized as follows. The system model is presented in Section 2. In Section 3, the proposed DSND scheme based on GA to overcome the effects of MU is illustrated. Numerical results are shown in Section 4. Finally, the paper is concluded in Section 5.

## 2. System Model

To improve the sensing performance, we consider a CRN scenario in which all SUs are searching for a common PU in their coverage area and report the channel status to the FC as shown in Figure 1. Based on the spectrum sensing results received from normal SUs and MUs, the FC makes a more



precise and authentic global decision on the PU channel availability.

The received signal energy is used to decide  $H_0$  and  $H_1$  hypothesis in a particular spectrum as

$$x_j = \begin{cases} H_0, & w_j(k) \\ H_1, & h_j s(k) + w_j(k) \end{cases}, \quad (1)$$

where  $H_0$  is the hypothesis that the PU spectrum is free and  $H_1$  represents that the PU channel is occupied,  $x_j$  is the  $j^{\text{th}}$  SU observed signal in the  $k^{\text{th}}$  sensing slot,  $w_j(k)$  is the additive white Gaussian noise (AWGN) experienced by  $j^{\text{th}}$  SU,  $h_j$  is the channel gain between the PU and the  $j^{\text{th}}$  SU, and  $s(k)$  is the PU transmitted signal in the  $k^{\text{th}}$  sensing slot.

It assumed that an energy detector is used by all SUs due to its simplicity and no requirements of any prior information of PU power. The energy received at the  $i^{\text{th}}$  sensing interval is

$$E_j(i) = \begin{cases} \sum_{k=k_i}^{k_i+S-1} |w_j(k)|^2, & H_0 \\ \sum_{l=l_i}^{l_i+S-1} |h_j s(k) + w_j(k)|^2, & H_1 \end{cases}, \quad (2)$$

where  $S$  is the number of samples in the  $i^{\text{th}}$  interval. According to the central limit theorem (CLT), sufficient number of samples provides the energy distribution to be Gaussian distribution under both the  $H_0$  and  $H_1$  hypotheses, given by [26]

$$E_j \sim \begin{cases} N(\mu_0 = S, \sigma_0^2 = 2K), & H_0 \\ N(\mu_1 = S(v_j + 1), \sigma_1^2 = 2S(v_j + 1)), & H_1 \end{cases}, \quad (3)$$

where  $v_j$  is the SNR. Similarly,  $(\mu_0, \sigma_0^2)$  and  $(\mu_1, \sigma_1^2)$  denote the means and variance values of the received energy when either  $H_0$  or  $H_1$  hypothesis is true.

### 3. Proposed Methodology

In this section, we discuss the proposed methodology in detail. The FC applies the DSND technique as part of the GA for identifying abnormalities, and then with aid of crossover and mutation, sensing observations with high fitness are selected for the reporting SUs. The selected fitness is also used to assign weights to the received soft energy statistics of individual SUs. In the soft combination, reliability of the user report is guaranteed by determining weights against each SU information. All MUs receive lower weights than normal SUs' energy information in the SDF scheme. In the proposed DSND algorithm, history log is developed against the reporting SUs at the FC to filter out any abnormal SU from the global decision by computing the distance of each SU with its neighbors. The fitness function is based on the absolute sum of the Hamming distances of the individuals with the sensing reports provided by all other SUs. At the end of selected iterations, sensing observation with the minimum differences amongst neighbors is considered as the true sensing facts. In the next measurement, MV-HDF

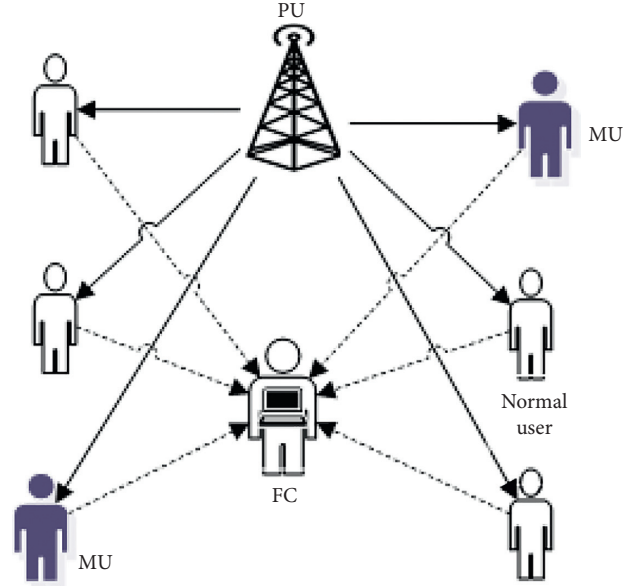


FIGURE 1: Conventional cooperative network.

and weighted SDF schemes are employed to announce the global decision on the existence of PU. The impact of including MUs in the resultant CSS has a minor effect on the final decision at the FC.

**3.1. Local Spectrum Decisions.** The proposed sensing model is shown in Figure 2. In this model, cooperative SUs sense the PU channel and compare the received signal energy with a threshold to send a binary report to FC as

$$y_j(i) = \begin{cases} 1, & E_j(i) \geq \lambda_j \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where  $E_j(i)$  is the expected energy of the  $j^{\text{th}}$  SU in the  $i^{\text{th}}$  sensing interval and  $\lambda_j$  is the set threshold point against the  $j^{\text{th}}$  SU. As cooperative SUs in given CSS environment sense the PU channel at different locations, they experience different fading and shadowing effects. Therefore, the threshold setup at each user is not the same. If received energy of the  $j^{\text{th}}$  SU is higher than threshold, a binary decision "1" is forwarded to the FC indicating an occupied channel. Similarly, if the energy is less than the threshold, a binary decision "0" is reported to depict the channel as free.

The FC collects the local spectrum decisions  $Z_j(i)$  from all  $P$  SUs for the  $N$  history intervals and forms a history reporting matrix against all SUs as

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1P} \\ y_{21} & y_{22} & \cdots & y_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NP} \end{bmatrix}, \quad (5)$$

where  $\mathbf{Y}$  is the  $N \times P$  population matrix which consists of the accumulated spectrum sensing notifications at the FC for the  $P$  SUs in  $N$  total reports. The information is collected for

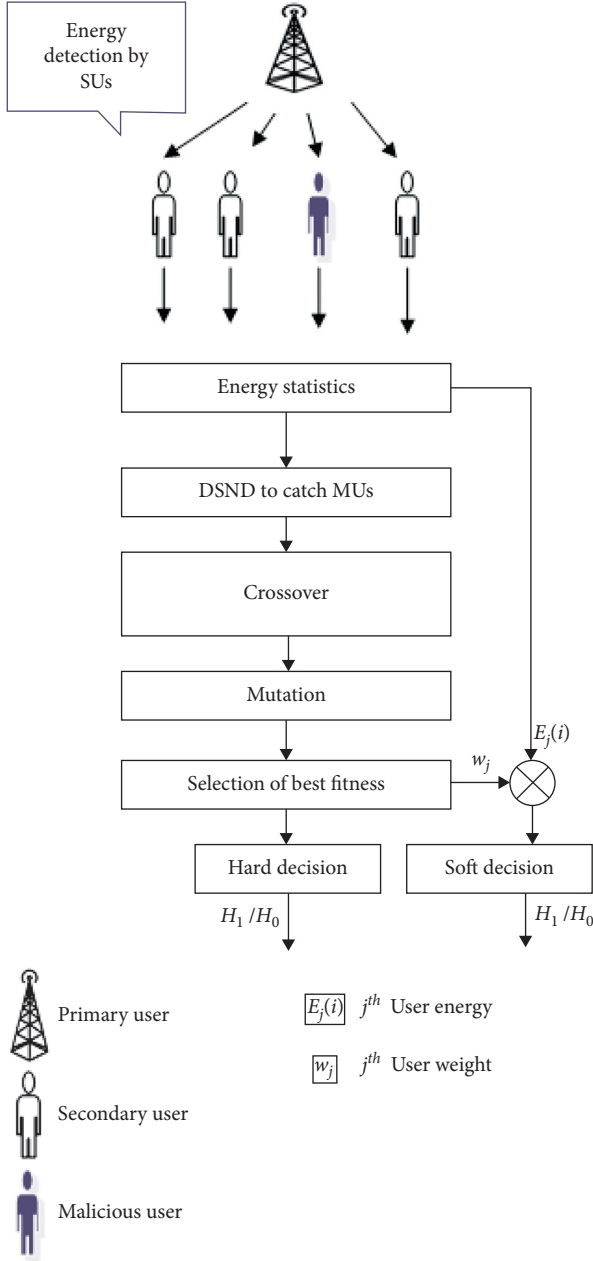


FIGURE 2: Proposed sensing model.

both the SUs and MUs. The CSS can make the system secure against falsification effects of various different MUs' policies such as always yes malicious user (AYMU), always no malicious user (ANMU), opposite malicious user (OMU), and random opposite malicious user (ROMU) by employing the following methodology. As the AYMU policy always reports nonavailability of the PU channel, therefore, the presence of AYMU in CSS leads to an increase in misdetection probability of the system that results in low detection probability at the FC. Similarly, the ANMU policy reports an always free state of the PU channel and results in increasing false alarm probability of the system. The reports of AOMU and ROMU policies negate actual condition of the PU activity by reporting high energy states when the channel

is free and low energy states when the PU is occupying the channel, hence leading to an increase in both false alarm and misdetection probabilities.

**3.2. Double-Sided Neighbor Distance (DSND) for Catching Malicious Users.** The DSND algorithm is employed to determine outliers by their sensing reports, which is away from the other SUs in the history table. Based on the received sensing notifications of all SUs in the  $N$  intervals, FC is able to recognize any outlier MU with the DSND algorithm.

The FC first receives local spectrum observations from individual SUs. When FC collects  $N$  sensing reports from all  $P$  SUs as in (5),  $J_1$  and  $J_2$  indices are selected such that  $J_1 < J_2$ . Similarly, the selections of  $J_1$  and  $J_2$  indices must satisfy  $M < J_1 \ll P$  and  $M \ll J_2 < P$ , where  $J_1$  and  $J_2$  are the gauges for MUs detection, when the total number of MUs consideration is  $M$  in the  $P$  cooperative SUs. As the DSND algorithm compares history reports of the SUs, therefore, the inter-SU distance smaller than  $J_1$  or larger than  $J_2$  declares the SU as MU. An SU cannot be considered as malicious with the detection of both  $J_1$  and  $J_2$  gauges. As the DSND algorithm is applied to the sensing history of the SUs, therefore, the more information the system collects about the reporting SUs, the more precisely this algorithm works to identify abnormality.

The distance in the sensing reports of the  $j^{\text{th}}$  SU with all other SUs is determined in (6). This measurement is the dissimilarity in the reported bits of the  $j^{\text{th}}$  SU with all other SUs:

$$b_{ij} = \sum_{k=1}^P |y_{ij} - y_{ik}|, \quad i \in 1 \dots N, j \in 1 \dots P, \quad (6)$$

where  $b_{ij}$  is the total absolute distance measurement of the  $j^{\text{th}}$  SU sensing with all  $P$  users in the  $i^{\text{th}}$  sensing period.

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1P} \\ b_{21} & b_{22} & \dots & b_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \dots & b_{NP} \end{bmatrix}. \quad (7)$$

The matrix  $\mathbf{B}$  is the sensing difference collections against each cooperative SUs in all sensing iterations. Matrix  $\mathbf{B}$  is sorted and the result is used to set limits for the detection of abnormalities as

$$L = \mu \pm C \times \sigma^2. \quad (8)$$

In (8),  $\mu$  and  $\sigma^2$  are the mean and the variance measurements of  $\mathbf{B}$  matrix, respectively,  $C$  is a constant with value  $10/N$  for  $N$  total reports representing history of the sensing information on behalf of all cooperative SUs. The upper and lower limits are defined as

$$L_u = \mu + \frac{10\sigma^2}{N}, \quad (9)$$

$$L_l = \mu - \frac{10\sigma^2}{N}, \quad (10)$$

where  $L_u$  is selected as the upper and  $L_l$  as lower limit. After the selection of  $J_1^{\text{st}}$  and  $J_2^{\text{st}}$  entries based on  $\mathbf{B}$  sorted results, if  $J_1^{\text{st}}$  entry of the SU is greater than  $L_u$ , the SU is declared as MU in  $J_1$  sense and if  $J_2^{\text{st}}$  entry of the user is less than  $L_l$ , the SU is declared as MU in  $J_2$  sense.

$$\text{MU} = \left\{ \begin{array}{ll} j^{\text{th}}, & \text{if } (J_1^{\text{st}} > L_u \text{ or } J_2^{\text{st}} < L_l) \\ 0, & \text{otherwise} \end{array} \right\}. \quad (11)$$

The intuition of the DSND is that if SU history is too farther from other SUs or too close to other SU's histories, its behavior is probably abnormal, hence representing an MU. Due to the double detection thresholds, the DSND is not only able to detect the attackers with their reports largely varying from the MUs, i.e., AYMU, ANMU, and OMU users, but it can also identify the attackers with their reports supported by the honest SUs and performing malicious act occasionally such as ROMUs.

**3.3. Production of New Population.** Referring to the GA population, the  $N$  rows are the representations of the chromosomes which consist the reported sensing data from the  $P$  SUs denoting the genome values.

The fitness function is selected based on the Hamming distances of each SU with its neighbor SUs in (7) as

$$F_i = \sum_{j=1}^P (b_{ij}), \quad i \in 1, \dots, N, \quad j \in 1, \dots, P, \quad (12)$$

$$F = [F_1 \ F_2 \ \dots \ F_N]^T. \quad (13)$$

The fitness function is used to check the suitability of sensing observations at each sensing interval. The fitter chromosomes with high regularity in the sensing data are able to pass through inheritance, while the detrimental chromosomes with inconsistent MUs data are postponed due to survival of the fittest theory.

The fitter chromosomes with high regularity in the SUs reported data and minimum disturbances from any MU which are allowed to pass to the next generation. The fitness score is utilized in ascending order to sort the population.

Based on (13), the top two chromosomes comprising minimum Hamming distance with the neighbors which are selected as the parent chromosomes for the next population, and crossover procedure is carried out in the rest to find out fresh juvenile.

The crossover practice is repeated for the offsprings to take advantage of the best behavior of the individual chromosomes by mixing them in a bid to raise the chances of finding a more suitable candidate. A random locus point is selected and the subsequences, prior to and following the locus in the parent chromosomes, are exchanged to build new children pairs. This operator randomly selects a locus and exchanges the subsequences between two parent chromosomes to build a pair of children. Selection of the crossover point is random in the proposed work.

Mutation alters the selected genome status randomly which shows the modification in sensing data of the

designated user in this work. Mutation is applied to the sensing reports of the detected abnormalities. The reports from the detected MUs in (11) are randomly inverted by changing the genome bits.

After the crossover and random mutations of the detected MUs data, a new population  $\mathbf{Y}$  is formed which leads to the formation of a new neighbor distance matrix  $\mathbf{B}'$  as follows:

$$\mathbf{B}' = \begin{bmatrix} b'_{11} & b'_{12} & \dots & b'_{1P} \\ b'_{21} & b'_{22} & \dots & b'_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ b'_{N1} & b'_{N2} & \dots & b'_{NP} \end{bmatrix}. \quad (14)$$

The new fitness function values are determined as

$$F'_i = \sum_{j=1}^P (b'_{ij}). \quad (15)$$

Fitness scores in (15) are arranged in ascending order and the one with minimum Hamming distance measurement is elected as the best fitness. In matrix  $\mathbf{Y}$ , sensing reports with similar index number to the best fitness is selected as the final recommendation of the DSND-based GA scheme. The recommended sensing observations are used in the following section by the MV-HDF scheme to get to the final assessment about the PU activity.

The results of the Hamming distance are normalized for assigning weights to each SU decision as

$$w_j = \left( \frac{(1/b'_{1j})}{\sum_{j=1}^P (1/b'_{1j})} \right), \quad j \in 1, \dots, P. \quad (16)$$

The SUs with abnormal behavior obtain lower weights in comparison with MUs from the result in (16).

A detailed flow chart diagram of the proposed CSS with stepwise operation from individual spectrum sensing to the final global decision using MV-HDF, and weighted SDF is illustrated in Figure 3.

**3.4. Global Decision.** Based on the weighed results for the authenticity of each SU sensing information as in (16), the global decision  $G_B(i)$  at the FC is formulated as

$$G_B(i) = \left\{ \begin{array}{ll} H_1, & \sum_{j=1}^P w_j(i) \times E_j(i) \geq \varepsilon \\ H_0, & \text{otherwise} \end{array} \right\}, \quad (17)$$

$$i \in 1, \dots, N, \quad j \in 1, \dots, P,$$

where  $w_j$  is the weight assigned to the  $j^{\text{th}}$  SU energy in the data fusion and  $\varepsilon$  is the threshold value for detection of the PU. The SUs with malicious behavior at the FC are charged with lower weights compared with the normal SUs which receive higher weights. All MUs including AYMU, ANMU, OMU, and ROMU are easily identified by the proposed scheme with their behavior. The MUs have higher  $b_{ij}$  results because they have less inconsistency with the reported information of other SUs. The MUs receive lower weights

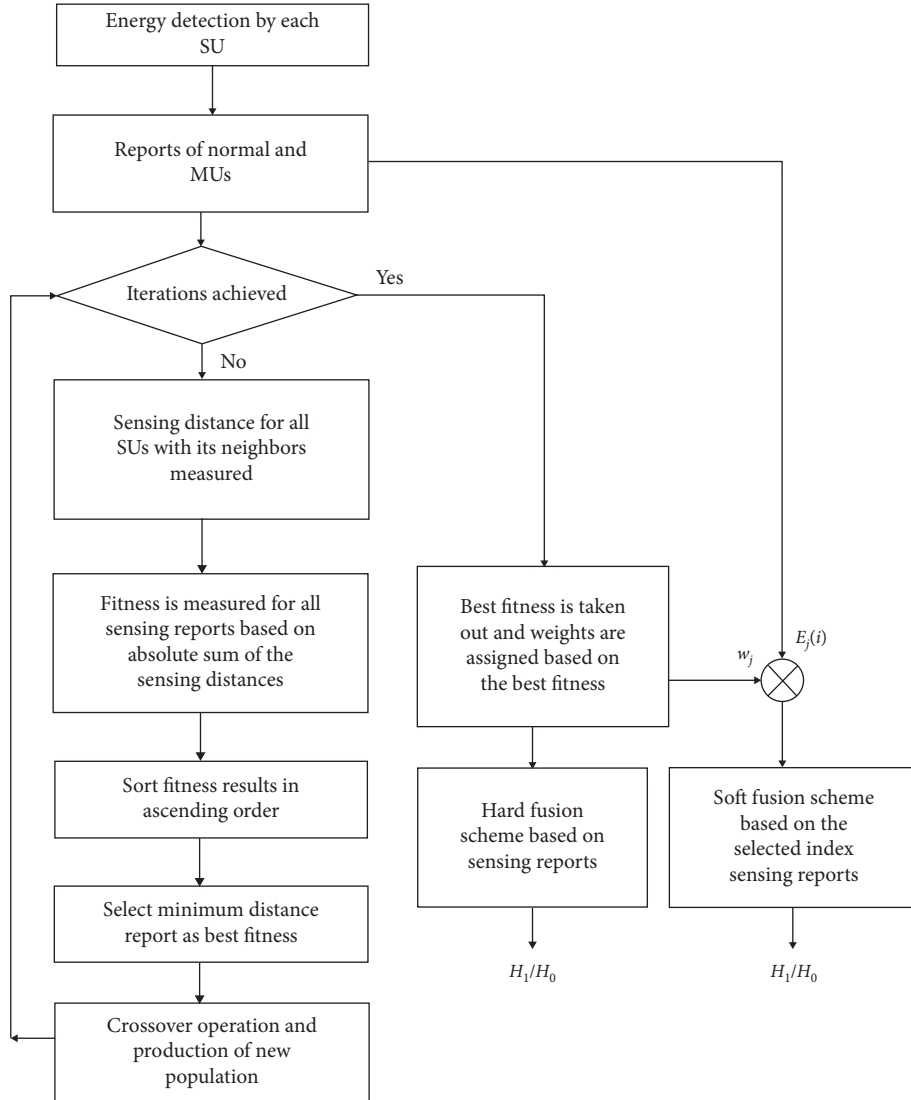


FIGURE 3: Proposed CSS flowchart.

because the information provided by MUs deviates more significantly from that of the other SUs.

The three most commonly used HDF schemes are MV-HDF, OR-HDF, and AND-HDF schemes. After identifying the abnormal users by the DSND algorithm, GA is used to make the final decision at the FC to further improve its accuracy. The sensing selections of the DSND-based GA technique is utilized by the MV-HDF to get more accurate information of the PU channel with minimum impact of the selfish users in the final combination.

The MV-HDF scheme takes unanimous decision of the PU presence if  $Z$  out of  $P$  cooperative users states the PU detection. Similarly, if the detection reports received from the SUs are less than  $Z$  then decision is made in favor of  $H_0$  to state the channel as free of the PU. For the MV-HDF scheme, the voting criteria are selected with  $Z = P/2$  as a special case below:

$$G_B(i) = \begin{cases} H_1 & \sum_{j=1}^P y_j(i) \geq Z \\ H_0 & \text{otherwise} \end{cases}, \quad (18)$$

where  $P$  is the total number of SU reports reaching the fusion center for PU detection,  $y_j(i)$  is the local decision of the  $j^{\text{th}}$  SU in the  $i^{\text{th}}$  period, and  $G_B(i)$  is the global decision made by the MV-HDF scheme in the  $i^{\text{th}}$  period.

#### 4. Numerical Results and Evaluation

In this section, we present the numerical results of the proposed scheme in comparison with the other existing schemes. CRN setting is made with total  $P$  (10 to 20) SUs. All SUs are located randomly to sense the existence of the PU. Out of these  $P$  SUs, four of the SUs were assigned the malicious responsibilities of AYMU, ANMU, OMU, and ROMU. The MUs in this work are tested under low average SNR compared with normal SUs, i.e., MUs have low SNR of the channel compared with normal SUs. The simulation results were observed for the proposed scheme under varying SNR and increased ratio of cooperating SUs. The sensing period of each SU is taken as 1 ms which is divided into  $S = 270$  samples. The number of sensing iterations is

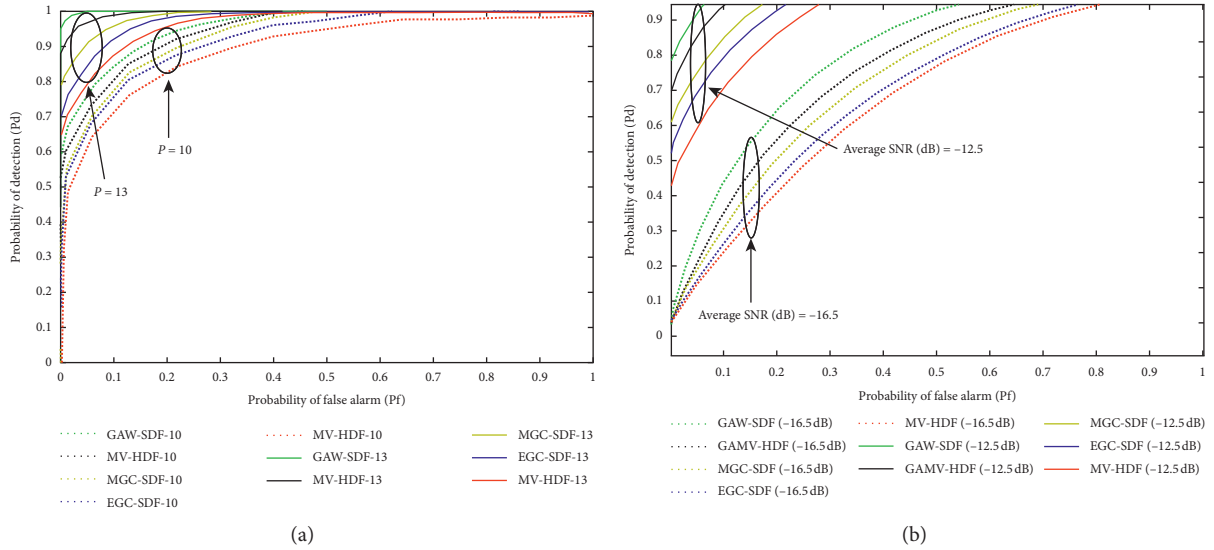


FIGURE 4: (a) ROC curves, when  $P = 10$  and  $P = 13$ . (b) ROC curves, when SNR =  $-12.5$  dB and SNR =  $-16.5$  dB.

further selected as 100. The ROMU user performs malicious act randomly in these 100 iterations.

For the GA, the total number of chromosomes is taken as 16 containing the sensing information of the  $P$  cooperating SUs with random crossover point selection from 1 to  $P - 1$ . The crossover and mutation operations performed for 10 cycles and best fitness results are selected.

The results shown in Figures 4(a) and 4(b) illustrate the region of convergence (ROC) curve of the GAW-SDF, GAMV-HDF, MV-HDF, EGC-SDF, and MGC-SDF schemes. The detection and false alarm probabilities under varying SNR for each cooperating SU are shown in Figure 4(a). The sensing SUs are  $P = 10$  and  $P = 13$ .

Figure 4(a) shows improved results of the detection probability for a given false alarm probability as the SUs are increased from  $P = 10$  to  $P = 13$ . A similar result of the detection probability is obtained for a given false alarm probability for different ratios of SUs, keeping average SNR as  $-16.5$  dB and  $-12.5$  dB in Figure 4(b). The probability of detection results in Figure 4(b) at increased SNR improves with the increasing SNR from  $-16.5$  dB to  $-12.5$  dB. Comparing the results in Figure 4(a) and 4(b), cooperative schemes are able to give effective ROC results in Figure 4(a) under increased SNR compared with increased number of cooperative SUs in Figure 4(b).

Both Figures 4(a) and 4(b) compare the proposed GAMV-HDF and GAW-SDF schemes with the simple MV-HDF, EGC-SDF, and MGC-SDF schemes. The result shows that the proposed soft and hard fusion combinations using prior identification of MUs with DSND algorithm followed by the crossover and mutation operation produce sophisticated PU detections against simple MV-HDF, EGC-SDF, and MGC-SDF schemes. In both Figures 4(a) and 4(b), the proposed scheme outperforms existing conventional MV-HDF, MGC-SDF, and EGC-SDF schemes.

The probability of detection against the SNRs is drawn in Figure 5(a) for varying numbers of the cooperating SUs. Figure 5(a) shows that, by increasing the number of SUs from 10 to 13, the detection performance of all cooperative mechanisms is significantly improved. Similarly, in Figure 5(b), the detection performance results are achieved at different number of SUs. The result demonstrates an improvement in the detection results when the number of SUs increases from 10 to 20. Figure 5(b) also shows that the detection performance for a given number of SUs improves rapidly, when the SNR value increases for the number of SUs. Both the results in Figures 5(a) and 5(b) demonstrate that the detection performance of the proposed soft and hard fusion schemes is producing best detection results in comparison with EGC-SDF, MV-HDF, and MGC-SDF schemes. The proposed scheme detection results are followed by the MGC-SDF scheme while the simple MV-HDF scheme gives worst performance.

The probability of error  $P_e$  is plotted against SNR for different number of SUs in Figures 6(a) and 6(b). The results show that by increasing the average SNR and total number of SUs, the error in sensing the PU channel reduces considerably. The results in Figures 6(a) and 6(b) show that the proposed schemes are intelligent in generating less probability of error in comparison with other soft and hard fusion schemes such as MGC-SDF, EGC-SDF, and MV-HDF.

It is clear from the simulations that the DSND-based GA followed by the soft and hard fusion combination schemes make the CSS performance more reliable and accurate in the presence of different variations of MUs, i.e., AYMU, ANMU, ROMU, and OMU. The numerical results of the proposed hard and soft decision schemes such as MV-HDF, MGC-SDF, and EGC-SDF confirm that SUs' cooperation provides high reliability and precision in sensing PU activity. The proposed scheme is able to identify and eliminate MUs in order to make the sensing process reliable.

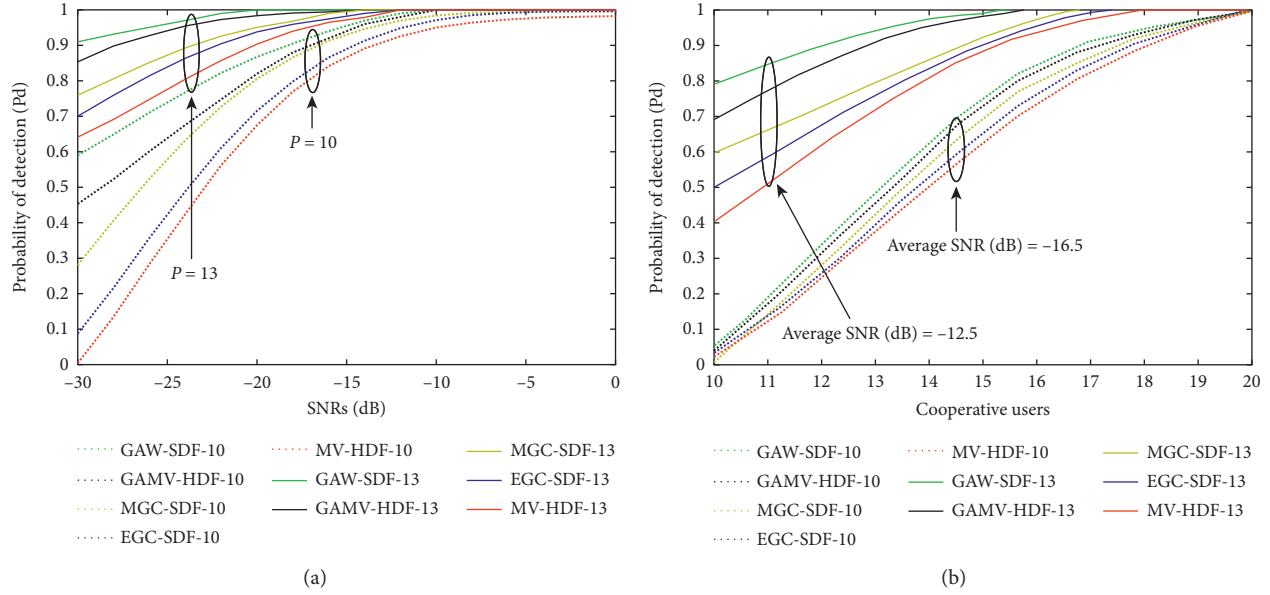


FIGURE 5: (a) Probability of detection vs. SNR, when  $P = 10$  and  $P = 13$ . (b). Probability of detection vs. cooperative users, when SNR =  $-12.5$  dB and SNR =  $-16.5$  dB.

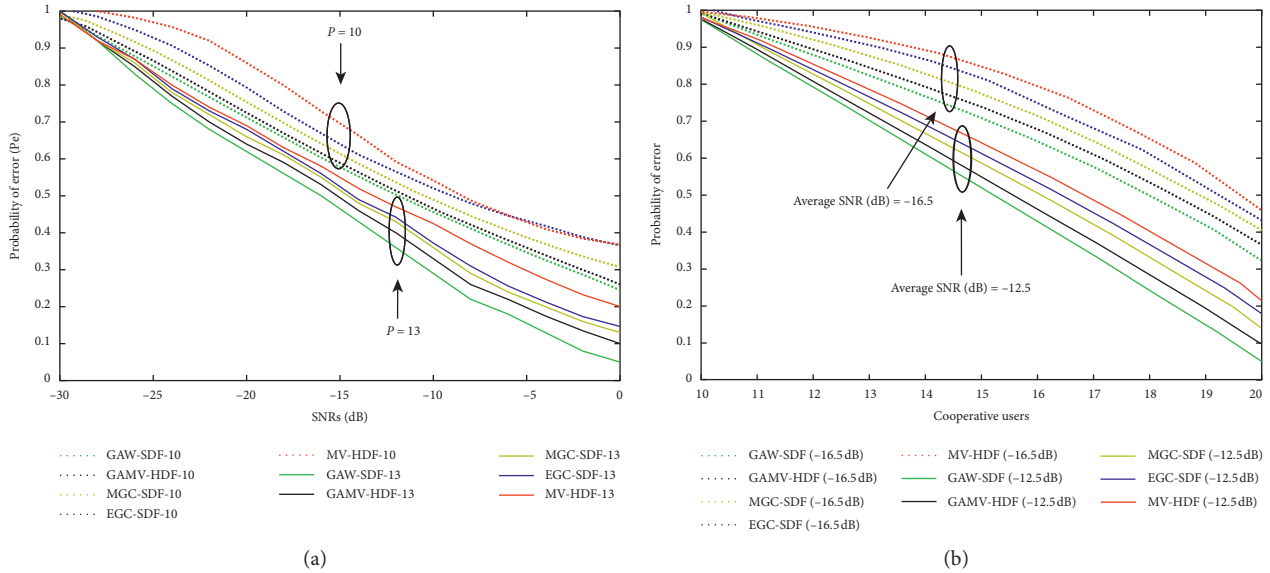


FIGURE 6: (a) Probability of error vs. SNR, when  $P = 10$  and  $P = 13$ . (b) Probability of error vs. secondary error, when SNR =  $-12.5$  dB and SNR =  $-16.5$  dB.

## 5. Conclusions

The false sensing data of MUs reduce effectiveness of CSS. It is therefore essential to evade any confusion in sensing. This paper focuses on improving the existing soft and majority-voting hard fusion combination schemes using GA in the presence of MUs. GA employed DSND for detecting MUs and used crossover and mutation to get precise and reliable sensing results at the FC. The FC used weighted SDF and MV-HDF schemes to take global decision of PU spectrum

occupancy. MUs of different natures are considered, i.e., AYMU, ANMU, ROMU, and OMU, to intensify the harshness of the environment. The numerical results demonstrated that the proposed scheme greatly improves the system performance including sensing accuracy.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) Support Program (IITP-2020-2018-0-01426) supervised by the IITP (Institute for Information and Communication Technology Planning and Evaluation) and in part by the National Research Foundation (NRF) funded by the Korea Government (MSIT) (no. 2019R1F1A1059125).

## References

- [1] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, 2008.
- [2] H. Zhang, C. Jiang, X. Mao, and H. Chen, "Interference-limited resource optimization in cognitive femtocells with fairness and imperfect spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1761–1771, 2016.
- [3] B. Khaled, B. Letaief, and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of IEEE*, vol. 97, no. 5, pp. 878–893, 2009.
- [4] E. Axell, G. Leus, E.G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: state-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.
- [5] R. Chen, J.-M. Park, Y. Thomas Hou, and J. H. Reed, "Toward secure distributed spectrum sensing in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 50–55, 2008.
- [6] S. Mishra, A. Sahai, and R. Brodersen, "Cooperative sensing among cognitive radio," in *Proceedings of IEEE International Conference on Communications*, IEEE, Istanbul, Turkey, June 2006.
- [7] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
- [8] D. Lee, "Adaptive random access for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 831–840, 2014.
- [9] Y. He, S. Member, J. Xue, T. Ratnarajah, M. Sellathurai, and F. Khan, "On the performance of cooperative spectrum sensing in random cognitive radio networks," *IEEE Systems Journal*, vol. 12, no. 1, pp. 881–892, 2016.
- [10] H. Zhang, C. Jiang, N.C. Beaulieu, X. Chu, and X. Wang, "Resource allocation for cognitive small cell networks: a cooperative bargaining game theoretic approach," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3481–3493, 2015.
- [11] S.-S. Kim, S. McLoone, J.-H. Byeon, S. Lee, and H. Liu, "Cognitively inspired artificial bee colony clustering for cognitive wireless sensor networks," *Cognitive Computation*, vol. 9, no. 2, pp. 207–224, 2017.
- [12] M. Zheng, W. Liang, H. Yu, and M. Song, "SMCSS: a quick and reliable cooperative spectrum sensing scheme for cognitive industrial wireless networks," *IEEE Access*, vol. 4, pp. 9308–9319, 2016.
- [13] M. Nabil, W.E. Sayed, and M. Elnainay, "A cooperative spectrum sensing scheme based on task assignment algorithm for cognitive radio networks," in *Proceedings of 2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, Nicosia, Cyprus, August 2014.
- [14] H. Li and Z. Han, "Catch me if you can: an abnormality detection approach for collaborative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3554–3565, 2010.
- [15] H. Guo, W. Jiang, and W. Luo, "Linear soft combination for cooperative spectrum sensing in cognitive radio networks," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1573–1576, 2017.
- [16] M. S. Khan, M. Jibrán, I. Koo, S. M. Kim, and J. Kim, "A double adaptive approach in cognitive radio networks to tackle malicious user," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 2350694, 9 pages, 2019.
- [17] A. A. Sharifi, M. Sharifi, and M. J. Musevi Niya, "Collaborative spectrum sensing under primary user emulation attack in cognitive radio networks," *IETE Journal of Research*, vol. 62, no. 2, pp. 205–211, 2015.
- [18] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," *IEEE Transaction on Wireless Communications*, vol. 9, no. 8, pp. 2488–2497, 2010.
- [19] V.-V. Hiep and I. Koo, "A robust cooperative spectrum sensing based on Kullback-Leibler divergence," *IEICE Transactions on Communications*, vol. E95.B, no. 4, pp. 1286–1290, 2012.
- [20] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4502–4507, 2008.
- [21] H. Guo, N. Reisi, W. Jiang, and W. Luo, "Soft combination for cooperative spectrum sensing in fading channels," *IEEE Access*, vol. 5, pp. 975–986, 2016, <https://search.crossref.org/?q=Soft+combination+for+cooperative+spectrum+sensing+in+fading+channels%2C+IEEE+Access%2C+>
- [22] D. Hamza, S. Aïssa, and G. Aniba, "Equal gain combining for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4334–4345, 2014.
- [23] S. Hunt, Q. Meng, C. Hinde, and T. Huang, "A consensus-based grouping algorithm for multi-agent cooperative task allocation with complex requirements," *Cognitive Computation*, vol. 6, no. 3, pp. 338–350, 2014.
- [24] YL. Lee, WK. Saad, AA. El-Saleh, and M. Ismail, "Improved detection performance of cognitive radio networks in AWGN and Rayleigh fading environments," *Elsevier Journal of Applied Research and Technology*, vol. 11, no. 3, pp. 437–446, 2013.
- [25] Y.-S. Ong and A. Gupta, "Evolutionary multitasking: a computer science view of cognitive multitasking," *Cognitive Computation*, vol. 8, no. 2, pp. 125–142, 2016.
- [26] N. Gul, A. Naveed, A. Elahi, T. Khattak, and I. M. Qureshi, "A combination of double sided neighbor distance and genetic algorithm in cooperative spectrum sensing against malicious users," in *14th International Bhurban Conference on Applied Sciences & Technology (IBCAST)*, IEEE, Islamabad, Pakistan, January 2017.

- [27] P. S. Z. Aizaz and P. Sinha, "A survey of cognitive radio reconfigurable antenna design and proposed design using genetic algorithm," in *Proceedings of Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, IEEE, Bhopal, India, March 2016.
- [28] M. S. Khan, N. Gul, J. Kim, I. M. Qureshi, and S. M. Kim, "A genetic algorithm-based soft decision fusion scheme in cognitive IOT networks with malicious users," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 2509081, 10 pages, 2020.
- [29] A. A. El-Saleh and K. Hussain, "Cognitive radio engine model utilizing soft fusion based genetic algorithm for cooperative spectrum optimization," *International Journal of Computer Networks & Communications (IJCNC)*, vol. 2, pp. 169–173, 2013.
- [30] N. Gul, I. M. Qureshi, A. Omar, A. Elahi, and M. S. Khan, "History based forward and feedback mechanism in cooperative spectrum sensing including malicious users in cognitive radio network," *PLoS One*, vol. 12, pp. 1–21, 2017.
- [31] N. Gul, I. M. Qureshi, A. Elahi, and I. Rasool, "Defense against malicious users in cooperative spectrum sensing using genetic algorithm," *International Journal of Antennas and Propagation*, vol. 2018, pp. 1–11, 2018.



## Research Article

# Provenance Transmission through a Two-Dimensional Covert Timing Channel in WSNs

Qinbao Xu,<sup>1</sup> Li Liu,<sup>1</sup> Rizwan Akhtar,<sup>2</sup> Muhammad Asif Zahoor Raja,<sup>3</sup>  
and Changda Wang<sup>1</sup> 

<sup>1</sup>School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China

<sup>2</sup>School of Electronics and Information, Jiangsu University of Science and Technology, No. 2 Mengxi Road, Zhenjiang 212003, China

<sup>3</sup>Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Attock, Pakistan

Correspondence should be addressed to Changda Wang; changda@ujs.edu.cn

Received 15 March 2020; Revised 14 June 2020; Accepted 30 June 2020; Published 23 July 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Qinbao Xu 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.

Provenances, which record the history of data acquisition and transmission, are hard to be transmitted in resource-tightened wireless sensor networks (WSNs) due to their drastic size expansion with the increase in packet transmission hops. To ease the burden caused by the provenance transmission, we first designed a two-dimensional covert timing channel (2dCTC) and then applied it to provenances transmission in WSNs. Based on Cantor Expansion, 2dCTC uses pseudo packet IDs permutation and packet sizes variation together to form a two-dimensional communication medium. Both theoretical analysis and experimental results show that 2dCTC not only has a much higher channel capacity than those of most of the known CTCs, but also conserves more energy for provenance transmission in WSNs. Furthermore, 2dCTC provides a new way to increase CTCs channel capacity and stealthiness through multi-dimensional approaches.

## 1. Introduction

In the context of wireless sensor networks (WSNs), the provenance of a data item refers to where the item is produced and how it is delivered, i.e., forwarded and/or aggregated to the base station (BS) [1]. Provenance plays an important role in data trust evaluations. Because the size of provenance grows rapidly when packet transmission hop increases, it is then critical to efficiently transmit provenance in resource-tightened WSNs [2]. As a result, several lightweight provenance schemes have been proposed [2–6].

Originally, in a multilevel security system, a covert channel is a mechanism by which a user with high security level can violate the system's security policy to leak sensitive information to a user with lower security level [7]. Now it has been extended to various communication networks and generally defined as the following: if a sender and a receiver use a medium that is not originally designed as the

communication medium for the overt channel, it is a covert channel. As a result, a covert channel has two interesting characteristics: (1) as a side channel it can enlarge its overt channel's capacity without consuming extra energy on signals transmission; (2) its channel capacity is much smaller than that of its overt channel in general. Although the first characteristic is fascinating for provenance transmission through covert channel in WSNs, the second characteristic limits such a usage due to the fact that the channel capacity is too small.

In a packet-switched network, according to the applied communication mediums, covert channels can be roughly categorized as covert storage channels (CSCs) and covert timing channels (CTCs). CSC uses the shared storage in a packet as the communication mediums, e.g., the reserved bits in a packet head; CTC uses the timing characteristics relevant to packet transmissions as the communication mediums, e.g., packet sending frequencies, inter-packet delays, etc. Due to the mediums' deference, CSC can be

eliminated by a network firewall through traffic normalization [8], whereas CTC is hardly to be removed thoroughly. Many CTC schemes such as [9–12] are then proposed.

The inspiration of the paper is to build a CTC which has much higher channel capacity for provenance transmission in WSNs. We then propose a two-dimensional CTC (2dCTC) scheme which uses pseudo packet IDs permutation and packet sizes variation together as the communication medium. Because the two-dimensional communication medium can carry more information, 2dCTC has a much higher channel capacity than the known traditional CTCs.

The main contributions of this paper are as follows:

- (1) We propose a 2dCTC which encodes covert messages into multiple dimension spaces. 2dCTC overwhelms most of the known CTCs with respect to both channel capacity and channel stealthiness.
- (2) We devise the message encoding and decoding algorithms for 2dCTC through Cantor Expansion, which is the key to build a two-dimensional communication medium.
- (3) We apply 2dCTC to the provenance transmissions in resource-tightened WSNs, which saves both energy and channel capacity.

The remainder of this paper is arranged as follows: Section 2 provides the related works. Section 3 presents 2dCTC's design and implementation. Section 4 shows 2dCTC's performance and corresponding experimental results. Section 5 gives the practice of provenance transmission through 2dCTC. Section 6 concludes the paper.

## 2. Related Works

Generally, CTCs adopt the timing behaviour of an entity to transmit covert messages in overt network communication.

Among the entities, inter-packet delays (IPDs) are the most common one that are modulated to encode covert messages. Berk et al. [10] proposed encoding messages through the intervals between adjacent packet transmissions, which avoids the time synchronization requirement that may threaten the channel's concealment. In [11], a CTC is built through mimicking the inter-packet delays (IPDs) of the normal packet traffic flow, by which to implement a detect-resisting CTC. In addition to the IPDs, packet order can also be used to establish CTC, in which the covert messages are represented as reorderings of packets. El-Atawy et al. [12] proposed a packet-reordering channel which uses the packet sequence disorder in transmission as the communication medium. Such a CTC simulates the phenomenon of naturally occurring packet reordering over networks, which has higher channel capacity than those of CTCs based on the fixed time windows and the IPDs. Zhang et al. [13] proposed a method for establishing a VoLTE CTC through packet re-orderings. To further improve the robustness of such a CTC, Gray code is employed to encode the covert message for the purpose of alleviating the packet loss and packet out-of-order. Liang et al. [14] proposed a payload-dependent packet rearranging CTC for mobile VoIP

traffic. Such a CTC can deal with the traffic with more complicated packet distributions such as that in the mobile VoIP environments. In contrast to the aforementioned packet re-ordering methods, we use pseudo packet IDs permutation to encode messages, which can gain more flexibility. There are also some studies using packet length information to build CTC. Liang et al. [15] proposed a packet length covert channel for mobile VoIP traffics, in which the packet length distribution was partitioned and such partitions were mapped to data symbols. The main concept of such a CTC is to send covert messages through transmitting packets of corresponding size. Our method is inspired by such a concept. There is also a category of CTCs using the number of packets transmitted within a time slot to encode/decode messages. Cabuk et al. [9] proposed the Simple Timing Covert Channel (STC), in which the sender divides the timeline into a series of smaller time slots with fixed length; the binary number 1 or 0 is then encoded based on whether a packet is sent within a given time slot. However, such a method requires the clock synchronization between the sender and receiver, which is hard to achieve especially in large-scale networks.

Because each of the CTCs mentioned above uses only one communication medium, all of them are one-dimensional CTCs. To drastically raise the CTCs' capacity, in addition to applying any hardware-based methods, we propose the concept of multi-dimensional CTCs. As a first step for multi-dimensional CTCs' practice, we design and implement a two-dimensional CTC named 2dCTC in the paper.

Among the existing provenance schemes in WSNs, Probabilistic Provenance Flow (PPF) scheme [16] as a block provenance scheme probabilistically appends the node IDs on the packet path to the provenance, and therefore each packet only carries a block of the provenance, i.e., a connected subgraph of a packet transmission path, to the BS. Similarly, Probabilistic Provenance Mark (PPM) scheme [17] probabilistically incorporates node ID to the packet and each packet only contains one node ID. As to provenance transmission through covert channels, to the best of our knowledge, only one paper can be found; viz., in [18], Sultana et al. use the IPDs (inter-packets delays) based CTC for provenance transmission, in which the original purpose is to increase the concealment of the transmission, but objectively saves both energy and channel capacity in WSNs. As a one-dimensional CTC, the IPDs based CTC has very limited channel capacity; the steady packet flows are then required for provenance transmission in [18].

## 3. 2dCTC's Design and Implementation

The 2dCTC proposed in this paper uses pseudo packet IDs permutation and packet sizes variation together as the communication medium. Like the works in [18], the relatively stable data packets flow is required. To facilitate understanding our two-dimensional CTC scheme, we first provide the message encoding and decoding in two one-dimensional mediums, viz., messages encoding and decoding through pseudo packet IDs permutation and packet sizes variation, respectively.

**3.1. Pseudo Packet IDs Permutation as the Medium.** In packet-switched networks, the packet ID disorder rate in transmission is between 0.1% and 3% roughly [19], which provides few packets to form a CTC by the packet IDs permutation. We thus propose the concept of pseudo packet ID that is a data block with a unique value appended to a packet. Unlike packet ID that resided in packet-header, the pseudo packet ID resided in the payload area. Figure 1 shows the working principle of a CTC using the pseudo packet IDs permutation as the communication medium.

At the beginning, the message is divided into  $N$  binary blocks, i.e.,  $\{s_1, s_2, s_3, \dots, s_i, \dots, s_N\}$ , and each block contains 8 bits. The corresponding decimal number of  $s_i$  is  $S_i$ . Let  $\{sid_i \mid sid_n \in R^+, i = 1, 2, \dots, n\}$  represent the set of pseudo packet IDs; the main steps of the message encoding through the pseudo packet IDs permutation are as follows.

- (1) With the number of bits in  $s_i$ , the number of packets  $n$  that satisfies  $2^L \leq n!$  is chosen. So, each  $s_i$  keeps 8 bits and  $n = 6$ .
- (2) With the value of  $S_i$ , a pseudo packet IDs permutation generated from  $\{sid_1, sid_2, sid_3, \dots, sid_n\}$  is processed by Cantor Expansion inverse operation [20], which provides a bijection between a Cantor value  $X$  and a permutation. If there are  $n$  packets, a pseudo packet IDs permutation of  $a[i]$  ( $1 \leq i \leq n$ ), where a Cantor value  $X$  can be derived through the following equation:

$$X = a[n](n-1)! + a[n-1](n-2)! + \dots + a[1]0! \quad (1)$$

- (3) Each generated pseudo packet ID is appended to the payload area of the sending packets in a stream manner.

Note that, compared to the message encoding and decoding through a mapping table whose time complexity is  $O(nlgn)$ , the time complexity of our Cantor Expansion based scheme is  $O(n)$ .

After the CTC receiver filtrates the required packets, the pseudo packet IDs are rearranged according to the packet's arrival time and then the messages can be retrieved through Cantor Expansion by equation (1).

To better understand the approach in this subsection, we provide an example in here. Assume that  $s_i$  is 00001011 (the corresponding decimal number  $S_i$  is equal to 11). The Cantor value  $X$  is then equal to 11 and the pseudo packet IDs are  $\{1, 2, 3, 4, 5, 6\}$ . According to the inverse form of Cantor Expansion, the process is as follows: 11 divided by  $5!$  equals 0 with remainder 11; therefore  $a[6] = 0$ ,  $sid_1 = 1$ ; 11 divided by  $4!$  equals 0 with remainder 11; therefore  $a[5] = 0$ ,  $sid_2 = 2$ . Following the same process,  $a[4] = 1$ ,  $sid_3 = 4$ ;  $a[3] = 2$ ,  $sid_4 = 6$ ;  $a[2] = 1$ ,  $sid_5 = 5$ ;  $a[1] = 0$ ,  $sid_6 = 3$ . As a result, the order of the pseudo packet IDs of 11 is  $\{1, 2, 4, 6, 5, 3\}$ . We append  $\{1, 2, 4, 6, 5, 3\}$  to the sending packets' payload areas. After the CTC receiver filtrates such packets, the pseudo packet IDs permutation  $\{1, 2, 4, 6, 5, 3\}$  whose Cantor value  $X = 11$  is retrieved;  $s_i = 00001011$  is then decoded.

**3.2. Packet Sizes Variation as the Medium.** Using packet sizes variation to encode and decode messages has several obvious advantages. For instance, such a coding method cannot be easily affected by the channel noise such as packet transmission delays and jitters. The working principle of a packet sizes variation based CTC is illustrated in Figure 2. By adopting such a CTC, the message can be encoded through the following steps:

- (1) A histogram model  $\{M, B, X\}$  of packet size is established, in which  $M$ ,  $B$ , and  $X$  denote the number of packets of each group, the group distance, and the sample data sequence, respectively. The statistical function  $M = Hg_{B,R}(X)$ , in which  $R$  sets the packet sizes range for each group, is used to calculate the value of  $M$ , i.e., the number of packets in each group.
- (2) A mapping table is built to represent the correlation between the packet sizes barrel, i.e., a packet size group, and the corresponding binary blocks. Obviously, if a packet size barrel represents  $\alpha$  bits, the number of packet size barrels will be equal to  $2^\alpha$ .
- (3) The message  $s_i$  in a binary representation is encoded into the sending packets based on the mapping table built in the previous step.

After the receiver filtrates the corresponding packets, the messages can be retrieved by looking up the mapping table.

A simple example is provided here for better understanding such a coding method. Assume the message  $s_i$  to be sent is represented in binary as 00001011. There are 9 packets, i.e.,  $p_1, p_2, \dots, p_9$  with different sizes, i.e.,  $l_1, l_2, \dots, l_9$ . We suppose to classify these 9 packets into two packet size barrels,  $B_1$  and  $B_2$  according to the packet size threshold  $l$ ; i.e., packets whose sizes are less than  $l$  are associated with  $B_1$ ; otherwise,  $B_2$ . Assume that  $p_1, p_3, p_4, p_5, p_8, p_6, p_2, p_9$  belong to  $B_1$  and others belong to  $B_2$ . In this example,  $\alpha$  is equal to 1 and the number of packet size barrel is 2. Then,  $s_i$  can be encoded into packet transmission order:  $p_1, p_3, p_4, p_5, p_8, p_6, p_2, p_9$ . After the receiver filtrates the packet size as  $l_1, l_3, l_4, l_5, l_8, l_6, l_2, l_9$ , it can decode  $s_i$  as 00001011 by looking up the mapping table.

**3.3. Two Mediums Are Used Together.** To transmit a message consisting of  $L$  bits, the message needs to be organized as two parts. The first part ( $K$  bits) is encoded through packet sizes variation and the second part ( $L - K$  bits) is encoded through pseudo packet IDs permutation. Figure 3 shows the working principle of 2dCTC. The main steps are shown as follows.

- (1) Calculate  $n$ , the number of packets needed in communication, by

$$\alpha n + \log n! \geq L, \quad (2)$$

where  $\alpha$  denotes the number of bits represented by one packet size in the mapping table. As a result,  $K$  bits are the first  $\alpha n$  bits of the message counting from the left.

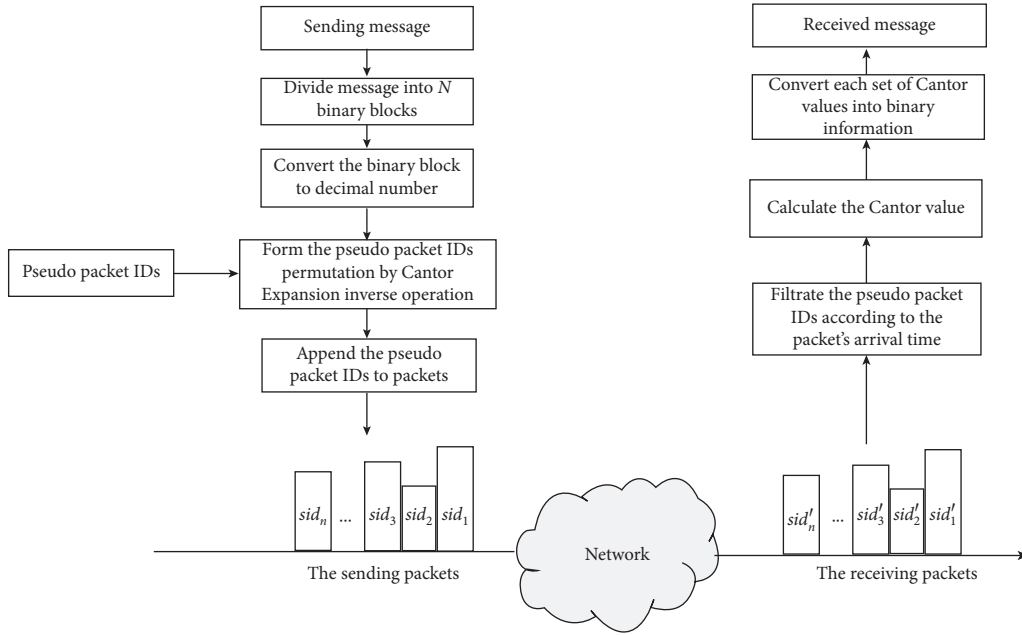


FIGURE 1: Message encoding and decoding through pseudo packet IDs permutation.

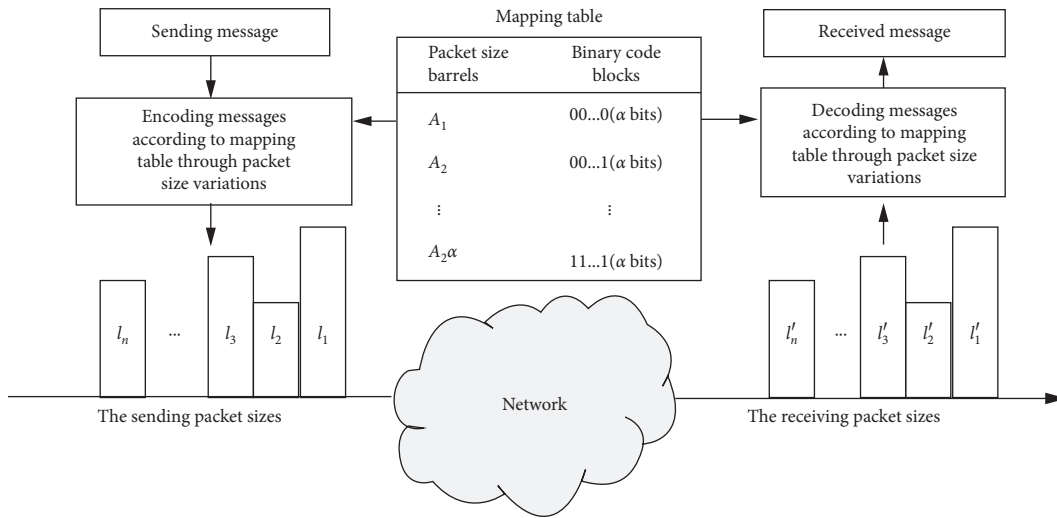


FIGURE 2: Message encoding and decoding through packet size variations.

- (2) Encode  $K$  bits through packet sizes variation and  $(L - K)$  bits through pseudo packet IDs permutation.

Algorithms 1 and 2 are messages encoding and decoding, respectively.

To better understand the approach in this subsection, we provide an example in here. Assume that  $s_i$  is equal to 00001011;  $\alpha$  is equal to 1; the packet size variation satisfies  $l_1 < l_3 < l_4 < l_5 < l < l_2$ ; and the set of the pseudo packet IDs is  $\{1, 2, 3, 4\}$ . According to equation (2),  $n = 4$ ,  $K = 0000$ , and  $L - K = 1011$ . The first part  $K$  bits are encoded as the packet sending order as follows: 1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup>, and the second part  $L - K$  bits are encoded as the pseudo packet IDs permutation  $\{2, 4, 3, 1\}$ . Therefore, the pseudo packet IDs, viz., 2, 4, 3, 1, are appended to the sending packets. At the receiver, the

packet sizes variation  $l_1 < l_3 < l_4 < l_5 < l$  and the pseudo packet IDs permutation  $\{2, 4, 3, 1\}$  can be retrieved. Thereafter,  $K = 0000$  can be decoded by looking up the mapping table. Furthermore,  $L - K = 1011$  can be decoded through Cantor Expansion.  $s_i$  is then successfully decoded as 00001011.

#### 4. Provenance Transmission through 2dCTC

To transmit provenance through 2dCTC, a new provenance scheme 2dCTCP (2dCTC provenance scheme) is devised.

*4.1. Provenance Encoding.* In the continuous data flow environment of WSNs, it is assumed that the network topology is relatively stable, which is the basis for the provenance

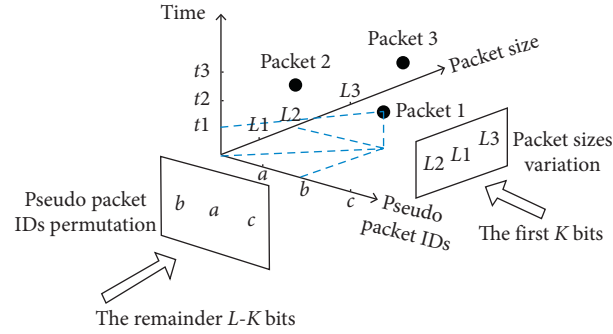


FIGURE 3: The working principle of 2dCTC.

```

Input: packets, message
Output: packet flow
packet size  $\leftarrow \{l_i | l_i \in R, i = 1, 2 \dots n\}$ 
pseudo packet IDs  $\leftarrow \{sid_i | sid_i \in R, i = 1, 2 \dots n\}$ 
message  $\leftarrow \{s_1, s_2, s_3, \dots, s_i, \dots, s_N\}$ 
FOR each  $s_i$  DO
    calculate the number of packet  $n$ 
     $L \leftarrow$  the number of bits in  $s_i$ 
     $K \leftarrow$  the first  $n$  bits of message from  $s_i$ 
    cache packets before sending
     $L - K \leftarrow$  the number of remainder bits
     $D_{L-K} \leftarrow$  the decimal number of the  $L - K$  bits
    new packet sequence ( $M_i$ ) = encoding through packet sizes variation ( $K$ )
    send pseudo packet ID ( $sid_1, sid_2, \dots, sid_n$ ) =  $\text{uncantor}(n, D_{L-K}, \text{pseudo package IDs})$ 
    FOR each packet ( $M_n$ ) DO
        append pseudo packet IDs ( $sid_1, sid_2, \dots, sid_n$ ) into ( $M_1, M_2, \dots, M_n$ )
    END FOR
END FOR
send packets in a stream manner
    
```

ALGORITHM 1: 2dCTC encoding.

```

Input: packet flow
Output: message
assigne  $n$  packets into a group
FOR each group DO
    add pseudo packet IDs to generate permutation
     $D_{L-K} \leftarrow$   $\text{cantor}(\text{pseudo packet IDs permutation})$ 
     $L - K \leftarrow$  the binary of the  $L - K$  bits
    FOR each packet size DO
         $K \leftarrow$  looking up mapping table
        decode the first part  $K$  bits
    END FOR
    decode the second part  $L - K$  bits
END FOR
    
```

ALGORITHM 2: 2dCTC decoding.

transmission method based on 2dCTC proposed in this paper. 2dCTCP is a segmented scheme, which probabilistically incorporates the provenance at each node on the packet path into a series of packets provenance blocks.

In this paper, we consider a node-level provenance; i.e., the node IDs on the path the packet traversed are encoded as

provenance. For the formal network model of the WSN we considered and provenance model, one can refer to [3–5].

The main steps of provenance transmission by 2dCTC are as follows.

- (1) Set the hash value to group the provenance blocks.

In order to identify the packets that have the same provenance, we calculate the hash value for the packet path at each node through

$$H(n_i) = H(H(n_{i-1}) + n_i), \quad (3)$$

where  $n_i$  and  $H(n_{i-1})$  denote  $i^{\text{th}}$  node's ID and the hash value on the  $(i-1)^{\text{th}}$  node, respectively. Therefore, the packets that encoded the different part of the same provenance share the same hash value.

- (2) Determine the number of packets needed to encode provenance.

Assume that the length of the maximum ID is  $L$  bits; the number of packets  $n$  then satisfies

$$\alpha n + \log n! \geq L. \quad (4)$$

- (3) Update the provenance.

If the random probability  $p_i$  generated at the current node is larger than the preset probability threshold  $P$ , the provenance and hash value will be updated; otherwise, only the hash value is updated.

- (4) Encode the provenance to the sending packets.

Algorithm 3 shows provenance encoding through 2dCTCP.

**4.2. Provenance Decoding.** When the BS receives the packets, the main steps of provenance decoding are as follows:

- (1) The BS classifies these packets according to the hash values and assigns  $n$  packets into a group
- (2) In each group, the BS gets the packet sizes and decodes partial provenance through looking up the mapping table; thereafter, the BS retrieves the remainder provenance part according to the Cantor value formed by the pseudo packet IDs permutation

Algorithm 4 shows the provenance decoding through 2dCTCP. In the related works, the only known provenance transmission through CTC uses the IPDs based one-dimensional CTC [18], which was designed mainly to improve the concealment of provenance transmission. Compared to such a method, our 2dCTC provenance scheme can conserve more energy and channel capacity in WSNs.

## 5. Evaluation

**5.1. 2dCTC Performance Analysis.** The performance of 2dCTC is analysed and the corresponding experimental results are provided.

**5.1.1. Channel Capacity.** Note that  $n$  packets can represent (1)  $n!$  bits through pseudo packet IDs permutation and (2)  $m^n$  bits through packet sizes variation, where  $m$  is the number of packet size differences. If  $L$  bits are encoded by  $n$  packets,  $L$ ,  $n$ , and  $m$  should satisfy the following equation:

$$L = \log n! + n \log m. \quad (5)$$

**Input:** *packets, pseudo packet IDs*  
**Output:** *provenance*  
**FOR** *each packet* **DO**  
     *hash\_value = hash(hash\_value + n<sub>i</sub>)*  
**END FOR**  
*choose n packets with the same hash value*  
**IF** *p<sub>i</sub> < P* **THEN**  
     *encode provenance with Algorithm 1*  
**END IF**  
*send packets in a stream manner*

ALGORITHM 3: 2dCTCP provenance encoding.

As a result, the upper bound of the channel capacity is as follows:

$$C = \frac{L}{(n-1)T} = \frac{\log n! + n \log m}{(n-1)T}. \quad (6)$$

**5.1.2. Channel Error Rate.** The 2dCTC's channel error rate can be caused: (1) the noise that spoils the order of packets in transmission, e.g., packet transmission jitters and delays; (2) the noise that spoils the number of packets in transmission, i.e., packet loss, packets aggregation, packet division, and dummy packet padding.

In our previous work [21], the negative influence of those noises has been thoroughly discussed for one-dimensional CTCs. Here, we used part of the conclusions from [21] to derive 2dCTC's channel error rate.

As to the error rate caused by the packet transmission delays and jitters, the inter-packet delay  $T_r$  at the receiver can be calculated by

$$\begin{aligned} T_r &= t_{k+1} + T_d + j_{k+1} - (t_k + T_d + j_k) \\ &= T + (j_{k+1} - j_k) \\ &= T + j_k^{(1)}, \end{aligned} \quad (7)$$

where  $t_k$  and  $t_{k+1}$  denote the sending moments of the  $k^{\text{th}}$  and  $(k+1)^{\text{th}}$  packets, respectively;  $T_d$  denotes the transmission expectation time;  $j_k$  and  $j_{k+1}$  denote the transmission jitters of the  $k^{\text{th}}$  and  $(k+1)^{\text{th}}$  packets, respectively; and  $j_k$  and  $j_{k+1}$  are normal distribution random variables.

As a result, to keep the order of packets in transmission,  $\Delta + j_k^{(1)} > 0$  must be satisfied. Since  $n$  packets in transmission form  $n-1$  delays, the channel error rate is then as the following [22]:

$$P_e = 1 - \left[ P_{\Delta + j_k^{(1)} > 0} \right]^{n-1} = 1 - \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \frac{\Delta}{2\sigma} \right) \right]^{n-1}, \quad (8)$$

where  $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = (2/\sqrt{\pi}) \int_x^{+\infty} \exp(-y^2) dy$ .

To decrease the channel error rate caused by packet transmission jitters and delays, the interval between adjacent packets sending should be enlarged.

As to the channel error rate caused by packet loss, packets aggregation, packet division, and dummy packet

```

Input: packets, pseudo packet IDs, packet sizes
Output: provenance
FOR each packet DO
  IF packets have the same hash value THEN
    every  $n$  packets are assigned into a group
  END IF
  FOR each group DO
    get packet sizes and pseudo packet IDs
     $L - K = \text{cantor}(\text{pseudo packet IDs})$ 
    provenance_remainder = the binary of the  $L - K$  bits
    provenance = message corresponding to packet sizes obtained from the mapping table
  END FOR
END FOR

```

ALGORITHM 4: 2dCTCP provenance decoding.

padding, without loss of generality, assuming  $\lambda$  denotes the probability of packet loss,  $\mu$  denotes the probability of a packet aggregated with its following packet,  $\nu$  denotes the probability of a dummy packet insertion, and  $\omega$  denotes the probability of a packet division. The expectation for the channel error rate under those kinds of noise is then

$$\varphi = 1 - (1 - \lambda)(1 - \mu)(1 - \nu)(1 - \omega). \quad (9)$$

The physical meaning of  $\varphi$  is that the probability of at least one of those kinds of noise has happened.

To mitigate the negative influence caused by packet loss, packets aggregation, packet division, and dummy packet padding, the redundant information should be added, i.e., sending the same message  $K$  times under a noisy 2dCTC, where  $K \geq 1$  and  $k \in \mathbb{N}^+$ .

**5.2. 2dCTC Experiments.** In order to verify the correctness and effectiveness of 2dCTC, we used Python to implement the covert communication between two hosts. The IP addresses of the two hosts were 112.24.29.117 and 10.3.11.180, respectively, where TCP is used as the communication protocol. In the experiment, packets are generated through the Scapy library. A 400-byte text file is selected as the message. The intervals between packets are selected from 5 ms to 40 ms. We compare the total time consumption and capacity of 2dCTC with those of two one-dimensional CTCs, where the unit of capacity is Bps, i.e., the number of bytes transmitted in 1 s. The first one-dimensional CTC is packet rearrangement CTC, which uses different packet IDs permutation to represent the message. The other one-dimensional CTC is packet rearrangement CTC that applies the packet sizes variation to represent the message. Packet rearrangement CTC represents 8 bits by 6 packets, and the other packet rearrangement CTC uses each different packet size to represent 1 bit, viz., 8 packets bearing 8 bits. The 2dCTC uses 4 packets to represent 8 bits. The experimental results are shown in Figures 4(a) and 4(b), respectively, in which 2dCTC has the smallest time consumption and the higher channel capacity than those of the two one-dimensional CTCs.

**5.3. 2dCTCP Simulations.** We used TinyOS 2.1.2 TOSSIM as the simulator to evaluate the performance of the 2dCTCP scheme. The energy consumption is measured by POW-ERTOSSIMz [23]. We compared the performance of our scheme with those of segment based provenance schemes, i.e., Probabilistic Provenance Mark (PPM) scheme [17] and Probabilistic Provenance Flow (PPF) scheme [16]. The sensor network of 121 nodes with IDs 0 through 120 is deployed. The node with ID 0 is set as the BS. The maximum network diameter is 12, the communication protocol is CTP (Collection Tree Protocol) [24], and the data stream was generated by TinyOS through setting the packets sending interval.

**5.3.1. Performance Metrics.** The main performance metrics are as follows:

(A) Average Provenance Size (APS). The APS is defined as follows [4]:

$$\text{APS} = \frac{\sum_{i=1}^m PS_i}{m}, \quad (10)$$

where  $PS_i$  is the provenance length of the  $i^{\text{th}}$  packet and  $m$  is the total number of packets received by the BS.

(B) Total Energy Consumption (TEC). The TEC is defined as follows [4]:

$$\text{TEC} = \sum_{i=1}^N EC_{n_i}, \quad (11)$$

where  $EC_{n_i}$  is the energy consumed by the node  $n_i$  and  $N$  is the total number of nodes in the WSN.

**5.3.2. Simulation Results.** We simulated the PPM, PPF, and 2dCTCP schemes under the same simulation environment and the results are shown in Figures 5(a) and 5(b), respectively.

Figure 5(a) shows the APS for the PPM, PPF, and 2dCTCP schemes with respect to packet transmission hops. The APS in our scheme does not increase as the number of

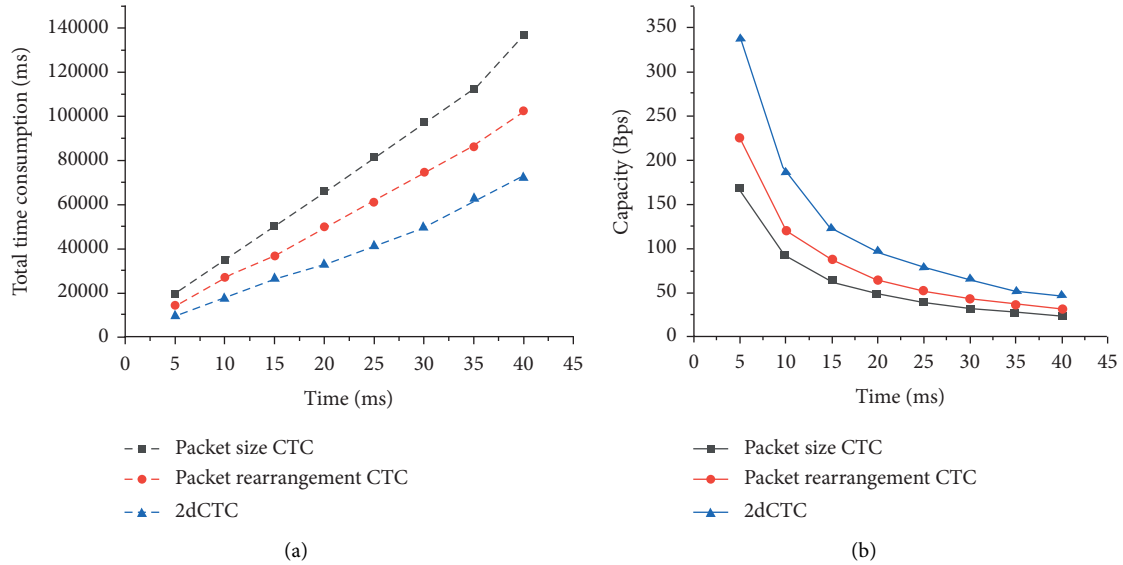


FIGURE 4: (a) Total time consumption for sending a 400-byte file with different time intervals. (b) Channel capacities with different time intervals.

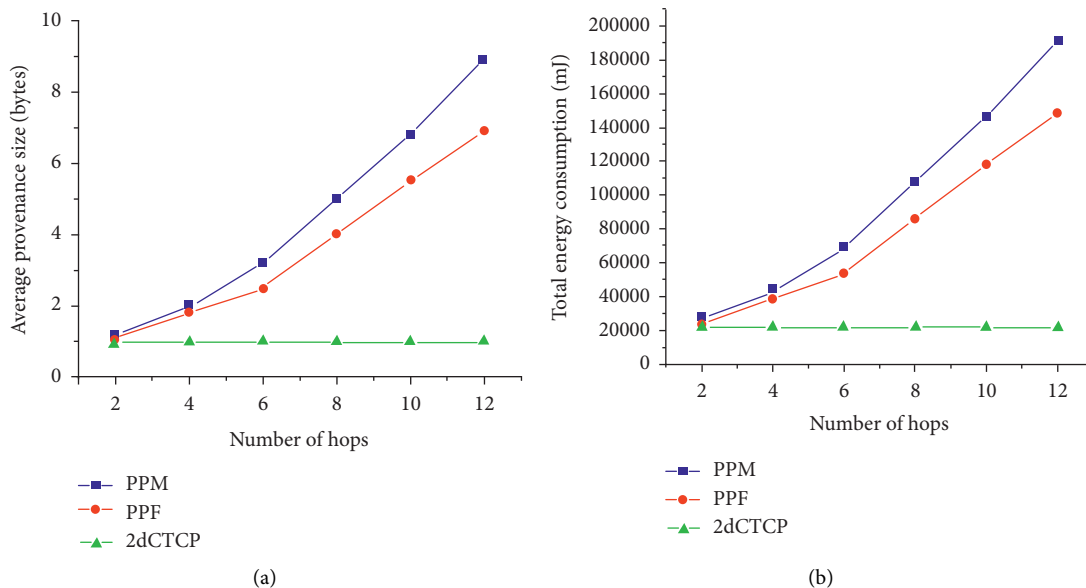


FIGURE 5: (a) The APS generated from a WSN with 121 nodes in which each source node generates about 360 packets. (b) The TEC generated from a WSN with 121 nodes in which each source node generates about 360 packets.

hops increases and remains constant at around 1 byte, whereas for PPM and PPF schemes, APS increases with the increases of packet transmission hops. In the 2dCTCP scheme, the provenances were encoded and transmitted in the timing channel but not in the packets. Although the packets are required to carry pseudo packet IDs, the size of packets is not expanded further according to the provenance's expansion. Hence, our scheme has much better performance than the PPM and PPF schemes with respect to provenance size.

Figure 5(b) shows the relationship between the number of packet transmission hops and TEC of the PPM, PPF, and 2dCTCP schemes. The trend of the curves in Figure 5(b) is

closely consistent with that of the curves in Figure 5(a). As a result, under the same condition, the 2dCTCP scheme is more efficient than that of the PPM and PPF schemes regarding energy consumption.

## 6. Conclusion

In the paper, we propose 2dCTC, a two-dimensional CTC. By using both pseudo packet IDs permutation and packet sizes variation as the communication medium, 2dCTC can dramatically increase the channel capacity compared to the one-dimensional CTC. To ease the burden of provenance transmission, we apply 2dCTC to provenance transmission



in resource constrained WSNs. We analysed the performance of the 2dCTC and validated the benefits of our method through experiments. The simulation results show that using 2dCTC for provenance transmission can conserve more energy than that of PPM and PPF, which further confirms the efficiencies of our method.

## Data Availability

No data are associated with this study.

## Conflicts of Interest

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

## Authors' Contributions

Qinbao Xu and Li Liu contributed equally to the paper.

## Acknowledgments

This work was supported by the National Science Foundation of China under grant 61672269, the National Key Research Project under grant 2017YFB1400703, and the Jiangsu Provincial Science and Technology Projects under grant BK20180860.

## References

- [1] C. Wang, W. Zheng, and E. Bertino, "Provenance for wireless sensor networks: a survey," *Data Science and Engineering*, vol. 1, no. 3, pp. 189–200, 2016.
- [2] S. Sultana, G. Ghinita, E. Bertino, and M. Shehab, "A lightweight secure scheme for detecting provenance forgery and packet DropAttacks in wireless sensor networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 3, pp. 256–269, 2015.
- [3] S. R. Hussain, C. Wang, S. Sultana, and E. Bertino, "Secure data provenance compression using arithmetic coding in wireless sensor networks," in *Proceedings of the 2014 IEEE International Performance Computing and Communications Conference (IPCCC)*, pp. 1–10, Austin, TX, USA, December 2014.
- [4] C. Wang and E. Bertino, "Sensor network provenance compression using dynamic bayesian networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 1, pp. 1–32, 2017.
- [5] C. Wang, S. R. Hussain, and E. Bertino, "Dictionary based secure provenance compression for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 405–418, 2016.
- [6] B. Shebaro, S. Sultana, S. R. Gopavaram, and E. Bertino, "Demonstrating a lightweight data provenance for sensor networks," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, pp. 1022–1024, Raleigh, NC, USA, October 2012.
- [7] S. Cabuk, "Network covert channels: design, analysis, detection, and elimination," Ph D. thesis, Purdue University, West Lafayette, Indiana, 2006.
- [8] L. Yao, X. Zi, L. Pan, and J. Li, "A study of on/off timing channel based on packet delay distribution," *Computers & Security*, vol. 28, no. 8, pp. 785–794, 2009.
- [9] S. Cabuk, C. E. Brodley, and C. Shields, "Ip covert timing channels: design and detection," in *Proceedings of the 11th ACM Conference on Computer and Communications Security*, pp. 178–187, Washington, DC, USA, October 2004.
- [10] V. Berk, A. Giani, G. Cybenko, and N. Hanover, "Detection of covert channel encoding in network packet delays," Tech. Rep. TR536, Université de Dartmouth, Hanover, New Hampshire, 2005.
- [11] S. Gianvecchio, H. Wang, D. Wijesekera, and S. Jajodia, "Model-based covert timing channels: automated modeling and evasion," in *Lecture Notes in Computer Science*, pp. 211–230, Springer, Berlin, Germany, 2008.
- [12] A. El-Atawy, Q. Duan, and E. Al-Shaer, "A novel class of robust covert channels using out-of-order packets," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 2, pp. 116–129, 2017.
- [13] X. Zhang, L. Zhu, X. Wang, C. Zhang, H. Zhu, and Y. Tan, "A packet-reordering covert channel over volte voice and video traffics," *Journal of Network and Computer Applications*, vol. 126, pp. 29–38, 2019.
- [14] C. Liang, X. Wang, X. Zhang, Y. Zhang, K. Sharif, and Y. Tan, "A payload-dependent packet rearranging covert channel for mobile voip traffic," *Information Sciences*, vol. 465, pp. 162–173, 2018.
- [15] C. Liang, Y. Tan, X. Zhang, X. Wang, J. Zheng, and Q. Zhang, "Building packet length covert channel over mobile voip traffics," *Journal of Network and Computer Applications*, vol. 118, pp. 144–153, 2018.
- [16] S. M. I. Alam and S. Fahmy, "A practical approach for provenance transmission in wireless sensor networks," *Ad Hoc Networks*, vol. 16, pp. 28–45, 2014.
- [17] M. T. Goodrich, "Probabilistic packetmarking for large-scale IP traceback," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 15–24, 2008.
- [18] S. Sultana, M. Shehab, and E. Bertino, "Secure provenance transmission for streaming data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 8, pp. 1890–1903, August 2013.
- [19] J. C. R. Bennett, C. Partridge, and N. Shtetman, "Packet reordering is not pathological network behavior," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 789–798, 1999.
- [20] C. S. Calude, L. Staiger, and K. Svozil, "Randomness relative to cantor expansions," *Communications in Nonlinear Science and Numerical Simulation*, vol. 10, no. 8, pp. 921–930, 2005.
- [21] H. Jin and C. Wang, "Robustness of the packet delay channels," in *Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA*, pp. 260–267, Tianjin, China, August 2016.
- [22] C. Wang, Y. Yuan, and L. Huang, "Base communication model of ip covert timing channels," *Frontiers of Computer Science*, vol. 10, no. 6, pp. 1130–1141, 2016.
- [23] E. Perla, A. O. Cathain, R. S. Carbajo, M. Huggard, and C. M. Goldrick, "PowerTOSSIM z: realistic energy modelling for wireless sensor network environments," in *Proceedings of the 3rd ACM International Workshop on Performance Monitoring, Measurement, and Evaluation of Heterogeneous Wireless and Wired Networks (PM2HW2N'08)*, pp. 35–42, Vancouver, Canada, October 2008.
- [24] O. Gnawali, R. Fonseca, K. Jamieson, M. Kazandjiev, D. Moss, and P. Levis, "CTP: an efficient, robust, and reliable collection tree protocol for wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 10, no. 1, pp. 1–49, 2013.

## Research Article

# Towards the Design of Context-Aware Adaptive User Interfaces to Minimize Drivers' Distractions

Inayat Khan and Shah Khusro 

*Department of Computer Science, University of Peshawar, Peshawar 25120, Pakistan*

Correspondence should be addressed to Shah Khusro; [khusro@uop.edu.pk](mailto:khusro@uop.edu.pk)

Received 11 March 2020; Revised 28 May 2020; Accepted 4 June 2020; Published 27 June 2020

Academic Editor: Fawad Zaman

Copyright © 2020 Inayat Khan and Shah Khusro. 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 usage of a smartphone while driving is a pervasive problem and has been acknowledged as a significant source of road accidents and crashes. Several solutions have been developed to control and minimize risky driving behavior. However, these solutions were mainly designed from the perspective of normal users to be used in a nondriving scenario. In a driving scenario, any deviation from these assumptions (e.g., touching or taping interfaces and looking to visual items) could impact driving performance. In this research paper, we aimed to design and develop a context-aware adaptive user interface framework to minimize driver distraction. The proposed framework is implemented in Android platform, namely, "DriverSense," which is capable of adapting smartphone user interfaces based on contextual factors including driver preferences, environmental factors, and device usage in real time using adaptation rules. The proposed solution is evaluated both in real time using AutoLog application and through an empirical study by collecting data from 93 drivers through a mixed-mode survey using a questionnaire. Results obtained from AutoLog dataset show that performing activities on smartphone native interfaces while driving leads to abrupt changes in speed and steering wheel angle. However, minimal variations have been observed while performing activities on DriverSense interfaces. The results obtained from the empirical study show that the data are found to be internally consistent with 0.7 Cronbach's alpha value. Furthermore, an Iterated Principal Factor Analysis (IPFA) retained 60 of a total of 61 measurement items with lower uniqueness values. The findings show that the proposed solution has significantly minimized the driver distractions and has positive perceptions in terms of usefulness, attitude, learnability and understandability, and user satisfaction.

## 1. Introduction

Smartphone-distracted driving is one of the main concerns in road safety, which is evident from the fact that 1.25 million deaths and 50 million injuries are reported each year [1]. The usage of a smartphone while driving has made driving more complex by requiring fine-grained cognitive, physical, and psychological skills to perform concurrent executions [2]. Despite known catastrophes, people are habitual of using a smartphone while driving. For example, 0.66 million drivers are using smartphones at a particular instant of time while driving [3]. In reality, the status of a driver while driving is different from a person not driving. In other words, in nondriving scenarios, a person is free to be engaged and performs smartphone activities in almost every situation. However, in driving scenarios, a driver can

somehow say to be a special person due to having limitations to perform smartphone activities. These limitations are due to excessive physical and visual interaction as well as cognitive overload. One of the main reasons for physical and mental engagement in performing smartphone activities is the complex nature and rich interfaces of smartphone platforms. The existing interfaces (i.e., handheld and dock-mounted smartphone interfaces) are typically designed with the assumption that they may be used by the normal users (e.g., a user who has perceptual and cognitive abilities, who can interact for the maximum duration, and who is sitting in comfortable environments) [4]. In a driving scenario, any deviation from these assumptions (e.g., touching or taping the interfaces via hands, looking to the visual items, and cognitive overload) could impact the driving performance.

Furthermore, the interaction of the driver with smartphone applications is not suitable due to complex interfaces as each activity is time-consuming, redundant, and repetitive and has complex navigational structure, requires much cognitive power, and needs a long route to follow [5]. For example, typing and reading text messages require several steps, which could seriously affect eyes movements, reaction time, lane positioning, stimulus detection, speed, and headway while driving [6]. A driver consumes about 12.4 seconds while interacting with a smartphone for dialing calls and an average of 36.4 seconds for performing a texting activity [7]. Moreover, using a smartphone for sending or receiving a text message diverts eyes off the road for an average of 23 seconds [7]. It means that a text message sent or received can divert a driver's eyes off the road for more than half a kilometre while driving at the speed of 90 km/h [7]. Similarly, safe driving requires full attention and loss of focus due to taking eyes off the road for 2 seconds could increase the chances of accidents to twenty-four times [8].

Therefore, there is a strong need to balance the safety and usability of the smartphone while keeping in mind the drivers' status. One way to improve the usability is to change the interaction between the drivers and smartphone, using an adaptive user interface. The adaptive user interfaces use a context-awareness approach and generate new interfaces according to the change in environment, user preferences, and device usage [9, 10]. This approach will help drivers in the personalization of their smartphone user interfaces irrespective of their visual, physical, and cognitive limitations. To achieve a considerably improved driver-friendly user interface design, it requires moderate revisions in the existing interfaces to meet the driver's needs and requirements. This may require a framework supporting an adaptation mechanism to address the drivers' needs, capabilities, and context-of-use to ensure a high degree of acceptability and usability [11].

In this paper, we propose a multimodal smartphone context-aware adaptive user interface framework for drivers. The proposed framework aims to accommodate the user interface requirements of drivers based on the evaluation of different driving and environmental contexts. The proposed framework is implemented on Android platform, namely, "DriverSense," which makes effective use of smartphone and vehicular sensing capabilities to capture and identify different driving contexts (e.g., number of people in the vehicle, road status, weather status, traffic status, speed, noise, vehicle dynamics, and drivers' interests and preferences) to adjust smartphone user interface automatically. The context-dependent simplified interface can be generated using adaption rules and will improve driver safety by minimizing visual, manual, and mental interactions. In this research work, the available researches and best practices from the other domains (e.g., applications of ICT for naturally disabled people) have been borrowed/reused with different levels of details and have come up with a more flexible and adaptive solution for the drivers to ensure their safe journey on the road.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 introduces the proposed

framework, and its implementation is presented in Section 4. Section 5 illustrates the experimental evaluation. The results and discussion are presented in Section 6. Finally, Section 7 concludes the paper.

## 2. Related Work

With the rapid development in vehicular technologies, Intelligent Transportation System, Advanced Driver Assistance Systems, and vehicle handling stability have been promoted since the past century [12]. However, a growing problem of driver distractions, especially usage of smartphone, still exists. The driver's distraction by smartphone, such as texting, phone calling, and using a navigation system, can divert attention away from the primary task, which is one of the main contributors to the road traffic accidents [13]. The usage of a smartphone while driving contributes to nearly one thousand crashes or near-crashes per year, which is a challenging hurdle for road safety [3]. The researchers have tried to minimize driver engagement with a smartphone with the help of some adaptive technologies. These technologies aimed to limit the interactions or provide simplified interactions to the drivers. The existing adaptive technologies focus on three basic principles: blocking of smartphone features, changing the nature of interactions, and simplifying smartphone functionalities (e.g., with the help of shortcuts to the apps) [14, 15].

Several solutions have been designed to reduce drivers' interactions with their smartphones while driving [16]. These solutions recommend blocking off some of smartphone features/functions, including texting, web browsing, and phone calls [14, 15]. Although the blocking approach is encouraging by considering the leading cause of accidents and crashes [14, 17, 18], the approach of blocking smartphone features is not a viable solution to fully mitigate the issues as it is against the will of smartphone users [19]. In addition, researchers from Australia and USA have reported in their studies that blocking of smartphone features has low acceptability among drivers as it is against the adoption of the technology [20–22].

The other approach used by the researchers to minimize drivers' distractions is to change the nature of the interactions between drivers and smartphones by using text-to-speech and speech-to-text metaphors instead of visual-manual interactions [23, 24]. This is an emerging concept and has shown comparatively distinct advantages over the visual-manual interfaces [25, 26]. However, the researchers have suggested that drivers could still face numerous challenges while driving as it requires visual-manual demands, interior glance time, and higher mental demand than a baseline drive [15]. In addition, cognitive demands are high for tasks using voice-based interfaces [27]. Similarly, voice-commands-based interfaces are difficult to comprehend properly in a noisy environment and could have language barriers as most of the system supports only a few natural languages, including English [28, 29]. Privacy is another issue in a driving scenario due to the presence of other commuters in the vehicle, which may restrict the use of smartphones. The privacy issues for auditory interactions

can be resolved by using headphones. However, this will lead to compromised safety due to blocking important background sounds and increase cognitive overload [27]. Moreover, the interaction between driver and smartphone can also be minimized using Head-Up-Displays (HUDs) (i.e., Android Auto, CarPlay, etc.) [30]. These devices can be paired with a smartphone using Bluetooth or physical interfacing. The aim of these devices is to keep eyes on the road and hands on the steering wheel when performing common activities on a smartphone. However, there is a probability to lose focus off the road when looking into HUDs for necessary operations. In addition, using external hands-free systems is often a barrier due to usability, cost, and lack of practicality [14]. Although hands-free systems could reduce visual-manual interaction, they would not reduce cognitive overload [31–33].

A third and emerging approach is the simplification of smartphone functionalities [14, 15]. This approach is aimed at reducing visual interactions by simplifying driver interactions with smartphone applications. Following the idea, several solutions have been developed, which aim to simplify the interactions between drivers and smartphones with the help of shortcuts for apps and voice commands for interactions [34]. However, these solutions can result in excessive cognitive overload due to voice commands, as discussed earlier, off-road visual engagement, and navigational complexity [35]. Furthermore, the latest study [14] found no empirical evidence of these applications regarding minimizing the risk of crashes. Similarly, performing common activities on smartphone and other technologies are tedious and risky tasks for the drivers; even people in normal daily life routines consume about 66% of their efforts and time in correcting and editing text in automatic speech recognizing devices [5]. Various high-quality applications have been introduced but were washed out from the market due to their complex, inefficient, unattractive, static, and confusing user interfaces [36]. These nonadaptive effects of user interfaces can create frustration, which can impact usability and performance among the end users [37]. Therefore, adaptive user interfaces can provide a significant assistance to overcome these usability barriers. The researchers from different domains have emphasized on the development of adaptive user interfaces and have designed easy-to-use, user-friendly, and accessible interfaces according to the HCI guidelines to solve real-world problems in the different domains [5].

Similarly, various tools and methodologies have been used to generate user interfaces in real time by the researchers automatically. A system called “Supple System” [4] generates user interfaces for the users based on their tasks, preferences, and cognitive abilities. The findings have shown that novice users can complete a complex task in less than 20 minutes using the proposed user interface. Multipath user interface systems are developed, which use XML to generate user interfaces on the basis of current contexts [38]. The Egoki system is a user interface generator system designed for people with disability [39]. The purpose of the system was to recommend appropriate user interfaces for the selection of multimedia contents to the users based on their needs.

The MARIA system proposed a model-based user interface description language to automatically generate and customize user interfaces for the different devices in runtime [40]. The ODESeW system is a semantic web portal using the WebODE platform and an ontology application to generate a knowledge portal of interests automatically [41]. For example, it generates different menus based on the users’ interests and adjusts the visibility of contents according to the users’ needs. A generic interface infrastructure has been presented in the MyUI system, which aims to increase accessibility through an adaptive user interface [42]. The MyUI provides a runtime adaptation to user preferences, device usages, and work conditions. An XML-based pervasive multimodal user interface framework is proposed, which helps the designer to design a wide range of platforms that support multiple languages [43]. The aim was mainly how to change the monomodal web-oriented environment of simplified interface for the variety of platforms. A context-aware framework called ViMos has been proposed to provide adapted information to the users through devices embedded in the environment [44]. The system is composed of a set of available widgets to render different data patterns on various visualization techniques to adapt and customize visual layouts in the available area. A conceptual framework has been designed for Intelligent Adaptive Interfaces (IAIs) to guide interface design with the help of a user-centred design approach and proactive use of adaptive intelligent agents (AIAs). These AIAs provide interface aids to minimize the workload and increase awareness. Similarly, the framework will enable the researchers to design knowledge-based systems such as uninhabited aerial vehicle using the IAI models [45].

Researchers have proposed numerous tools for designing creative adaptive UIs for the heterogeneous domains. An adaptive UI has been designed by the researcher [46] to prevent and block the phone calls and messages during the distracted condition. However, blocking the smartphone features is against the will of drivers and is strongly discouraged by the driver as discussed earlier. Furthermore, the researchers investigated the limited adaptive effects like the speed of the car and the angle of the steering wheel. ICCS [46, 47] is an in-car communication system intended to minimize driver distraction when the drivers engages with their cell phones with the help of speech input and output. However, this system is not widely adopted because it does not use the vehicle contextual information for generating automatic UI.

Researchers have proposed different adaptation techniques related to user interface features, such as content, layout optimization, navigation, and modality. These existing adaptation techniques still have limitations and gaps as they merely focus on design-time feature minimization rather than the runtime. Similarly, these adaptations cannot be effectively applied to generate user interfaces needed for the drivers while driving. Most of them are using pre-identified UI feature set based on context at design time. However, they lack recommending the different mode of interactions, which is essential for the contextual changes in driving scenarios. To the best of our knowledge, no attention

has been given to proposing a system that automatically generates user interfaces based on the driver history and profile, with different varying contexts such as speed, road status, noise, and weather.

### 3. Proposed Framework

To provide cellular connectivity to drivers and avoid distractions caused by smartphone usage has the prime focus of researchers. A number of solutions with varying capabilities and strengths have been presented over the years; however, each has its own shortcomings and limitations. In addition, the solutions are developed by the researchers, academia, and organization using their self-developed methodologies with no common understandings and consensuses, therefore resulting in separate islands, which is the wastage of potentials, resources, and time. The context-aware adaptive user interfaces paradigm can potentially solve the distractions and would result in increased usability of a smartphone while driving. Therefore, a context-aware adaptive user interface framework is proposed. The proposed framework is aimed to be adaptive, flexible, workable, and context-aware in different driving scenarios. The framework architecture is pluggable, where external services may be plugged in in a seamless fashion. The framework will make effective use of smartphone and vehicular sensing capabilities to capture and identify different driving contexts (e.g., number of people in the vehicle, road status, weather status, traffic status, speed, noise, vehicle dynamics, and drivers' interests and preferences) to adjust a smartphone user interface dynamically. The context-dependent simplified user interface will improve driver safety by minimizing visual and manual interactions and reduce physical and mental distractions. The framework architecture is a layered architecture consisting of three layers (as shown in Figure 1): data curation layer, processing layer, and UI layer. However, the schematic diagram of the proposed framework and the flow of information between the components to materialize the context-aware adaptive user interface for a driver is depicted in Figure 2. The layered architecture and schematic diagram are explained in the following subsections.

**3.1. Data Curation Layer.** The data curation layer is responsible for obtaining data from multiple sources for processing and use by the upper layers. The data curation layer is divided into various modules, including interaction module, sensory module, data acquisition, and pre-processing module. In the beginning, the driver input could be captured through voice commands, touches, or gestures and stored in user interactions-log for further operations. The speech input of the driver can be captured using a smartphone microphone, car internal infotainment system, or a hand-free Bluetooth device. Sensory input from smartphone sensors as well as vehicle sensors could also be collected. For example, information can be obtained from various sensors, including the Global Position System (GPS), accelerometer, light, noise, and gyroscope. The GPS is used to find the location, altitude, direction, and speed of the car.

Information from the online sources (i.e., web services) could also be used to obtain weather information, temperature, speed of wind, humidity, and so forth. The status of a road can be detected using accelerometer data. The vehicular data could be obtained from the Controller Area Network (CAN) using the standard Onboard Diagnostic (OBD-II) port [48]. Similarly, the data regarding steering angle, brake pressure, and accelerator could be obtained using a Bluetooth scanner. However, the captured data will be processed to obtain meaningful contextual information using contextual values to devise a new mode of interaction for the drivers while driving.

**3.2. Processing Layer.** The processing layer is the core layer of the proposed system, which is responsible for processing and storing the contextual information received from the data curation layer. The reception of contextual information, identification of user context, user information models, and transformation of the user interface into an appropriate layout is the responsibility of this layer. To simplify the operations of this layer, it has been divided into three main modules: information model building, adaptation rule manager, and transformation.

**3.2.1. Information Models Building.** This module is focused on the development of different models based on the creation of adaption rules in online and offline phases. These models include driver model, vehicle model, device model, and context model. The main classes of the models are shown in Figure 3. These models and associated rules could be considered the baseline requirements for the context-aware adaptive user interface generations. The driver model stores information about driver demographics, cognition, sensing power, and experience. The driver's demographics information is all about his/her driving skills, education, age, and cognition including driver attention, learning ability, perception, and concentration. The driver's sensory information is modelled as driver's hearing, sight, and touch sensitivity that directly affects his/her interactions with the system. The experiences are modelled as the level of satisfaction of the user interface after changing according to the context.

The vehicle model stores information about vehicle data (i.e., type of vehicle, type of transmission, capacity, safety features, types of telematics, etc.). The type of vehicle information includes a company of the vehicle model and so forth and transmission system involved automatic or manual gear system, which will also affect the interaction with the system.

The capacity can be modelled by the number of maximum passengers in a vehicle. The safety features include brake assist, automatic emergency braking, and adaptive cruise control. The device information could be stored in the device model (e.g., device type (i.e., smartphone, smartwatch, or other infotainment systems), screen size, screen resolutions, display type, interaction mode, input/output capabilities, connectivity, etc.). This information is essential for the efficient adaptation of the user interface.

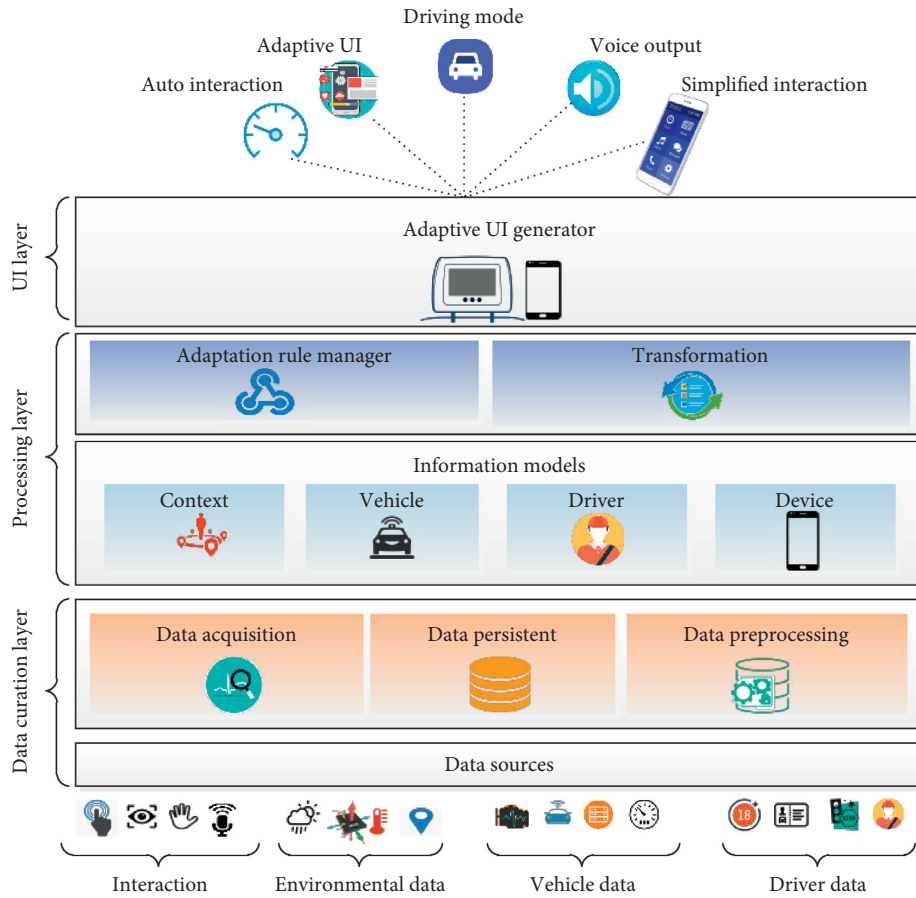


FIGURE 1: Framework of the proposed context-aware adaptive user interface.

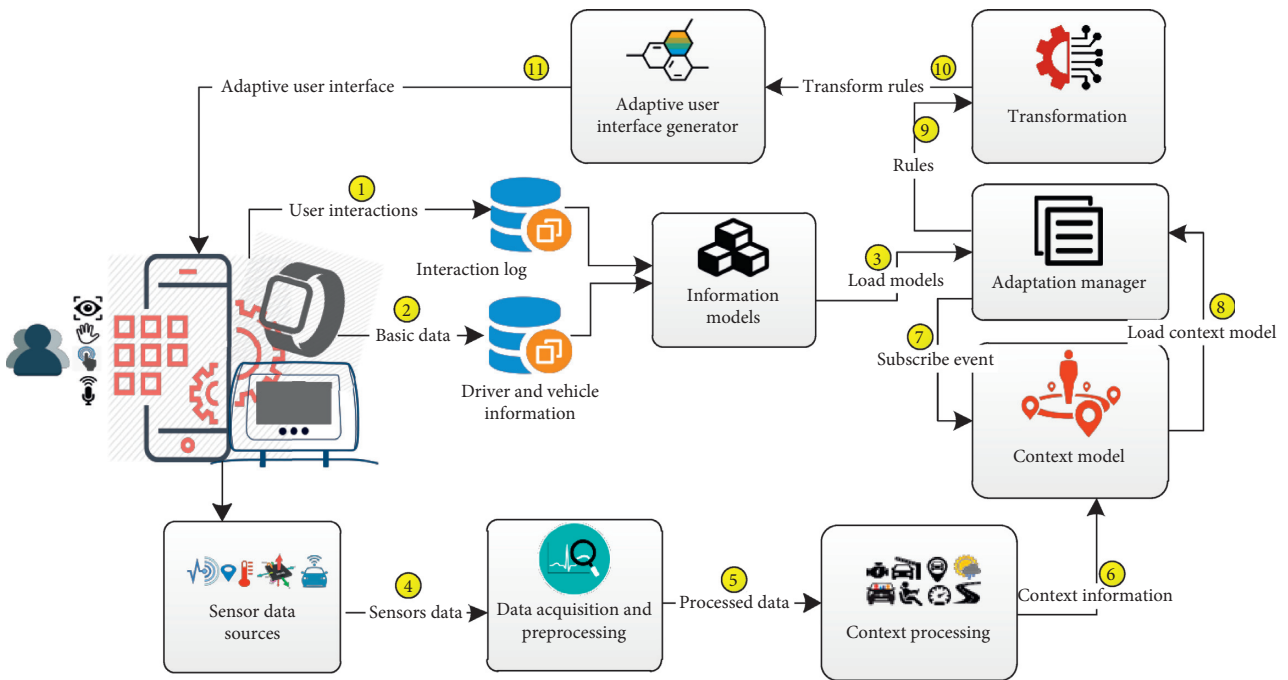


FIGURE 2: Schematic diagram of the proposed system.

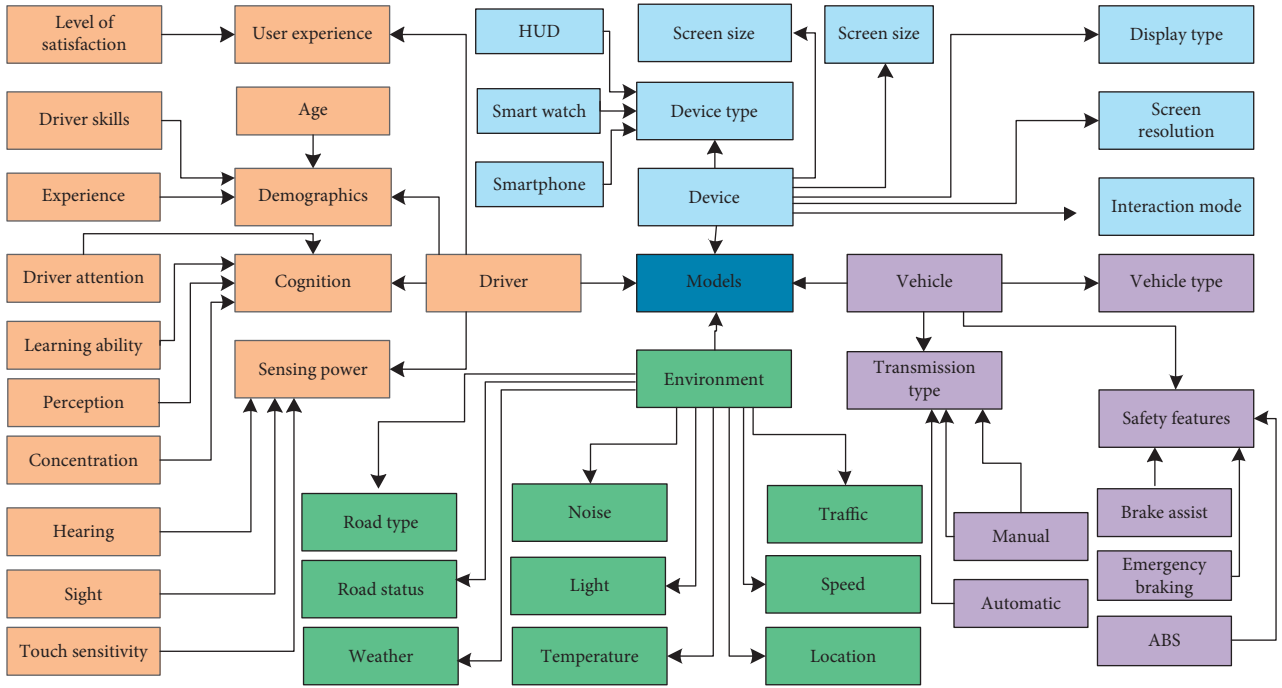


FIGURE 3: Proposed models for driver, device, vehicle, and environmental contexts.

Furthermore, the user-preferred mode of interaction also contributes to the better user interface adaptation. The context model stores information about the environments and context (e.g., road condition, weather, noise, light, temperature, location, time, speed, traffic condition, etc.). The context model is composed of a user, platform, vehicle, and environment (as shown in Figure 4). Once the models are built, they will be passed to the adaptation rule manager.

3.2.2. *Adaptation Manager.* The information models are input to the adaptation rule manager, where the concepts are selected from these models that are associated with different contextual dimensions. The adaptation rules can be specified in the form of *events*, *conditions*, and *actions* [49]. This approach has been extensively used in [50, 51] to provide adaptive UIs. The *event* part of the rule should be composed of the associated event whose manifestation activates the evolution of the rules. The *condition* part is composed of a Boolean condition, which needs to be satisfied to execute the action part.

The *action* part may lead to one or more simple actions containing indications of how the description of the proposed UI should be changed to perform the adaptation process. The rules can be triggered due to contextual cues, which can be dependent on various aspects (i.e., user preferences, environmental changes, etc.). The UI or mode of interaction can be changed according to adaptation rules (e.g., change user interface from vocal to graphical in case the environment is noisy). The proposed adaptation rules for the generation of the context-aware adaptive UI for drivers have been depicted in Table 1 and their threshold values are described in Table 2.

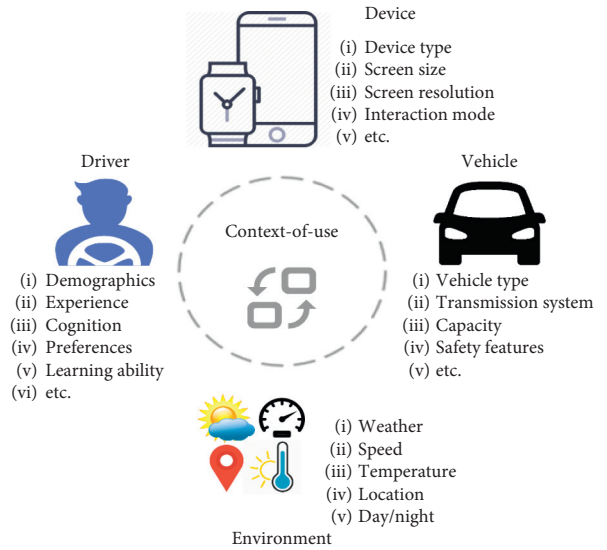


FIGURE 4: Context-of-use for adaptive user interface.

3.2.3. *Transformation.* This module ensures the transformation of a personalized user interface to drivers while driving. The user information model and context model are input to the transformation module through adaptation rules and generate the appropriate user interface to the driver. The contexts and preferences of the drivers are changing with the passage of time in such case when the adaptation rules manager automatically fires the rules to generate a new instance of a user interface or mode of interaction to the driver at runtime. The automatic user interface transformation identifies and transforms the common interface elements/feature into specific interface

TABLE 1: The proposed list of adaptation rules used to generate an adaptive user interface for drivers.

Term	Event	Condition	Action(s)
Driving status	Driving is detected	Smartphone has a default user interface.	The interface will be changed to the driving mode. Divide user interface into sections. Assign priorities to each activity. High-priority activity will be highlighted (high-contrast color and large font size).
Speed	Low speed	The smartphone is in driving mode.	Change the smartphone to moderate modality. Change the default keyboard to the simplified layout. Allow reading short messages. Avoid reading lengthy messages. Divide the SMS reply into categories and choose an option. (i) Standard reply (I'm driving). (ii) Personal reply (you have an option to write a message by yourself). (iii) Fun reply (friends-may be skipped). Auto-reply of SMS for unknown contact. Display contact list in large font size. Use navigation by voice. Dialer digits in large font size. Change the smartphone to severe modality mode. Divide the SMS reply into categories and choose an option. (i) Standard reply (I'm driving). (ii) Personal reply (you have an option to write a message by yourself). (iii) Fun reply (friends-may be skipped). Auto-cancelled lengthy messages.
	Medium speed	Smartphone mode of interaction is in low-speed modality.	SMS/e-mail and WhatsApp reading through voice if no noise and other occupants are detected. Making calls, searching call log, and search contact number through voice. Allow listening to audio (e.g., music). Block watching the videos. Change the smartphone into profound modality mode.
	High speed	Smartphone mode of interaction is in high-speed modality.	(i) SMS: auto reply. (ii) Navigation: voice (top priority in case the route is unknown); however, not loading the navigation activity on a familiar route. (iii) An auto reply of SMS and e-mail to a nonfamily number and those who are not specified in the list. (iv) Allow listening to audio (e.g., music, etc.). (v) Stop audio tuning and selection. (vi) Block watching the videos. (vii) Stop Internet/web browsing.
Noisy environment	Environment is noisy	The vocal modality used for interaction.	The applications change to the graphical modality. Allow reading short messages with a large font size. Digits in dialer will be in large font size. Display contact list in large font size.
Location	Familiar place	The place has been visited for the last five times.	Hide navigation activity.
	Nonfamiliar/little familiar places	First visited location.	Show navigation activity.
Light	Low light	The interface is in normal mode.	The interface should change to night mode.
	High light	The interface is in night mode.	Normal mode.
Interaction	Interaction problem	User is not maintaining attention.	Only contents of one application will be displayed in the user interface at a time.



TABLE 2: Snippet of different contexts and threshold values.

Term	Context	Threshold value
Velocity	High speed	Speed greater or equal to 80 km/h.
	Low speed	Speed is less or equal to 30 km/h.
	Medium speed	Speed is between 30 km/h and 80 km/h.
Location	Familiar location	More than five-time visit on this road.
	Little familiar	More than one-time visit last time.
	Nonfamiliar	First visited road.
Noisy	Environment is noisy	If noise is 25 decibels.
Text messages	Short messages	If the length of the message is less than 30 characters.
	Lengthy messages	If the length of the message is greater than 30 characters.
Contact number	Known number	If saved in the contact list.
	Unknown number	If not saved in the contact list.
Driving status	Is driving	If speed is $\geq 10$ km/h or D gear is detected.
	No driving	If speed is =0 km/h or P gear is detected.

through a series of adaptation rules. These rules constitute a knowledge base system for drivers and the transformation module hinders drivers to be not visually, mentally, and physically distracted while using a smartphone during driving.

**3.3. Adaptive User Interface Generator.** The adaptive user interface generator is communicating with the transformation module to receive the information in real time in order to visualize the appropriate user interface according to the contextual information and adaption rules. The adaptive user interface generator module implements the action part of the adaptation rule depending upon the content received from the transformation module. It can either transform the new simplified user interface or indicate some changes in the exiting interface accordingly. The generated user interfaces could be multimodal (e.g., voice-based, gesture-based, and tactile-based) and will be changed dynamically according to the contexts.

## 4. Implementation

The proposed framework is implemented on an Android platform. Figure 5 shows the snapshots of DriverSense application. The DriverSense app is basically developed for smartphones; however, it can be deployed on any other platform (e.g., infotainment system, etc.) if the required technologies (e.g., libraries and APIs) and resources (e.g., sensors) are available. The DriverSense app is developed while keeping in view all of the design considerations (e.g., privacy and security, battery power consumption, and accessibility). The app is flexible to accommodate and support the new upcoming technologies, especially those related to accessibility. On startup, the main user interface will be divided into subsections yet in a simplified interface whenever the vehicle status is changed to driving mode. The app will take an assessment of a driver's behavior based on the interactions with the user interface. The app will automatically adjust the icons on the main screen, font size, and alert volume, based on the context and the driver's responses. The front screen will contain the selection of most frequently used applications automatically. Furthermore, the settings would be adjusted

according to the contexts: if the noise level is detected, the option for a graphical user interface would be initiated.

Text messaging is found to be the most distracting activity while driving, which can divert eyes off the road and could lead to accidents and crashes. The DriverSense app will handle the text messaging process according to the different driving contexts (i.e., speed, road condition, etc.). For example, if lower speed is detected, such as 30 km/h or less, text messages with a length of less than or equal to 30 characters will be allowed to be read with maximum adjustable font size, whereas lengthy text messages with a length of more than 30 characters will be placed on reading later queue. The auto-reply message will be generated for the SMSs from unknown contacts. The DriverSense app is provided to divide the SMS reply into categories, and driver will choose an option. For example, an SMS reply could be shown in three parts (i.e., standard reply (I'm driving), personal reply (you have an option to write a short message or auto-reply), and fun reply (gossip-type message from friends, which may be skipped)).

Likewise, emails and WhatsApp messages could be managed similar to SMSs. The DriverSense app will also effectively manage a driver's phone calling activities based on the driving scenarios. When the DriverSense app detects a vehicle's driving mode, the simplified user interface for managing phone calls will be launched. The phone calls activities have been classified into simplified and easy-to-access modes including simplified dialer, missed calls, dialed calls, received calls, favorite contacts, and contact list. The activities can be performed using simple touches or using voice commands in case of no external noise. The dialer activity will be automatically sent into the background, and the mode of interaction with the interface will be changed into voice mode when a vehicle's medium speed is detected. Similarly, only the favorite contact list will be made visible, and other activities will be hidden when high speed is detected. Furthermore, the DriverSense app also manages to receive calls activity in the different driving contexts. For example, receiving call option will be displayed for every call if low speed is detected and an option of auto-reply SMS will be made accessible along with receiving call option if medium speed is detected (a driver may swipe the received call option or simply touch the auto-reply SMS to caller) and incoming calls from the unknown

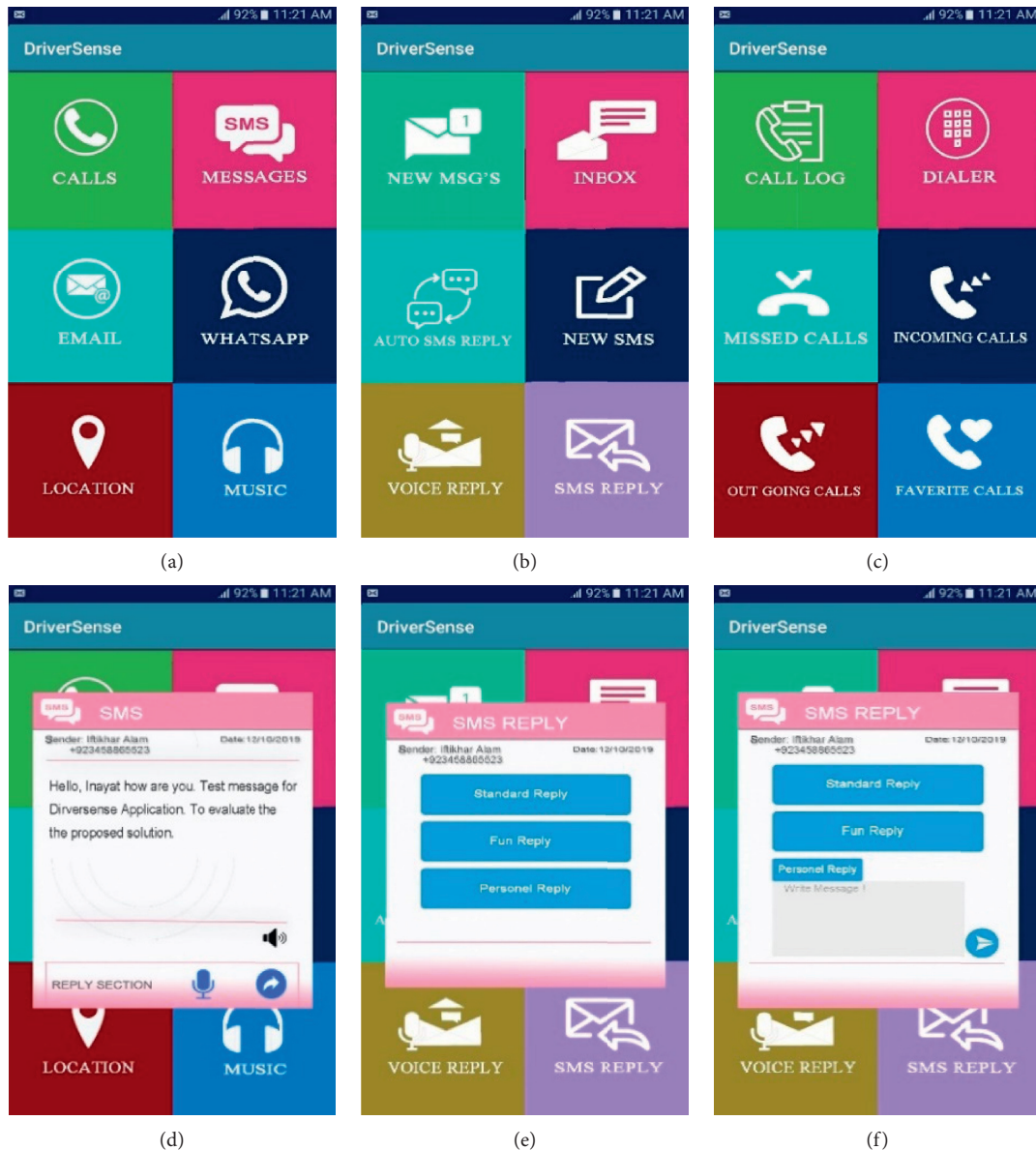


FIGURE 5: DriverSense user interface. (a) Main simplified user interface. (b) SMS activity. (c) Calls activity. (d) New SMS activity. (e) Personalized reply with option activity. (f) SMS personal reply activity.

number will be automatically cancelled with auto-reply SMS if high speed is detected. The DriverSense app also manages the navigation activity. The activity will be on the top in case of unknown routes. If the visited place is familiar (the place visited for five times), the navigation activity will be automatically hidden from the main user interface. For unknown routes, the navigational activity will inform drivers about their current locations on request as well as automatically after some time interval based on their speeds. The DriverSense app will automatically announce the points saved by the drivers and public points of interests through voice. Furthermore, the DriverSense app will automatically send the web-browsing activity into the background whenever vehicle motion is detected. In

addition, the DriverSense app will automatically block video watching in any driving scenario.

### 5. Experimental Evaluation

To the best of our knowledge, the DriverSense app is the first attempt to demonstrate context-aware adaptive user interfaces for drivers to minimized distractions. Therefore, there are no widely agreed evaluation techniques proposed by the researchers. The DriverSense app is tested using basic research-oriented technique and user-based evaluation to demonstrate its effectiveness, accuracy, and usability. In addition, the evaluation is aimed at investigating the systematic understanding of user experiences in using smartphone applications on DriverSense

and measuring the reductions in visual interactions, physical interactions, and cognitive overloads; for the evaluation, the following hypotheses were made:

$$\begin{array}{l}
 H1 \left\{ \begin{array}{l} H_0: \text{The DriverSense does not improve user satisfaction.} \\ H_1: \text{The DriverSense will improve user satisfaction.} \end{array} \right\} \\
 H2 \left\{ \begin{array}{l} H_0: \text{The participants have not a positive attitude towards the usage of DriverSense.} \\ H_1: \text{The participants have positive attitude towards the usage of DriverSense.} \end{array} \right\} \\
 H3 \left\{ \begin{array}{l} H_0: \text{The DriverSense will not minimize cognitive overload.} \\ H_1: \text{The DriverSense will minimize cognitive overload.} \end{array} \right\} \\
 H4 \left\{ \begin{array}{l} H_0: \text{The DriverSense will not minimize visual interaction.} \\ H_1: \text{The DriverSense will minimize visual interaction.} \end{array} \right\} \\
 H5 \left\{ \begin{array}{l} H_0: \text{The DriverSense will not minimize physical interaction.} \\ H_1: \text{The DriverSense will minimize physical interaction.} \end{array} \right\}
 \end{array}$$

*5.1. Evaluation Parameters.* The evaluation process of DriverSense has been carried out through an empirical study on drivers. The usability methods have a common smartphone usage over time for evaluating the usability of applications. Among the others, the most commonly used usability evaluation includes heuristic evaluation, end-user-usability test, and survey and cognitive modelling [52]. Similarly, numerous alternative methods have been used for usability, user experience, and accessibility evaluation, which include automated checking of conformance to guidelines and standards, evaluation using model and simulations, the evaluation conducted by experts, evaluation through users, and evaluation through collected data using keystroke analysis [53]. The DriverSense app is evaluated through the already established set of methods, metrics, and usability parameter suggested by Human-Computer Interaction (HCI) (i.e., ease of use, perceived usefulness, intention to use, operability, understanding and learnability, minimal memory load, system usability scale, consistency, and user satisfaction).

*5.2. Participants Recruitment.* To conduct the empirical evaluation, a sample of 93 participants (79 males and 14 females) are selected voluntarily from the different professional and casual sectors including truck drivers, taxi drivers, students, businessmen, and employees. However, the participants were filtered with conditions of (1) having a valid driving license and more than two years of postlicense driving experience and (2) having experience of using smartphone while driving for more than a year at least. The participants are briefly addressed regarding the purpose of the study and research and expressed their willingness. Table 3 depicts the details of the participants' information in terms of demographic profile, educational background, and

gender. The DriverSense app is installed on the participant smartphones, and initial training has been provided to the participants about its usage.

*5.3. Evaluation Criteria.* The three types of experiments (i.e., user satisfaction, user experience assessment, and perceived usability) are performed in the evaluation process. User satisfaction has been assessed by using a questionnaire for user interaction satisfaction, which measures the overall satisfaction of the system in terms of nine user interface (UI) factors [54]. Similarly, the user experience has been assessed using User Experience Questionnaire (UEQ) [55], which allows a quick assessment of the user experience by getting impressions, user feelings, and attitude after using DriverSense. The UEQ measures both the user experience aspect and the classical usability aspect. Finally, for achieving the perceived usability, the most widely used measure of System Usability Scale (SUS) has been used. The findings obtained from the SUS are more accurate as compared to the Post-Study System Usability Questionnaire (PSSUQ) and Computer System Usability Questionnaire (CSUQ) when the sample size is greater than 8. We are interested in finding the user experience, perceived usefulness, and user satisfaction of the drivers by performing common activities on the interface shown by the DriverSense app. The effectiveness of each activity was evaluated through a set of usability parameters, including the degree of easiness, navigational complexity, consistency, and persistency.

*5.4. Evaluation Process.* User evaluation of the proposed methodology has been performed using the real-world DriverSense and AutoLog [56] android applications. The DriverSense and AutoLog applications are installed on the participants' smartphones. We have instructed the

TABLE 3: Demographic information of the participants who participated in the evaluation.

Variable	Group	Number of participants	Percentage
Gender	Female	14	15.06
	Male	79	84.94
Age	22 to 35 years	61	65.59
	36 to 45 years	18	19.35
	46 to 56 years	14	15.06
Background	Educated	75	80.64
	Literate	18	19.36
Valid driving license	2 to 3 years	53	56.99
	4 to 5 years	14	15.06
	6 to 7 years	13	13.98
	8 to 9 years	10	10.75
	More than 9 years	03	03.22
Upper limb usage	Right hand	46	49.46
	Left hand	18	19.36
	Both hand	29	31.18

participants that the AutoLog application will be running in the background to record the drivers' smartphone activities (e.g., time for activity and activity completion time), vehicle dynamics (e.g., speed, steering angle, brake status, accelerator status, and engine RPM), and environmental data (e.g., location, traffic status, road condition, weather information, temperature, and light intensity) [56]. The participants are ensured that the data will automatically be anonymized before being stored in the database to protect the privacies of the participants. Furthermore, the AutoLog application will automatically stop logging the data whenever the drivers stop driving. The participants are instructed that the logged data will only be used for the evaluation purpose to compare the activities performed in native smartphone interfaces with the activities performed on DriverSense. After completing the exercise of three months, the participants are asked to fill the questionnaire to investigate DriverSense user satisfaction, perceived usability, user experience, and efficiency.

## 6. Results and Discussion

The data were collected through both the questionnaire and AutoLog application and used for performing two types of analysis: empirical analysis and dataset-based analysis. The purpose of these two types of analysis is to find out the significance of the DriverSense application.

We have carried out different tests in this study and analyzed the statistical data using different software like STATA, SPSS, AMOS, and Excel. In our case, we have used descriptive tabulation reporting frequencies and percentages of the categories of the variables. After that, a cross-tabulation is performed with cell percentages and cell likelihood ratio Chi-squared tests. The complied results have significant importance in a way that they give us two-way ( $2 \times 2$ ) cell frequencies count and cell percentages along with the measures of association of measurement items. To check the variable's scales reliability, Cronbach's alpha test has been carried out. Furthermore, we have also performed factor analysis in which Iterated Principal Factor Analysis (IPFA) was found to be better as compared to others. The purpose of

these tests is to investigate the relationship between the user experience attributes of DriverSense user interfaces on attitude, perceived usefulness, ease of use, intention to use, understandability and learnability, minimal memory load, minimal visual interaction, minimal physical interaction, etc. Finally, structural models have been estimated to test the study hypothesis.

*6.1. Descriptive Test Statistics.* The results in Table 4 are self-explanatory, showing descriptive statistics of frequencies and percentages of the categorical indicators of all the variables. For attitude, 60.22% of the respondents chose "very probably" to use DriverSense and 31% chose "Definitely" to use DriverSense. In terms of intention to use DriverSense, 49% and 29% chose "very probably" and "Definitely". A higher 69% of the respondents agreed and 12.90 strongly agreed in terms of perceived usefulness. It shows that almost above 80% perceived its usefulness well.

For the understandability and learnability, more than 90% found DriverSense understandable and easy to learn. About 80% were satisfied with the operation of the DriverSense, and 77% agreed that it is easy to use. In terms of system usability, 74% were in agreement with the software system usability, while 18% chose "probably."

About minimal memory load, 88% moderately agreed, less 7% strongly agreed, and only 4% slightly agreed. It shows that more than 90% were in agreement in terms of minimal memory load, which means that DriverSense requires significantly minimum memory load. For minimal visual interaction, the results were similar as 63% moderately agreed, 14% strongly agreed, and 22% slightly agreed. In minimizing the physical interaction, the results show that more (48%) agreed moderately, less 19% strongly agreed, a good 30% slightly agreed, and negligible 2% disagreed moderately. It shows that majority of the respondents agreed that DriverSense did not require as much memory load and visual and physical interactions. Finally, 66% were very satisfied, 21% were extremely satisfied, and lowly 12% were moderately satisfied with the DriverSense usefulness.

TABLE 4: Descriptive statistics.

Usability parameters	Variables (questions)	Scales	Frequency	Percentages
Attitude	ATTD1, ATTD2, ATTD3, ATTD4	Definitely	29	31.18
		Very probably	56	60.82
		Probably	07	7.53
		Possibly	01	1.08
		Definitely not	—	—
Intention to use	ITU1, ITU2, ITU3, ITU4	Definitely	27	29.03
		Very probably	46	49.46
		Probably	19	20.43
		Possibly	01	1.08
		Definitely not	—	—
Perceived usefulness	PDU1, PDU2, PDU3, PDU4, PDU5, PDU6	Strongly agree	12	12.90
		Agree	65	69.89
		Probably	16	17.20
		Possibly	—	—
		Strongly disagree	—	—
Understandability and learnability	UAL1, UAL2	Much higher	40	43.01
		Higher	45	48.39
		About the same	08	8.60
		Lower	—	—
		Much lower	—	—
Operability	OPT1, OPT2, OPT3, OPT4, OPT5, OPT6, OPT7, OPT8, OPT9, OPT10, OPT11	Very satisfied	06	6.45
		Satisfied	75	80.65
		Neither	12	12.90
		Dissatisfied	—	—
		Very dissatisfied	—	—
Ease of use	EOU1, EOU2, EOU3, EOU4, EOU5, EOU6, EOU7, EOU8	Strongly agree	09	9.68
		Agree	72	77.42
		Probably	12	12.90
		Possibly	—	—
		Strongly disagree	—	—
System usability scale (SUS)	SUS1, SUS2, SUS3, SUS4, SUS5, SUS6, SUS7	Strongly agree	07	18.28
		Agree	69	74.19
		Probably	17	18.28
		Possibly	—	—
		Strongly disagree	—	—
Minimal memory load	MML1, MML2, MML3, MML4, MML5, MML6, MML7, MML8	Agree strongly	07	7.53
		Agree moderately	82	88.17
		Agree slightly	04	4.30
		Disagree moderately	—	—
		Disagree strongly	—	—
Minimal visual interaction	MVI1, MVI2, MVI3, MVI4	Agree strongly	13	13.98
		Agree moderately	59	63.44
		Agree slightly	21	22.58
		Disagree moderately	—	—
		Disagree strongly	—	—
Minimal physical interaction	MPI1, MPI2	Agree strongly	18	19.35
		Agree moderately	45	48.39
		Agree slightly	28	30.11
		Disagree moderately	02	2.15
		Disagree strongly	—	—
User satisfaction	US1, US2, US3, US4, US5	Extremely satisfied	20	21.51
		Very satisfied	62	66.67
		Moderately satisfied	11	11.83
		Slightly satisfied	—	—
		Not at all satisfied	—	—

In Table 5, cross-tabulation of cell percentage and LR Chi-squared test statistics are presented. These results are of significant importance in a way that they give us two-way ( $2 \times 2$ ) cell frequencies count and cell percentages along with the measures of association of measurement items. Cell frequencies and cell percentages give us more exact values of how much each category of factors contributes to the category of the second factor. Also, we have calculated cell test statistics, which gives us the measure of association of each category cell contribution to LR Chi-square of both factors. The significant coefficients of cell LR Chi-squared test statistics are marked with asterisk (\*) at different levels of significance.

**6.2. Data Reliability and Factor Analysis.** Cronbach's alpha tests have been carried out to measure the reliability or, more specifically, internal consistency of the scales of the measurement items [57, 58]. The alpha is measured for each measurement item (factor), and the alpha score represents the expected squared correlation of one scale (also called test) of an item with all other scales (correlation among observed and true value). Here the coefficient of scale reliability is 0.68 ( $\approx 0.7$ ), which is good, and alpha score for each item ranges from 0.67 to 0.68. This shows that our scale items are reliable and internally consistent. For reference, the alpha value of 0.70 and above is considered good, and 0.60 is acceptable [57, 59]. However, a good alpha score varies with the nature of the study and scales of the measurement items. In Table 6, the observations in the alpha column show the number of nonmissing values of the measurement items, while the sign shows the direction of scales correlation. The item-test coefficient shows the strength of correlation of each item with the scales of all other items, while a more robust rest-item coefficient (Corrected Item Total correlation) shows the strength of correlation with the scales of all other 60 items only. The higher the item-test and item-rest correlation coefficients are, the better fit the items are. The average interitem (between measurement items) correlation shows the average correlation between the items. Scales reliability of the measurement items has a theoretical relationship with the factor analysis, as it is assumed that the factor loadings contribute to almost the same/equal information about the score [60]. We have carried out all types of factor analysis of the measurement items (Principal Factor (PFA), Principal Component Factor (PCFA), Iterated Factor (IFA), and Maximum Likelihood Analysis (MLE)) but reported the Iterated Principal Factor Analysis because it retained 60 factors out of 61 factors (measurement items).

Several studies demonstrated that PCFA is the best factor analysis and the most commonly used. The reason why we preferred IPFA over PCFA is the lower uniqueness values of the former over the latter, and there is not much significant difference of factor's retention between the two analyses despite the fact that the PCFA retained all 61 factors. Secondly, the PCFA assumes uniqueness of "0," but here they were all higher than IPFA. In factor model analysis, uniqueness shows the variance of a particular factor that is

not explained by other factors in the model. The results are presented in Table 7. Higher uniqueness values show higher measurement error or a variable with higher uniqueness values means that the latent variable is not well explained by the factor model. For comparison, PCFA has higher uniqueness than IPFA.

In terms of interpretation of the FA results, the eigenvalues show the amount of variation (variance) explained by a particular factor in total variation. In IPFA, 60 factors out of 61 contributed to total variance as all of these factors' eigenvalues are above 0 (positive eigenvalues). But the first 23 factors are stronger than the rest because of their values being above 1. The difference shows the difference between one eigenvalue and the next. But here the proportion is important to be discussed as it shows the proportion of the explained variation to the total variation of a particular factor. Finally, the LR test for the factor model is significant, showing low factor saturation, which is good.

Based on the nature of our variables, we have estimated Kendall's tau-b rank correlation coefficient. In Table 8, we can see that there is no multicollinearity issue in the data (responses to the scales of the variables). The Kendall's tau-b correlation coefficients show the independence of the responses of the factor's scales, which is good in terms of analysis. The values having asterisk (\*) in Table 8 show that the correlation is significant. The results of IPFA in terms of independence versus the saturated model are similar to correlation matrix results.

**6.3. Model Summary and Fitness.** The measurement model had 61 items for 8 latent variables and estimated the absolute and relative, parsimony, and noncentrality fit indices, i.e., Chi-square/d.f., Comparative Fit Index (CFI), Normed Fit Index (NFI), Increment Fit Index (IFI), Tucker-Lewis Index (TLI), Parsimonious Comparative Fit Index (PCFI), Parsimonious Normed Fit Index (PNFI), Relative Fit Index (RFI), and RMSEA for model's assessment. The results show good model fitness with Chi-square/d.f. = 1.227, CFI = 0.543, NFI = 0.84, IFI = 0.825, TLI = 0.5, PCFI = 0.5, RFI = 0.15, PNFI = 0.542, and RMSEA = 0.05. The model estimates the measurement items with their standard errors and probability values.

These measurements indicate that the estimated covariance metrics of the proposed model, as well as the observed model, are found to be significant and satisfactory. Figure 6 shows the final structural model generated from the relationship of latent variables, and Table 9 shows the model estimates of measurement items with their standard errors and probability values.

In respect of Hypotheses such as  $H_1$ ,  $H_2$ ,  $H_3$ ,  $H_4$ , and  $H_5$ , we reject the null hypotheses as a structural model has significant positive estimates as shown in Table 10. The structural model gives  $p$  values less than 0.05, which means that the DriverSense app will minimize mental, visual, and physical interaction and will significantly improve user satisfaction. In terms of attitude, we have significant positive estimates with  $p = 0.030$ , which shows that respondents have a positive attitude towards the usage of DriverSense.

TABLE 5: Categorized cross test statistics.

Variables	Cell percentages																Cell likelihood $\chi^2$																																																																							
	Variables categories				Definitely not				Possibly				Definitely				Very probably				Probably				Possibly				Definitely not				Definitely				Very probably				Probably				Possibly				Definitely																																							
	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M																																								
Attitude	ATTD1	2.15	0.00	2.15	6.45	1.08	18.28	4.30	26.88	5.38	33.33	7.6**	0.00	2.0*	-1.5*	-2.0**	3.6***	-0.7	0.7	0.7	0.7	3.33	33.33	7.6**	0.00	2.0*	-1.5*	-2.0**	3.6***	-0.7	0.7	0.7	0.7	5.38	33.33	7.6**	0.00	2.0*	-1.5*	-2.0**	3.6***	-0.7	0.7	0.7	0.7	26.88	4.30	26.88	5.38	33.33	7.6**	0.00	2.0*	-1.5*	-2.0**	3.6***	-0.7	0.7	0.7	0.7	4.30	26.88	5.38	33.33	7.6**	0.00	2.0*	-1.5*	-2.0**	3.6***	-0.7	0.7	0.7	0.7	4.30	26.88	5.38	33.33	7.6**	0.00	2.0*	-1.5*	-2.0**	3.6***	-0.7	0.7	0.7	0.7
	ATTD2	1.08	2.15	1.08	4.30	3.23	20.43	8.60	34.41	1.08	23.66	1.6*	-1.0	0.6	-0.5	-0.6	0.6	4.5***	-3.8***	6.5***	2.3**	23.66	1.6*	-1.0	0.6	-0.5	-0.6	0.6	4.5***	-3.8***	6.5***	2.3**	1.08	23.66	1.6*	-1.0	0.6	-0.5	-0.6	0.6	4.5***	-3.8***	6.5***	2.3**	1.08	23.66	1.6*	-1.0	0.6	-0.5	-0.6	0.6	4.5***	-3.8***	6.5***	2.3**	1.08	23.66	1.6*	-1.0	0.6	-0.5	-0.6	0.6	4.5***	-3.8***	6.5***	2.3**																				
	ATTD3	2.15	3.23	0.00	10.75	3.23	27.96	1.08	40.86	8.60	40.86	4.8***	-2.1**	0.0	1.0	2.6**	-2.0**	0.8	-1.0	1.0	3.0**	40.86	4.8***	-2.1**	0.0	1.0	2.6**	-2.0**	0.8	-1.0	1.0	3.0**	1.08	40.86	8.60	40.86	4.8***	-2.1**	0.0	1.0	2.6**	-2.0**	0.8	-1.0	1.0	3.0**	1.08	40.86	8.60	40.86	4.8***	-2.1**	0.0	1.0	2.6**	-2.0**	0.8	-1.0	1.0	3.0**																												
	ATTD4	0.00	1.08	0.00	2.15	15.05	4.30	27.96	8.60	38.71	0.0	3.0	0.3	0.0	0.7	-0.7	0.8	-1.0	1.0	3.0**	27.96	8.60	38.71	0.0	3.0	0.3	0.0	0.7	-0.7	0.8	-1.0	1.0	3.0**	4.30	27.96	8.60	38.71	0.0	3.0	0.3	0.0	0.7	-0.7	0.8	-1.0	1.0	3.0**	4.30	27.96	8.60	38.71	0.0	3.0	0.3	0.0	0.7	-0.7	0.8	-1.0	1.0	3.0**																											
Intention to use	ITU1	—	—	2.15	13.98	4.30	18.28	5.38	22.58	3.23	30.11	—	—	-0.5	0.5	1.9*	-1.6*	2.4**	-2.1**	1.0	3.0**	3.23	30.11	—	—	-0.5	0.5	1.9*	-1.6*	2.4**	-2.1**	1.0	3.0**	4.30	18.28	5.38	22.58	3.23	30.11	—	—	-0.5	0.5	1.9*	-1.6*	2.4**	-2.1**	1.0	3.0**	4.30	18.28	5.38	22.58	3.23	30.11	—	—	-0.5	0.5	1.9*	-1.6*	2.4**	-2.1**	1.0	3.0**																							
	ITU2	0.00	1.08	1.08	2.15	6.45	24.73	5.38	31.18	2.15	25.81	0.0	0.3	1.6	-1.0	3.8**	-3.2**	-0.2	0.2	2.7**	4.0***	25.81	0.0	0.3	1.6	-1.0	3.8**	-3.2**	-0.2	0.2	2.7**	4.0***	5.38	31.18	2.15	25.81	0.0	0.3	1.6	-1.0	3.8**	-3.2**	-0.2	0.2	2.7**	4.0***	5.38	31.18	2.15	25.81	0.0	0.3	1.6	-1.0	3.8**	-3.2**	-0.2	0.2	2.7**	4.0***																												
	ITU3	0.00	1.08	2.15	4.30	3.23	13.98	7.53	29.03	2.15	36.56	0.0	0.3	3.2*	-1.9**	1.3	-1.2	4.4***	-3.6**	4.1***	7.2***	36.56	0.0	0.3	3.2*	-1.9**	1.3	-1.2	4.4***	-3.6**	4.1***	7.2***	7.53	29.03	2.15	36.56	0.0	0.3	3.2*	-1.9**	1.3	-1.2	4.4***	-3.6**	4.1***	7.2***	7.53	29.03	2.15	36.56	0.0	0.3	3.2*	-1.9**	1.3	-1.2	4.4***	-3.6**	4.1***	7.2***																												
	ITU4	—	—	3.23	1.08	4.30	10.75	3.23	32.26	4.30	40.86	—	—	9.6***	-2.4**	5.1**	-3.5***	-3.0**	4.1***	4.1***	7.2***	32.26	4.30	40.86	—	—	9.6***	-2.4**	5.1**	-3.5***	-3.0**	4.1***	4.1***	7.2***	4.30	40.86	—	—	9.6***	-2.4**	5.1**	-3.5***	-3.0**	4.1***	4.1***	7.2***	4.30	40.86	—	—	9.6***	-2.4**	5.1**	-3.5***	-3.0**	4.1***	4.1***	7.2***																														
Perceived usefulness	PDU1	0.00	2.15	1.08	4.30	1.08	16.13	6.45	41.94	6.45	20.43	0.0	0.7	0.6	-0.5	-1.8*	3.0**	-1.5*	1.6*	5.6**	4.8***	41.94	6.45	20.43	0.0	0.7	0.6	-0.5	-1.8*	3.0**	-1.5*	1.6*	5.6**	4.8***	6.45	20.43	0.0	0.7	0.6	-0.5	-1.8*	3.0**	-1.5*	1.6*	5.6**	4.8***	6.45	20.43	0.0	0.7	0.6	-0.5	-1.8*	3.0**	-1.5*	1.6*	5.6**	4.8***																														
	PDU2	0.00	2.15	1.08	5.38	2.15	20.43	5.38	33.33	6.45	38.71	0.0	0.7	0.6	-0.5	-1.4	1.8*	0.7	-0.7	1.2	4.2**	33.33	6.45	38.71	0.0	0.7	0.6	-0.5	-1.4	1.8*	0.7	-0.7	1.2	4.2**	5.38	33.33	6.45	38.71	0.0	0.7	0.6	-0.5	-1.4	1.8*	0.7	-0.7	1.2	4.2**	5.38	33.33	6.45	38.71	0.0	0.7	0.6	-0.5	-1.4	1.8*	0.7	-0.7	1.2	4.2**																										
	PDU3	1.08	2.15	3.23	16.13	3.23	18.28	3.23	30.11	4.30	33.33	2.4	-1.1	1.7*	-1.4	1.0	-0.9	-2.0**	2.5**	4.7***	0.6	18.28	3.23	18.28	3.23	30.11	4.30	33.33	2.4	-1.1	1.7*	-1.4	1.0	-0.9	-2.0**	2.5**	4.7***	0.6	16.13	3.23	18.28	3.23	30.11	4.30	33.33	2.4	-1.1	1.7*	-1.4	1.0	-0.9	-2.0**	2.5**	4.7***	0.6																																	
	PDU4	0.00	8.60	3.23	11.83	0.00	13.98	5.38	11.83	6.45	38.71	0.0	2.6**	2.1**	-1.7*	0.0	4.2***	7.3***	-4.7***	0.6	13.98	5.38	11.83	6.45	38.71	0.0	2.6**	2.1**	-1.7*	0.0	4.2***	7.3***	-4.7***	0.6	11.83	6.45	38.71	0.0	2.6**	2.1**	-1.7*	0.0	4.2***	7.3***	-4.7***	0.6																																										
	PDU5	0.00	1.08	0.00	10.75	6.45	21.51	4.30	22.58	4.30	29.03	0.0	0.3	0.0	3.3**	5.1**	-4.0**	0.5	-0.5	-1.2	1.4	21.51	4.30	22.58	4.30	29.03	0.0	0.3	0.0	3.3**	5.1**	-4.0**	0.5	-0.5	-1.2	1.4	4.30	22.58	4.30	29.03	0.0	0.3	0.0	3.3**	5.1**	-4.0**	0.5	-0.5	-1.2	1.4																																						
	PDU6	1.08	0.00	1.08	3.23	3.23	18.28	5.38	29.03	4.30	34.41	3.8	0.0	1.0	-0.7	-0.0	0.0	0.4	-0.4	-2.4**	2.9**	18.28	5.38	29.03	4.30	34.41	3.8	0.0	1.0	-0.7	-0.0	0.0	0.4	-0.4	-2.4**	2.9**	5.38	29.03	4.30	34.41	3.8	0.0	1.0	-0.7	-0.0	0.0	0.4	-0.4	-2.4**	2.9**																																						
U&L	UAL1	—	—	0.00	2.15	4.30	22.58	7.53	35.48	3.23	24.73	—	—	0.0	0.7	0.5	-0.5	2.1	-1.9*	1.9*	22.58	7.53	35.48	3.23	24.73	—	—	0.0	0.7	0.5	-0.5	2.1	-1.9*	1.9*	4.30	22.58	7.53	35.48	3.23	24.73	—	—	0.0	0.7	0.5	-0.5	2.1	-1.9*	1.9*																																							
	UAL2	—	—	4.30	2.15	7.53	9.68	2.15	36.56	1.08	36.56	—	—	11.9***	-3.7***	14.9***	-7.4***	-4.0**	7.2**	9.1***	9.68	2.15	36.56	1.08	36.56	—	—	11.9***	-3.7***	14.9***	-7.4***	-4.0**	7.2**	9.1***	2.15	7.53	9.68	2.15	36.56	1.08	36.56	—	—	11.9***	-3.7***	14.9***	-7.4***	-4.0**	7.2**	9.1***																																						
Operability	OPT1	0.00	3.23	2.15	9.68	4.30	17.20	3.23	32.26	5.38	22.58	0.0	1.0	0.8	-0.7	2.3**	-1.9*	-3.0**	4.1***	2.4**	17.20	3.23	32.26	5.38	22.58	0.0	1.0	0.8	-0.7	2.3**	-1.9*	-3.0**	4.1***	2.4**	4.30	17.20	3.23	32.26	5.38	22.58	0.0	1.0	0.8	-0.7	2.3**	-1.9*	-3.0**	4.1***	2.4**	4.30	17.20	3.23	32.26	5.38	22.58	0.0	1.0	0.8	-0.7	2.3**	-1.9*	-3.0**	4.1***	2.4**																								
	OPT2	0.00	2.15	1.08	8.60	1.08	22.58	8.60	34.41	4.30	27.50	0.0	0.7	2.0	0.6	0.7	-2.4**	4.9***	4.5***	-3.8***	-1.9*	22.58	8.60	34.41	4.30	27.50	0.0	0.7	2.0	0.6	0.7	-2.4**	4.9***	4.5***	-3.8***	-1.9*	1.08	22.58	8.60	34.41	4.30	27.50	0.0	0.7	2.0	0.6	0.7	-2.4**	4.9***	4.5***	-3.8***	-1.9*																																				
	OPT3	0.00	2.15	2.15	6.45	4.30	21.51	7.53	34.41	1.08	20.43	0.0	0.7	2.0	0.6	0.7	-2.4**	4.9***	4.5***	-3.8***	-1.9*	21.51	7.53	34.41	1.08	20.43	0.0	0.7	2.0	0.6	0.7	-2.4**	4.9***	4.5***	-3.8***	-1.9*	4.30	21.51	7.53	34.41	1.08	20.43	0.0	0.7	2.0	0.6	0.7	-2.4**	4.9***	4.5***	-3.8***	-1.9*																																				
	OPT4	—	—	2.17	8.70	4.35	15.22	5.43	32.61	3.26	28.26	—	—	1.1	1.1	-0.9	3.0**	-2.4**	-0.6	0.7	2.9**	15.22	5.43	32.61	3.26	28.26	—	—	1.1	1.1	-0.9	3.0**	-2.4**	-0.6	0.7	2.9**	4.35	15.22	5.43	32.61	3.26	28.26	—	—	1.1	1.1	-0.9	3.0**	-2.4**	-0.6	0.7	2.9**																																				
	OPT5	—	—	1.09	2.17	1.09	13.04	10.87	32.61	2.17	36.96	—	—	1.6*	-1.0	-1.0	-1.4	2.0**	9.9***	-7.4***	7.3***	13.04	10.87	32.61	2.17	36.96	—	—	1.6*	-1.0	-1.0	-1.4	2.0**	9.9***	-7.4***	7.3***	1.09	13.04	10.87	32.61	2.17	36.96	—	—	1.6*	-1.0	-1.0	-1.4	2.0**	9.9***	-7.4***	7.3***																																				
	OPT6	0.00	1.08	0.00	3.23	6.45	12.90	4.30	41.94	4.30	25.81	0.0	0.3	0.0	1.0	1.0	9.5***	-5.8***	-3.9***	5.1***	0.4	12.90	4.30	41.94	4.30	25.81	0.0	0.3	0.0	1.0	1.0	9.5***	-5.8***	-3.9***	5.1***	0.4	6.45	12.90	4.30	41.94	4.30	25.81	0.0	0.3	0.0	1.0	1.0	9.5***	-5.8***	-3.9***	5.1***	0.4																																				
	OPT7	0.00	1.10	0.00	13.19	4.40	15.38	6.59	28.57	3.30	27.47	0.0	0.3	0.0	3.7***	3.5***	-2.7**	3.3**	-2.8**	1.7*	2.0**	15.38	6.59	28.57	3.30	27.47	0.0	0.3	0.0	3.7***	3.5***	-2.7**	3.3**	-2.8**	1.7*	2.0**	4.40	15.38	6.59	28.57	3.30	27.47	0.0	0.3	0.0	3.7***	3.5***	-2.7**	3.3**	-2.8**	1.7*	2.0**																																				
	OPT8	1.08	6.45	1.08	16.13	4.30																																																																																		

TABLE 5: Continued.

Variables	Cell percentages																		Cell likelihood $\chi^2$																																																																																																																																	
	Variables Definitely not									Variables Probably									Very probably						Definitely																																																																																																																											
	F		M		F		M		F		F		M		F		M		F		M		F		M		F		M																																																																																																																							
System usability scale (SUS)	SUS1	1.08	2.15	3.23	13.98	2.15	15.05	5.38	22.58	3.23	31.18	1.6*	-1.0	1.3	-1.2	-0.7	0.8	2.4**	-2.1**	-2.8**	3.8**	SUS2	0.00	2.15	2.15	17.20	1.08	6.45	1.08	9.68	10.75	49.46	0.0	0.7	-1.2	1.5*	-0.1	0.1	-0.8	1.0	3.4**	-3.1**	SUS3	1.08	1.08	2.15	9.68	3.23	15.05	4.30	21.51	4.30	37.63	2.4	-1.1	0.8	-0.7	1.0	-0.9	0.8	-0.8	-3.1**	3.8**	SUS4	1.08	1.08	1.08	7.53	8.60	17.20	2.15	30.11	2.15	29.03	2.4	-1.1	-0.4	0.4	12.7***	-7.8***	-3.3**	5.3**	-3.1**	5.0**	SUS5	1.08	3.23	2.15	12.90	3.23	22.58	0.00	10.75	8.60	35.48	1.0	-0.7	-0.2	0.2	-1.1	1.2	0.0	3.3***	4.2***	-3.6***	SUS6	1.08	5.38	0.00	5.38	4.30	10.75	5.38	29.03	4.30	34.41	0.2	-0.2	0.0	1.6*	5.1***	-3.5***	0.4	-0.4	-2.4**	2.9**																						
	Minimal memory load	SUS7	-	-	0.00	3.23	4.30	20.43	5.38	32.26	5.38	29.03	-	-	0.0	1.0	1.2	-1.1	-0.5	0.5	0.4	-0.4	MML1	0.00	1.08	1.08	11.83	6.45	19.35	3.23	21.51	4.30	31.18	0.0	0.3	-1.2	1.7*	6.1***	-4.5***	-0.9	0.9	-1.7*	2.0**	MML2	0.00	2.15	1.08	5.38	3.23	17.20	2.15	31.18	8.60	29.03	0.0	0.7	0.2	-0.2	0.3	-0.3	-3.4**	5.6***	6.7***	-5.2***	MML3	0.00	2.15	1.08	7.53	4.30	15.05	3.23	26.88	6.45	33.33	0.0	0.7	-0.4	0.4	3.1**	-2.5**	-2.0**	2.5**	0.9	-0.9	MML4	-	-	1.08	7.53	3.23	18.28	4.30	20.43	6.45	38.71	-	-	-0.4	0.4	-0.0	0.0	1.2	-1.1	-0.6	0.6	MML5	-	-	2.15	3.23	3.23	22.58	3.23	31.18	6.45	27.96	-	-	3.9***	-2.1**	-1.1	1.2	-2.8**	3.8**	2.6*	-2.3**	MML6	-	-	0.00	6.45	6.45	17.20	6.45	31.18	2.15	30.11	-	-	0.0	2.0**	7.1***	-5.0***	1.6	-1.4	-3.3**	5.3***
		Minimal visual interaction	MML7	1.08	1.08	0.00	5.38	5.38	18.28	1.08	30.11	7.53	30.11	2.4	-1.1	0.0	1.6*	4.1***	-3.2**	-2.9**	7.2***	4.0***	-3.4**	MML8	0.00	1.08	1.08	6.45	2.15	11.83	5.38	29.03	6.45	36.56	0.0	0.3	-0.1	0.1	0.1	-0.1	0.4	-0.4	-0.0	0.0	MV11	0.00	2.15	1.08	9.68	2.15	15.05	9.68	23.66	2.15	34.41	0.0	0.7	-0.8	1.0	-0.7	0.8	11.8***	-7.9***	-3.8***	6.6***	MV12	-	-	1.08	5.38	3.23	21.51	8.60	37.63	2.15	20.43	-	-	0.2	-0.2	-0.9	0.9	3.4**	-3.0**	-1.8*	2.4**	MV13	-	-	3.23	12.90	4.30	17.20	3.23	33.33	4.30	21.51	-	-	1.7	-1.4	2.3**	-1.9	-3.2**	4.4***	0.8	-0.8																																									
			MPI	MV14	-	-	0.00	3.23	4.30	31.18	8.60	30.11	2.15	20.43	-	-	0.0	1.0	-1.7*	2.0**	6.2***	-4.9***	-1.8**	2.4**	MPI1	0.00	1.08	4.30	8.60	5.38	23.66	4.30	25.81	1.08	25.81	0.0	0.3	6.4***	-3.9***	2.1**	-1.8*	-0.4	0.4	-2.7**	5.9***	MPI2	0.00	7.53	1.08	11.83	4.30	25.81	7.53	19.35	2.15	20.43	0.0	2.3**	-1.2	1.7*	-0.4	0.4	8.7***	-6.0***	-1.8*	2.4**																																																																																		
				US1	1.08	0.00	1.08	6.45	5.38	21.51	5.38	35.48	2.15	21.51	3.8	0.0	-0.1	0.1	2.8**	-2.4**	-1.3	1.5*	-2.0**	2.7**	US2	-	-	2.15	4.30	3.23	21.51	5.38	26.88	4.30	32.26	-	-	3.2**	-1.9*	-0.9	0.9	1.0	-1.0	-2.0**	2.3**	US3	-	-	0.00	25.81	4.30	4.30	6.45	16.13	4.30	38.71	-	-	0.0	1.3	2.7**	-2.2**	-0.6	0.6	-0.4	0.4	US4	-	-	1.08	1.08	3.23	17.20	3.23	27.96	7.53	38.71	-	-	2.4**	-1.1	0.3	-0.3	2.8**	1.1	-1.0	US5	-	-	2.15	3.23	1.08	10.75	4.30	30.11	7.53	40.86	-	-	3.9***	-2.1**	-1.0	1.4	-1.5*	1.7*	0.5	-0.5																																									

\*, \*\*, and \*\*\* refer to significance level of likelihood Chi-squared test statistic at 10%, 5%, and 1%.



TABLE 6: Data reliability test (Cronbach's alpha).

Measurement items	Observations	Sign	Item-test correlation	Item-rest correlation	Average interitem correlation	Cronbach's alpha
ATTD1	93	+	0.4502	0.3883	0.0317	0.6630
ATTD2	93	+	0.1809	0.1082	0.0338	0.6773
ATTD3	93	+	0.2464	0.1753	0.0333	0.6739
ATTD4	93	-	0.0678	0.0061	0.0347	0.6830
ITU1	93	+	0.5422	0.4868	0.0310	0.6577
ITU2	93	+	0.2894	0.2197	0.0330	0.6717
ITU3	93	+	0.2562	0.1853	0.0332	0.6734
ITU4	93	+	0.2598	0.1890	0.0332	0.6732
PDU1	93	-	0.1753	0.1025	0.0338	0.6776
PDU2	93	+	0.0650	0.0090	0.0347	0.6832
PDU3	93	+	0.3073	0.2382	0.0328	0.6707
PDU4	93	-	0.1690	0.0960	0.0339	0.6780
PDU5	93	+	0.3526	0.2855	0.0325	0.6683
PDU6	93	+	0.4834	0.4236	0.0315	0.6611
UAL1	93	+	0.3182	0.2496	0.0328	0.6702
UAL2	93	+	0.3504	0.2832	0.0325	0.6684
OPT1	93	+	0.2179	0.1460	0.0335	0.6754
OPT2	93	+	0.1529	0.0797	0.0340	0.6788
OPT3	93	+	0.1217	0.0482	0.0343	0.6803
OPT4	92	+	0.2932	0.2239	0.0329	0.6715
OPT5	92	+	0.0354	0.0381	0.0349	0.6845
OPT6	93	-	0.0368	0.0371	0.0349	0.6846
OPT7	91	+	0.1371	0.0639	0.0341	0.6794
OPT8	93	-	0.1884	0.1158	0.0337	0.6770
OPT9	93	+	0.4510	0.3892	0.0317	0.6629
OPT10	92	-	0.1312	0.0582	0.0342	0.6798
OPT11	93	-	0.1546	0.0814	0.0340	0.6787
EOU1	93	+	0.1236	0.0500	0.0342	0.6802
EOU2	93	-	0.1411	0.0678	0.0341	0.6794
EOU3	93	+	0.1426	0.0693	0.0341	0.6793
EOU4	93	+	0.2362	0.1647	0.0334	0.6745
EOU5	93	+	0.2029	0.1306	0.0336	0.6762
EOU6	93	-	0.1325	0.0590	0.0342	0.6798
EOU7	93	+	0.1484	0.0751	0.0341	0.6790
EOU8	93	+	0.2544	0.1834	0.0332	0.6735
SUS1	93	+	0.3642	0.2976	0.0324	0.6677
SUS2	93	-	0.1246	0.0510	0.0342	0.6802
SUS3	93	+	0.0575	0.0164	0.0347	0.6835
SUS4	93	+	0.2450	0.1738	0.0333	0.6740
SUS5	93	-	0.1081	0.0343	0.0344	0.6810
SUS6	93	+	0.2057	0.1335	0.0336	0.6761
SUS7	93	+	0.3498	0.2825	0.0325	0.6685
MML1	93	+	0.1521	0.0788	0.0340	0.6788
MML2	93	+	0.1069	0.0332	0.0344	0.6811
MML3	93	+	0.1309	0.0574	0.0342	0.6799
MML4	93	+	0.1866	0.1139	0.0338	0.6771
MML5	93	-	0.1242	0.0506	0.0342	0.6802
MML6	93	-	0.0467	0.0272	0.0348	0.6841
MML7	93	-	0.2135	0.1414	0.0336	0.6757
MML8	93	+	0.2316	0.1600	0.0334	0.6747
MVI1	93	+	0.3433	0.2758	0.0326	0.6688
MVI2	93	+	0.1661	0.0931	0.0339	0.6781
MVI3	93	+	0.3659	0.2993	0.0324	0.6676
MVI4	93	+	0.3062	0.2369	0.0328	0.6708
MPI1	93	+	0.1574	0.0842	0.0340	0.6785
MPI2	93	-	0.3388	0.2711	0.0326	0.6690
US1	93	+	0.3073	0.2383	0.0328	0.6707
US2	93	+	0.2948	0.2253	0.0329	0.6714
US3	93	+	0.3582	0.2913	0.0324	0.6680
US4	93	+	0.1360	0.0625	0.0341	0.6796
US5	93	+	0.3066	0.2374	0.0328	0.6708
Test scale					0.0335	0.6788

TABLE 7: Iterated Principal Factor Analysis.

Measurement items	Eigenvalues	Difference	Proportion	Cumulative	Uniqueness
Factor1	4.16775	0.86173	0.0733	0.0733	0.0332
Factor2	3.30602	0.41087	0.0582	0.1315	0.0332
Factor3	2.89515	0.11932	0.0509	0.1824	0.1042
Factor4	2.77584	0.34843	0.0488	0.2313	0.1598
Factor5	2.42741	0.12902	0.0427	0.2740	-0.0303
Factor6	2.29839	0.13477	0.0404	0.3144	-0.0068
Factor7	2.16362	0.13926	0.0381	0.3525	0.1618
Factor8	2.02436	0.14853	0.0356	0.3881	0.1298
Factor9	1.87583	0.03791	0.0330	0.4211	0.0975
Factor10	1.83793	0.07342	0.0323	0.4535	0.0898
Factor11	1.76450	0.01773	0.0310	0.4845	0.0954
Factor12	1.74677	0.08717	0.0307	0.5153	0.0334
Factor13	1.65960	0.05695	0.0292	0.5445	0.0013
Factor14	1.60264	0.04585	0.0282	0.5727	0.0075
Factor15	1.55680	0.08512	0.0274	0.6000	0.0056
Factor16	1.47167	0.12395	0.0259	0.6259	0.0437
Factor17	1.34772	0.03929	0.0237	0.6497	0.0255
Factor18	1.30844	0.10513	0.0230	0.6727	0.0973
Factor19	1.20330	0.03614	0.0212	0.6938	-0.0078
Factor20	1.16716	0.01158	0.0205	0.7144	0.1079
Factor21	1.15558	0.02960	0.0203	0.7347	0.0711
Factor22	1.12598	0.11679	0.0198	0.7545	0.1533
Factor23	1.00920	0.04908	0.0178	0.7723	0.1003
Factor24	0.96012	0.05031	0.0169	0.7892	0.1117
Factor25	0.90981	0.05021	0.0160	0.8052	0.0324
Factor26	0.85960	0.03158	0.0151	0.8203	0.2479
Factor27	0.82801	0.04933	0.0146	0.8349	0.1175
Factor28	0.77868	0.03654	0.0137	0.8486	0.0953
Factor29	0.74214	0.04227	0.0131	0.8616	0.0880
Factor30	0.69987	0.07413	0.0123	0.8740	0.0720
Factor31	0.62574	0.01323	0.0110	0.8850	0.1018
Factor32	0.61251	0.05363	0.0108	0.8957	0.1097
Factor33	0.55888	0.05524	0.0098	0.9056	0.1659
Factor34	0.50364	0.02172	0.0089	0.9144	0.0001
Factor35	0.48192	0.02075	0.0085	0.9229	0.1346
Factor36	0.46117	0.02582	0.0081	0.9310	0.0962
Factor37	0.43535	0.07850	0.0077	0.9387	0.0038
Factor38	0.35685	0.00565	0.0063	0.9450	0.0978
Factor39	0.35120	0.01275	0.0062	0.9512	0.1304
Factor40	0.33845	0.04914	0.0060	0.9571	0.1338
Factor41	0.28931	0.01496	0.0051	0.9622	-0.0255
Factor42	0.27435	0.02779	0.0048	0.9670	0.0676
Factor43	0.24656	0.00400	0.0043	0.9714	-0.0185
Factor44	0.24255	0.03557	0.0043	0.9756	0.1242
Factor45	0.20699	0.01929	0.0036	0.9793	0.0242
Factor46	0.18770	0.02451	0.0033	0.9826	0.0465
Factor47	0.16319	0.01043	0.0029	0.9854	0.0497
Factor48	0.15275	0.01490	0.0027	0.9881	0.1494
Factor49	0.13786	0.02172	0.0024	0.9906	-0.0380
Factor50	0.11614	0.01123	0.0020	0.9926	0.0230
Factor51	0.10491	0.03082	0.0018	0.9945	-0.0580
Factor52	0.07408	0.01385	0.0013	0.9958	0.0987
Factor53	0.06023	0.01138	0.0011	0.9968	0.1319
Factor54	0.04885	0.00722	0.0009	0.9977	0.0197
Factor55	0.04163	0.00121	0.0007	0.9984	-0.0502
Factor56	0.04042	0.00934	0.0007	0.9991	-0.0417
Factor57	0.03108	0.01773	0.0005	0.9997	0.1320
Factor58	0.01335	0.00820	0.0002	0.9999	0.0725
Factor59	0.00515	0.00398	0.0001	1.0000	0.0407
Factor60	0.00117	0.00175	0.0000	1.0000	0.0763

TABLE 7: Continued.

Measurement items	Eigenvalues	Difference	Proportion	Cumulative	Uniqueness
Factor61	0.00058		-0.0000	1.0000	0.0961
Number of observations	88				
Retained factors with min. eigenvalue (0)	60				
Number of parameters	1830				
Likelihood ratio Chi-squared test (1830)	2188***				

TABLE 8: Kendall's tau correlation matrix.

	ATTD	ITU	PDU	UAL	OPT	EOU	SUS	MML	MVI	MPI	US
<b>ATTD</b>	0.5402										
<b>ITU</b>	0.1510*	0.6360									
<b>PDU</b>	0.0659	0.1050*	0.4703								
<b>UAL</b>	0.0683	0.1316*	0.0468	0.5797							
<b>OPT</b>	0.0972*	0.0208	0.0136	-0.0500	0.3324						
<b>EOU</b>	0.1092*	0.0180	0.0806*	0.0771*	0.0853*	0.3787					
<b>SUS</b>	0.0194	-0.0023	0.0341	0.0547	0.0042	0.0252	0.4149				
<b>MML</b>	0.0103	-0.0026	0.0187	-0.0402	0.0400	-0.0171	0.0421	0.2174			
<b>MVI</b>	0.1192*	0.0079	0.0561	-0.0037	0.0570	-0.0021	0.0210	0.0215	0.5327		
<b>MPI</b>	0.0561	0.0250	0.1094*	0.0327	0.0785*	0.0136	0.0171	0.0526	-0.0376	0.6442	
<b>US</b>	0.1384*	0.1087*	0.0187	0.1185*	0.0767*	0.0222	0.0465	0.0140	0.1019*	-0.0795	0.5007

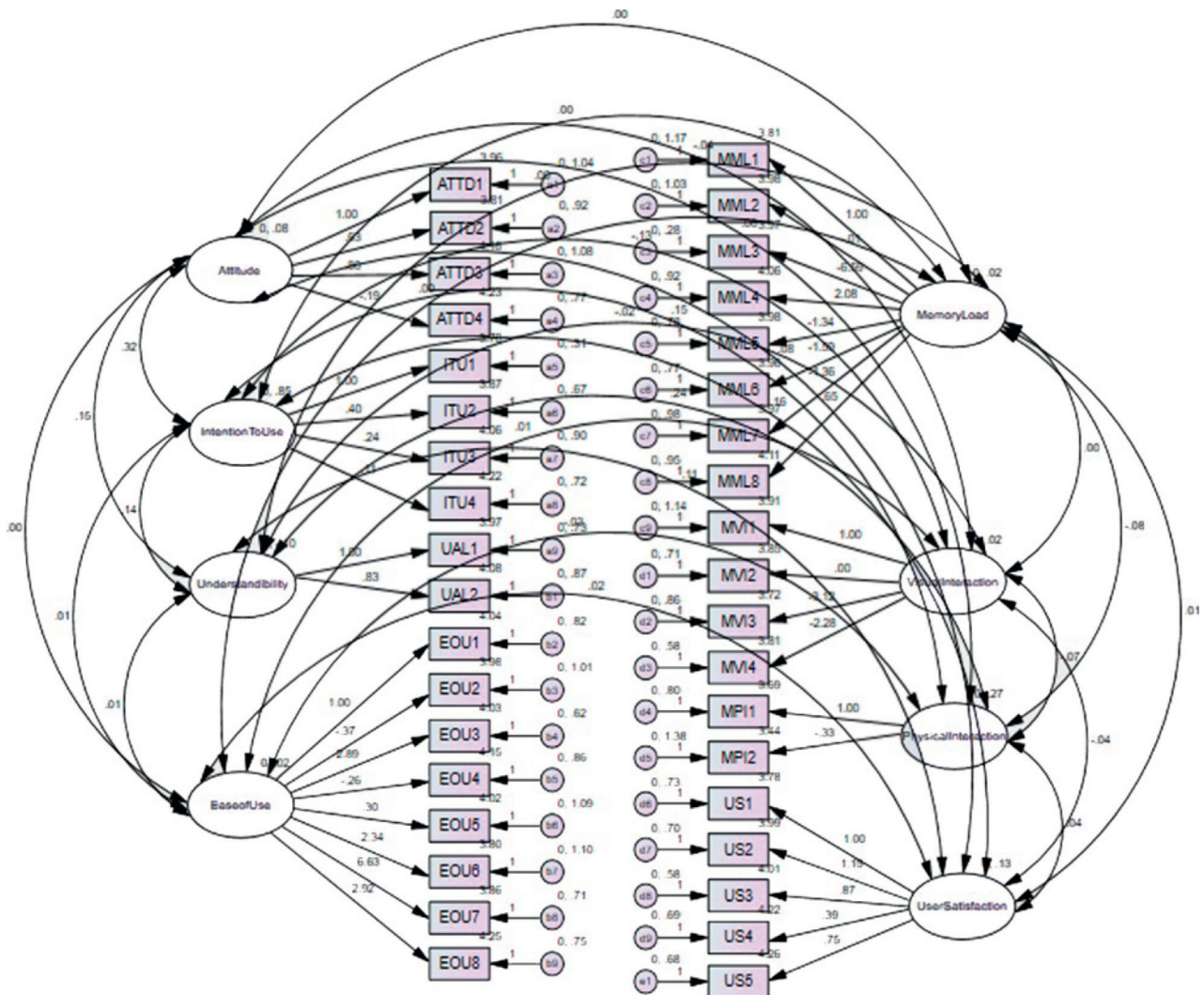


FIGURE 6: Structural model for investigating relationship between the latent variables.

TABLE 9: Model fit indices of the measurement and structural model.

Fit index	Structural model	Recommended values
CMIN/DF	1.227	$\leq 3.00$
CFI	0.543	$\leq 3.00$
NFI	0.85	$\geq 0.90$
IFI	0.825	$\geq 0.90$
TLI	0.5	$\geq 0.50$
PCFI	0.5	$\geq 0.50$
RFI	0.15	$\leq 1.00$
PNFI	0.542	$\geq 0.50$
RMSEA	0.05	$\leq 0.08$

TABLE 10: Model testing/hypothesis testing.

Hypothesis	Unstandardized coefficients	Standardized coefficients	Standard errors	<i>P</i>
<i>H1</i> →User satisfaction	0.151	0.441	0.485	0.018
<i>H2</i> →Attitude	0.795	0.206	0.366	0.030
<i>H3</i> →Minimal memory load	0.588	0.455	0.235	0.027
<i>H4</i> →Minimal visual interaction	0.811	0.375	0.255	0.023
<i>H5</i> →Minimal physical interaction	0.327	0.143	0.171	0.046

Similarly, understandability and learnability and intention to use the app have significant positive estimates, showing positive perceptions of the proposed solution.

**6.4. Analysis through AutoLog Dataset.** The AutoLog application is used for logging data about drivers' interactions with common smartphone applications [56]. The logged data contain information about different operations carried out by smartphone application such as number of activities used to perform tasks and number of input taps. The common smartphone applications include calls, SMS, e-mail, WhatsApp, Navigation, and Weather. As discussed earlier, the applications and their interfaces are designed from the perspective of a normal user as the number of activities is either redundant or repetitive and has a complex structure, long route to follow, and so forth. The logged data obtained from smartphone native interfaces are analyzed and compared with the data obtained from DriverSense for using common smartphone activities. The DriverSense interfaces are found to be less complex and have minimum activities and input taps. The comparison is shown in Table 11. To investigate the performance of DriverSense, the AutoLog data generated from the DriverSense app during the normal operations performed by the participants have been analyzed and compared with the AutoLog dataset generated from the smartphone native interfaces. After analyzing the data from both datasets, the findings, as shown in Figure 7, indicate that DriverSense requires comparatively less visual and physical attention to perform smartphone activities while driving as compared to native interfaces. It is due to the fact that DriverSense interfaces are simplified, adaptive, and consistent, having a minimum number of activities and input taps. Since most of the activities can be performed automatically based on a context, it will minimize the drivers' interactions. The

results obtained after the analysis are discussed in the following sections.

**6.4.1. Automatic Response.** Since the operations of DriverSense user interfaces change according to drivers' context, most of the activities are automatically performed. Analyzing the dataset, the activities automatically performed by the DriverSense are auto-reply, auto-skipping lengthy and unknown SMSs, auto-reply for unknown calls during high speed, and so forth. The operations automatically performed by DriverSense user interface are compared with smartphone native interfaces and other technologies (i.e., Android Auto, CarPlay, etc.) and results are shown in Table 12.

**6.4.2. Steering Wheel Control Variations.** The datasets also captured steering wheel control variation while driving. The control of the steering wheel has been analyzed while the driver performed smartphone activities in both smartphone native interfaces and DriverSense. Comparatively high steering wheel variations have been observed when drivers performed common activities such as SMS and phone calls using smartphone native interfaces. However, significantly minimum steering wheel variations have been observed when the drivers performed the same activities on DriverSense. A comparison of the steering wheel control variations while receiving voice call is depicted in Figure 8.

**6.4.3. Speed Variations.** The speed variations data are also captured while performing activities like attending the call, reading, and replying to text messaging using both smartphone native interfaces and DriverSense interfaces. The significant speed variations are observed when the drivers attended calls and read and replied to text messages on smartphone native interfaces. The speed is found to be degraded from approximately 80 km/h to 50 km/h. On the

TABLE 11: Comparison of smartphone native interfaces with DriverSense interfaces.

No.	Applications	Smartphone native interfaces			DriverSense interfaces		
		Activity	No. of input taps	Adaptation	Activity	No. of input taps	Adaptation
1	Call	13	43	No	7	15	Yes
2	SMS	11	35	No	6	13	Yes
3	E-mail	25	25	No	8	11	Yes
4	Navigation	14	26	No	6	11	Yes
5	Weather	5	14	No	3	8	Yes
6	Music	6	16	No	3	10	Yes

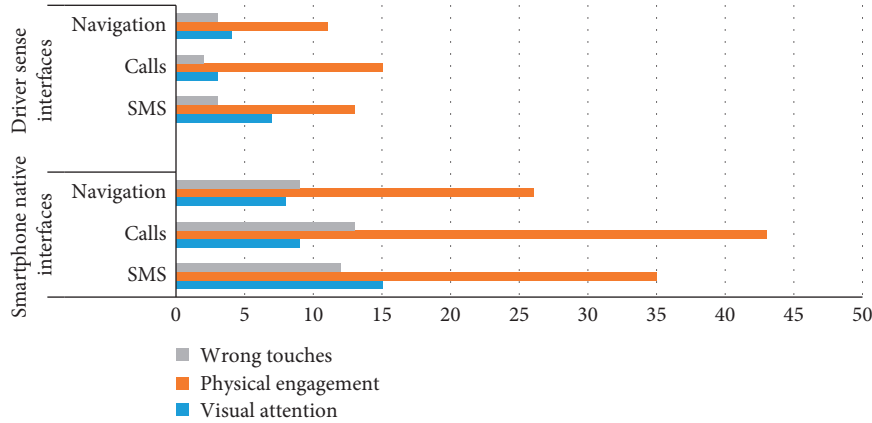


FIGURE 7: Comparison of DriverSense and smartphone native interfaces in terms of wrong touches, physical engagement, and visual attention when performing common smartphone activities while driving.

TABLE 12: Comparison of automatic responses generated by DriverSense and other technologies.

Responses	DriverSense	Smartphone native apps	Other technologies
Auto-cancel lengthy SMS	Yes	No	No
Auto-reply at higher speed	Yes	No	No
Auto-reply to unknown number calls	Yes	No	No
Auto-switching from text-to-speech if no noise detected	Yes	No	No
Auto-hide navigation activity for known routes	Yes	No	No
Auto-block videos	Yes	No	Yes
Interface auto-switched to night mode	Yes	No	Yes
Interface auto-switched to a driving mode	Yes	No	Yes

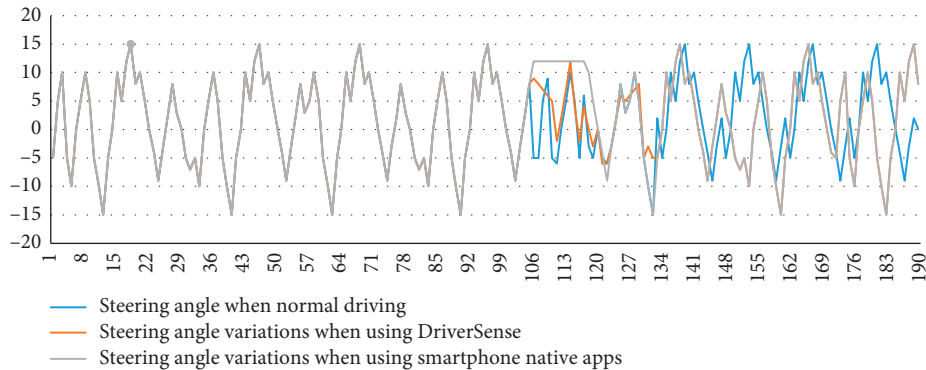


FIGURE 8: Comparison of steering wheel control variations during receiving voice calls.

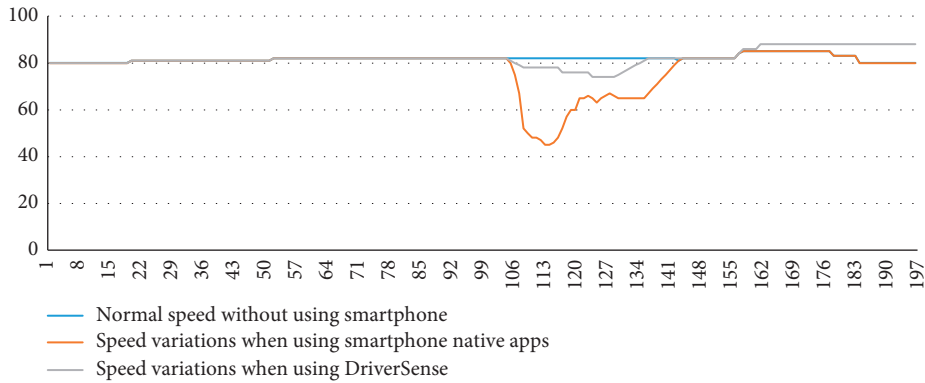


FIGURE 9: Comparison of speed variations due to phone calls and text messaging.

other hand, the data extracted from DriverSense dataset have shown less speed variations as compared to smartphone native interfaces. A comparison of the speed variations is depicted in Figure 9.

## 7. Conclusions

The usage of a smartphone is a global phenomenon and has been acknowledged as a major source of accidents and crashes. Using a smartphone while driving requires much visual interaction, physical interaction, and mental workload, which cannot be afforded by the drivers as eyes off the road for two seconds increases the chances of accidents to twenty-four times. The researchers have tried to minimize the visual, physical, and mental distractions of the drivers with the help of supportive technologies. However, the available solutions are not designed with the assumption that the drivers have certain limitations such as physical limitations, visual limitations, and cognitive limitations. These limitations can vary due to different driving contexts.

In this research paper, we have designed and developed a context-aware adaptive user interfaces framework named DriverSense for the drivers to minimize distractions and subsequent catastrophes. The proposed framework uses contextual and models information to minimize the drivers' distractions by providing an adaptive, semantically consistent, simplified, context-sensitive, and task-oriented user interface design. The efficiency of the proposed solution with respect to the adaptive user interface is considered to be significant and acceptable in terms of usability and user satisfaction. The users' experiences after using DriverSense are measured through questionnaire and evaluated in different dimensions such as driver attitude for DriverSense usage, intention to use the app, perceived usefulness, understandability and learnability, operability, ease of use, system usability scale, minimal memory load, minimal physical and visual interaction, and user satisfaction. The results have indicated that DriverSense has significantly reduced the drivers' distractions caused by cognitive overload, visual interactions, and physical interactions. Furthermore, the results have also shown that DriverSense is more robust, adaptable, and easy to use as compared to the other infotainment solutions.

## Data Availability

The data that support the findings of this study are available upon request from the first author, Mr. Inayat Khan (inayat\_khan@uop.edu.pk).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] WHO, *Mobile Phone Use: A Growing Problem of Driver Distraction*, WHO, Geneva, Switzerland, 2011.
- [2] K. Young, M. Regan, and M. Hammer, "Driver distraction: a review of the literature," *Distracted Driving*, vol. 2007, pp. 379–405, 2007.
- [3] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin, "Sensing vehicle dynamics for determining driver phone use," in *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 41–54, Taipei, Taiwan, 2013.
- [4] K. Z. Gajos, J. J. Long, and D. S. Weld, "Automatically generating custom user interfaces for users with physical disabilities," in *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility-Assets'06*, pp. 243–244, Portland, OR, USA, October 2006.
- [5] A. Khan, S. Khusro, and I. Alam, "Blindsense: an accessibility-inclusive universal user interface for blind people," *Engineering, Technology & Applied Science Research*, vol. 8, pp. 2775–2784, 2018.
- [6] J. K. Caird, K. A. Johnston, C. R. Willness, M. Asbridge, and P. Steel, "A meta-analysis of the effects of texting on driving," *Accident Analysis & Prevention*, vol. 71, pp. 311–318, 2014.
- [7] G. M. Fitch, S. A. Soccolich, F. Guo et al., *The Impact of Hand-Held and Hands-free Cell Phone Use on Driving Performance and Safety-Critical Event Risk*, National Highway Traffic Safety Administration, Washington, DC, USA, 2013.
- [8] R. T. Frankle, "Nutrition education in the medical school curriculum: a proposal for action: a curriculum design," *The American Journal of Clinical Nutrition*, vol. 29, no. 1, pp. 105–109, 1976.
- [9] P. A. Akiki, A. K. Bandara, and Y. Yu, "Adaptive model-driven user interface development systems," *ACM Computing Surveys*, vol. 47, no. 1, pp. 1–33, 2015.
- [10] I. A. Doush, I. Damaj, M. A. Al-Betar et al., "A survey on accessible context-aware systems," in *Technological Trends in*

- Improved Mobility of the Visually Impaired*, pp. 29–63, Springer, Berlin, Germany, 2020.
- [11] D. Weld, C. Anderson, P. Domingos et al., “Automatically personalizing user interfaces,” in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, Aca-pulco, MX, USA, 2003.
- [12] H. Guo, D. Cao, H. Chen, C. Lv, H. Wang, and S. Yang, “Vehicle dynamic state estimation: state of the art schemes and perspectives,” *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 2, pp. 418–431, 2018.
- [13] R. Tian, K. Ruan, L. Li, J. Le, J. Greenberg, and S. Barbat, “Standardized evaluation of camera-based driver state monitoring systems,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 716–732, 2019.
- [14] O. Oviedo-Trespalacios, M. King, A. Vaezipour, and V. Truelove, “Can our phones keep us safe? A content analysis of smartphone applications to prevent mobile phone distracted driving,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 60, pp. 657–668, 2019.
- [15] G. Albert, O. Musicant, I. Oppenheim, and T. Lotan, “Which smartphone’s apps may contribute to road safety? An AHP model to evaluate experts’ opinions,” *Transport Policy*, vol. 50, pp. 54–62, 2016.
- [16] S. Siuhi and J. Mwakalonge, “Opportunities and challenges of smart mobile applications in transportation,” *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 6, pp. 582–592, 2016.
- [17] National Highway Traffic Safety Administration, *Visual-manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Devices*, National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT), Washington, DC, USA, 2016.
- [18] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Washington, “Understanding the impacts of mobile phone distraction on driving performance: a systematic review,” *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 360–380, 2016.
- [19] A. Bianchi and J. G. Phillips, “Psychological predictors of problem mobile phone use,” *CyberPsychology & Behavior*, vol. 8, no. 1, pp. 39–51, 2005.
- [20] C. Delgado, S. Batista, M. Canales, J. R. Gállego, J. Ortín, and M. Cesana, “An implementation for dynamic application allocation in shared sensor networks,” in *Proceedings of the 2018 11th IFIP Wireless and Mobile Networking Conference (WMNC)*, pp. 1–8, Prague, Czech Republic, 2018.
- [21] M. K. Delgado, C. C. McDonald, F. K. Winston et al., “Attitudes on technological, social, and behavioral economic strategies to reduce cellphone use among teens while driving,” *Traffic Injury Prevention*, vol. 19, no. 6, pp. 569–576, 2018.
- [22] G. Ponte, M. Baldock, and J. Thompson, *Examination of the Effectiveness and Acceptability of Mobile Phone Blocking Technology Among Drivers of Corporate Fleet Vehicles*, The National Academies of Sciences, Engineering, and Medicine, Washington, DC, USA, 2016.
- [23] D. P. Chiang, A. M. Brooks, and D. H. Weir, “Comparison of visual-manual and voice interaction with contemporary navigation system HMIs,” *SAE Transactions*, pp. 436–443, 2005.
- [24] J. Shutko, K. Mayer, E. Laansoo, and L. Tijerina, *Driver Workload Effects of Cell Phone, Music Player, and Text Messaging Tasks with the Ford SYNC Voice Interface Versus Handheld Visual-Manual Interfaces*, SAE International, Harrisburg, PA, USA, SAE Technical Paper 0148-7191, 2009.
- [25] B. Reimer, B. Mehler, J. Dobres, and J. Coughlin, *The Effects of a Production Level “Voice-command” Interface on Driver Behavior: Summary Findings on Reported Workload, Physiology, Visual Attention, and Driving Performance*, MIT AgeLab, Cambridge, MA, USA, 2013.
- [26] V. E.-W. Lo and P. A. Green, “Development and evaluation of automotive speech interfaces: useful information from the human factors and the related literature,” *International Journal of Vehicular Technology*, vol. 2013, Article ID 924170, 13 pages, 2013.
- [27] J. M. Cooper, H. Ingebreetsen, and D. L. Strayer, “Mental workload of common voice-based vehicle interactions across six different vehicle systems,” Technical Report, AAA Foundation for Traffic Safety, Washington, DC, USA, 2014.
- [28] Y. Fukatsu, B. Shizuki, and J. Tanaka, “No-look flick: single-handed and eyes-free Japanese text input system on touch screens of mobile devices,” in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 161–170, Munich, Germany, August 2013.
- [29] T. C. Lansdown and A. N. Stephens, “Couples, contentious conversations, mobile telephone use and driving,” *Accident Analysis & Prevention*, vol. 50, pp. 416–422, 2013.
- [30] N. E. Boudette, *Biggest Spike in Traffic Deaths in 50 Years? Blame Apps*, New York Times, New York, NY, USA, 2016.
- [31] N. S. Council, “Understanding the distracted brain: why driving while using hands-free cell phones is risky behavior (White Paper),” 2012, <https://www.nsc.org/Portals/0/Documents/DistractedDrivingDocuments/Cognitive-Distracted-White-Paper.pdf>.
- [32] P. Tchankue, J. Wesson, and D. Vogts, “Are mobile in-car communication systems feasible?: a usability study,” in *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, pp. 262–269, Pretoria, South Africa, October 2012.
- [33] E. Alepis and C. Patsakis, “Monkey says, monkey does: security and privacy on voice assistants,” *IEEE Access*, vol. 5, pp. 17841–17851, 2017.
- [34] J. Hindy: Best driving apps, 2018.
- [35] B. Adipat and D. Zhang, “Interface design for mobile applications,” in *Proceedings of the AMCIS 2005*, p. 494, Omaha, NE, USA, August 2005.
- [36] P. A. Akiki, A. K. Bandara, and Y. Yu, “Adaptive model-driven user interface development systems,” *ACM Computing Surveys (CSUR)*, vol. 47, no. 1, pp. 1–33, 2014.
- [37] M. W. Iqbal, N. Ahmad, S. K. Shahzad, I. Feroz, and N. A. Mian, “Towards adaptive user interfaces for mobile-phone in smart world,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 11, 2018.
- [38] Q. Limbourg, J. Vanderdonckt, B. Michotte, L. Bouillon, and V. López-Jaquero, “USIXML: a language supporting multi-path development of user interfaces,” in *Proceedings of the IFIP International Conference on Engineering for Human-Computer Interaction*, pp. 200–220, Hamburg, Germany, July 2004.
- [39] B. Gamecho, R. Minón, A. Aizpurua et al., “Automatic generation of tailored accessible user interfaces for ubiquitous services,” *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 612–623, 2015.
- [40] F. Paterno, C. Santoro, and L. D. Spano, “MARIA: a universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 16, p. 19, 2009.

- [41] O. Corcho, A. Gómez-Pérez, A. López-Cima, V. López-García, and M. C. O. Suárez-Figueroa, *Automatic Generation of Knowledge Portals for Intranets and Extranets. LNCS 2870*, Springer-Verlag, Berlin, Germany, 2003.
- [42] M. Peissner, D. Häbe, D. Janssen, and T. Sellner, “MyUI: generating accessible user interfaces from multimodal design patterns,” in *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pp. 81–90, Copenhagen, Denmark, June 2012.
- [43] F. Paterno, C. Santoro, J. Mantyjarvi, G. Mori, and S. Sansone, “Authoring pervasive multimodal user interfaces,” *International Journal of Web Engineering and Technology*, vol. 4, no. 2, pp. 235–261, 2008.
- [44] R. Hervás and J. Bravo, “Towards the ubiquitous visualization: adaptive user-interfaces based on the Semantic Web,” *Interacting with Computers*, vol. 23, no. 1, pp. 40–56, 2011.
- [45] M. Hou, H. Zhu, M. Zhou, and G. R. Arrabito, “Optimizing operator-agent interaction in intelligent adaptive interface design: a conceptual framework,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, pp. 161–178, 2010.
- [46] P. Tchankue, J. Wesson, and D. Vogts, “The impact of an adaptive user interface on reducing driver distraction,” in *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 87–94, Salzburg, Austria, November 2011.
- [47] P. Tchankue, J. Wesson, and D. Vogts, “Designing a mobile, context-aware in-car communication system,” in *Proceedings of SATNAC 2012 on Limited Range Communication*, pp. 1–6, Fancourt in George, Western Cape, South Africa, 2012.
- [48] I. Khan, M. A. Khan, S. Khusro, and M. Naeem, “Vehicular lifelogging: issues, challenges, and research opportunities,” *Journal of Information Communication Technologies and Robotics Applications*, vol. 8, 2017.
- [49] S. Ali, S. Khusro, I. Ullah, A. Khan, and I. Khan, “Smartontosensor: ontology for semantic interpretation of smartphone sensors data for context-aware applications,” *Journal of Sensors*, vol. 2017, Article ID 8790198, 26 pages, 2017.
- [50] S. Bongartz, Y. Jin, F. Paternò, J. Rett, C. Santoro, and L. D. Spano, “Adaptive user interfaces for smart environments with the support of model-based languages,” in *Proceedings of the International Joint Conference on Ambient Intelligence*, pp. 33–48, Pisa, Italy, November 2012.
- [51] J. Hussain, A. Ul Hassan, H. S. Muhammad Bilal et al., “Model-based adaptive user interface based on context and user experience evaluation,” *Journal on Multimodal User Interfaces*, vol. 12, no. 1, pp. 1–16, 2018.
- [52] C. Rohrer, “When to use which user experience research methods,” *Jakob Nielsen’s Alertbox*, 2008.
- [53] H. Petrie and N. Bevan, “The evaluation of accessibility, usability, and user experience,” in *The Universal Access Handbook*, pp. 1–16, CRC Press, Boca Raton, FL, USA, 2009.
- [54] K. L. Norman, B. Shneiderman, B. Harper, and L. Slaughter, *Questionnaire for User Interaction Satisfaction*, University of Maryland, College Park, MD, USA, 1989.
- [55] B. Laugwitz, T. Held, and M. Schrepp, “Construction and evaluation of a user experience questionnaire,” in *Proceedings of the Symposium of the Austrian HCI and Usability Engineering Group*, pp. 63–76, Graz, Austria, November 2008.
- [56] I. Khan, S. Khusro, and I. Alam, “Smartphone distractions and its effect on driving performance using vehicular lifelog dataset,” in *Proceedings of the 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–6, Istanbul, Turkey, 2019.
- [57] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
- [58] R. Likert, “A technique for the measurement of attitudes,” Columbia University, Ph. D. Dissertation, New York, NY, USA, 1932.
- [59] J. M. Cortina, “What is coefficient alpha? An examination of theory and applications,” *Journal of Applied Psychology*, vol. 78, no. 1, pp. 98–104, 1993.
- [60] R. E. Zinbarg, W. Revelle, I. Yovel, and W. Li, “Cronbach’s  $\alpha$ , Revelle’s  $\beta$ , and McDonald’s  $\omega$ H: their relations with each other and two alternative conceptualizations of reliability,” *Psychometrika*, vol. 70, no. 1, pp. 123–133, 2005.



## Research Article

# Access and Use of Mobile Phone in Daily Life Activities by Rural Women of Gilgit-Baltistan, Pakistan

**Sabit Rahim** , **Sadrudin Bahadur Qutoshi**, **Syeda Abida**, **Faqeer Muhammad**,  
and **Imtiaz Hussain**

*Karakoram International University, Gilgit-Baltistan, Pakistan*

Correspondence should be addressed to Sabit Rahim; [sabit.rahim@kiu.edu.pk](mailto:sabit.rahim@kiu.edu.pk)

Received 13 March 2020; Revised 19 May 2020; Accepted 25 May 2020; Published 13 June 2020

Academic Editor: Fawad Zaman

Copyright © 2020 Sabit Rahim 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.

This study aims to investigate the access to a mobile phone, usage pattern, and its impact on mountainous rural women of two districts, i.e., Hunza and Nagar districts of Gilgit-Baltistan, Pakistan. To attain the objective of the research, the researchers have employed various statistic methods, and data were collected through a questionnaire from 190 respondents in the study area (200 respondents were selected). Initially, Kaiser–Meyer–Olkin (KMO) and Bartlett’s tests were used for sampling adequacy, and factor analysis technique was used to explain correlations among multiple outcomes. The results revealed that 80% of women in the Hunza and Nagar districts own mobile phones (access) and 63% have good skills of mobile phone usage for a variety of purposes. Moreover, 56.4% of women use mobile phones in their daily life activities; however, 23.6% disagreed with the statements. On the contrary, the results show that 71.8% women use mobile phones for security purposes. Therefore, the study recommends that effective use of the mobile phone in daily activities of mountainous rural women can be one of the effective strategies to boost their confidence level and feeling of security. Finally, socioeconomic development of the area is possible by providing technical skills related to mobile phone business to the unskilled women of the two districts.

## 1. Introduction

Information and communication technology (ICT) in today’s world comprise applications and communication tools, such as social media [1], digital information repositories either online or offline, and digital photography and video, and among them, the mobile phone is becoming the most prominent tools of technology [2–5]. As the mobile phone is refashioning the lives of its users regardless of their socioeconomical and gender differences, the rural mountainous women cannot be excluded [5, 6]. The mobile phone, as one of the mostly used tools of ICT, has been recognized worldwide as an effective and efficient tool used for multiple purposes [6, 7]. This technological tool has a positive impact on the lives of its users especially people living in rural [7] and deprived areas like mountainous regions by connecting them with the information society [8]. Thus, it has revolutionized the life of its users in the field of development.

On the contrary, according to the GSMA, two billion women in developing countries do not own mobile phones. Likewise, women are 26% less likely than men to own a mobile phone. The situation is even worst in South Asia, where women on average are 38% less likely to own a mobile phone than men [9]. According to the World Bank data, Pakistan falls under the lower-middle-income group with a population of 212 million of which 60% live in rural areas (Pakistan Census Report, 2017). However, in 2018, it was reported that the percentage of fixed-line phone subscribers was 1.3%, while the mobile phone subscribers were 74% [10–12].

Thus, in Pakistan the access to mobile phone services in its primitive phase and limited to provide only voice communication facilities within cities. However, over time, this technology has revolutionized the lives of people [13] in one way or the other. When we look back in history, the mobile phone culture came to Pakistan in 1994 with the introduction of some cellular networks. Now, out of two, the

second person owns a mobile phone in the world, and one-third of the population in Pakistan possesses a mobile phone [14]. The recent forms of mobile phone services enabled mankind to enjoy a diverse range of facilities such as text messages, multimedia messages, online payment, banking, online shopping, web surfing, financial transactions, and the diverse kind of social engagements [15, 16].

These opportunities opened new doors to both males and females to interact with each other and engage in socio-economic activities such as how to help rural women to improve their socialization and economic development [8]. Studies revealed that these technological tools have enabled rural women to use technology in business activities [17]. At the same time, access and usage have increased very rapidly in developed as well as in the developing countries for the last fifteen years. This technology has brought huge changes in the lives of people, more quickly than any previous technology [18] along with several challenges as well.

People in developing countries, including Pakistan, face several developmental challenges due to economic and cultural barriers and rural and urban gaps [19]. Likewise, low literacy, poor healthcare facilities, low per capita income, a high degree of poverty, and poor infrastructure are common in most developing countries including Pakistan [16]. In the rural and mountainous areas, the mobile phone has emerged as an important development tool. For instance, people access health facilities, economic development (publicize their products), and transfer amount using mobile banking [20]. It is seen as a device that has the potential to break the rural-urban digital divide and the developmental gap by delivering information on a variety of economic and social issues. In Hunza and Nagar districts, the landline phone is not used to the extent that the mobile phone has revolutionized with extraordinary growth as one of the most accessible and affordable ICT tools. Today, all over the world, its expansion in a very short period provided multiple opportunities for its users to access each other [16]. However, access to mobile phones and its impact on rural mountainous people, especially in the context of Hunza and Nagar, has not been substantiated empirically.

Therefore, this paper aims to explore the socioeconomic impact of mobile phones on rural women of the two districts, i.e., Hunza and Nagar. This region has geopolitical and strategic importance due to the China-Pakistan Economic Corridor (CPEC) which connects China and Pakistan. To this end, how the women population of this region has been exposed to this technological tool plays a vital role to predict the future development associated with the CPEC. This study provides an opportunity and policy recommendations to stakeholders on the role of mobile phone technology in women's development in mountainous areas.

*1.1. Objectives of the Study.* The objectives of this study are as follows:

- (1) To explore the access to the mobile phone by rural women across the two districts of the Gilgit-Baltistan province [1, 21, 22]

- (2) To examine the differences in the usage pattern of the mobile phone by rural mountainous women of the two districts of the Gilgit-Baltistan province [4, 23, 24]

## 2. Review of Relevant Literature

*2.1. Access to Mobile Phone and Preferences.* The study of the UN's International Telecommunication Union (ITU) reports that men users of mobile phones are greater than the women in most countries including Pakistan [12, 25]. Most of the women in developing countries do not have mobile phones because of cultural restrictions and a male-dominated society [26]. However, Pakistan has the highest gender gap of mobile ownership with women 37% less likely than men to own a mobile phone and Bangladesh was the second highest at 33% [9].

However, the mobile phone provides greater facilities in life such as connecting with friends and family, taking selfies and photos, getting financial independence [17], seeking employment opportunities, finding family health and educational venues [27], accessing the Internet and government services, feeling secure, and accessing information [28, 29]. As access to mobile phones and the Internet grows, ICTs are playing an ever-stronger role in efforts aimed at improving the lives of people, especially women [13, 30]. There are some psychological and other barriers such as gender discrimination [31], lack of confidence [32], language difficulties, low literacy [33], and lack of time and money which prevent women and young girls from taking full advantages of the technology [34].

Thus, along with these barriers, technology adoption in rural Pakistan appears to be a challenging situation to promote ICT skills and women's access to this technology in rural areas [35]. However, some steps need to be taken towards ICT adaptation such as educating women on ICTs, recruitment drive (equal opportunity for males and females), mentoring schemes, and equal opportunities to business (pay, respect, leadership, no discrimination, and flexible work schedules) [8]. In doing so, women can improve their access to ICTs and use these tools to move towards reaching their full potential [36]. The limited access to technology for the women of rural areas of Pakistan, for many reasons as mentioned above, appears to be different in the case of [37] Hunza and Nagar districts.

*2.2. Mobile Phone Supports Rural Women in Their Daily Activities.* Studies support the view that the mobile phone empowers women to participate in socioeconomic services [19], report familial violence, consult family planning agencies [38], access education, health care, and financial services [39], and enable them to develop their value of life and that of their families [9]. Unlike other ICT devices, the mobile phone does not require literacy or sophisticated skills. That may be one of the reasons why women can easily use a mobile phone to manage, uphold, change, and contribute to the affairs of the socioeconomic life [40]. In every society, the mobile phone has introduced new ways for

economical activities [15, 20, 34] due to its easiest and quickest flow of information. For instance, it offers quick access to educational, cultural, social, sports, and financial events all over the world at any time in any place [41, 42].

In some of the developing countries, women face problems to own mobile phones for many reasons including cultural, gender, and attitudinal factors [39, 43]. However, in the case of Hunza and Nagar districts, the women use a mobile phone for accessing information and sharing ideas, for business purposes and other socioeconomic activities that are different as compared to other parts of the developing world [44].

The literature argues that effective use of mobile phones can change the lives of the people in general and rural mountainous women in particular in the areas of business (MMS, 3G, 4G, and portability) and other activities [16, 45]. Thus, in the context of Hunza and Nagar districts, women's use of mobile phones for business activities cannot be ignored [46, 47]. However, there are many social benefits of mobile phones such as peace of mind, protection, increased personal security, better social life, the aftereffect of a calamity, and productive tool at work [48]. However, there are many challenges along with the new opportunities coming at the doors of these rural mountainous women. CPEC is one of those opportunities which can also bring many challenges as well, if not well prepared to face those challenges and convert them into meaningful life-enhancing opportunities.

*2.3. Gender and Age-Wise Differences of Mobile Phone Usage for Safety and Security Purpose.* Mobile phone is an empowering technological tool that offers multiple opportunities to both males and females equally, depending upon the level of individuals' exposure, sociocultural, and other barriers on the way to access this technology. For example, rural mountainous women, within their own limits, can use it for doing small businesses, to keep themselves secure from possible risks, to create a learning environment, and to eliminate gender income variation gaps [49]. It is believed that the use of a mobile phone can play a vital role in education, safety, and security [23]. Thus, the accessibility of a mobile phone, especially by young women (e.g., university-level girls and women doing jobs), is now becoming a priority matter to achieve their objectives [50] such as connecting, accessing knowledge [7], doing business, and improving safety and security matters. [6] Moreover, the advancement in the mobile phone technology brought challenges and difficulties for the elder users as compared to younger users [51]. In the same line, the authors in [5] listed the mobile phone users in groups and studied their behaviors towards mobile phones such as cost consciousness, safety conscious, mobile-dependent users, sophisticated users, and practical uses.

Both male and female students of GB are taking advantage of the use of mobile phones like mutual activities through SMS, listening lectures, sharing information, and educational applications (like a mobile dictionary) [52]. Students easily share thoughts and information with each other [53, 54]. Students use mobile phone devices to upload

and post their academic data on course websites, and each student gets access to their course [21, 22]. Many boys and girls use a mobile phone as a symbol of identity that help them to be independent in their interactions within and outside the family and to reinforce their identity [55].

#### 2.4. Research Questions

- (1) Do the women of Hunza and Nagar districts have access to the mobile phone and have enough skills to use it?
- (2) What are the different usage patterns of the mobile phone adopted by the women of Hunza and Nagar?

### 3. Methodology

*3.1. Theoretical Framework.* The following theoretical framework is the concept map that provides a guideline for researchers to follow accordingly. Based on the previous literature, we developed a theoretical framework as shown in Figure 1. Four major categories were identified such as mobile phone (MP) access, skills to use MP, use in daily life activities, and use for safety and security purposes due to prevailing security threats in Pakistan.

*3.2. Sample and Procedure.* A survey was carried out to collect the data from female research participants from Hunza and Nagar districts of GB, Pakistan. For this purpose, the researchers developed a questionnaire consisting of 31 items with two major sections. Further, Section A consisted of two subsections: the first subsection contains demographic data (age, the region of origin, and qualification) and the second subsection contains the access to the mobile phone, types of mobile phones, mobile phone brand, and service provider. Before the distribution of questionnaires to the participants, 5 students and 2 faculty members from the university (i.e., the research site) were asked to review the questionnaires. The purpose of this pilot testing was to find the level of difficulty in understanding the questions by the respondents. As a result of this testing, few items were rephrased for a better understanding of questionnaires. A reliability test was performed, and Cronbach's alpha value is 0.972 which is an acceptable level of internal consistency.

200 female participants were selected using a random sampling procedure from the two districts of GB and 190 responded. All participants were asked to fill the consent forms that helped the researchers to use these consent forms as alternatives to research approval from the Ethics Committee of the university. Because there does not exist such a kind of body in the university (Table 1 shows the result).

In the second subsection of A, the participants were also asked questions regarding access to the mobile phone, kinds of mobile phone (simple mobile or smartphone), mobile phone brand, and service providers they use (results are shown in Table 2).

Section B: the section B has been divided into three parts such as skills to use the mobile phone, using the mobile phone in daily life activities, and using the mobile phone for

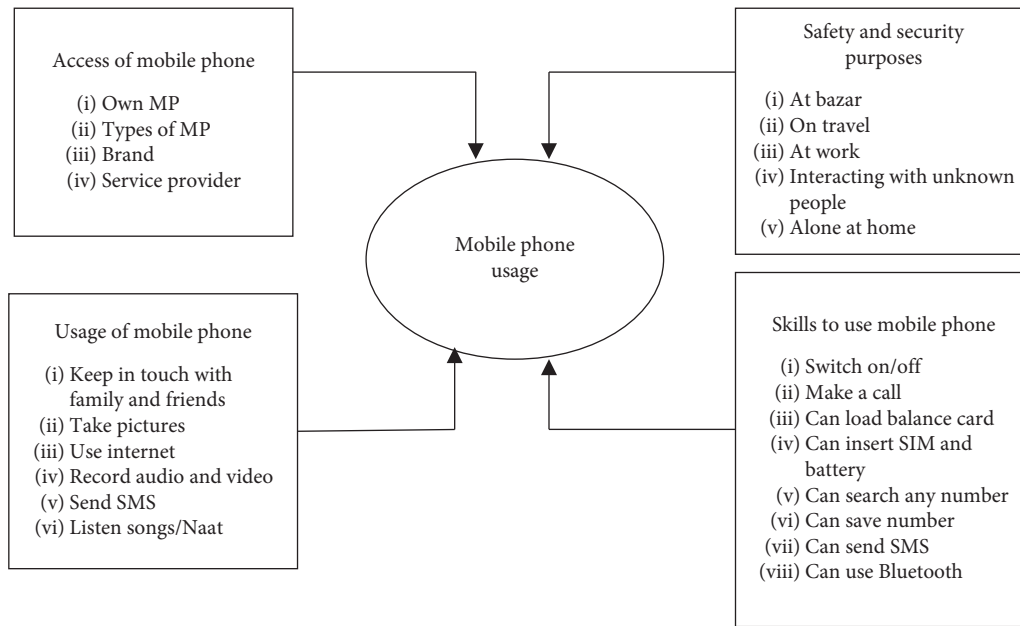


FIGURE 1: Theoretical framework [1-3].

TABLE 1: Respondents' profile.

Age of respondents	Hunza		Nagar	
	N	%	N	%
20 to 25	13	7	17	9
26 to 30	34	18	22	12
31 to 35	31	16	40	21
36 to 40	17	9	16	8
Qualification of respondents				
No education	54	28	55	29
Below middle	4	2	3	2
Below 10	9	5	5	3
Below 12	13	7	18	9
Bachelors	10	5	12	6
Masters and above	5	3	2	1

safety and security purposes by the women of Hunza and Nagar (shown in Table 3).

This part indicates that how the women of Hunza and Nagar use the mobile phone in their daily life activities. A 5-point Likert scale ranging between 1 = strongly disagree (SDA), 2 = disagree (DA), 3 = somehow agree (SA), 4 = agree (A), and 5 = strongly agree (SA) were used to indicate the respondents agreement on the statements (results are shown in Table 4).

In Section B, the respondents were asked 5-point Likert scale questions regarding the use of the mobile phone for safety and security purposes (results are shown in Table 5).

#### 4. Results

The Statistical Package of the Social Sciences (SPSS 23.0) was used to analyze the data. For sections A and B initially, descriptive statistics were used to analyze the data to find a percentage. Similarly, Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were used for sampling adequacy, and factor analysis technique was used to explain correlations among multiple

TABLE 2: Percentage of access of the mobile phone by the women of Hunza and Nagar.

Questions (1-3)	Percentages	
	Hunza	Nagar
Do you have mobile phone?		
No	10.5	9.5
Yes	39.5	40.5
What kind of mobile phone do you have?		
Simple	31.6	31.1
Smart	7.9	9.5
How many smartphones do you have in your family?		
Only one	16.1	12.5
2 to 4	13	8.9
5 to 7	1	0
Questions (3-4)		
Mobile phone brand		
Nokia	3.2	7.4
Samsung	0.5	2.1
Qmobile	19.5	15.8
BlackBerry	0	1.6
1 and 3	11.6	10.5
2 and 3	3.2	1.6
1 and 2	1.6	1.6
Which is your service provider?		
SCOM	2.6	2.6
Telenor	9.5	14.7
Zong	15.3	10.5
More than 1	12.1	12.6

outcomes such as mobile phone access, use in daily life activities, skills to use, and use for safety and security purposes.

**4.1. Demographic.** In this research study, 200 female participants were selected by employing a random sampling procedure from the two districts of GB and 190 responded.

TABLE 3: Percentage of mobile phone usage skills of the women of Hunza and Nagar.

Questions (1-5)	Percentages	
	Hunza	Nagar
Can switch on/off mobile phone		
No mobile	10.5	9.5
Yes	39.5	40.5
Make a call without help		
No	6.3	1.6
Yes	33.2	38.9
Can load balance card		
No	16.3	11.1
Yes	23.2	29.5
Can insert and change SIM card		
No	7.9	2.6
Yes	31.6	37.9
Can charge my mobile phone		
No	1.1	0.5
Yes	38.4	40
Questions (6-10)		
Can insert mobile battery		
No	5.8	3.2
Yes	33.7	37.4
Can enter and save number of any person to my contact list		
No	21.1	13.7
Yes	18.4	26.8
Can send SMS		
No	23.7	21.1
Yes	15.8	19.5
Can use Bluetooth or infrared to transfer pictures or anything and can use to increase my vocabulary		
No	24.7	25.3
Yes	14.7	15.3
Can search number of specific persons by name		
No	11.1	3.2
Yes	28.4	37.4

The average age of participants was 25 years which ranged from 20 to 40 years. 57% women have no education from Hunza and Nagar such as 20 to 25 (only 1.1% out of 16%); 26 to 30 (12.1% out of 29%); 31 to 35 (31.1% out of 37%), and 36 to 40 (13.1% out of 17%) (as shown in Figure 2).

4.2. *Access to Mobile Phone.* This section addresses the first part of the research question one. The participants were asked questions regarding access to the mobile phone, and 80% of women said that they own mobile phones and 20% responded that they do not have a mobile phone. The participants responded on the kinds of mobile phones such as 62.7% said that they have a simple mobile phone and 17.4% said that they have a smartphone. On the question of which is the service provider, 5.2% said that they use the SCOM service, 24.2% said that they use Telenor, 25.8% said that they use Zong, and 24.7% said that they use more than one service provider (detail is shown in Figure 3).

4.3. *Skills to Use a Mobile Phone.* This section addresses the second part of the research question one. The women of Hunza and Nagar districts were asked questions regarding skills to use a mobile phone. The result showed that 80% of

TABLE 4: Use of mobile phone in daily life activities.

Questions (1-3)	Scale	Region			
		Hunza		Nagar	
		N	%	N	%
Keep in touch with family and friends	SA	0	0	1	1
	A	19	10	22	12
	SA	56	29	54	28
Help you to take picture	SDA	1	1	0	0
	DA	15	8	16	8
	SA	18	9	20	11
Help you to send SMS	A	26	14	28	15
	SA	15	8	13	7
	SDA	2	1	4	2
	DA	26	14	19	10
Questions (4-6)	SA	14	7	20	11
	A	20	11	23	12
	SA	13	7	11	6
Record audio/video	SDA	3	2	6	3
	DA	29	15	29	15
	SA	21	11	17	9
Mobile phone has given quick access to Internet use	A	17	9	19	10
	SA	5	3	6	3
	SDA	18	9	30	16
	DA	30	16	23	12
Help to listen Naat/Ginans/songs	SA	5	3	7	4
	A	15	8	10	5
	SA	7	4	7	4
	SDA	2	1	2	1
Help to listen Naat/Ginans/songs	DA	6	3	8	4
	SA	23	12	22	12
	A	33	17	31	16
	SA	11	6	14	7

women have skills to switch on and off the mobile phone; can make a call without help (72%); can load the balance card (52%); can insert and change the SIM card (69.5%); can charge the mobile phone (78.4%); can insert the mobile battery (71%); can enter and save contact numbers (45.5%); can send SMS (35.3%); can use Bluetooth or infrared to transfer pictures or anything (30%), and can search a number of a specific person by name (65.8%).

4.4. *Mobile Phone Usage Pattern.* The research question two has been addressed in this section. The data were analyzed using the KMO and Bartlett's tests for sampling adequacy and factor analysis technique to explain correlations among multiple outcomes. The estimated results of KMO and Bartlett's response to mobile phone access (MPA), use in daily life activities, skills to use, and use for safety and security purposes are presented in Table 6.

It is figured out that all the values of KMO are greater than 0.5 or all the values are greater than 0.8. These results indicate that samples are adequate and applicable to factor analysis. Also, the estimated probability values of Bartlett's test are presented in the last column of Table 6. All the *p* values are less than 0.05 in all the aspects of mobile phone usage. In brief, both KMO and Bartlett's tests suggested that results are statistically significant and reliable for factor analysis.

TABLE 5: Use of the mobile phone for security purposes.

Questions (1-3)	Region					
	Hunza		Nagar			
	N	%	N		%	
Mobile phone has increased the sense of protection while going to bazar						
DA	8	4	0		0	
SA	10	5	1		1	
A	29	15	31		16	
SA	28	15	45		24	
Mobile phone has increased the sense of protection while travelling in public transport						
DA	8	4	1		1	
SA	8	4	4		2	
A	40	21	37		19	
SA	19	10	35		18	
Mobile phone has increased the sense of protection while going for work						
SDA	1	1	0		0	
DA	9	5	2		1	
SA	23	12	15		8	
A	28	15	37		19	
SA	14	7	23		12	
Questions (4-5)						
Mobile phone has increased the sense of protection while interacting with unknown people						
SDA	0	0	5		3	
DA	21	11	11		6	
SA	27	14	22		12	
A	22	12	25		13	
SA	5	3	14		7	
Mobile phone has increased the sense of protection while alone at home or somewhere else						
SDA	1	1	3		2	
DA	4	2	4		2	
SA	11	6	7		4	
A	41	22	32		17	
SA	18	9	31		16	

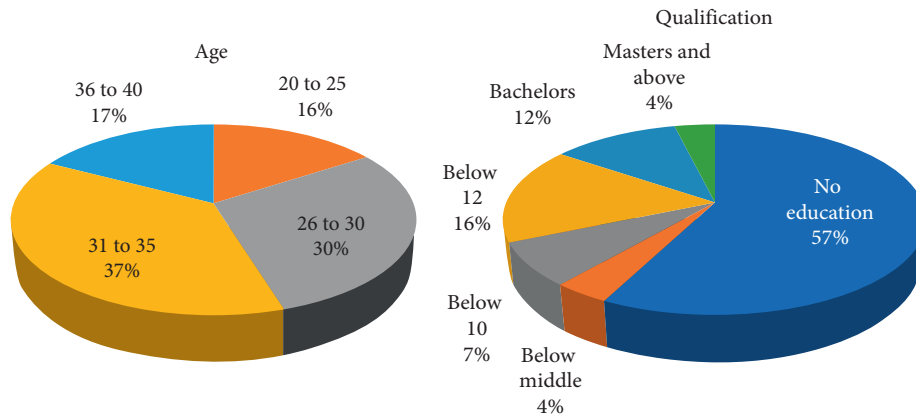


FIGURE 2: Participants' characteristics: age and qualification.

Table 7 shows the estimated results of factor loads and the reliability of scales. Factor analysis has extracted 6 factors of mobile phone access (MPA). We found a higher contribution of the statement “Do you have a mobile phone,” and its factor load is .958 in this component. It indicates that the majority of respondents have a mobile phone. The lowest contribution of the statement “How many smartphones do you have in your family?” has also been found to be 0.607. It also indicates that people are not much sensitive towards more than one smart mobile phone.

We also found a higher factor load of the skills of respondents in mobile phone usage because the entire factor load is greater than 0.8 and each statement is contributing high in this factor. On the other hand, the values of Cronbach's alpha in the second column of Table 8 suggested that the subscales are highly reliable and significant. The highest factor load is .950 of “Making a call without any help from other family members” and lowest factor load is 0.908 which is for “Can use Bluetooth or infrared to transfer pictures or anything”. Every component has the highest

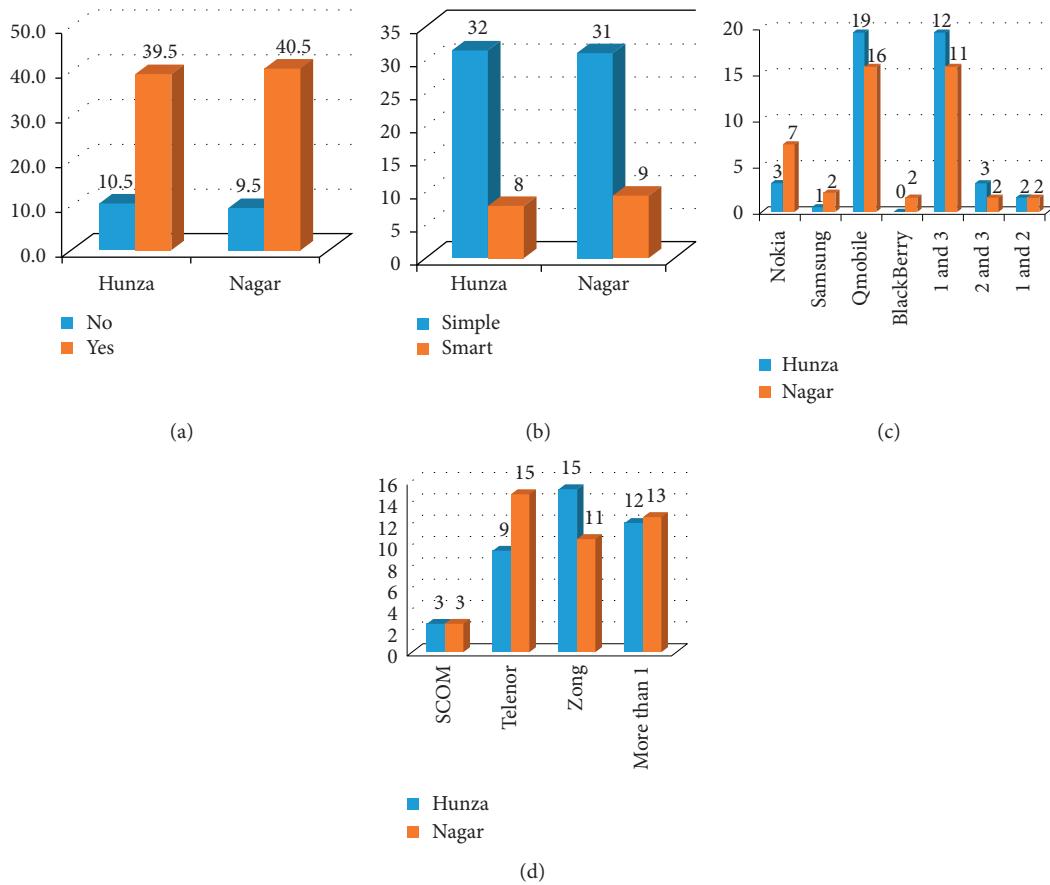


FIGURE 3: (a) Access of the mobile phone, (b) kinds of the mobile phone, (c) brand of the mobile phone, and (d) the service provider.

TABLE 6: KMO and Bartlett’s test.

Variables	KMO	Bartlett’s test of sphericity	Sig
MP_Access	0.910	1607.687	0.000
MPU_Daily_Life_Activities	0.907	1664.807	0.000
MPU_Security	0.899	1309.195	0.000
MP_Skills	0.954	4295.132	0.000

TABLE 7: Mobile phone access and use in daily life activities.

Variables	Factor loads	Cronbach $\alpha$
Mobile phone access		
Do you have mobile phone?	0.958	0.831
What kind of mobile phone do you have?	0.864	
How many mobile phones do you have in your family?	0.876	
How many smartphones do you have in your family?	0.607	
Mobile phone brand	0.794	
Which is your service provider?	0.735	

contribution towards the skills of mobile phone usage (detail is shown in Table 8).

There are 8 factors extracted for mobile phone usage in daily life activities (MPU\_Daily\_Life\_Activities), and the statement contributing the most in this component is “Keep in touch with family” having a factor load of .942. These results indicate that all the statements contribute almost

higher than .8, so mobile phone usage has greater importance in their daily life activities. Additionally, the statement contributing higher in mobile phone usage in daily activities is “Help you to send SMS” having a factor load of .937. It indicates that the majority of respondents are using SMS services in their daily life activities (results are shown in Table 9). The estimated coefficients of reliability (Cronbach’s

TABLE 8: Mobile phone usage skills of the women of Hunza and Nagar.

Variables	Factor loads	Cronbach $\alpha$
Mobile phone usage skills		
Can switch on/off mobile phone	0.948	
Make a call without help	0.950	
Can load balance card	0.918	
Can insert and change SIM card	0.947	
Can charge my mobile phone	0.949	0.983
Can search number of specific persons by name	0.947	
Can insert mobile battery	0.935	
Can enter and save a number of any person to my contact list	0.920	
Can send SMS	0.909	
Can use Bluetooth or infrared to transfer pictures or anything	0.908	

TABLE 9: Use of mobile phone in daily life activities.

Variables	Factor loads	Cronbach $\alpha$
Mobile phone daily life activities		
Keep in touch with family	0.942	
Keep in touch with friends	0.938	
Help you to take picture	0.920	0.973
Help you to send SMS	0.937	
Record audio video	0.922	
Mobile phone has given quick access to the Internet and use Internet	0.856	
Help to listen Naat/Ginans/songs	0.886	

alphas) of all the variables are presented in the last column of Table 9. These values suggested that subscales of mobile phone usage are highly reliable.

In the section, there are 5 factors extracted for mobile phone usage for safety and security purposes, and the statement contributing the most in this component is “Mobile phone has increased the sense of protection while going to the bazar” having a factor load of .963 and the lowest is .928 which is “Mobile phone has increased the sense of protection while interacting with unknown people”. All the statements contribute almost higher than 0.9 (detail is shown in Table 10).

## 5. Discussion

The potential of mobile phones cannot be ignored, which improves access to information and services especially in low-resourced settings [21, 22]. The use of conventional mobile phone technologies to support daily life activities such as phone calls, SMS, wake up rings, and access to the Internet is found to be very useful for rural women [49]. This study revealed that young women’s effective use of mobile phones helped them to ensure self-security in existing security threats in the Pakistani society which has been widely acknowledged in the literature [48]. Besides business activities, mobile phones have also affected human life. For example, rural women started using mobile phones in their daily life and developed their skills to use it for multiple purposes. Most of the women from both districts (Hunza and Nagar) own mobile phones. In this context, however, there is no barrier from the part of their families as compared to other parts of the country. But, very few respondents showed concern about

misuse of mobile phones such as receiving unethical messages, wrong calls, and people’s misperception of using mobile phones which appeared to be some of the barriers in using mobile phones freely to engage themselves with boosting business activities in the region.

The survey of the research shows that 57% of women are illiterate while they have the skills to use a mobile phone. In addition, they frequently and properly use mobile phones for business purposes at different levels. Hence, the usage of mobile phones is mainly for productive activities, i.e., for income generation and livelihood activities. On the contrary, the study results show that a huge number of unskilled workforce, i.e., about 57% in the study area, would be directly affected in the future due to the China-Pakistan Economic Corridor (CPEC) which is expected to change the dynamics of the existing business. Therefore, future research needs to focus on the various skills required for modern business. Therefore, all the mobile phone users in the region should focus on enhancing the technical skills required for an online business, for example, freelancing and e-commerce. In addition, there is a need to increase the different skills of women to boost up business activities through the use of mobile phones which would ultimately promote developmental activities in the region. However, in Pakistan, the service providers have already established the mobile phone infrastructure both in urban and rural remote areas of Gilgit-Baltistan that would help the people especially females to access mobile phones [56]. Lastly, new innovative business ventures and enhancement of business activities are possible by providing proper training on skill development and effective use of the mobile phone to 57% unskilled women in the study area.



TABLE 10: Mobile phone use for security purposes.

Variables	Factor loads	Cronbach $\alpha$
MPU_Security		
Mobile phone has increased the sense of protection while going to bazar	0.963	
Mobile phone has increased the sense of protection while travelling in public transport	0.953	0.97
Mobile phone has increased the sense of protection while going for work	0.947	
Mobile phone has increased the sense of protection while interacting with unknown people	0.928	
Mobile phone has increased the sense of protection while alone at home or somewhere else	0.940	

**5.1. Limitation and Future Direction of Research.** This study has provided some significant empirical shreds of evidence regarding the access to mobile phones and skills and use of the mobile phone for business purposes [57] by rural mountainous women of GB. There are some limitations such as the current study only focused on the usage of mobile phones for business purposes [58], security, communication, and access to information. However, the use of mobile phone in areas such as health, education, sociocultural, knowledge sharing, and professional development, and in sum all fields of life, is highly dependent upon the effective use of mobile phones [59]. Similarly, there are many other devices such as iPads, Tablets, PCs, Palm devices, and laptops which can also be used for these activities that this study did not focus. In the future, the research may focus on the use of mobile phones in a diverse range of its applications and its impact on the overall quality of life. Future research needs to focus on psychological issues and the negative impact of the mobile phone usage among the women of Gilgit-Baltistan.

## 6. Conclusion

The present study focuses on exploring the access and usage of a mobile phone in the daily life of rural women of two districts, i.e., Hunza and Nagar of Gilgit-Baltistan, Pakistan. For this purpose, the data are collected from 200 women through a questionnaire. The results of the study revealed that most of the women in the study area possess a mobile phone, and they use it for productive purposes. Furthermore, 37% women of age 31–35 years and 30% of age 26–30 years have access to mobile phones. However, the lowest users of mobile phones are the women aged 36–40 (17%) and women aged 20–25 years (16%). This shows that higher-aged women use mobile phones more than the youngsters. The women of rural Hunza and Nagar use mobile phones for various purposes, i.e., security, entertainment, communication, and sharing of knowledge. Moreover, the outcomes of the study reveal that women feel more comfortable by having a mobile phone which enhanced their level of confidence. On the contrary, the findings also show that 57% of the female population of the region is illiterate though they can use a mobile phone for everyday activities such as communication. However, fewer participants misuse the mobile phone, which needs to be discouraged, and they should focus on using this technology for improving the quality of life and productive purpose.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] S. Chatterjee, "A sociological outlook of mobile phone use in society," *International Journal of Interdisciplinary and Multidisciplinary Studies (IJIMS)*, vol. 4, pp. 59–63, 2014.
- [2] J. E. Katz and M. Aakhus, *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*, Cambridge University Press, Cambridge, UK, 2002.
- [3] J. E. Katz, "Mobile phones as fashion statements: the Co-creation of mobile communication's," in *Magic in the Air*, pp. 79–100, Routledge, Abingdon, UK, 2017.
- [4] J. E. Katz and S. Sugiyama, "Mobile phones as fashion statements: the co-creation of mobile communication's public meaning," in *Mobile Communications*, pp. 63–81, Springer, Berlin, Germany, 2005.
- [5] K. Aoki and E. J. Downes, "An analysis of young people's use of and attitudes toward cell phones," *Telematics and Informatics*, vol. 20, no. 4, pp. 349–364, 2003.
- [6] L. Fortunati, "The mobile phone: towards new categories and social relations," *Information, Communication & Society*, vol. 5, no. 4, pp. 513–528, 2002.
- [7] T. Kreutzer, "Assessing cell phone usage in a South African township school," *International Journal of Education and Development Using ICT*, vol. 5, pp. 43–57, 2009.
- [8] UNESCO, *Mobile Phones & Literacy, Empowerment in Women's Hands*, UNESCO Publishing United Nations, Paris, France, 2015.
- [9] GSMA, *The Mobile Gender Gap Report*, GSMA, London, UK, 2019, <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/GSMA-The-Mobile-Gender-Gap-Report-2019.pdf>.
- [10] PTA, "Annual report," 2019, <https://www.pta.gov.pk>.
- [11] WorldBank, "Annual Report," 2014, <https://www.worldbank.org>.
- [12] G. Sylvester, *Use of Mobile Phones by the Rural Poor: Gender Perspectives from Selected Asian Countries*, IDRC, Ottawa, ON, CA, 2016.
- [13] A. S. Nurullah, "The cell phone as an agent of social change," *Rocky Mountain Communication Review*, vol. 6, pp. 19–25, 2009.
- [14] M. Javid, M. A. Malik, and A. A. Gujjar, "Mobile phone culture and its psychological impacts on students' learning at the university level," *Language in India*, vol. 11, 2011.

- [15] UNCTAD, "United nations conference on trade and development," *Review of Maritime Transport*, UNCTAD, Geneva, Switzerland, 2014.
- [16] A. Bhavnani, R. W.-W. Chiu, S. Janakiram, P. Silarszky, and D. Bhatia, "The role of mobile phones in sustainable rural poverty reduction," *Retrieved November*, vol. 22, 2008.
- [17] S. Wyche and J. Olson, "Gender, mobile, and mobile Internet| Kenyan women's rural realities, mobile internet access, and "africa rising"" *Information Technologies & International Development*, vol. 14, no. 15, 2018.
- [18] C. Z.-W. Qiang, "Mobile telephony: a transformational tool for growth and development," *Private Sector & Development*, vol. 4pp. 7–8, 2009.
- [19] ADB, "Sustainable development challenges," *World Economic and Social Survey 2013*, ADB, Mandaluyong, Philippines, 2013.
- [20] Z. Laizu, J. Armarego, and F. Sudweeks, "Cognitive change in women's empowerment in rural Bangladesh," in *Proceedings of the 13th International Conference on Computer and Information Technology (ICCIT)*, pp. 277–282, Dhaka, Bangladesh, December 2010.
- [21] J. Gikas and M. M. Grant, "Mobile computing devices in higher education: student perspectives on learning with cellphones, smartphones & social media," *The Internet and Higher Education*, vol. 19, pp. 18–26, 2013.
- [22] S. Misra, L. Cheng, J. Genevie, and M. Yuan, "The iPhone Effect," *Environment and Behavior*, vol. 48, no. 2, pp. 275–298, 2016.
- [23] G. Demombynes and A. Thegeya, *Kenya's Mobile Revolution and the Promise of Mobile Savings*, The World Bank, Washington, DC, USA, 2012.
- [24] L. Caronia and A. H. Caron, "Constructing a specific culture: young people's use of the mobile phone as a social performance," *Convergence: The International Journal of Research Into New Media Technologies*, vol. 10, no. 2, pp. 28–61, 2004.
- [25] ITU, *UN's International Telecommunication Union (ITU)*, ITU, Geneva, Switzerland, 2019, <https://www.itu.int/en/publications/Pages/default.aspx>.
- [26] H. E. Chew, P. Vigneswara Ilavarasan, and M. R. Levy, "A latency effect for mobile phone investments by micro-entrepreneurs," *Media Asia*, vol. 39, no. 2, pp. 99–108, 2012.
- [27] S. Bailur, S. Masiero, and J. Tacchi, "Gender, mobile, and mobile internet| gender, mobile, and development: the theory and practice of empowerment—Introduction," *Information Technologies & International Development*, vol. 14, no. 9, 2018.
- [28] Cherie Blair Foundation, *Women & Mobile: A Global Opportunity, A Study on the Mobile Phone Gender Gap in Low and Middle-Income Countries*, GSMA, London, UK, 2017.
- [29] S. A. Asongu, J. C. Nwachukwu, and A. Aziz, "Determinants of mobile phone penetration: panel threshold evidence from sub-Saharan Africa," *Journal of Global Information Technology Management*, vol. 21, no. 2, pp. 81–110, 2018.
- [30] W. M. Olatokun, "Availability, accessibility and use of ICTs by Nigerian women academics," *Malaysian Journal of Library & Information Science*, vol. 12, pp. 13–33, 2017.
- [31] E. Buja, "Hofstede's dimensions of national cultures revisited: a case study of South Korea's culture," *Acta Universitatis Sapientiae, Philologica*, vol. 8, no. 1, pp. 169–182, 2016.
- [32] A. A. Shaikh and H. Karjaluo, "Mobile banking adoption: a literature review," *Telematics and Informatics*, vol. 32, no. 1, pp. 129–142, 2015.
- [33] E. L. Slade, M. D. Williams, and Y. K. Dwivedi, "Mobile payment adoption: classification and review of the extant literature," *The Marketing Review*, vol. 13, no. 2, pp. 167–190, 2013.
- [34] Z. Laizu, J. Armarego, and F. Sudweeks, "The role of ICT in women's empowerment in rural Bangladesh," in *Proceedings of the Cultural Attitudes Towards Communication and Technology 2010*, pp. 217–230, Murdoch, Australia, June 2010.
- [35] M. A. Islam and K. M. G. Hoq, "Community Internet access in rural areas: a study on community information centres in Bangladesh," *Malaysian Journal of Library & Information Science*, vol. 15, pp. 109–124, 2017.
- [36] K. B. Leahy and I. Yermish, "Information and communication technology: gender issues in developing nations," *Informing Science*, vol. 6, pp. 143–155, 2003.
- [37] UNESCO, *Shaping the Future Education through Mobile Phone*, Nokia and UNESCO Pakistan, Islamabad, Pakistan, 2017.
- [38] B. Reema and S. Sambargi, "Evaluation of personal innovativeness and perceived expertise on digital marketing adoption by women entrepreneurs of micro and small enterprises," *International Journal of Research and Analytical Reviews*, vol. 6, no. 1, pp. 338–351, 2019.
- [39] D. Potnis, "Culture's consequences: economic barriers to owning mobile phones experienced by women in India," *Telematics and Informatics*, vol. 33, no. 2, pp. 356–369, 2016.
- [40] E. R. Mbise, A. F. Kapinga, and C. S. Montero, "Mobile marketing application for entrepreneurship development: codesign with women entrepreneurs in Iringa," *EJISDC*, vol. 85, no. 2, 2018.
- [41] A. Doron, "Mobile persons: cell phones, gender and the self in North India," *The Asia Pacific Journal of Anthropology*, vol. 13, no. 5, pp. 414–433, 2012.
- [42] A. R. Madni, G. Ali, M. Abdullah, and S. Batool, "Mobile phone—need or status sybmol:(exploring usage & liking of mobile connections and packages by university students)," *Global Media Journal*, vol. 7, 2014.
- [43] G. Porter, K. Hampshire, A. Abane et al., "Mobile phones, gender, and female empowerment in sub-saharan Africa: studies with African youth," *Information Technology for Development*, vol. 26, no. 1, pp. 1–14, 2019.
- [44] J. C. Aker, C. Ksoll, and T. J. Lybbert, "ABC, 123: the impact of a mobile phone literacy program on educational outcomes," *SSRN Electronic Journal*, 2010.
- [45] E. E. Baro and B.-e. C. Endouware, "The effects of mobile phone on the socio-economic life of the rural dwellers in the Niger Delta region of Nigeria," *Information Technology for Development*, vol. 19, no. 3, pp. 249–263, 2013.
- [46] S. Malik, I. S. Chaudhry, and Q. Abbas, "Socio-economic impact of cellular phones growth in Pakistan: an empirical analysis," *Pakistan Journal of Social Sciences (PJSS)*, vol. 29, pp. 23–37, 2009.
- [47] S. Mittal, S. Gandhi, and G. Tripathi, "Socio-economic impact of mobile phones on Indian agriculture," Working Paper. 246, Indian Council for Research on International Economic Relations, New Delhi, India, 2010.
- [48] K. R. A Bairagi, T. S. Polin, and A. Terrence, "Socio-economic impacts of mobile phone in rural Bangladesh: a case study in batiaghata thana, khulna district anupam," M.Sc thesis, The University of Hong Kong, Pokfulam, Hong Kong, 2011.
- [49] S. A. Asongu, "How has mobile phone penetration stimulated financial development in africa?" *Journal of African Business*, vol. 14, no. 1, pp. 7–18, 2013.
- [50] R. Ling and T. Bertel, "Mobile communication culture among children and adolescents," *The Routledge International Handbook of Children, Adolescents and Media*, pp. 127–133, Routledge, Abingdon, UK, 2013.
- [51] T. Page, "Touchscreen mobile devices and older adults: a usability study," *International Journal of Human Factors and Ergonomics*, vol. 3, no. 1, pp. 65–85, 2014.

- [52] J. P. Rossing, W. Miller, A. K. Cecil, and S. E. Stamper, "iLearning: the future of higher education? student perceptions on learning with mobile tablets," *Journal of the Scholarship of Teaching and Learning*, vol. 12, no. 2, pp. 1–26, 2012.
- [53] M. Hakoama and S. Hakoyama, "The impact of cell phone use on social networking and development among college students," *The American Association of Behavioral and Social Sciences Journal*, vol. 15, p. 20, 2011.
- [54] A. D. Alanazi, *The Use of the Smartphones as a Resource for News Among Saudi Arabian Students in the United States*, Indiana University of Pennsylvania, Indiana, PA, USA, 2014.
- [55] M. Sánchez-Martínez and A. Otero, "Factors associated with cell phone use in adolescents in the community of madrid (Spain)," *CyberPsychology & Behavior*, vol. 12, no. 2, pp. 131–137, 2009.
- [56] J. O. T. A. Watkins, J. Goudge, F. X. Gómez-Olivé, and F. Griffiths, "Mobile phone use among patients and health workers to enhance primary healthcare: a qualitative study in rural South Africa," *Social Science & Medicine*, vol. 198, pp. 139–147, 2018.
- [57] L. F. Motiwalla, "Mobile learning: a framework and evaluation," *Computers & Education*, vol. 49, no. 3, pp. 581–596, 2007.
- [58] D. Raftery, "Ubiquitous mobile use: student perspectives on using the VLE on their phone," *Irish Journal of Technology Enhanced Learning*, vol. 3, no. 2, pp. 47–57, 2018.
- [59] C. Kaatz, C. Brock, and L. Figura, "Are you still online or are you already mobile? - predicting the path to successful conversions across different devices," *Journal of Retailing and Consumer Services*, vol. 50, pp. 10–21, 2019.

## Research Article

# A Novel Fuzzy Logic-Based Medical Expert System for Diagnosis of Chronic Kidney Disease

**Jimmy Singla** <sup>1</sup>, **Balwinder Kaur**,<sup>1</sup> **Deepak Prashar**,<sup>1</sup> **Sudan Jha** <sup>1</sup>,  
**Gyanendra Prasad Joshi** <sup>2</sup>, **Kyungyun Park** <sup>3</sup>, **Usman Tariq**,<sup>4</sup> and **Changho Seo** <sup>3</sup>

<sup>1</sup>*School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India*

<sup>2</sup>*Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea*

<sup>3</sup>*Department of Convergence Science, Kongju National University, Gongju 32588, Republic of Korea*

<sup>4</sup>*College of Computer Science and Engineering, Prince Sattam bin Abdulaziz University, Saudi Arabia*

Correspondence should be addressed to Gyanendra Prasad Joshi; [joshi@sejong.ac.kr](mailto:joshi@sejong.ac.kr) and Changho Seo; [chseo@kongju.ac.kr](mailto:chseo@kongju.ac.kr)

Received 7 March 2020; Accepted 24 April 2020; Published 5 June 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Jimmy Singla 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.

Chronic kidney disease is a life-threatening complication. Primary diagnosis and active control avoid its progression. To increase the life span of a patient, it is necessary to detect such diseases in early stages. In this research paper, design and development of a fuzzy expert system (FES) to identify the current stage of chronic kidney disease is proposed. The proposed fuzzy rule-based expert system is developed with the help of clinical practice guidelines, database, and the knowledge of a team of specialists. It makes use of input variables like nephron functionality, blood sugar, diastolic blood pressure, systolic blood pressure, age, body mass index (BMI), and smoke. The normality tests are applied on different input parameters. The input variables, i.e., nephron functionality, blood sugar, and BMI have more impact on the chronic kidney disease as shown by the response of surface analysis. The output of the system shows the current stage of patient's kidney disease. Totally 80 tests were performed on the FES developed in this research work, and the generated output was compared with expected output. It is observed that this system succeeds in 93.75% of the tests. This system supports the doctors in assessment of chronic kidney disease among patients. The detection of chronic kidney disease is a serious clinical problem that comprises imprecision, and the use of fuzzy inference system is suggested to overcome this issue. The proposed FES is implemented in the MATLAB.

## 1. Introduction

Chronic kidney disease or chronic renal disorders arise due to the lack of functionality of nephrons in the kidney. It basically occurs when the kidney stops its functionality like the maintenance of pH value, water, and salt in the blood. The kidney is the natural filter of blood and it discharges the wastage from the body in the form of urine. This disease can also lead to the damage of other neighbor organs of a kidney.

Nowadays, due to increase in population and unhealthy lifestyle, number of patients with renal disorder is increasing [1–3]. It is tough for individual to recover from every kidney disease [5]. The detection of kidney disease and disorder can be done by studying the features extracted from an acquired image. The image-processing techniques are applied on

magnetic resonance images of total kidney volume [6]. The registration is performed on the selected portion of an image to increase the accuracy [7]. The optimal path is detected during the segmentation of magnetic resonance images of kidneys for diagnosis of Chronic Kidney Disease (CKD) [8]. Computer aided diagnostic systems are used to examine the kidney features for the early identification of kidney diseases [9, 10]. Certain test, like F-test, is applied to find out the dependency of features on the kidney area [11]. Ultrasonography images are also used for detection of stages of CKDs. The stages of CKD are predicted by using machine learning techniques [12–15]. The comparisons have been made between the performances of different systems developed by using various algorithms of machine learning [16]. Intelligent models are designed by using different

methods of artificial neural networks, such as generalized feed forward neural network, back-propagation neural network, and modular neural network [17]. The Naive-Bayes method is also used to classify the various stages of CKD [18]. An approach is designed for detection of disease at initial stage by training the neural network with some amount of training data [19–21]. Further, the data mining algorithms and techniques are used for CKD detection [22, 23]. Ensemble learning is used to clean the data that are acquired and to fill the values which are missing in that particular dataset to recalculate the different stages of CKD, and hence, the performance of the system is increased with this step [24]. The decision support systems are also built by using the information or dataset of CKD [25].

Early detection of the disease permits the doctor to suggest the required measure so that the risk may reduce [26]. The ranks are given to all the risk factors that extremely affect the kidney [27, 28]. Renal illness is a public health problem across the nation. In Taiwan, there are over 1,100,000 patients living with CKD and this number is increasing gradually. Moreover, the data from United States Renal Data System (USRDS) stated that Taiwan has extreme occurrence of end stage kidney disease across the world [29, 30].

In India, the death rate is increasing rapidly due to various chronic diseases, and CKD is one of the major diseases among them. The average age of a person that suffers from CKD is from 18 to 98 years in which the number of male patients is more than females [31]. The CKD is not detected easily till it grows due to its asymptomatic nature, and it results in decrease in the opportunities for prevention of the disease [32, 33]. It is very important to detect and treat CKD in early stages to avoid kidney failure [34]. It is reported that this problem can be mitigated by developing fuzzy logic-based expert system [35].

The main objective of this work is to develop a medical expert system for diagnosis of chronic kidney disease by using fuzzy logic. The aim is to assist doctors in diagnosis of patients suffering from chronic kidney disease. Fuzzy logic is a rule-based logic that deals with imprecise concepts. It is a many-valued logic that can combine human heuristics into computer-assisted decision-making. The expert system is a computer program which is derived from the artificial intelligence and can be used in real-time applications [36]. Over 70%, this domain is used to solve the real-life applications [37]. The expert systems are used for diagnosis of various diseases. The main idea behind expert system is that the knowledge from the experts of a specific disease will be transferred to the computer system [38]. The knowledge stored in the computer system can be used by the user for specific task. After that, the system will conclude the result by doing various estimations.

The fuzzy logic-based system is having three blocks and the fuzzy logic-based system used here is Mamdani type fuzzy model, all three blocks are discussed as follows:

- (1) Fuzzification: It is the process of transforming the crisp values into the fuzzy inputs. The crisp values are inputs measured by the sensors or the values given by the users [39].

- (2) Inference mechanism: In the inference mechanism, the system maps rules satisfy input given by user and it gives an output.

- (3) Defuzzification: This is the last step in the system. In defuzzification process, a single number from the output of the aggregated fuzzy set is obtained.

Response surface methodology is a set of precise and numerical procedures that are supportive for exhibiting and analysing issues wherever an interest response obtains the impact of various parameters, and also the aim is to maximise such response [39–41].

The research works discussed above provide fruitful information about the applications of expert systems for medical diagnosis and especially for the diagnosis of chronic kidney disease. The primary risk factors or causes of the chronic kidney disease have been used as the input variables in the development of a proposed medical expert system. After the processing of all blocks of fuzzy logic, it generates an output. In this developed medical expert system, the output is obtained according to the health of a patient who is suffering from this disease.

The main limitations of this system are that the number of rules and input variables can be increased to get the more accurate result and also this system does not have capability to adapt itself according to the environment or to train itself from new examples.

The rest of the paper is organized as follows.

In Section 2, the related work is explained briefly. In Section 3, some established theories about fuzzy logic and notations used in the proposed system are discussed. In Section 4, the methodology for the proposed work including architecture of input-output parameters is described. The normality tests performed on different input variables are also shown in this section. The performance of the proposed system in terms of sensitivity, specificity, precision, and accuracy is evaluated in Section 5. In Section 6, the fuzzy expert system developed in this work along with ideas about future work is discussed. Finally, the work is concluded in Section 7.

## 2. Related Work

Kubota et al. [5] proposed a methodology by using image-processing technique in which the detection of the kidney failure of a patient has been done in the early stage based on information of kidney region. The required information is gathered from the CT image of kidney. The image of kidney is segmented into number of regions to collect correct and accurate information from it. Tang et al. [7] projected a method of registration known as intrasubject registration by using the local MI maximization. In this method, the registration is not done on the unified volumes that were acquired from the abdomen region of the patient's body. It will be accomplished only on the selected areas of kidney, which are affected by the disease. This method also increases the accuracy of performing registration. Bommanna Raja and Madheswaran [8] developed a computer aided system that examined the various features which are unconstrained and

also independent of kidney's area. These examined features are used in the developed system to examine the disease. Rovența and Roșu [37] developed a medical expert system to diagnose twenty-seven different kidney diseases from nine different categories. The identification of kidney disease is done by taking the symptoms that can be seen in clinical examination and the result evaluated in the laboratory test. The system was developed by using Visual Prolog 5.2 and helps the doctor to diagnose the disease more accurately.

Shen et al. [29] developed a low cost method for the patients of dialysis. This method is used to identify the risks of cardiovascular disease (CVD) on the chronic kidney disease (CKD). ECG features and heart rate variability of a patient are considered in this system. It uses the decision-based neural network structure for the feature fusion. The overall accuracy of developed method for cardiovascular disease on chronic kidney disease is 71.07%. Adam and Hashim [19] used the artificial neural network technique for the detection of kidney problems in the initial stage. This method predicts various symptoms of kidney problem and then compares these symptoms with the mental behaviour of an individual. The artificial neural network is first trained by giving training samples, and then the testing procedure is done for the same. Chw et al. [17] designed intelligence models by using the various technologies of artificial neural network such as back-propagation network (BPN), generalized feed forward neural networks (GRNN), and modular neural network (MNN). These methodologies are developed to detect the chronic kidney disease at initial stages. These three models are compared with each other based on accuracy and effectiveness and also deployed on the Google Cloud platform. The constructed models help to avoid the chronic renal disorders in the beginning of the disease or in the introductory stages. Ahmed et al. [42] proposed a fuzzy expert system for the diagnosis of chronic kidney disease. The seven inputs were considered, and by using the values of these inputs, the membership functions are generated. The rules are further generated, and Matlab software is used to implement this research work. The output is predicted by expert system and it gave similar results as the human expert can give.

### 3. Notations and Preliminaries

The notations and preliminaries used in this work are as follows [43].

**3.1. Crisp Sets.** The crisp set or the classical set is defined as the assemblage or group of elements  $x \in X$  that can be finite. For example, to define the set if elements are larger than 5, the membership function can be defined as follows:

$$X = \{x | x > 5\}, \quad (1)$$

where  $X$  is the set of positive integers.

**3.2. Fuzzy Sets.** If  $X$  is the set of elements denoted by  $x$ , then a fuzzy set  $A$  in  $X$  is the set of ordered pair and is represented by

$$A = \frac{(x, \mu_A(x))}{x \in X}, \quad (2)$$

where  $\mu_A(x)$  = membership function of  $x$  in  $A$ .

**3.3. Fuzzy Numbers.** A fuzzy number  $U$  is convex. A normalized fuzzy set  $U$  of the real line  $R$ , such that

$$\mu_U(x): R \longrightarrow [0, 1], \forall x \in R, \quad (3)$$

where  $\mu_U$  = membership function of fuzzy set.

There are many fuzzy numbers that exist. But in this work, only trapezoidal fuzzy numbers are used.

**3.4. Trapezoidal Fuzzy Number (TrFN).** Assume an arbitrary trapezoidal fuzzy number  $\tilde{U} = (p, q, r, s)$ . The membership function of  $\tilde{U}$  is represented by  $\mu_{\tilde{U}}$  and is given by

$$\mu_{\tilde{U}}(x) = \begin{cases} 0, & x \leq p, \\ \frac{x-p}{q-p}, & p \leq x \leq q, \\ 1, & q \leq x \leq r, \\ \frac{s-x}{s-r}, & r \leq x \leq s, \\ 0, & x \geq s. \end{cases} \quad (4)$$

## 4. The Proposed Methodology

The methodology applied in this research work is shown in Figure 1.

The expert system has one layer. As shown in Figure 2, this layer detects the stage of CKD based on various input variables. The proposed expert system is explained mathematically as follows:

$$\mu_{RC, \text{layer1}} = \text{MFIS}[\mu_{NF}, \mu_{BS}, \mu_{DBP}, \mu_{SBP}, \mu_{Age}, \mu_{BMI}, \mu_{Smoke}]. \quad (5)$$

**4.1. Selection of Input and Output Parameters.** The attributes should be selected carefully so that classification of patients can be done perfectly [44, 45]. The details of seven input variables used in this research work are described as follows.

**4.1.1. Nephron Functionality (NF).** The nephron functionality is the proportion of active nephron in the kidney [42]. The nephron functionality is in safe zone if the value is greater than 0.47. However, the functionality is moderately risky if the value is in between 0.3 to 0.5 and it is very-risky if it is below 0.35.

**4.1.2. Blood Sugar (BS).** The amount of glucose available in the blood refers to the blood sugar. It is composed of three

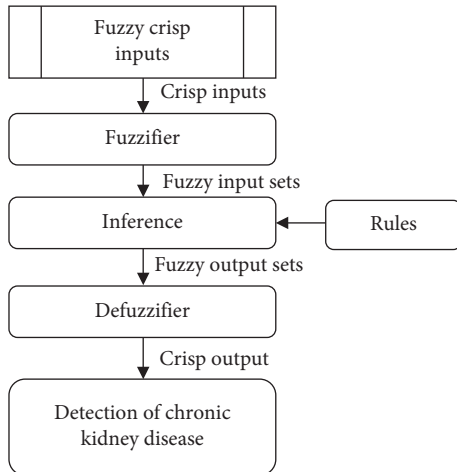


FIGURE 1: Methodology of the proposed expert system.

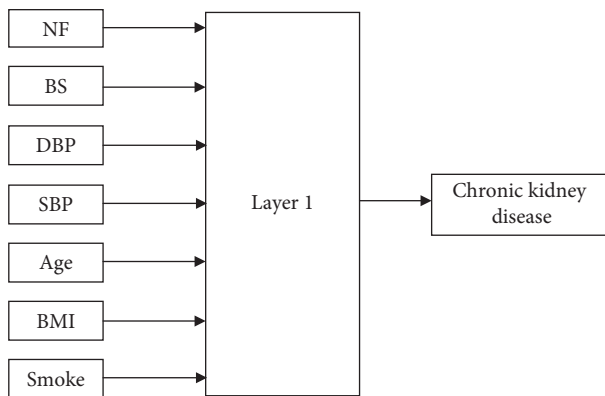


FIGURE 2: Proposed fuzzy expert system.

fuzzy sets. If the level of blood sugar is below 104, then it is safe, whereas it is at borderline if it is between 104 to 150 and if the level is greater than 140 then the existence level of glucose in blood is high.

**4.1.3. Diastolic Blood Pressure.** The diastolic blood pressure (DBP) is the function of congesting the heart with blood in between the muscle convulsion. It consists of 4 fuzzy sets. If the range is less than 89, then the diastolic blood pressure is low. If the range is between 87 to 110, then it is medium and if the range is between 106 to 121, then it is high. Nevertheless, if the range is above 118, then the diastolic blood pressure is extremely high.

**4.1.4. Systolic Blood Pressure.** When the heart contracts, the systolic blood pressure (SBP) is stimulated. It is peculiarly the utmost arterial pressure during shrinkage of heart's left ventricle. The time at which ventricular contraction occurs is called systole. If the range of systolic blood pressure is below 134, then it is low; if the value is between 127 to 153, then it is medium; if the range is between 142 to 172, then it is high; and if the range is more than 162, then it is extremely high.

**4.1.5. Age.** Patient's kidneys are devilishly affected by age. It includes three variables. If the age is under 36, then the patient is young; if age is within 33 to 66, the patient is considered as the mid-age person; and if the age is greater than 52, the patient is classified as old.

**4.1.6. Body Mass Index (BMI).** BMI is a simple and clinically useful data for CKD diagnosis. It is calculated by dividing the person's weight with square of height. It has four variables. A person is underweight if BMI is less than 19, normal weight if it is between 18.5 to 24.9, overweight if the value is 24.6 to 30, and obese if it is 29.5 or more.

**4.1.7. Smoke.** Smoking plays a major role in disturbing the kidney. It has three fuzzy sets. It is low if the range is less than 2.64, medium if the range lies between 1.8 to 9.5, and high if the range is greater than 8.5.

The output variable evaluates the condition of patient's kidney according to the behaviour of the input variables. The range of the output is from 0 to 10. The value 0 means that there is no hiccup in the patient's kidney and 10 means the patient is abundantly sick. This variable consists of six fuzzy sets.

- (i) **Healthy:** The patient is healthy if the value of the output variable is less than 1.7.
- (ii) **Concerning:** The condition of the patient's kidney is concerning if the value of the output variable is within 1.5 to 3.26.
- (iii) **Very concerning:** If the range of output variable falls between 2.84 to 4.5, then the condition of kidney is considered as "very concerning".
- (iv) **Sick:** Sick class is assigned to the patients who get output value between 4 to 6.
- (v) **Very sick:** The patient is considered as very sick if the range of output variable is within 5.5 to 7.98.
- (vi) **Extremely Sick:** The condition of the patient's kidney is extremely sick if the output value is above 7.5.

**4.2. Architecture for Input and Output Parameters.** There are two ways to develop the architecture of input and output parameters:

- (1) The first way is to determine the ranges of variables such as nephron functionality, blood sugar, and BMI used in fuzzy set using the confidence intervals (CI) [46].
- (2) The second way is to get them from the experts or Clinical Practice Guidelines (CPG).

**4.2.1. Development of the Parameters Using CIs.** CIs are used on the sample of individuals from the target population because it is not possible to perform a study on entire population [39]. Sample data of 102 patients are taken, and to further obtain some important information from the

given sample, it is crucial to apply tests. Figures 3–5 depict the normality tests applied on the dataset with Minitab 18 software. The normality tests used for this analysis are Anderson Darling and Ryan–Joiner normality test. To accept the null hypothesis, the  $P$ -value must be greater than 0.05, whereas value of correlation coefficient must be near to 1.

After the normality test, the next step is to apply confidence intervals to these three variables, i.e., Nephron Functionality, Blood sugar, and BMI. For example, in Table 1 the case of nephron functionality is mentioned, and the various parameters like mean are described as per interest.

4.2.2. *Ranges of Input and Output Parameters.* An efficient analysis of each variable range determines the parameters corresponding to each variable. Figure 6 shows the membership function of the nephron functionality. Figure 7 shows the membership function of blood sugar by CIs and Figure 8 shows the membership function for BMI by CIs. With the information of the nephron functionality variable, the mathematical representation of the interval used in this model can be described as follows:

$$\mu_{\text{very-risky}}(x) = \begin{cases} 1, & x < 0.28, \\ \frac{0.35 - x}{0.07}, & 0.28 < x \leq 0.35, \\ 0, & x > 0.35, \end{cases}$$

$$\mu_{\text{moderately-risky}}(x) = \begin{cases} 0, & x < 0.3, \\ \frac{x - 0.3}{0.07}, & 0.3 \leq x \leq 0.37, \\ 1, & 0.37 \leq x \leq 0.42, \\ \frac{x - 0.42}{0.08}, & 0.42 \leq x \leq 0.5, \\ 0, & x > 0.5, \end{cases} \quad (6)$$

$$\mu_{\text{Safe-zone}}(x) = \begin{cases} \frac{x - 0.47}{0.07}, & 0.47 \leq x \leq 0.54, \\ 1, & x > 0.54. \end{cases}$$

4.3. *Knowledge Base (IF-THEN Rules Formulation).* In this phase, a group of rules is related to the fuzzy variables with the output classification. A knowledge base contains these fuzzy rules. The precedent-consequential way is followed for these rules. In this model, the Mamdani fuzzy rule-based model is applied, and these rules are stated as follows:

*IF* . . . . . *Clinical variable 1 AND Clinical variable 2. . .*  
*THEN* . . . . . *Output.*

*Example 1, inference rule no. 1.*

*If Nephron Functionality is moderately risky,*  
*And Blood sugar is high,*  
*And DBP is high,*  
*And SBP is high,*  
*And Age is mid age,*

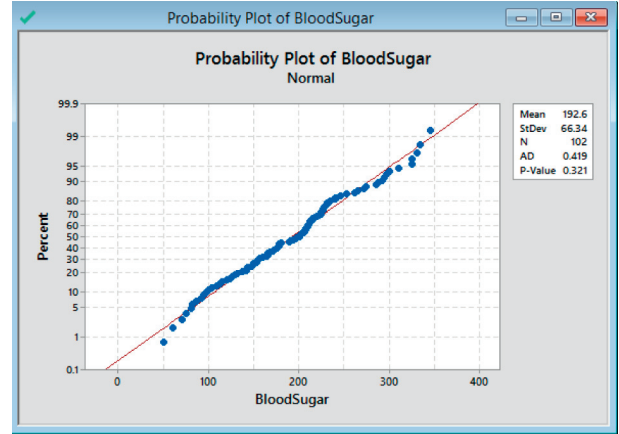


FIGURE 3: Normality test for input variable Blood Sugar (mmol/L).

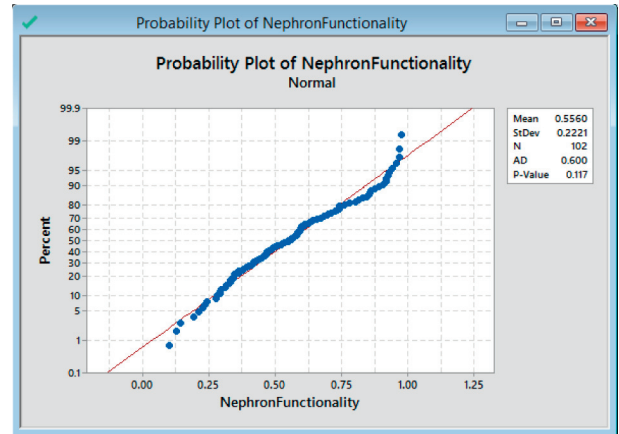


FIGURE 4: Normality test for input variable “Nephron Functionality” (ml/min).

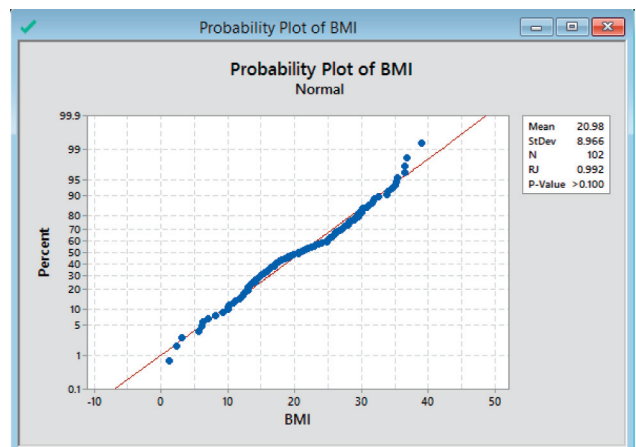


FIGURE 5: Normality test for input variable “BMI” (kg/m<sup>2</sup>).

*And BMI is underweight,*  
*And Smoke is Medium,*  
*THEN the patient’s kidney condition is very concerning.*  
*Example 2, inference rule no. 81.*



TABLE 1: Parameters of nephron functionality.

Linguistic labels	Sample size	Mean	$\sigma$	$\alpha$	$\tau\alpha, n-1$	$Z\alpha=$	Lower limit	Unilateral lower limit	Unilateral upper limit	Upper limit
Safe-zone	36	0.71	0.14	0.1		1.65	0.478	0.538	0.548	1.00
Moderately risky	30	0.40	0.05	0.1	2.045		0.3	0.37	0.42	0.5
Very-risky	14	0.23	0.06	0.1	2.160		-1	0.1	0.28	0.35

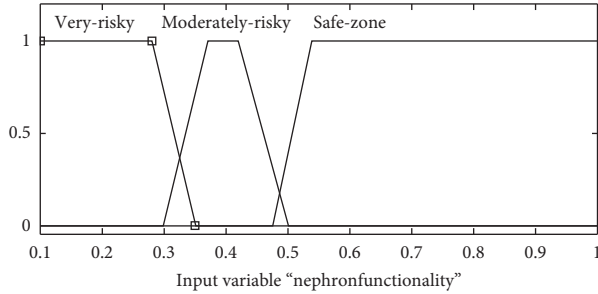


FIGURE 6: Membership function for nephron functionality by CIs.

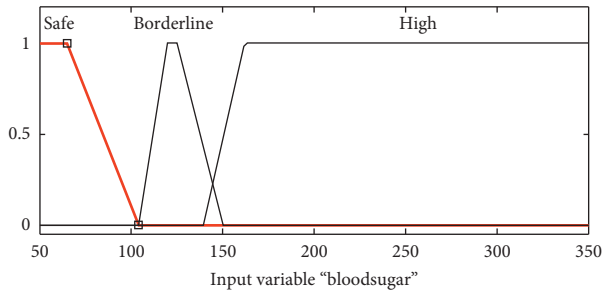


FIGURE 7: Membership function for blood sugar by CIs.

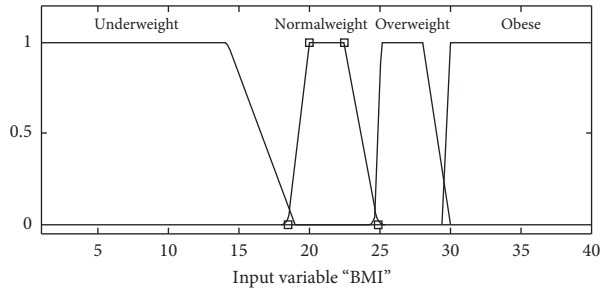


FIGURE 8: Membership function for BMI by CIs.

*IF Nephron functionality is Very-Risky,  
 And Blood sugar is High,  
 And DBP is Extremely High,  
 And SBP is Extremely High  
 And age is Old,  
 And BMI is obese,  
 And smoke is high,  
 THEN patient's kidney is extremely sick.*

Table 2 shows the variables in FES, and the abovementioned rules are interpreted into fuzzy rules as presented in Table 3.

In the fuzzy inference system, the input-output rules play an important role. The efficiency of a proposed system

is directly proportional to the I/O rules generated and stored in the knowledge base of expert system. Figure 9 presents the framework of knowledge base in which the rules are stored.

In the proposed model, there are seven variables; hence, it gives 5184 rules, but it is difficult to handle so many rules. Therefore, in this work, 83 rules are used. These rules have no omission because each of them has probability of its occurrence. The variables are chosen in such a way that the redundancy in rules can be avoided and impeded the dependency of one variable on other so that it will not affect the inference mechanism. The rules are also revised to find out the inconsistent rules, and the result of analysis showed positive outcomes. The abovementioned seven input variables are mixed and according to the reference rules the output is determined.

$$\text{IRT} = \text{nephron functionality (3)} \times \text{blood sugar (3)}$$

$$\times \text{diastolic blood pressure (4)}$$

$$\times \text{systolic blood pressure (4)} \times \text{age (3)}$$

$$\times \text{BMI (4)} \times \text{smoke (3)} = 5,184 \text{ inference rules.}$$

(7)

where IRT stands for Inference rules total.

The defuzzification method computes the centre of gravity of images that is driven at the breakdown time of every input linguistic parameter. It implies that this method inspects the relationship between the degrees of membership of two joint sets and enumerates the outside of each subsequent image. The classic structure is utilized in the centroid method corresponding to the abscissa axis. The centre of gravity is in the middle of its base, in rectangles. Further, for triangles, it is set at the third part of its base as indicated by inverse side of the point induced by hypotenuse and base. Finally, to obtain the total centroid, the summation of surface product of every figure is divided by its centroid into the total surface. [46].

**4.4. Tests.** Here, the need was to interview the patients suffering from CKD. The proposed fuzzy expert system is validated. This validation is done by various specialist doctors and experts from the nephrology and urology department. To evaluate confidence of the system, eighty tests are performed. Results are of a high certainty degree.

The main aim of the evaluation of system is to identify the error percentage and imperfections which could be possibly found in the fuzzy expert system. It also helped to demonstrate the expectations that will be fulfilled by system and the possible error which further helped to calculate

TABLE 2: Variables in FES.

Input variables	Linguistic label	Membership function	Interval
Nephron functionality	Very-risky	Trapezoidal	(-1, 0.1, 0.28, 0.35)
	Moderately risky	Trapezoidal	(0.3, 0.37, 0.42, 0.5)
	Safe-zone	Trapezoidal	(0.47, 0.53, 0.54, 1)
Blood sugar	Safe	Trapezoidal	(-58, 4, 65, 104)
	Borderline	Trapezoidal	(104, 120, 125, 150)
	High	Trapezoidal	(140, 162, 362, 458)
Diastolic blood pressure	Low	Trapezoidal	(65, 78.33, 82, 89)
	Medium	Trapezoidal	(87, 95, 100, 110)
	High	Trapezoidal	(106, 110, 115, 121)
Systolic blood pressure	Extremely high	Trapezoidal	(118, 123, 131.7, 144.9)
	Low	Trapezoidal	(29.6, 74.4, 118, 134)
	Medium	Trapezoidal	(127, 137, 143, 153)
Age	High	Trapezoidal	(142, 157, 162, 172)
	Extremely high	Trapezoidal	(162, 180, 240, 250)
	Young	Trapezoidal	(-18.8, 6.8, 28, 36)
BMI	Mid-age	Trapezoidal	(33, 44, 55, 66)
	Old	Trapezoidal	(52, 61, 93.2, 118.8)
	Underweight	Trapezoidal	(-46.66, 1, 14.18, 19)
Smoke	Normal weight	Trapezoidal	(18.5, 20, 22.5, 24.9)
	Overweight	Trapezoidal	(24.6, 25.1, 28.1, 30)
	Obese	Trapezoidal	(29.5, 30, 41, 55)
Output variable	Low	Trapezoidal	(-7.25, -0.85, 1.5, 2.64)
	Medium	Trapezoidal	(1.8, 5.2, 5.8, 9.5)
	High	Trapezoidal	(8.5, 11, 20.8, 27.2)
Result	Healthy	Trapezoidal	(-3.57, -0.34, 1.1, 1.7)
	Concerning	Triangular	(1.5, 2.4, 3.26)
	Very-concerning	Triangular	(2.84, 3.67, 4.5)
	Sick	Triangular	(4, 5, 6)
	Very sick	Triangular	(5.5, 6.74, 7.98)
	Extremely sick	Trapezoidal	(7.5, 8.4, 300)

TABLE 3: The system knowledge base.

Rule no.	Nephron functionality	Blood sugar	DBP	SBP	Age	BMI	Smoke	Conclusion
2	Extremely risky	High	High	Extremely high	Mid age	Obese	Low	Very concerning
11	Safe-zone	Borderline	Medium	Medium	Young	Overweight	High	Healthy
18	Moderately risky	Borderline	High	High	Young	Underweight	Low	Healthy
29	Very-risky	Safe	Medium	Medium	Young	Obese	High	Concerning
36	Moderately risky	Borderline	Extremely high	Extremely high	Old	Underweight	Low	Very concerning
48	Safe-zone	High	Extremely high	Extremely high	Mid-age	Under-weight	Medium	Very concerning
54	Very-risky	Safe	Extremely high	Extremely high	Young	Obese	Low	Very sick
64	Very-risky	Borderline	High	High	Mid-age	Overweight	High	Very-sick
73	Moderately risky	High	Extremely high	Extremely high	Young	Overweight	High	Extremely sick
80	Moderately risky	High	Extremely high	Extremely high	Old	Obese	High	Extremely sick

performance of system and where the performance measure was lacking behind, it helped to improve that problem.

The various tests that were done with this fuzzy expert system are given in Table 4. In these tests, doctors evaluated suffering patients by this fuzzy expert system with various input variables that are used in the system, and then the obtained results were compared with the actual results. In Test 1, Nephron functionality is equal to 0.458, blood sugar is 132.2, Diastolic Blood Pressure is equal to 105.5, systolic Blood Pressure equal to 128.9, age is 38, BMI have the value 29.55, and smoke value is 2.037, and then control assessment

by expert system for that particular patient's kidney condition is concerning. The result as per doctor is also that the kidney of the patient is in concerning phase.

**4.5. Response Surface Analysis.** The system also gives the three-dimensional plot which describes the influence of input variables on output and output can also be calculated at particular inputs from this three-dimensional plot. These surfaces are helping to depict the level of clinical variables that influence results. It can be seen as if the

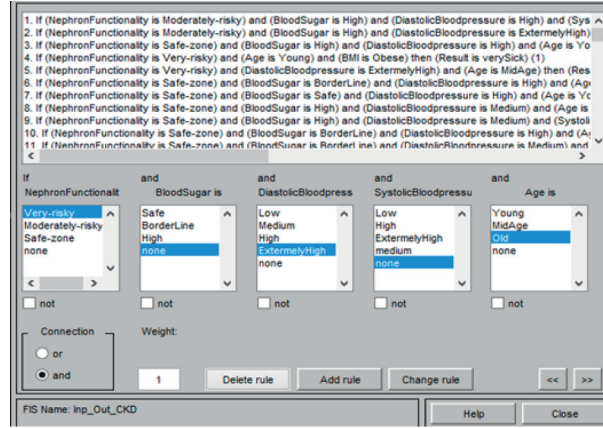


FIGURE 9: Input-output rules for the proposed expert system.

combinations of variables are in light colour, then it means the optimal zone, whereas if the colour of combination is dark, then the variable levels activate a nonoptimal zone, which indicates a possibility of having CKD to a particular patient.

## 5. Experimental Results and Analysis

Eighty tests were carried out to analyse CKD using developed fuzzy expert system. The team of doctors also diagnosed each patient and gave result for same. After that, the results given by experts are compared with obtained results from the fuzzy expert system. By correlating the two results, 75 out of 80 tests were correctly classified, and hence the system showed 5 errors. A sample of tests is shown in Table 4.

In Table 5, the confusion matrix is given. 80 cases of different patients are carried out. In the first column, 15 cases of healthy patients are considered and all of them are classified into correct class, i.e., healthy class. In the second column, the data of 10 patients at the concerning level is considered, but as a result out of 10 cases, 8 cases are exactly classified into concerning class and rest of 2 are classified into sick class. Similarly, in the third column, 16 cases very concerning are carried out, and from those 16 cases 15 are classified correctly and one is classified into class sick. In the fourth column, out of 11 diagnosed patients, 1 is classified incorrectly and 10 are in exact class. Likewise, fifth column illustrates that 14 diagnosed patients of class very sick are classified in correct class. At last, the sixth column shows that there are 14 cases of extremely sick patient, but out of 14, 1 is misclassified.

$$\text{confidence indicator} = \left( \frac{\text{success number}}{\text{total number of tests}} * 100 \right), \quad (8)$$

where number of successes = 75 and total number of tests = 80.

It was examined that 93.75% of the results given by fuzzy expert system are classified correctly by calculating the confidence indicator using above formula. Therefore, by observing the confidence indicator, it can be said that the

developed fuzzy expert system can work as the supporting tool for doctors in diagnosis of CKD.

The results show that the categorization by fuzzy inference system for CKD is 06.25% wrong. Now, the first three classes that is healthy, concerning, and very concerning are taken as “no”. Similarly, the rest of classes sick, very sick, and extremely sick are taken as “yes”. Hence, the confusion matrix shown in Table 4 is reduced to 2\*2 matrix and shown in Table 6.

The performance of developed medical expert system is evaluated by considering various parameters like accuracy, precision, specificity, and sensitivity.

In case of CKD, from Table 5,

- (i) TP: true positive is 37
- (ii) FN: false negative is 02
- (iii) FP: false positive is 03
- (iv) TN: true negative is 38

$$\text{sensitivity} = \frac{(TP)}{(TP + FN)} = \frac{37}{37 + 02} = 94.87\%,$$

$$\text{specificity} = \frac{(TN)}{(TN + FP)} = \frac{38}{38 + 03} = 92.68\%,$$

$$\text{precision} = \frac{(TP)}{(TP + FP)} = \frac{37}{37 + 03} = 92.5\%,$$

$$\begin{aligned} \text{classification accuracy} &= \frac{(TP + TN)}{(TP + FP + TN + FN)} \\ &= \frac{37 + 38}{37 + 03 + 38 + 02} = 93.75\%. \end{aligned} \quad (9)$$

These calculated parameters are shown in the form of graph in Figure 10.

The main aim to develop this expert system is that it will help doctors to identify the CKD in patients. This medical

TABLE 4: A sample of tests performed with developed fuzzy expert system.

Test	Nephron functionality	Blood sugar	DBP	SBP	Age	BMI	Smoke	CADT	CAFES	Success
1	0.458	132.1	105.5	128.9	38	29.55	2.037	Concerning	Concerning	Yes
2	0.226	165	113	155	64	3	10	Extremely sick	Extremely sick	Yes
3	0.1	300	120	212	26	35	20	Extremely sick	Very sick	No
4	0.4083	75.47	89.72	134.2	20	25.92	8.333	Healthy	Healthy	Yes
5	0.875	177.4	107.3	152.6	59	17.25	16.48	Concerning	Concerning	Yes
6	0.3292	81.13	121	180.7	24	30.07	2.075	Very sick	Very sick	Yes
7	0.5917	137.7	115.6	144.7	28	26.24	18.33	Healthy	Healthy	Yes
8	0.6417	205.7	120.3	176.4	47	13.64	6.481	Very sick	Very concerning	No
9	0.9417	86.79	93.43	139.4	58	10.03	8.333	Healthy	Healthy	Yes
10	0.3583	120.8	124	192.3	27	27.36	12.41	Concerning	Concerning	Yes
11	0.4583	86.79	104.5	139.4	64	25.92	9.444	Concerning	Concerning	Yes
12	0.2083	109.4	115.6	160.6	56	35.31	11.67	Extremely sick	Extremely sick	Yes
13	0.975	324.5	111	184.3	28	27.36	9.381	Healthy	Healthy	Yes
14	0.425	177.4	109.2	155.3	68	12.19	6.481	Very concerning	Very concerning	Yes
15	0.5917	137.7	115.6	144.7	28	26.64	18.33	Healthy	Healthy	Yes
16	0.4622	178	122.1	180.7	61	29	17.32	Extremely sick	Extremely sick	Yes
17	0.3292	81.13	121	180.7	24	30.07	2.075	Very sick	Very sick	Yes
18	0.3462	154.7	108.6	150	56	15.35	3.208	Sick	Sick	Yes
19	0.4991	284.9	120.1	176.4	40	30.07	1.698	Sick	Very concerning	No
20	0.4841	148.8	87.93	136.3	30	17.65	3.659	Concerning	Concerning	Yes
21	0.55	86.79	91.79	134.2	62	16.08	5.849	Healthy	Healthy	Yes
22	0.4311	160.4	85.19	128.9	90	3.22	19.06	Sick	Sick	Yes
23	0.3972	132.1	88.02	134.2	36	25.65	1.321	Healthy	Healthy	Yes
24	0.55	149.1	108.8	147.4	54	17.56	10.38	Concerning	Concerning	Yes
25	0.3462	81.13	85.19	126.2	42	30.8	10	Healthy	Very concerning	No
26	0.55	171.7	97.45	102.5	31	27	14.91	Healthy	Healthy	Yes
27	0.4651	154.7	105.9	102.5	28	10.02	8.491	Healthy	Healthy	Yes
28	0.3292	149.1	123.9	181.7	24	31.54	10	Extremely sick	Extremely sick	Yes
29	0.343	147.1	120.7	172	60	31.03	10.6	Extremely sick	Extremely sick	Yes
30	0.361	111.8	107.9	163.7	39	24.79	2.2	Healthy	Healthy	Yes
31	0.487	111.8	107.9	155.5	34	13.09	2.2	Concerning	Healthy	No
32	0.343	141.2	117.7	155.5	31	17.77	1	Healthy	Healthy	Yes

DBP: diastolic blood pressure, SBP: systolic blood pressure, CADT: control assessment by doctors' team, and CAFES: control assessment by fuzzy expert system.

TABLE 5: Confusion matrix for CKD.

Healthy	Concerning	Very concerning	Sick	Very sick	Extremely sick	Class names
15	00	00	00	00	00	Healthy
00	08	00	00	02	00	Concerning
00	00	15	01	00	00	Very concerning
01	00	00	10	00	00	Sick
00	00	00	00	14	00	Very sick
00	00	01	00	00	13	Extremely sick

expert system is able to diagnose the disease and can help specialist to provide an appropriate and suitable treatment for disease.

To portray the behaviour of two given variables in 3-dimensional surface, the response surface is used. It explains the probability of risk with which a person suffers on two considered variables. Three plots are considered in this work. First in Figure 11, it is the plot between BMI and nephron functionality. The response surface plot of variable BMI and nephron functionality is analysed by test 2 in which the value of BMI is 35 and similarly value of nephron functionality is 0.1. From plot, it can be estimated easily that at these values, the result is high i.e. extremely sick.

TABLE 6: Matrix with reduced dimensionality.

Yes	No	Class name
37	02	Yes
03	38	No

Furthermore, by Figures 12 and 13, the output can be estimated by considering blood sugar with nephron functionality and BMI with blood sugar, respectively. With a complete analysis of all the potential instances of test two response surface plots, and with grouping of seven parameters, it is found that the parameters with most extreme impact in diagnosis of CKD are nephron functionality, blood sugar, and BMI. These three variables have the highest effect

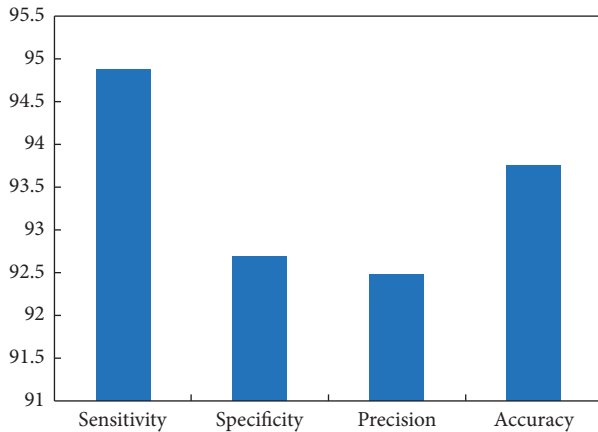


FIGURE 10: Performance of proposed system.

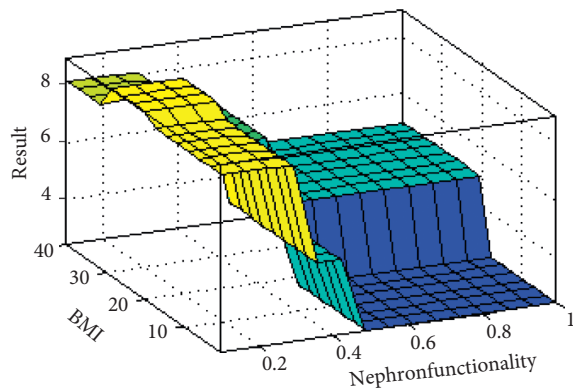


FIGURE 11: Response surface plot on test 2 with BMI and nephron functionality.

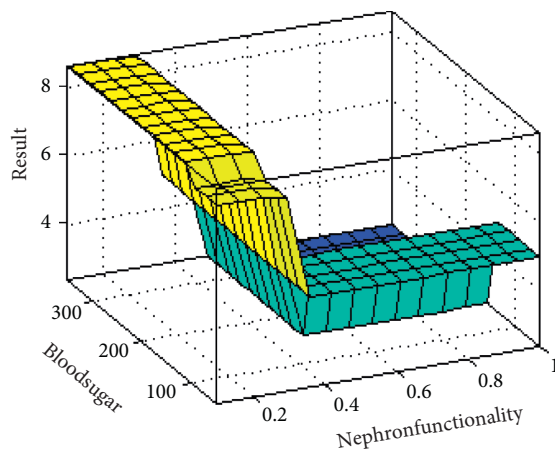


FIGURE 12: Response surface plot on test 2 with blood sugar and nephron functionality.

on patient's health in case of CKD. Therefore, doctors must give appropriate recommendation to patients to make their health stable and be cured in time from this CKD.

Figure 14 shows the input values of various input variables used in medical expert system for diagnosis of CKD and output corresponding to given inputs. Image-processing

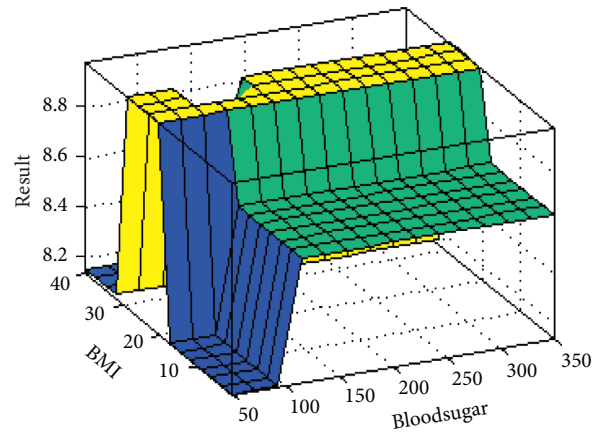


FIGURE 13: Response surface plot on test 2 with BMI and blood sugar.

techniques such as registration of an image, segmentation of an image, and extraction of features have also been applied on acquired magnetic resonance images and computed tomography images.

Additionally, the artificial neural network techniques such as backpropagation neural network and modular neural network are also used for the detection of CKD. Moreover, many systems have been proposed by using machine learning algorithms for the diagnosis of CKD such as  $k$  star and  $k$  nearest neighbor classifier to classify the stage of the disease into different classes according to the symptoms and risk factors.

All these abovementioned techniques deal with the crisp numbers. The developed system in this work deals with the fuzzy values, and it has also given more accuracy than these techniques.

## 6. Discussion

The developed fuzzy logic-based medical expert system is a supporting tool in decision-making about diagnosis of CKD. Although, specialist doctor's decision will be considered as final, the developed medical expert system can also make decision as a specialist doctor. Hence, it could be a very useful tool for assisting medical doctors for the robust decision regarding the CKD.

The fuzzy expert system can also be suitable in remote areas, such as villages, and small towns, as there is always scarcity of specialists in such places. This expert system can also help to train the new doctors who do not have enough experience to diagnose the CKD, because this fuzzy expert system has knowledge and information stored in knowledge base that was acquired from specialists of the disease. Hence, it can be used as a training tool as well.

The accuracy obtained from the developed system is 93.75%. It means the system is successful in identification of 93.75% of the total cases considered in this research work. Another performance parameter considered in this work is precision. The precision obtained from this developed system is 92.5%. It means the developed system is having high quality of giving exact and clear results. Similarly, the

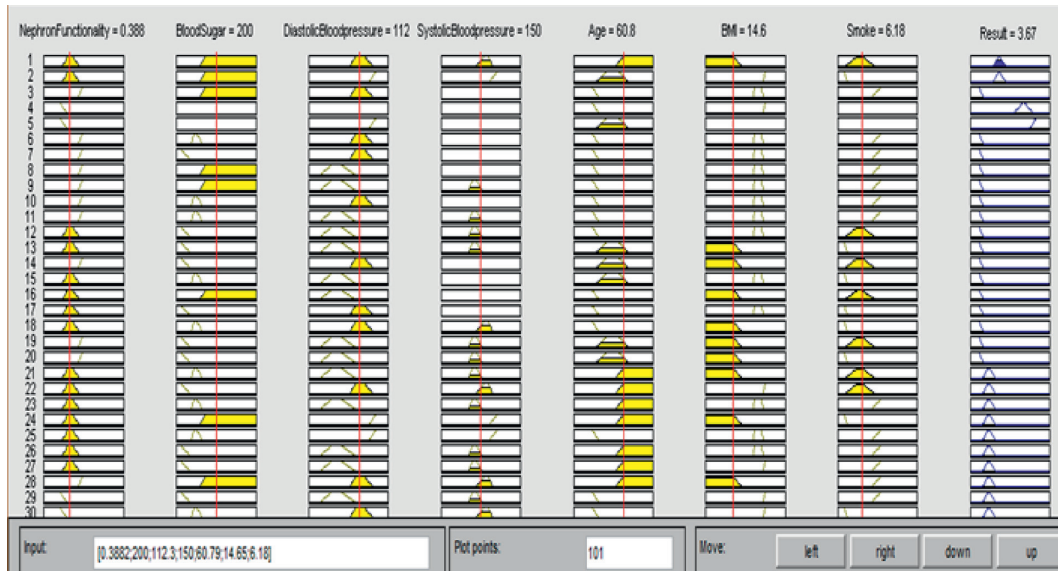


FIGURE 14: Mapping of rules for output.

sensitivity and specificity of developed system are 94.87% and 92.68%. It means developed system is capable of measuring the proportion of actual positives and negatives which are correctly determined as such.

The whole work done in this study is new and original as the clinical tests, and risk factors considered in this study have not been incorporated in the systems similar to it. The various research works are discussed in Introduction to understand more about risk factors, prevention, and treatment of CKD. Competitive advantage is that the proposed expert system is able to diagnose the disease with various variations in input variables. Variation states considered in this work are healthy, concerning, very concerning, sick, very sick and extremely sick.

The main contribution of this work is that it is a supporting tool which is capable to make decisions, or it supports the decision-making process according to the given inputs to system by the user to identify whether the patient is suffering from CKD or not. To improve the success rate of fuzzy expert system, the recommendation is to gather more knowledge about the disease from specialist doctor and make rules as enough as possible. One can also increase the number of variables used as input to recognize the CKD and make their range accordingly so that the result can be calculated more accurately.

### 7. Conclusion

The diagnosis of chronic kidney disease is a difficult task. There is always some possibility of misdiagnosis that leads to wrong treatment. This work proposes a fuzzy expert system that can help a doctor and nonspecialist to examine the chronic kidney patient.

The fuzzy expert system is based on clinical test ranges. It works on two types of information, that is, the knowledge acquired from specialist doctors and experts from the nephrology and urology departments. This study also gives

response surface plot of various input variables used to recognize the disease. This system can give correct results about the health status of a patient.

In the current state, it would be inappropriate to say that the developed fuzzy expert system will replace the specialist doctors or knowledge of the team of doctors. However, this could be a supporting tool for doctors that could assist in decision-making and does not replace their admirable work. In addition, only computer and software are required to implement this medical expert system. Therefore, this system can be used in hospitals, where there is lack of resources and can be used in those geographic areas where hospitals are not available.

### Data Availability

The data are available on request on demand through the first author jimmy.21733@lpu.co.in. or via phone +9779288330.

### 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 Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea Government (MSIT) (no.2018-0-01369).

### References

- [1] S. Varughese and G. Abraham, "Chronic kidney disease in India: a clarion call for change," *Clinical Journal of the American Society of Nephrology*, vol. 13, no. 5, pp. 802–804, 2018.

- [2] S. Kutia, S. H. Chauhdary, C. Iwendu, L. Liu, W. Yong, and A. K. Bashir, "Socio-technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems," *IEEE Access*, vol. 7, pp. 90777–90788, 2019.
- [3] G. P. Joshi, S. Acharya, C.-S. Kim, B.-S. Kim, and S. W. Kim, "Smart solutions in elderly care facilities with RFID system and its integration with wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 10, no. 8, Article ID 713946, 2014.
- [4] S. Varughese, "Chronic kidney disease in India a clarion call for change," *Clinical Journal of the American Society of Nephrology*, vol. 13, no. 5, pp. 802–804, 2018.
- [5] Y. Kubota, Y. Mitsukura, M. Fukumi, N. Akarnatsu, and M. Yasutomo, "Automatic extraction of a kidney region by using the q-learning," in *Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004*, pp. 536–540, IEEE, Seoul, South Korea, November 2004.
- [6] D. Turco, C. Corsi, S. Severi, and R. Mignani, "Assessment of kidney volumes in Polycystic Kidney Disease from Coronal and Axial MR Images," in *Proceedings of the 2013 8th International Symposium On Image And Signal Processing And Analysis (ISPA)*, pp. 528–532, IEEE, Trieste, Italy, September 2013.
- [7] H. Tang, J. L. Dillenseger, and L. M. Luo, "Intra subject 3D/3D kidney registration using local mutual information maximization," in *Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6379–6382, IEEE, Lyon, France, August 2007.
- [8] K. Bommana Raja and M. Madheswaran, "Determination of kidney area independent unconstrained features for automated diagnosis and classification," in *Proceedings of the 2007 International Conference on Intelligent and Advanced Systems*, pp. 724–729, Kuala Lumpur, Malaysia, 2007.
- [9] F. Iqbal, A. S. Pallewatte, and J. P. Wansapura, "Texture analysis of ultrasound images of chronic kidney disease," in *Proceedings of the 17th International Conference on Advances in ICT for Emerging Regions, ICTer 2017*, pp. 299–303, IEEE, Colombo, Sri Lanka, September 2017.
- [10] R. M. Pujari and V. D. Hajare, "Analysis of ultrasound images for identification of chronic kidney," in *Proceedings of the First International Conference on Networks and Soft Computing*, pp. 380–383, IEEE, Guntur, India, August 2014.
- [11] K. Bommana Raja and M. Madheswaran, "Determination of kidney area independent unconstrained features for automated diagnosis and classification," in *Proceedings of the 2007 International Conference On Intelligent And Advanced Systems, ICIAS*, pp. 724–729, IEEE, Kuala Lumpur, Malaysia, November 2007.
- [12] M. Ahmad, V. Tundjungsari, D. Widiandi, P. Amalia, and U. A. Rachmawati, "Diagnostic decision support system of chronic kidney disease using support vector machine," in *Proceedings of the 2nd International Conference on Informatics and Computing, ICIC*, IEEE, Kuala Lumpur, Malaysia, November 2017.
- [13] W. H. S. D. Gunarathne, K. D. M. Perera, and K. A. D. C. P. Kahandawaarachchi, "Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD)," in *Proceedings of the 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering, BIBE*, pp. 291–296, IEEE, Washington, DC, USA, October 2017.
- [14] M. S. Wibawa, I. M. D. Maysanjaya, and I. M. A. W. Putra, "Boosted classifier and features selection for enhancing chronic kidney disease diagnose," in *Proceedings of the 2017 5th International Conference on Cyber and IT Service Management, CITSM*, IEEE, Denpasar, Indonesia, August 2017.
- [15] U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier," in *Proceedings of the 2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC*, IEEE, Chennai, India, December 2016.
- [16] E. Avci, S. Karakus, O. Ozmen, and D. Avci, "Performance comparison of some classifiers on chronic kidney disease data," in *Proceedings of the 6th International Symposium on Digital Forensic and Security, ISDFS 2018*, IEEE, March 2018.
- [17] R. K. Chiw, R. Y. Chen, S. Wang, and S. Jian, "Intelligent systems on the cloud for the early detection of chronic kidney disease," in *Proceedings of the 2012 International Conference on Machine Learning and Cybernetics*, IEEE, Xian, China, July 2012.
- [18] V. Kunwar, K. Chandel, A. S. Sabitha, and A. Bansal, "Chronic kidney disease analysis using data mining classification," in *Proceedings of the Cloud System And Big Data Engineering (Confluence), 2016 6th International Conference*, IEEE, Noida, India, January 2016.
- [19] T. Adam, U. Hashim, and U. S. Sani, "Designing an artificial neural network model for the prediction of kidney problems symptom through patient's metal behavior for pre-clinical medical diagnostic," in *Proceedings of the 2012 International Conference on Biomedical Engineering, ICoBE 2012*, IEEE, Penang, Malaysia, February 2012.
- [20] L. Xun, N. Li, X. Wu, and T. Lou, "Application of radial basis function neural network to estimate glomerular filtration rate in Chinese patients with chronic kidney disease," in *Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, IEEE, Taiyuan, China, October 2010.
- [21] M. A. Venkatachalam, K. A. Griffin, R. Lan, H. Geng, P. Saikumar, and A. K. Bidani, "Acute kidney injury: a springboard for progression in chronic kidney disease," *American Journal of Physiology-Renal Physiology*, vol. 298, no. 5, pp. F1078–F1094, 2010.
- [22] E. Yudaningtyas, D. H. Santjojo, W. Djurianto, I. Siradjuddin, and M. R. Hidayatullah, "Identification of pulse frequency spectrum of chronic kidney disease patients measured at TCM points using FFT processing," in *Proceedings of the 2017 15th International Conference on Quality in Research (QIR): International Symposium on Electrical and Computer Engineering*, pp. 169–172, IEEE, Nusa Dua, Indonesia, July 2017.
- [23] C.-J. Lin, C.-Y. Chen, P.-C. Wu et al., "Intelligent system to predict intradialytic hypotension in chronic hemodialysis," *Journal of the Formosan Medical Association*, vol. 117, no. 10, pp. 888–893, 2018.
- [24] S. D. Arasu and R. Thirumalaiselvi, "A novel imputation method for effective prediction of coronary kidney disease," in *Proceedings of the 2017 2nd International Conference on Computing and Communications Technologies, ICCCT*, pp. 127–136, IEEE, Chennai, India, February 2017.
- [25] S. R. Raghavan, V. Ladik, and K. B. Meyer, "Developing decision support for dialysis treatment of chronic kidney failure," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 2, pp. 229–238, 2005.
- [26] C. Paper and N. Dey, "Hybrid modified cuckoo search-neural network in chronic kidney disease classification hybrid modified cuckoo search-neural network in chronic kidney

- disease classification,” in *Proceedings of the 2017 14th International Conference on Engineering of Modern Electric Systems (EMES)*, pp. 164–167, IEEE, Oradea, Romania, June 2017.
- [27] C. I. Bondor, I. M. Kacso, A. R. Lenghel, and A. Muresan, “Hierarchy of risk factors for chronic kidney disease in patients with type 2 diabetes mellitus,” in *2012 IEEE 8th International Conference on Intelligent Computer Communication and Processing*, pp. 103–106, IEEE, Cluj-Napoca, Romania, August 2012.
- [28] A. Salekin and J. Stankovic, “Detection of chronic kidney disease and selecting important predictive attributes,” in *Proceedings of the 2016 IEEE International Conference on Healthcare Informatics, ICHI*, pp. 262–270, IEEE, Chicago, IL, USA, October 2016.
- [29] T. Shen, T. Fang, Y. Ou, and C. Wang, “Low-cost detection of cardiovascular disease on chronic kidney disease and dialysis patients based on hybrid heterogeneous ECG features including T-wave alternans and heart rate variability,” in *Proceedings of the 2010 Computing in Cardiology*, IEEE, Belfast, UK, pp. 561–564, September 2010.
- [30] R. B. Fricks, A. Bobbio, P. Orientale, and K. S. Trivedi, “Reliability models of chronic kidney disease,” in *Proceedings of the 2016 Annual Reliability and Maintainability Symposium (RAMS)*, January 2016.
- [31] I. Council, S. Lanka, and A. Pradesh, “Prevalence of chronic kidney disease in India - where are we heading?” *Indian Journal of Nephrology*, vol. 25, no. 3, 2015.
- [32] A. Batra, U. Batra, and V. Singh, “A review to predictive methodology to diagnose chronic kidney disease,” in *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, New Delhi, India, March 2016.
- [33] A. Sobrinho, L. Dias Da Silva, M. E. Pinheiro, P. Cunha, A. Perkusich, and L. Medeiros, “Formal specification of a tool to aid the early diagnosis of the chronic kidney disease,” in *Proceedings of the 2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, pp. 173–178, IEEE, Santiago, Chile, October 2015.
- [34] Q.-L. Zhang and D. Rothenbacher, “Prevalence of chronic kidney disease in population-based studies: systematic review,” *BMC Public Health*, vol. 8, no. 1, 2008.
- [35] H. Ahmadi, M. Gholamzadeh, L. Shahmoradi, M. Nilashi, and P. Rashvand, “Diseases diagnosis using fuzzy logic methods: a systematic and meta-analysis review,” *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 145–172, 2018.
- [36] S. Sivathasan, F. Cecelja, and W. Balachandran, “ECG diagnosis using neural network and fuzzy expert system,” in *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference [Cat. No. 00CH37066]*, pp. 988–992, IEEE, Baltimore, MD, USA, May 2000.
- [37] E. Rovența and G. Roșu, “The diagnosis of some kidney diseases in a small prolog expert system,” in *Proceedings of the 2009 3rd International Workshop on Soft Computing Applications, SOFA*, pp. 219–224, IEEE, Arad, Romania, August 2009.
- [38] S. Das and P. K. Ghosh, “Hypertension Diagnosis: a comparative study using fuzzy expert system and neuro fuzzy system,” in *Proceedings of the 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2013.
- [39] R. Meza-palacios, A. A. Aguilar-lasserre, E. L. Ureña-Bogarín, C. F. Vázquez-Rodríguez, R. Posada-Gómez, and A. Trujillo-mata, “Development of a fuzzy expert system for the nephropathy control assessment in patients with type 2 diabetes mellitus,” *Expert Systems With Applications*, vol. 72, no. 1, pp. 335–343, 2017.
- [40] G. Ahmad, M. A. Khan, S. Abbas, A. Athar, B. S. Khan, and M. S. Aslam, “Automated Diagnosis of Hepatitis B Using Multilayer Mamdani Fuzzy Inference System,” vol. 2019, Article ID 6361318, 11 pages, 2019.
- [41] A. Ali, L. Abbas, M. Shafiq et al., “Hybrid fuzzy logic scheme for efficient channel utilization in cognitive radio networks,” *IEEE Access*, vol. 7, pp. 24463–24476, 2019.
- [42] S. Ahmed, T. Kabir, N. T. Mahmood, and R. M. Rahman, “Diagnosis of Kidney Disease Using Fuzzy Expert System,” in *Proceedings of the 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*, IEEE Dhaka, Bangladesh, December 2014.
- [43] S. Tapaswini, C. Mu, D. Behera, and S. Chakraverty, “Solving imprecisely defined vibration equation of large membranes,” *Engineering Computations*, vol. 34, no. 8, pp. 2528–2546, 2017.
- [44] N. Chetty, K. S. Vaisla, and S. D. Sudarsan, “Role of attributes selection in classification of chronic kidney disease patients,” in *Proceedings of the 2015 International Conference on Computing, Communication and Security (ICCCS)*, IEEE, Pamplemousses, Mauritius, December 2015.
- [45] A. K. Singh, Y. M. K. Farag, B.V. Mittal et al., “Epidemiology and risk factors of chronic kidney disease in India—results from the SEEK (screening and early evaluation of kidney disease) study,” *BMC Nephrology*, vol. 14, no. 1, pp. 1–10, 2013.
- [46] K. R. D. Hernández, A. A. A. Lasserre, R. P. Gómez, J. A. P. Guzmán, and B. E. G. Sánchez, “Development of an expert system as a diagnostic support of cervical cancer in atypical glandular cells, based on fuzzy logics and image interpretation,” vol. 2013, Article ID 796387, 17 pages, 2013.



## Research Article

# New Method for Forest Resource Data Collection Based on Smartphone Fusion with Multiple Sensors

Guangpeng Fan,<sup>1,2</sup> Yanqi Dong,<sup>1,2</sup> Danyu Chen,<sup>1,2</sup> and Feixiang Chen <sup>1,2</sup>

<sup>1</sup>School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China

<sup>2</sup>Engineering Research Center for Forestry-oriented Intelligent Information Processing, National Forestry and Grassland Administration, Beijing 100083, China

Correspondence should be addressed to Feixiang Chen; [bjfxchen@bjfu.edu.cn](mailto:bjfxchen@bjfu.edu.cn)

Received 4 February 2020; Revised 19 April 2020; Accepted 6 May 2020; Published 18 May 2020

Academic Editor: Sungchang Lee

Copyright © 2020 Guangpeng 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.

Tree parameter measurement is an important part of forest resource monitoring. Smartphones play an important role in forest resource surveys. Although sensors inside smartphones, such as gyroscopes and angle sensors, can meet the needs of the public for entertainment or games, the measurement accuracy in professional forest resource monitoring is slightly insufficient. In this paper, a method of collecting tree measurement factors based on personal smart space fusion with a variety of high-precision sensors is proposed. First of all, a high-precision attitude sensor measurement module and a laser ranging module are organically integrated and packaged in a black box. The smartphone is then connected to the sensor box using a magnet sheet, and the working personnel can determine key parameters in the forest stand by holding it. Finally, in order to verify the accuracy of the method, the measured values in this paper are compared with the reference values. The root mean square error (RMSE) of the tree position in the X and Y directions was 0.114 m and 0.147 m, the relative deviations (rBias) were 0.95% and 0.39%, and the average RMSE was 0.186 m. The RMSEs measured by tree height and diameter at breast height (DBH) were 0.98 m and 2.24 cm, the relative root mean square error (rRMSE) was 5.87% and 13.46%, and the relative deviations (rBias) were -1.40% and -1.06%, respectively. Therefore, the method of forest stand parameter measurement based on personal smart space fusion multitype sensors proposed in this paper can be effectively applied to forest resource data collection.

## 1. Introduction

Forests play an important role in maintaining ecosystem balance, protecting the environment, responding to climate change, sequestering carbon, maintaining soil and water, and providing forest products [1–3]. Mastering and understanding detailed forest information can help us strengthen forest protection. Some researchers took advantage of smartphones with various sensors such as angle sensors, orientation sensors, and gyroscopes and applied smartphones to forest resource monitoring [4, 5]. Some researchers proposed to measure key tree measurement factors such as tree species, diameter at breast height (DBH), tree height, and tree position based on smartphones and store the measured tree measurement factors in the database

of the smartphone [6, 7]. Some researchers developed tree height measurement software running on smartphones using the principle of trigonometric functions. These studies can theoretically measure accurate geometric lengths, but many smartphones have large deviations in sensor accuracy such as angle sensors and direction sensors [8–10]. In general, smartphones can meet the accuracy requirements of mass entertainment or game consumption. However, when applied to professional fields such as forest resource monitoring, the sensitivity of the sensors built in the smartphones will cause unstable measurement errors [11]. Some researchers have used the Global Navigation Satellite System (GNSS) function of smartphones in forest stand positioning, but such methods have significant limitations and are greatly affected by stand density [12–14]. Especially in some forest

areas where the GNSS signal is weak or there is no signal, the positioning accuracy of the smartphone is low or the positioning function cannot be used. Some researchers combined smartphones, laser rangefinders, and tripods to work in the forest [15]. Although this method can meet the functional requirements of forest resource survey, it can also lead to the problem of poor portability in practical work, especially in the face of hillside fields, which will affect the observation of workers.

At present, UAV, light detection and ranging (LiDAR), remote sensing, and photogrammetry have gradually become important technologies for forest resource monitoring [16–18]. A forest resource monitoring system integrating sky and ground and air has been formed and is applicable to different work scenarios [19]. Among the existing forest ground monitoring methods, in addition to the traditional tree measuring instruments, terrestrial photogrammetry, ground-based LiDAR, and other technologies have also emerged, which can not only obtain accurate tree location, tree height, DBH, and other basic tree measurement factors but also promote the computerization of forestry [20–22]. LiDAR technology also has some problems in forest resource monitoring, such as complicated operation, high investigation cost, poor portability, and large computation [23]. Photogrammetry is relatively cheap compared with LiDAR, but photogrammetry requires multiple independent digital cameras, and the measurement results need to be returned to the office for postprocessing [24], which cannot be solved in real time on work field. Therefore, LiDAR technology has many limitations in small-scale or low-cost forest surveys [25]. The tree position is necessary to accurately calculate and measure the characteristics of the forest and describe the changes in the forest [26]. Especially in the work of establishing a fixed plot, it is necessary to obtain accurate tree positions [27,28]. In the early investigation stage, the relative position of the trees was mainly determined by using the compass to obtain the azimuth and the rope to obtain the horizontal distance [29]. With the application of GNSS in forestry, GNSS-based stand positioning technology has gradually matured. Therefore, GNSS positioning technology based on smartphones will also produce large measurement errors in dense forest stands [30]. It is necessary to develop a relative positioning method to measure the position of trees based on smartphones. The DBH and tree height of standing trees are important contents of forest resource surveys and are also the basic factors for estimating forest accumulation and biomass [31, 32]. Taking advantage of the relatively low price of photogrammetry, it is necessary to develop a smartphone-based photogrammetry method to obtain the DBH and height of trees. In the current research on forest resource monitoring based on smartphones, there are problems of unstable measurement errors due to low accuracy of the sensor, as well as problems of single function or poor portability [4, 8, 33].

In the measurement of tree attributes, we need to obtain the key parameters of the standing tree (including factors such as position, DBH, and tree height). In order to make smartphones more convenient to obtain high-precision azimuth and horizontal distance, this paper proposes a

method for measuring tree attributes based on personal smart space fusion with multiple sensors. Considering the existence of hillside fields in forest resource monitoring, from the perspective of portability, a variety of high-precision sensors (including angle sensors, orientation sensors, gyroscopes, and laser ranging modules) are integrated into a small box that is easy to carry, and this paper defines it as “sensor box.” The “sensor box” is tightly connected to the back of the smartphone through a magnet sheet, which makes it easier for forest workers to carry around. The purposes of this paper are as follows. (1) In view of the lack of accuracy or sensitivity of the sensors of smartphones in forest resource monitoring, a method of measuring key parameters of forest stands by integrating a variety of high-precision sensors based on smart space is proposed. (2) The position, DBH, and the height of the tree are determined using a relative position positioning algorithm and a standing tree photo processing algorithm. In order to implement the method in this paper, smartphone-based forest stand measurement software was developed to verify the accuracy of the method and the applicability of the method was discussed.

## 2. Materials and Methods

*2.1. Sensor Box.* In order to improve the problem of low sensitivity of the sensor due to different hardware configurations of the smartphone, this paper has designed a multifunctional sensor called “sensor box” for forest resource monitoring. The sensor box has an attitude measurement function and a distance measurement function. It is mainly used to obtain high-precision 3D data (acceleration, gyroscope, Euler angle, and magnetic field data) and distance measurement data. It can send various types of sensor data to smartphones via Bluetooth. The sensor box overcomes the problem that the accuracy of the sensors built in many smartphones cannot meet the needs of investigation. It differs from the conventional laser rangefinder in the following two aspects. (1) Ordinary laser rangefinders focus on acquiring distance data and cannot acquire gyroscope attitude data. The sensor box designed in this article not only has the function of laser ranging but also has the function of attitude measurement. (2) The ordinary laser rangefinder cannot be physically connected to the smartphone and is considered as a whole. It is necessary for the workers to operate the rangefinder and the smartphone separately. The volume of the sensor box is  $80.3 \times 47 \times 30 \text{ mm}^3$ , which is 1/2 of the volume of a conventional laser rangefinder. It can be seamlessly connected to a smartphone through a magnet sheet, and their combination can be considered as a whole.

In this article, a variety of high-precision sensor modules on the market are organically integrated and packaged in a small black box (see Figure 1). The sensor box is mainly composed of an attitude measurement module and a laser ranging module. The attitude measurement module uses the gyro angle measurement module BWT901BLE5.0 provided by WitMotion ShenZhen Co., Ltd. Its X-axis and Y-axis angle measurement accuracy is  $0.01^\circ$ , and the Z-axis angle measurement accuracy is  $0.05^\circ$ . The operating frequency is

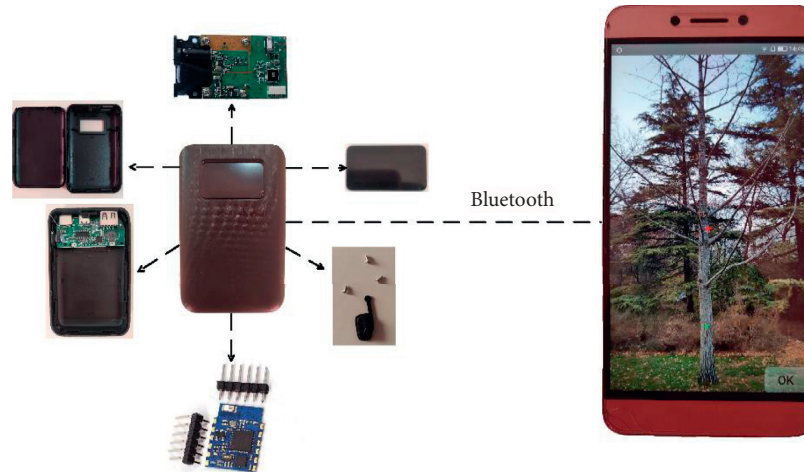


FIGURE 1: Sensor box and personal smartphone.

168 MHz, the acceleration is  $\pm 16g$ , the X-axis and Z-axis angle measurement range is  $-180^{\circ}\sim 180^{\circ}$ , and the Y-axis measurement range is  $-90^{\circ}\sim 90^{\circ}$ . The baud rate is 2400~921600 bps, and the magnetic field accuracy is 1 mG (milligauss). The return rate is 0.1~200 Hz. The model of the laser rangefinder module is L4-RS232, which has a measuring range of 0.02~100 meters, a measurement accuracy of  $\pm 2$  mm, and a laser wavelength and power of 620~690 nm. In Windows 10 operating system, based on Arduino 1.8.5 development environment, C++ programming language is used for serial debugging and secondary development. The sensor box uses a 3.3 v power supply and is equipped with a display to show sensor data.

**2.2. Development of Software for Measuring Tree Attributes Based on Personal Smart Spaces.** This article has developed software that runs on a smartphone to measure the stand. It can accept and process sensor data in the “sensor box” and can calculate the relative position, DBH, and tree height of standing trees. Under Windows 10 operating system, an application development environment based on Android Studio 3.3+ Android SDK (Java Development Kit)+ Java JDK 8 (Java Development Kit)+ ADT (Android Development Tool) is built. The stand measurement software was written in Java language. The software runs on a smartphone based on the Android operating system (Android 4.4 or higher). The application uses the distance and azimuth provided by the sensor box to calculate the relative position of the trees and uses the gyroscope data to calibrate the smartphone’s camera and obtain a fixed focal length. This paper is based on the principle of single image solution and uses the software’s “standing tree measurement” module to calculate the tree height and DBH. This article designs a way to connect the sensor box with a smartphone. The measurement personnel can carry out the forest parameter measurement by holding it.

**2.3. Research Area.** The research in this paper was conducted in an artificially regenerated forest located in

Haidian District ( $40^{\circ}0'40''N$ ,  $116^{\circ}20'20''E$ ), Beijing, China. It is planted with poplar, willow, and locust trees, with little or no undergrowth grass and shrubs. The study area is 48.7 m above sea level and the terrain is flat. It belongs to the temperate humid monsoon climate zone, with four distinct seasons, hot summers, cold winters, and low precipitation. As shown in Figure 2, this paper has verified the accuracy of the smart space-based method of fusing multiple sensors to measure stands in two temporary sample plots of  $40 \times 40$  m. The tree density of plantation stands is around 900 trees/ha. To verify the accuracy of the tree location measurement, a total of 67 standing trees were measured. To verify the accuracy of DBH and tree height measurements, a total of 96 standing trees were measured.

**2.4. Tree Key Parameter Measurement Principle Based on Smart Space.** The sensor box provides a variety of high-precision sensor data for the determination of tree position, DBH, and tree height in this paper. The relative position determination algorithm and standing tree image measurement algorithm are used to process the sensor data collected by the smartphone, and the position, DBH, and height of the standing tree are calculated, respectively. In forest areas with high stand density or severe canopy cover, the GNSS signal strength will be severely affected, resulting in low forest positioning accuracy or inability to meet actual work requirements. Since the GNSS signal of the smartphone into the forest will be severely weakened, it is impossible to directly obtain the absolute position of the trees. Under the guidance of traditional tree position measurement methods [29], this paper provides a new option for forestry investigators to determine the relative position of trees based on smart space fusion with multiple sensors. Based on the principle of close-range photogrammetry [34, 35], this paper uses the fixed focal length of a smartphone and the horizontal distance obtained by the sensor box. Then, the height and DBH of the tree can be measured by taking a picture containing the complete information of the standing tree.

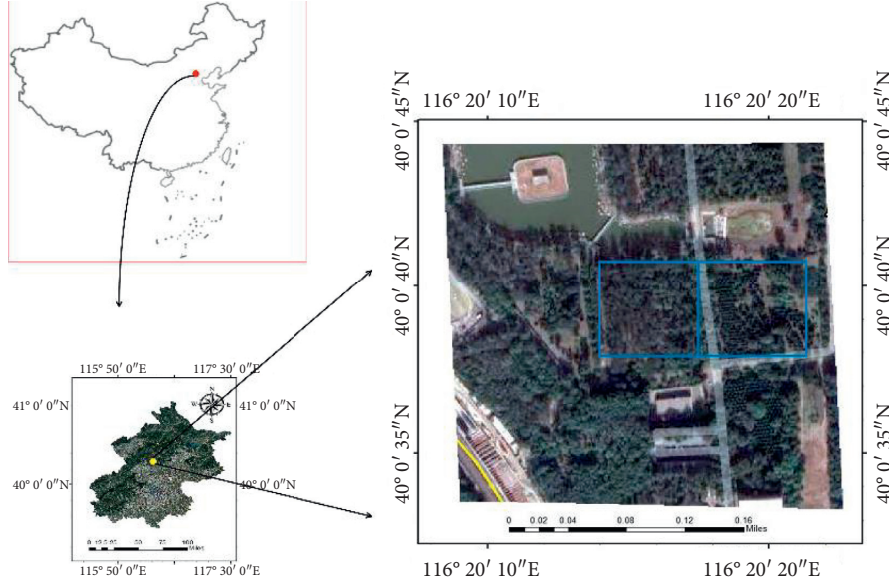


FIGURE 2: Research area map.

#### 2.4.1. Principle of Standing Tree Position Measurement.

First, the surveyor should set the position to the origin (the origin is the initial observation point, whose coordinates are known or assumed), and the location of the origin should be selected as far as possible to the place with wider surrounding field of view. Then, the surveyor opens the “forest stand measurement” software in the smartphone and enters the camera shooting interface of the “standing tree measurement” function and clicks the “Bluetooth” icon in the upper right corner to search and connect the sensor box. According to the actual requirements of the work, if there is a need to measure DBH and tree height, the laser emitted horizontally by the induction box can be projected to DBH (1.3 m) of the trunk by visual judgment. If you just need to measure the position of the tree, you can project the laser horizontally onto the tree trunk anywhere. When the laser point appears in the camera interface, click the “OK” button and keep the arm as stable as possible.

In Figure 3, the tree position coordinate system  $O-xy$  is established, and the origin  $O$  of the coordinate system is the first observation point. The red triangle represents the observation point, while  $O_x$  and  $O_y$  represent the coordinates of the observation point. The green dots represent trees, and  $x$  and  $y$  represent tree coordinates.  $L$  represents the distance from the observation point provided by the calculation box to the tree.  $\alpha$  is the azimuth calculated by the induction box. The coordinates of each tree can be calculated using trigonometric functions. For example, the plane coordinates of the first tree are calculated as follows, and the coordinates of other trees are calculated using a similar method.

$$\begin{cases} x_1 = L_1 * \cos(\pi - \alpha_0), \\ y_1 = L_1 * \sin(\pi - \alpha_0). \end{cases} \quad (1)$$

When measuring trees at the parcel level, surveyors may encounter situations in which trunks or treetops are blocked due to the density of trees, and some trees are partially visible

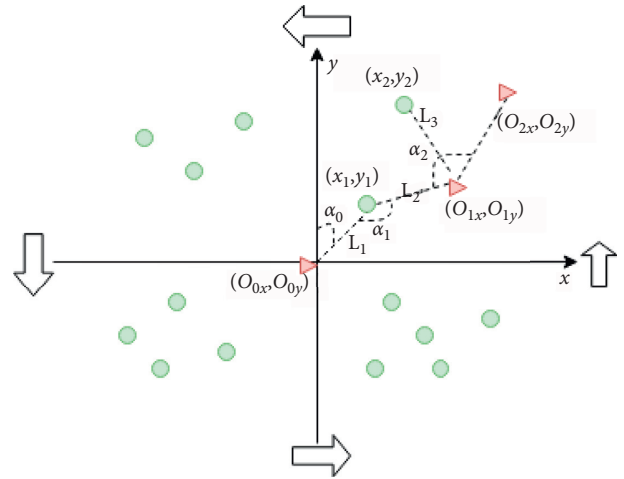


FIGURE 3: Schematic diagram of tree position observation scheme.

or completely invisible. In this paper, an observation mode of “observation point transfer” is designed for surveyors. The measured stands can be divided into two or four plots of similar size according to the density of trees or artificial segmentation markers. The observation principle is to transition from low-density plots to high-density plots. When dividing into two plots, the plot with less dense trees should be observed first. Each plot occupies a quadrant when divided into four cells. Firstly, the plot with the lowest density of trees was taken as the first quadrant, and the observation sequence of anticlockwise or clockwise was determined according to the density of trees in adjacent plots. When the entire stand is divided, the surveyor can start the determination of each stand. If the laser emitted by the sensor box cannot be projected onto the trunk, the surveyor can change the standing position and set up a second observation point. The selection of the second observation point is the same as the first observation point.

Using the distance data  $L$  provided by the sensor box, the coordinates of the second observation point can be inverted according to the coordinates of the previous tree or the coordinates of the measured tree nearby, and then continue the observation. According to the actual needs, the third or fourth observation points are set in the same way.

#### 2.4.2. DBH and Tree Height Measurement of Standing Trees.

DBH and tree height are the key parameters for stand determination. In this paper, the software's "measuring tree" function is used to calculate the chest diameter and tree height of the tree to be measured (see Figure 4). The observation scheme was similar to the observation scheme of tree position. Therefore, this paper realizes the synchronization of measuring the position, DBH, and height of the tree in the actual working process, and the sensor box can provide the data required for calculation at the same time. The tree height and DBH can be calculated based on the method of solving single-tall photos. This is a method of single photo in photogrammetry. After entering the "standing tree measurement" module of this software, the surveyor selects the "Image Capture" function. The surveyor will project the laser emitted by the sensor box to the chest diameter (1.3 m) of the trunk by visual judgment and click the "OK" button. The image of tree will be automatically loaded into the software interface and a red "+" symbol will appear in the center of the smartphone screen. To obtain the exact image plane coordinates of the tree top position, the image of the tree can be moved so that the position  $A_1$  of the tree is aligned with the "+" symbol, and click the "OK" button in the lower right corner. Then, the image is moved to align the bottom of the tree  $A_2$  with the "+" symbol. Click the "OK" button to get the image plane coordinates of the point in the image. At this time, the measured value of tree height will appear on the screen. The measurement process of DBH is similar to that of tree height. The location of DBH is determined by the position of laser point in the photo. Then, by zooming in the image and moving the red "+" symbol,  $A_3$  to the left of DBH and  $A_4$  to the right of DBH are determined to complete the acquisition of corresponding coordinates of the image plane. Finally, the DBH measurement value will also appear on the screen.

This algorithm is a special form of collinear equations in photogrammetry and is suitable for calculation of a single image [35–37]. This paper studies the application of this algorithm to smartphones. The measurement principle is shown in Figure 5.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R \begin{bmatrix} x \\ y \\ -f \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ -f \end{bmatrix}, \quad (2)$$

where  $x$  and  $y$  are the image plane coordinates of the image points,  $f$  is the focal length of the camera of the smartphone,  $X$ ,  $Y$ , and  $Z$  are the object space coordinates of the observation points, and  $a_i$ ,  $b_i$ , and  $c_i$  ( $i = 1, 2, 3$ ) are the 9 direction cosines composed of three external azimuth elements of the tree image.



FIGURE 4: Software interface of tree height and DBH measurement.

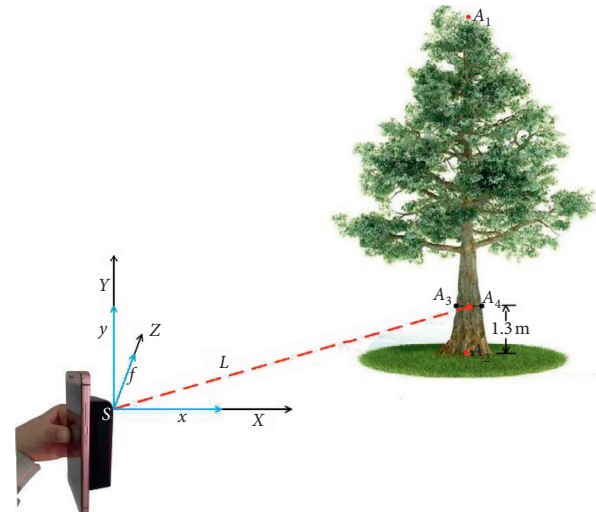


FIGURE 5: Schematic diagram of standing tree measurement.

In this scenario, the origin of the object space coordinate system and the origin of the image space coordinate system coincide, and the photography center  $S$ , the image space coordinate point, and the object space coordinate point are on a straight line to simplify the rotation matrix, as shown in the following formula:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \lambda_i \begin{bmatrix} x_i \\ y_i \\ f \end{bmatrix}. \quad (3)$$

The object space point coordinates of space point  $A_1$  is set as  $(X_A, Y_A, Z_A)$ . Formula (4) can be used to calculate the object-side coordinates of the image points on the standing tree image.

$$\begin{cases} X_i = \frac{Lx_i}{f}, \\ Y_i = \frac{Ly_i}{f}, \\ Z_i = L, \\ \lambda = \frac{L}{f}, \end{cases} \quad (4)$$

where  $(X_i, Y_i, Z_i)$  is the object space coordinates of the standing tree,  $L$  represents the horizontal distance between the observation point and the standing tree,  $x_i$  and  $y_i$  are image plane coordinates (2D), and  $f$  is the focal length of the camera. The tree height calculation formula is obtained by combining the following formula:

$$\begin{cases} \frac{X_i}{x_i} = \frac{Y_i}{y_i} = \frac{Z_i}{f} = \lambda, \\ \frac{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}{\text{height}} = \frac{L}{f}, \end{cases} \quad (5)$$

and the tree height is

$$\text{height} = \frac{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} * f}{L}, \quad (6)$$

where  $x_i, y_i$  corresponds to the image plane coordinates of DBH and height of the tree. The tree top  $A_1$  and the tree bottom  $A_2$  are selected, respectively, on the image of the tree to be measured. The corresponding image plane coordinates on the image are  $(x_1, y_1)$  and  $(x_2, y_2)$ , respectively at  $S$ .  $Z_i$  is the scale parameter  $(L/f)$ ,  $f$  is the focal length of the smartphone's fixed-focus camera, and  $L$  is the horizontal distance between the observation point and the tree to be measured.

Similarly, by selecting the left point  $A_3$  and right point  $A_4$  of DBH on the image, the corresponding image point coordinates (plane coordinates of the image) at  $s$  of the image, namely,  $(x_3, y_3)$  and  $(x_4, y_4)$ , are obtained, respectively. Then, we use the following formulas to calculate the DBH:

$$\begin{cases} \frac{X_i}{u_i} = \frac{Y_i}{v_i} = \frac{Z_i}{f} = \lambda, \\ \frac{\sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2}}{\text{DBH}} = \frac{L}{f}, \end{cases} \quad (7)$$

$$\text{DBH} = \frac{\sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2} * f}{L}. \quad (8)$$

In the study of measuring standing tree with smartphones, some researchers measured tree based on the principle of trigonometry. In this paper, photogrammetry algorithm was used to measure tree height and DBH. This algorithm requires a known external variable (horizontal distance  $L$  between the observation point and the tree) as a

“ruler,” and its advantages are reflected in the following two points. (1) The trigonometric function tree measurement algorithm needs to measure the observation point and the tree top and bottom, respectively. The angle of the smartphone's angle sensor is relatively low, and the angle error fluctuates greatly. The laser rangefinder only needs to measure the horizontal distance between the observation point and the stand. The accuracy of the distance measurement is very high and the error fluctuation is relatively small. (2) The algorithm in this paper can be used to measure the DBH and the tree height at the same time by using a smartphone to take a photo containing the complete information of the standing tree. However, the method based on the trigonometric function principle needs to find the tree height before using other algorithms to calculate the DBH.

**2.5. Precision Evaluation.** In order to verify the validity of this method [38, 39], the measurement accuracy of tree position, DBH, and tree height was analyzed in this paper. In terms of tree position measurement, we take the tree position measured with a total station (KTS-44R4LCN total station, South Surveying & Mapping Technology Co., Ltd) as a reference value. This paper uses the CGQ-1 direct reading altimeter and DBH tape used in forestry to measure the height and DBH of each tree and uses these data as the reference value of the method in this paper. This paper uses Bias, RMSE, rBias, and rRMSE to check the accuracy of tree position, tree height, and DBH.

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (y_i - y_{ri}),$$

$$\text{RMSE} = \sqrt{\frac{\sum (y_i - y_{ri})^2}{n}}, \quad (9)$$

$$\text{rBias\%} = \frac{\text{Bias}}{\bar{y}_r} \times 100\%,$$

$$\text{rRMSE\%} = \frac{\text{RMSE}}{\bar{y}_r} \times 100\%.$$

### 3. Results

**3.1. Analysis of Tree Position Results.** In this paper, 67 trees were measured to verify the accuracy of tree position measurements. During the measurement, the “observation point transfer” scheme in 2.4.1 section was used to solve the problem of trunk occlusion. The total station was used to measure the coordinates of 67 trees and took them as reference values. The position coordinates of trees measured by the fusion of multiple sensors based on personal smart spaces were taken as the measured values.

Figure 6(a) shows a schematic diagram of the distribution of trees under test in the sample plot. It can be clearly seen from Figure 6(b) that the measured value of tree position in this paper can better correspond to the reference

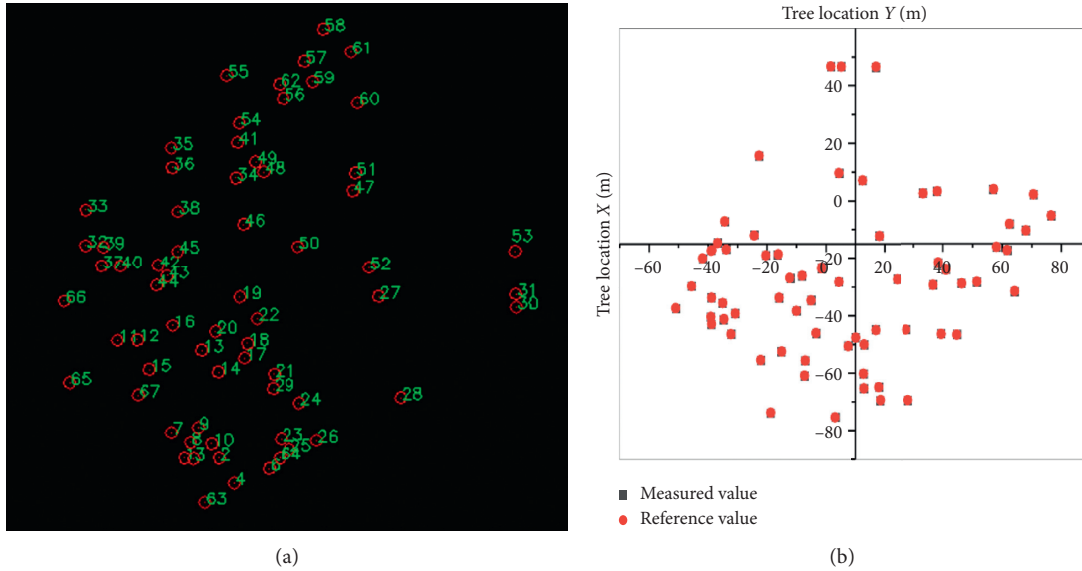


FIGURE 6: (a) The distribution of trees in the real world and (b) the comparison of tree position measurements.

value within the scope of sample land. Table 1 shows the Bias, rBias, RMSE, and rRMSE of X and Y, respectively.

The average RMSE of tree position in this experiment is 0.186 m, which can better express the distribution of tree position in forest resource monitoring. In practical work, in addition to external factors such as topography and magnetic field, the measurement accuracy of tree position will be affected by the way the surveyors operate. Therefore, the measurement operators should regulate the operation of the instrument, avoid bad measurement environment, and carry out accurate positioning and measurement according to the instructions for the use of the software. The experimental results show that this method can be used to measure the position of trees in forest resource monitoring.

**3.2. Analysis of DBH and Height of Standing Tree.** In order to verify the accuracy of this method in measuring the DBH and height of standing trees, 96 living standing trees were measured and these results were taken as measured values. The height and DBH of the standing tree were measured using the CGQ-1 direct reading altimeter and diameter tape, and these results were used as reference values. Figure 7 shows that the measured values of DBH and tree height are evenly distributed on both sides of the reference value, and the measurement effect is better overall.

The four precision indexes in Table 2 show the measurement results of tree height and DBH. The experimental results show that the stand measurement method based on smart space fusion of multiple sensors can meet the precision requirements of forest resource survey.

In terms of measurement accuracy, although the method in this paper has no obvious advantages compared with existing methods such as LiDAR and photogrammetry, it still provides valuable reference for forest resource monitoring. Therefore, the technology based on smart space fusion of multiple sensors can still be used as an effective

TABLE 1: The comparison of tree position measurement based on relative positioning and total station measurement.

Tree position	Bias	rBias (%)	RMSE	rRMSE (%)
X (m)	0.05	0.95	0.114	2.10
Y (m)	-0.11	0.39	0.147	-0.53

solution for the measurement of tree attributes. The measurement of tree height is directly related to stand density and canopy occlusion. When the stand density is small or the canopy is sparse, the determination accuracy of tree height is high. The accuracy of DBH measurement is mainly affected by picture shooting, finger movement, and other factors. When operating the software, accurately determining the DBH position is the key to improve the DBH measurement accuracy.

#### 4. Discussion

Intelligent, efficient, and accurate method for measuring tree attributes has always been the focus of forest resource monitoring. Nowadays, with the advantage of multiple sensors, smartphones not only provide a variety of entertainment and consumption-level functions for the public but also have gradually become an important tool in professional fields such as forest resource monitoring [14, 33, 39, 40]. However, the sensor accuracy of many smartphones can still meet the entertainment needs, while the accuracy of professional resource detection needs to be improved [39]. In this paper, a high-precision, small sensor box is designed, and it can be seamlessly connected to a smartphone. The sensor box integrates the attitude gyroscope module with the laser rangefinder module to provide high-precision attitude data and horizontal distance data for smartphones. The sensor box has two advantages over the conventional laser rangefinder. (1) The conventional laser rangefinder focuses on the acquisition of distance data but cannot acquire the

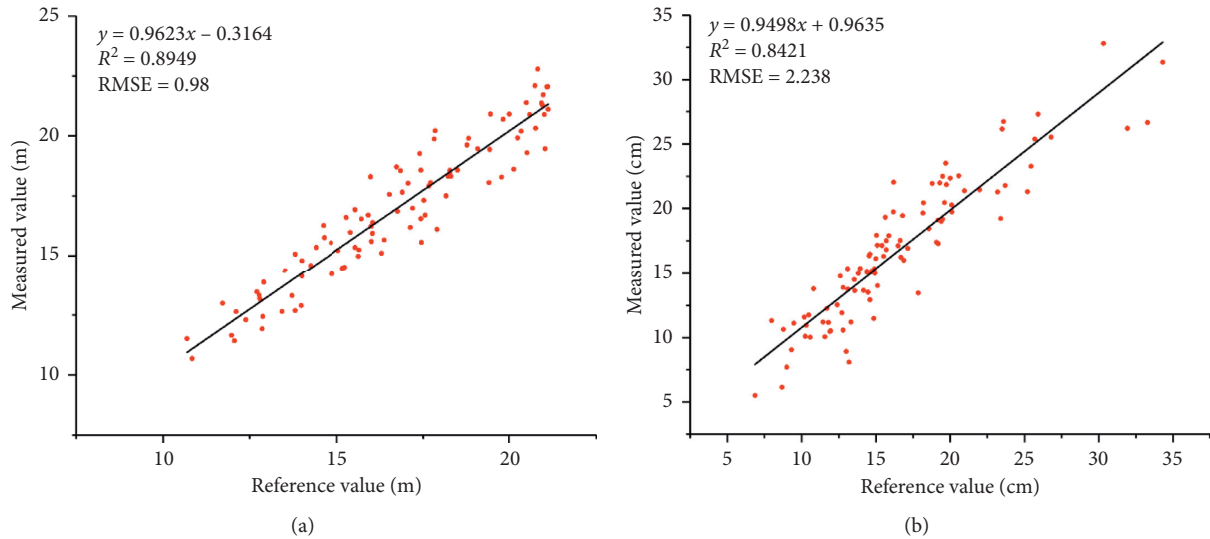


FIGURE 7: (a) The comparison of tree height measurement and (b) the comparison of DBH measurement.

TABLE 2: Comparison of tree measurement based on this method and traditional method.

Category	Bias	rBias (%)	RMSE	rRMSE (%)
Height (m)	-0.23	-1.40	0.98	5.87
DBH (cm)	-0.18	-1.06	2.24	13.46

attitude data. The sensor box designed in this paper has not only the function of laser ranging but also the function of attitude measurement. (2) The conventional laser rangefinders cannot be seamlessly connected to smartphones, so surveyors need separate smartphones and laser rangefinders. Such a combination cannot be taken as a whole. At  $80.3 * 47 * 30$  mm, the sensor box is half the size of a conventional laser rangefinder and can be seamlessly connected to a smartphone via a magnet sheet. Workers can combine the sensor box with the smartphone as a whole. It is easy to carry, easy to operate, and less affected by terrain.

In order to verify the effectiveness of this method, the position, DBH, and height of the tree were compared and analyzed. Although the method presented in this paper has no obvious advantage over LiDAR and photogrammetry in terms of accuracy [17, 23, 24], the measurement method is more flexible and less affected by topographic environment. LiDAR technology is greatly affected by the terrain, it also requires a lot of data processing in the office, and its calculation cost is high. This paper supports foresters to carry out preharvest measurements in the field. A method that combines a sensor box with a smartphone to measure tree properties can be used as an input parameter for a specific tree growth model and can automatically estimate wood volume or stand stock. Compared with the LiDAR technology which is more suitable for large-scale operation, this paper can provide a more flexible method for the investigation of rare tree protection. The detailed attributes of precious trees need to be recorded during the operation. This paper presented a method which measures the properties of

these precious trees and records the data in real time and stores it in the database of smartphones. In this paper, a scheme of “observation point transfer” is designed in the surveying work of plot level, but it takes a lot of space to describe the measurement of individual plant elevation. In terms of tree position measurement, mobile phone GNSS signal is greatly affected in some stands with high density of trees, resulting in low positioning accuracy. Also, the occluding GNSS signal of tree canopy is affected to varying degrees, and the absolute positioning technology based on GNSS signal may not be suitable for all stand positioning work [28, 30]. The relative positioning technology of tree position in this paper makes up the problem that the smartphone-based GNSS absolute positioning technology cannot locate in the forest with weak signal. This method shows high accuracy, and it is less affected by terrain and tree density. Due to the difficulty of absolute positioning with smartphones in stand measurement, the relative positioning with smartphones is rarely studied to determine the position of trees. In the measurement of tree height and DBH of single tree, although some researchers have also used smartphones to measure tree height, most of the phones have simple principles and single functions with poor universality. Some researchers have developed smartphone software to measure tree height or DBH based on trigonometric principles [5, 7, 9, 41], but this method uses the phone’s built-in angle sensor to get the angle of the observation point to the top and bottom of the tree. Because smartphone angle sensor sensitivity is different, the measurement error is unstable. In the actual work, because the sensor accuracy is insufficient, the result will have a large measurement error. In this paper, a high-precision sensor is attached to the smartphone to improve the measurement accuracy, and a single photo solution method of photogrammetry is applied to the monitoring of forest resources based on smartphone. Compared with other forest resource monitoring methods based on smartphones [26, 42–45], this method can directly measure the position, DBH, and height



of trees. The measurement accuracy will not be affected by different configurations of smartphones, so this method has the advantage of strong universality. Compared with the traditional method [46, 47], this method has computer power, does not need to touch trees, reduces the physical consumption of the surveyors, and saves the time and cost of the survey.

In order to give full play to the advantages of smartphones as microcomputers, this paper provides a low-cost and portable tool for current forest survey based on the method of measuring tree attributes based on smart space fusion of multiple sensors. Although surveyors can turn their smartphones into a professional tool for monitoring forest resources, there are still some problems that need to be further solved in this paper. The laser beam from the laser rangefinder module projects onto the tree trunk as a spot of light. When the light around the working environment is too weak or too strong, the visibility of the laser points in the tree trunk will be affected. In future research, this paper intends to change the geometry of the laser beam projected on the tree trunk from point to line. Finally, the laser beam is projected onto the tree trunk to form a laser line perpendicular to the tree trunk. Laser lines are more convenient for forest investigators than laser points. Compared with the way of setting up a smartphone by using a tripod in existing studies, the method in this paper is more convenient for the measurement personnel to carry and operate the smartphone, but it increases the instability of the attitude sensor in the sensor box to obtain the angle and direction data. During the collection process, the shaking of the arm will cause the measurement error, which will affect the accuracy of the tree position to some extent. In this paper, an image containing the complete information of standing trees is taken to calculate the DBH and tree height, which can complete the collection of sensor data in a very short time. The laser point is projected onto the tree trunk to minimize the error caused by the shaking of the arm. In this paper, on the premise of ensuring measurement accuracy, it is worthwhile to sacrifice less accuracy for the improvement of working efficiency and is more convenient to carry measurement tools. The sensor is still susceptible to electromagnetic interference, which reduces the measurement accuracy. The visibility of trees in forests remains the focus of further research in the future. Although we have designed an observation scheme of "observation point transfer" for the method in this paper, tree height cannot be directly measured for some totally closed stands. Therefore, the method of this paper is more suitable for small-scale, low-density, or medium-density forest resource monitoring with lower cost.

## 5. Conclusions

This paper presents a method of measuring key parameters of stand based on personal smart space fusion sensor. This paper designs a sensor box for personal smartphones that integrates various types of sensors and makes it easy to bind the two to each other. In order to verify the effectiveness of the method in this paper, the relative positioning algorithm and the standing tree image measurement algorithm can be

used to measure the position, height, and DBH of the tree. And this software was developed to process the data collected by the sensor. Finally, the experimental results were analyzed and discussed. This paper has solved the problems of poor universality, low precision, and single function in the previous forest resource monitoring methods based on smartphones, so it has higher flexibility and better portability. Forest investigators can turn their smartphones into professional forest resource monitoring tools in the work field, which provides valuable help for forest resource monitoring. In the future, more forest resource monitoring factors, such as volume, crown width, and the height of living crown, need to be determined based on personal smart spaces. In general, this paper provides a new forest resource monitoring method. Based on personal intelligent space fusion of various high-precision sensors, this method can basically meet the precision requirements of forest resource monitoring and can be popularized in forestry survey.

## Data Availability

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

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

We thank other team members for help with the experiment. This research was jointly supported by the Fundamental Research Funds for the Central Universities (TD2014-02).

## References

- [1] S. M. Pawson, A. Brin, E. G. Brockerhoff et al., "Plantation forests, climate change and biodiversity," *Biodiversity and Conservation*, vol. 22, no. 5, pp. 1203–1227, 2013.
- [2] Z. Qiu, Z. Feng, Y. Song, M. Li, and P. Zhang, "Carbon sequestration potential of forest vegetation in China from 2003 to 2050: predicting forest vegetation growth based on climate and the environment," *Journal of Cleaner Production*, vol. 252, Article ID 119715, 2020.
- [3] M. Arnold, B. Powell, P. Shanley, and T. C. H. Sunderland, "EDITORIAL: forests, biodiversity and food security," *International Forestry Review*, vol. 13, no. 3, pp. 259–264, 2011.
- [4] D. Han and C. Wang, "Tree height measurement based on image processing embedded in smart mobile phone," in *Proceedings of the in 2011 International Conference on Multimedia Technology*, pp. 3293–3296, Hangzhou, China, July 2011.
- [5] X. Wu, S. Zhou, A. Xu, and B. Chen, "Passive measurement method of tree diameter at breast height using a smartphone," *Computers and Electronics in Agriculture*, vol. 163, Article ID 104875, 2019.
- [6] A. Jaakkola, J. Hyyppä, A. Kukko et al., "A low-cost multi-sensoral mobile mapping system and its feasibility for tree measurements," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 514–522, 2010.

- [7] G. Fangli and X. U. Aijun, "Tree DBH measurement method based on smartphone and machine vision technology," *Zjnl dxxb*, vol. 35, no. 5, pp. 892–899, 2018.
- [8] D. Han, "Tree height measurement based on image processing with 3-points correction," in *Proceedings of the 2011 International Conference on Computer Science and Network Technology*, pp. 2281–2284, Harbin, China, December 2011.
- [9] K. Zhou, Y. Wang, J. LI, G. Jiang, and A. Xu, "Research and implementation of tree measuring system based on Android platform," *Journal of Nanjing Forestry University(Natural Sciences Edition)*, vol. 40, no. 4, pp. 95–100, 2016.
- [10] A. Kangas, J. Rasinmäki, K. Eyvindson, and P. Chambers, "A mobile phone application for the collection of opinion data for forest planning purposes," *Environmental Management*, vol. 55, no. 4, pp. 961–971, 2015.
- [11] M. Molinier, C. López-Sánchez, T. Toivanen et al., "Relasphome-mobile and participative in situ forest biomass measurements supporting satellite image mapping," *Remote Sensing*, vol. 8, no. 10, p. 869, 2016.
- [12] S. Korpilo, T. Virtanen, and S. Lehvävirta, "Smartphone GPS tracking-Inexpensive and efficient data collection on recreational movement," *Landscape and Urban Planning*, vol. 157, pp. 608–617, 2017.
- [13] M. Bakula, P. Przechodzinski, and R. Kazmierczak, "Reliable technology of centimeter GPS/GLONASS surveying in forest environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1029–1038, 2015.
- [14] J. Tomaščík, J. Tomaščík, Š. Saloň, and R. Piroh, "Horizontal accuracy and applicability of smartphone GNSS positioning in forests," *Forestry*, vol. 90, no. 2, pp. 187–198, 2017.
- [15] G. Fan, F. Chen, Y. Li, B. Liu, and X. Fan, "development and testing of a new ground measurement tool to assist in forest GIS surveys," *Forests*, vol. 10, no. 8, p. 643, 2019.
- [16] A. Jaakkola, J. Hyyppä, X. Yu et al., "Autonomous collection of forest field reference-the outlook and a first step with UAV laser scanning," *Remote Sensing*, vol. 9, no. 8, p. 785, 2017.
- [17] T. Mikita, P. Janata, and P. Surový, "Forest stand inventory based on combined aerial and terrestrial close-range photogrammetry," *Forests*, vol. 7, no. 12, p. 165, 2016.
- [18] L. Noordermeer, O. M. Bollandsås, H. O. Ørka, E. Næsset, and T. Gobakken, "Comparing the accuracies of forest attributes predicted from airborne laser scanning and digital aerial photogrammetry in operational forest inventories," *Remote Sensing of Environment*, vol. 226, no. 1, pp. 26–37, 2019.
- [19] G. D. Pearce, J. P. Dash, H. J. Persson, and M. S. Watt, "Comparison of high-density LiDAR and satellite photogrammetry for forest inventory," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 142, pp. 257–267, 2018.
- [20] D. R. Miller, C. P. Quine, and W. Hadley, "An investigation of the potential of digital photogrammetry to provide measurements of forest characteristics and abiotic damage," *Forest Ecology and Management*, vol. 135, no. 1–3, pp. 279–288, 2000.
- [21] K. Fankhauser, N. Strigul, and D. Gatzolis, "Augmentation of traditional forest inventory and airborne laser scanning with unmanned aerial systems and photogrammetry for forest monitoring," *Remote Sensing*, vol. 10, no. 10, p. 1562, 2018.
- [22] B. Talbot, M. Pierzchała, and R. Astrup, "Applications of remote and proximal sensing for improved precision in forest operations," *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, vol. 38, no. 2, pp. 327–336, 2017.
- [23] T. Sankey, J. Donager, J. Mcvay, and J. B. Sankey, "UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA," *Remote Sensing of Environment*, vol. 195, pp. 30–43, 2017.
- [24] S. Krause, T. G. M. Sanders, J.-P. Mund, and K. Greve, "UAV-based photogrammetric tree height measurement for intensive forest monitoring," *Remote Sensing*, vol. 11, no. 7, p. 758, 2019.
- [25] L. Wallace, A. Lucieer, Z. Malenovský, D. Turner, and P. Vopěnka, "Assessment of forest structure using two UAV techniques: a comparison of airborne laser scanning and structure from motion (SfM) point clouds," *Forests*, vol. 7, no. 12, p. 62, 2016.
- [26] R. d. Grote, "Estimation of crown radii and crown projection area from stem size and tree position," *Annals of Forest Science*, vol. 60, no. 5, pp. 393–402, 2003.
- [27] D. Alder and T. J. Synnott, *Permanent Sample Plot Techniques for Mixed Tropical Forest*, Oxford Forestry Institute, University of Oxford, Oxford, UK, 1992.
- [28] D. L. Evans, R. W. Carraway, and G. T. Simmons, "Use of global positioning system (GPS) for forest plot location," *Southern Journal of Applied Forestry*, vol. 16, no. 2, pp. 67–70, 1992.
- [29] P. Němec, "Comparison of modern forest inventory method with the common method for management of tropical rainforest in the Peruvian Amazon," *Journal of Tropical Forest Science*, vol. 27, no. 1, pp. 80–91, 2015.
- [30] R. R. Kennedy, *Use of Smartphone and GIS Technology for Sustainable Forestry in Eastern Ontario*, Morriest Hall, Ottawa, Canada, 2012.
- [31] J. Blackard, M. Finco, E. Helmer et al., "Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information," *Remote Sensing of Environment*, vol. 112, no. 4, pp. 1658–1677, 2008.
- [32] G. Chirici, R. E. McRoberts, S. Winter et al., "National forest inventory contributions to forest biodiversity monitoring," *Forest Science*, vol. 58, no. 3, pp. 257–268, 2012.
- [33] A. Villasante and C. Fernandez, "Measurement errors in the use of smartphones as low-cost forestry hypsometers," *Silva Fennica*, vol. 48, no. 5, p. 11, 2014.
- [34] E. M. Mikhail, J. Bethel, and J. C. McGlone, *Introduction to Modern Photogrammetry*, Wiley, Hoboken, NJ, USA, 2001.
- [35] K. B. Atkinson, *Close Range Photogrammetry and Machine Vision*, Whittles, Dunbeath Mill, UK, 1996.
- [36] D. Gaffrey, B. Sloboda, M. Fabrika, and Š. Šmelko, "Terrestrial single-photogrammetry for measuring standing trees, as applied in the Dobroc virgin forest," *Journal of Forest Science—UZPI (Czech Republic)*, vol. 47, pp. 75–87, 2001.
- [37] K. Kansanen, J. Vauhkonen, T. Lähivaara, and L. Mehtälö, "Stand density estimators based on individual tree detection and stochastic geometry," *Canadian Journal of Forest Research*, vol. 46, no. 11, pp. 1359–1366, 2016.
- [38] Z. Qiu, Z.-K. Feng, M. Wang, Z. Li, and C. Lu, "Application of UAV photogrammetric system for monitoring ancient tree communities in Beijing," *Forests*, vol. 9, no. 12, p. 735, 2018.
- [39] Z. Qiu, Z. Feng, J. Jiang, Y. Lin, and S. Xue, "Application of a continuous terrestrial photogrammetric measurement system for plot monitoring in the Beijing songshan national nature reserve," *Remote Sensing*, vol. 10, no. 7, p. 1080, 2018.
- [40] W. Xinmei, X. Aijun, and Y. Tingting, "Passive measurement method of tree height and crown diameter using a smartphone," *IEEE Access*, vol. 8, pp. 11669–11678, 2020.
- [41] D. Han, "Standing tree volume measurement technology based on digital image processing," in *Proceedings of the International Conference on Automatic Control and Artificial Intelligence (ACAI 2012)*, pp. 1922–1925, Xiamen, China, March 2012.

- [42] M. I. Marzulli, P. Raunonen, R. Greco, M. Persia, and P. Tartarino, "Estimating tree stem diameters and volume from smartphone photogrammetric point clouds," *Forestry: An International Journal of Forest Research*, vol. 93, no. 3, pp. 411–429, 2020.
- [43] M. Vastaranta, E. Latorre, V. Luoma, N. Saarinen, M. Holopainen, and J. Hyypä, "Evaluation of a smartphone app for forest sample plot measurements," *Forests*, vol. 6, no. 12, pp. 1179–1194, 2015.
- [44] J. Tomašík, Š. Saloň, D. Tunák, F. Chudy, and M. Kardo, "Tango in forests—an initial experience of the use of the new Google technology in connection with forest inventory tasks," *Computers & Electronics in Agriculture*, vol. 141, pp. 109–117, 2017.
- [45] E. González Latorre, "Evaluation of a mobilephone application for sample plot measurement in russia," Thesis, University of Helsinki, Helsinki, Finland, 2015.
- [46] D. R. Bower and W. W. Blocker, "Notes and observations: accuracy of bands and tape for measuring diameter increments," *Journal of Forestry*, vol. 64, no. 1, pp. 21–22, 1966.
- [47] J.-M. Binot, D. Pothier, and J. Lebel, "Comparison of relative accuracy and time requirement between the caliper, the diameter tape and an electronic tree measuring fork," *The Forestry Chronicle*, vol. 71, no. 2, pp. 197–200, 1995.

## Research Article

# A CRC-Based Classifier Micro-Engine for Efficient Flow Processing in SDN-Based Internet of Things

**Mahdi Abbasi** <sup>1</sup>, **Navid Mousavi**<sup>1</sup>, **Milad Rafiee**<sup>1</sup>, **Mohammad R. Khosravi**<sup>2,3</sup>  
and **Varun G. Menon**<sup>4</sup>

<sup>1</sup>Department of Computer Engineering, Engineering Faculty, Bu-Ali Sina University, Hamedan 65178-38695, Iran

<sup>2</sup>Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

<sup>3</sup>Telecommunications Group, Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran

<sup>4</sup>Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683582, Kerala, India

Correspondence should be addressed to Mahdi Abbasi; [abbasi@basu.ac.ir](mailto:abbasi@basu.ac.ir)

Received 14 December 2019; Accepted 3 April 2020; Published 18 May 2020

Academic Editor: Ali Kashif Bashir

Copyright © 2020 Mahdi Abbasi 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 Internet of things (IoT), network devices and mobile systems should exchange a considerable amount of data with negligible delays. For this purpose, the community has used the software-defined networking (SDN), which has provided high-speed flow-based communication mechanisms. To satisfy the requirements of SDN in the classification of communicated packets, high-throughput packet classification systems are needed. A hardware-based method of Internet packet classification that could be simultaneously high-speed and memory-aware has been proved to be able to fill the gap between the network speed and the processing speed of the systems on the network in traffics higher than 100 Gbps. The current architectures, however, have not been successful in achieving these two goals. This paper proposes the architecture of a processing micro-core for packet classification in high-speed, flow-based network systems. By using the hashing technique, this classifying micro-core fixes the length of the rules field. As a result, with a combination of SRAM and BRAM memory cells and implementation of two ports on Virtex®6 FPGAs, the memory usage of 14.5 bytes per rule and a throughput of 324 Mpps were achieved in our experiments. Also, the performance per memory of the proposed design is the highest as compared to its major counterparts and is able to simultaneously meet the speed and memory-usage criteria.

## 1. Introduction

Our world is connected by Internet of things (IoT). In the past few years, the considerable growth of network bandwidth and development of hardware technologies, especially in mobile communications, have led to a significant increase in the speed of communication lines of this worldwide network [1, 2]. That is, the speed of communication lines is reached to higher than “terabits per second.” The SDN paradigm aims to achieve good performance in managing networks by accelerating routers and switches to process the packets with the rate of network links [3, 4]. Making SDN flexible enough to satisfy the different requirements of heterogeneous IoT applications is desirable in terms of

software-defined IoT (SD-IoT) [5, 6]. For this purpose, network devices are equipped with a new mechanism, naming packet classification, which lets them to be flow-aware. That is, the network device, first classifies the incoming packets into predefined flows according to a set of filters, then any further processing is done accordingly. Therefore, a variety of packet processor devices including routers, firewalls, intrusion detection systems, account management systems, and network management systems use packet classification [1, 7–10]. That is, a number of important network management functions such as access control, quality of service provisioning, firewall, traffic policing, and policy-based switching make use of packet classification.

There are five fields in a typical classification rule including source and destination IP addresses (SA and DA), source and destination port numbers (SP and DP, respectively), and protocol (PT). SA and DA are address prefixes, SP and DP are number ranges, and the PT field may be either a specified value or a wildcard. The order of rules in a rule set determines their priority. The last rule is the default rule in which all the five fields are equal to the wildcard. If an incoming packet matches more than one rule, the action corresponding to the rule with the highest priority is performed. A description for packet classification algorithms is found in [1].

Classification of Internet packets in network devices is conducted through either software-based or hardware-based approaches. Considerable time overload of software-based methods makes them less popular among network equipment manufacturers [11]. On the other hand, there is a widespread tendency towards hardware-based methods and their higher throughput rate and lower delay [12]. Hardware-based implementation of packet classification algorithms may be categorized into two groups. The first group consists of algorithms based on parallel search in the content addressable memory (CAM) chips Z-TCAM [13], E-TCAM [14], and ZI-CAM [15]. In spite of their relatively high speeds, the use of ternary memories in these algorithms leads to disadvantages such as excessive power consumption, lower speed than other memory cells, lack of scalability, undue consumption of chip resources, and high prices. The second group consists of algorithms such as decision tree, decomposition tree, geometric space, field encryption, and similar methods which are realized on programmable hardware devices like ASIC or FPGA.

The main challenge in designing hardware-based methods is increasing the ratio of throughput to design cost. The throughput of a classifier is the number of packets that are classified in unit time. The required memory space is the main indicator of the system design cost. To reach this optimal point, we propose a micro-core that lowers memory consumption and simultaneously increases the classification throughput. The chief contributions of this paper are as follows:

- (1) The proposed micro-core uses SRAM and BRAM cells, which allow for dual-port implementations.
- (2) The proposed engine does not use any ternary content addressable memory. Instead, it encodes all of the prefixes by a cyclic redundancy check (CRC) code.
- (3) Implementing the proposed classifier on Virtex<sup>®</sup>6 FPGA shows that the memory cost reduces to 14.5 bytes per rule, and simultaneously the throughput of the classifier reaches 324 Mpps. This result confirms the superiority of the proposed architecture to its counterparts.

The rest of this article is organized as follows. In Section 2, the related works on hardware-based packet classification systems are reviewed. The proposed microclassifier architecture is explained in Section 3. The performance evaluation of the proposed architecture is presented in Section 4 after introducing the metrics. Finally, conclusions and directions for future research are discussed in Section 5.

## 2. Related Work

So far, a wide variety of hardware-based classifier architectures have been proposed for packet classification. All of them attempt to increase the throughput and decrease the memory usage. The CAM-based classifier architectures benefit from the parallel search property of CAM modules but suffer from high implementation costs and high levels of the consumption power. In [16, 17], two of the most recent architectures are proposed. They utilize a pipelined decision tree algorithm and a ternary memory, respectively. The architecture proposed in [16] has achieved a throughput of 103 Gbps, which is the highest among all the works mentioned here. However, depending on its hardware parameters like the number and length of pipelines on the distribution of the values of the classifier's rule fields, any updating requires reconfiguration of the architecture. On the other hand, the memory usage of this architecture is 63.5 bytes per packet. In [17], the researchers used ternary memories of the size 52\*144 and were able to reduce memory usage down to 18 bytes per rule; however, their maximum throughput was as low as 38 Gbps. The architectures proposed in [18, 19] could achieve a throughput of 100 Mpps while keeping the memory usage at 23.5 and 17.4 bytes, respectively. The focus of the architecture in [18] is on the rule search, and it does not address the issue of longest prefix matching (LPM). The architecture proposed in [19] adopts a TCAM-based approach to packet classification. Its major drawback is linear growth of TCAM usage is proportional to the number of rules that increases consumption of chip resources and power. Implementation of a merge algorithm based on decomposition tree in [20] achieves a throughput of 94 Gbps (amounting to 147 Mpps, given that each packet is 40 bytes). The study does neither provide the memory usage nor suggest any solution for updating rules.

Some classifiers like that presented in [21] use a special model to accelerate accessing the memory containing the rules, which in turn raises their memory consumption. Pipelined implementation of packet classification algorithms seeks appropriate solutions to reduce the number of pipeline stalls and the required memory space. For example, in the pipelined packet classifier of [22], the memory consumption varies from 16 to 24.5 bytes per rule.

To overcome the abovementioned disadvantages, we propose a packet-classifying micro-core with low memory consumption and high throughput. The proposed micro-core makes use of SRAM and BRAM cells, which allow for dual-port implementations. Processing based on cyclic redundancy check (CRC) codes in the internal structure of the micro-core without any need for ternary memories reduces the consumption of FPGA hardware resources and the time required for memory access in this classifier.

## 3. Proposed Architecture

This section explains the architecture of the proposed classifier which is aimed at increasing packet classification speed. Underlying the architecture are two principles: first,

making use of BCAM memory in prefix matching and, second, using hash codes to reduce memory usage.

In this classifier, a set of processing micro-cores act like a CAM, each one storing the information about one rule field of the rule set. The architecture is shown in Figure 1. In this architecture,  $n$  micro-cores are defined for each of the fields used for packet classification, where  $n$  is the number of rules in the classifier representing the number of micro-cores per field.

The incoming packets are classified as follows: First, the packets are transmitted through a shared bus to the micro-core unit. As soon as a packet enters a micro-core, the fields of source and destination IP addresses of the header are read by Parallel Hash Calculator. Next, in a parallel manner and proportional to the length of the value of the Prefix register, CRC-16 generation process is performed on the input address and the result is stored in the Temp register. Each micro-core has a control unit that, in addition to controlling the function of the micro-core, manages the generation process as well as the process of matching the hash code generated and stored in the Temp register against the hash code of the prefix field of the corresponding rule of the micro-core in the Hash-of-rule register. If the hash code of the incoming packet header matches the hash code in the micro-core, the Adder will add one to the variable stored in the Rank register. Moreover, in the case of correct matching, the one-bit flag Match and the corresponding bit of the micro-core in the  $n$ -bit register Packet Matching will be set. If matching fails, these bits will be reset by default. The Packet Matching register is used to record matches or mismatches in other micro-cores. In this register, any bit with a value of one denotes a match and any bit with a value of zero denotes a mismatch in the micro-core corresponding to a field that is being searched. After matching all fields, the results are written to the bits corresponding to each field in the Packet Matching register. Next, the result of logical AND operation on all Packet Matching registers is stored in Matched Vector. Finally, a prioritized decoder selects the matching rule with the highest priority.

As seen in Figure 1, the classifier consists of a set of processing micro-cores. In the following, we shall discuss the internal architecture of the micro-cores that is illustrated in Figure 2. Also, the length of each register of the micro-cores is shown in the bottom of Figure 2.

The main body of the micro-core is composed of two modules, i.e., CRC Calculator and Controller. The Controller module is responsible for management of all control lines, inputs, and outputs. Operations in this unit include management of selection lines, injection, and updating of registers. In fact, this unit is the decision-maker of the micro-core, which is separately controlled by the main controller of the classifier. In other words, the functioning of the micro-cores is not disrupted by updating and changing one of the micro-cores.

The second important module of the micro-core that bears a major part of its processing load is CRC Calculator. Figure 3 shows the function of this module. It receives the incoming packets and calculates their hash code in parallel (Line 1 of Algorithm 1). For this purpose, each IP address

along with the corresponding prefix which has been already stored in the Prefix register enters the module. Next, a hash code is computed using them and sent to the Controller for the purpose of matching. Implementation of this module consumes 40 out of 204000 LUTs on Virtex-6.

The micro-core has 7 input pins and 2 output pins. Table 1 lists the input and output ports of the micro-core along with their length in bits as well as their description. A micro-core is composed of registers, hash generator modules (CRC Calculator), and a controller.

One set of the input pins of the micro-core is named "Select," which determines the operation mode of the micro-core (Line 2 of Algorithm 1). In fact, this pin is responsible for management of the micro-core's functioning. It is connected to a two-bit bus which is used to address different modes of the micro-core function as described below (also, see Table 2):

Mode 0: the micro-core performs its main task, which is the packet classification (Lines 3–7 of Algorithm 1).

Mode 1: when the address from the Address port is identical with that in the Address register, the information in the selected micro-core is updated by means of the information from the Prefix and Rules ports (Lines 9–11 of Algorithm 1).

Mode 2: each micro-core sends the information related to its rank into a shared output bus. This operation is aimed at selecting the best candidate for being removed by the classifying controller. The central controller stores the number of the classifier with the lowest rank to update the rules (Lines 12–13 of Algorithm 1).

Mode 3: the majority of the existing classifiers do not offer a dynamic solution for updating rules and perform this task by reconfiguration of the chipset. However, given its low-rank feature, our proposed architecture enables us to update the rules under certain conditions. It is easy to inject new rules without changing the order of the existing rules. The last mode is Select(1, 1), which is used to initialize the micro-cores by injecting rules into each micro-core (Lines 14–17 of Algorithm 1).

Before injecting the packets into a processing micro-core, the rules are injected. In this step, the rules that have been converted to hash codes are stored in the Hash Code register. Prefix and Address registers keep the prefix length and the address of the processing core of the rules. Flag belongs to the internal controller which manages the input/output operation inside the micro-core so that the acts of processing the input packets would not overlap. Rank register has a length of 32 bits and holds the correct matches between the incoming packets and the matched field in the micro-core. For each correct match in the micro-core, one is added to the value of this register. This is a criterion used for updating and removing rules in other processing micro-cores. Thus, the micro-core with the lowest rank is selected for removing and updating. In fact, in this classifier, a lower rank is indicative of decreased use of the rule in the rule set.

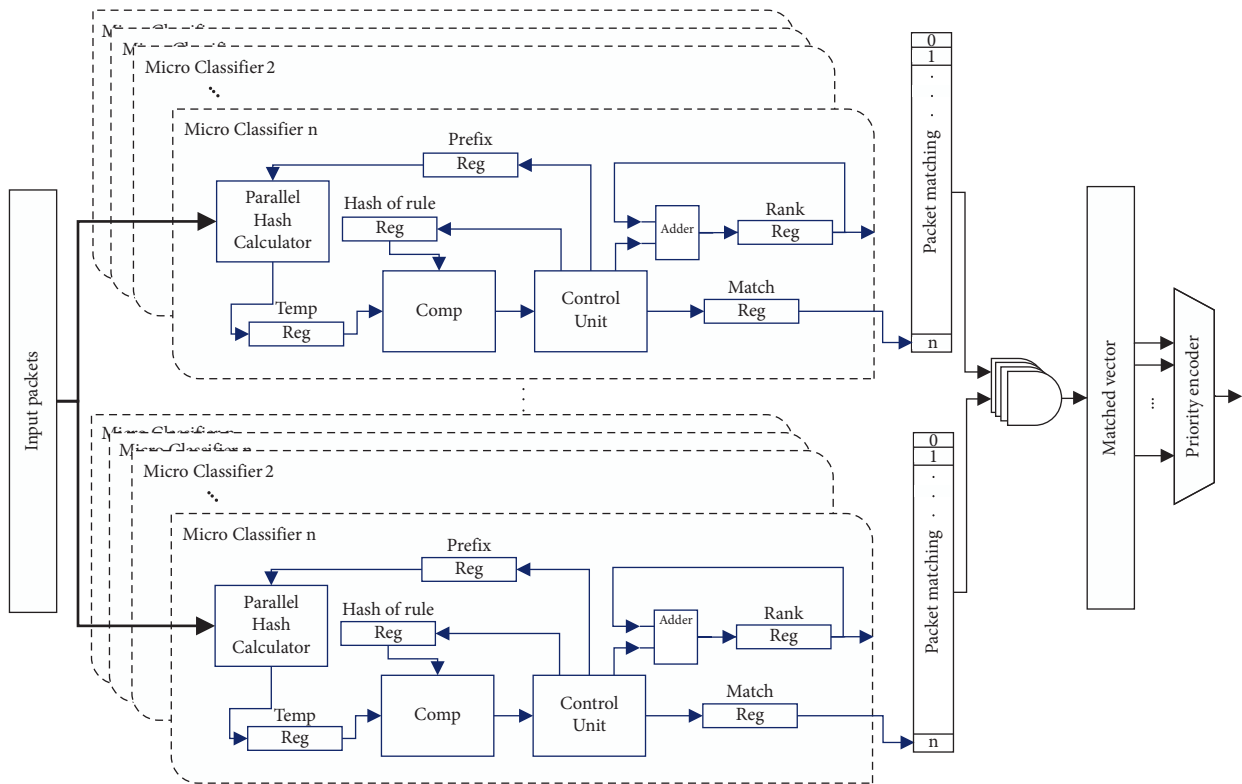


FIGURE 1: Proposed architecture for a packet classifier.

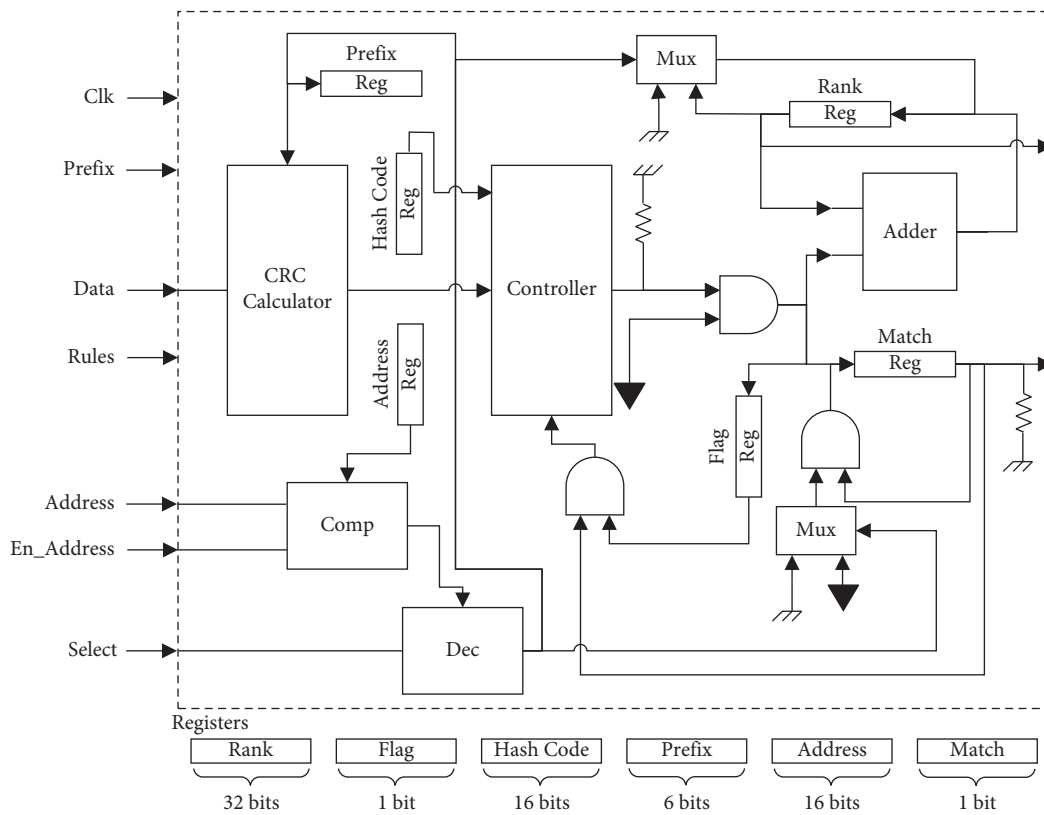


FIGURE 2: Proposed micro-core architecture.

```

Input: Data, Inputprefix, Rules, Select, InputAddress
Output: rankout, match
Registers: Rank, Flag, Hashcode, Prefix, Address, Match
Data: Packet P, CRC of Packet PCRC
(1)  $P_{CRC} \leftarrow \text{Calculate CRC (P, Hash}_{code})$ 
(2) Switch Select
(3) Case 00: //classify Operation
(4) if  $P_{CRC} == \text{Hash}_{code}$  then
(5) Match  $\leftarrow 1$ , Rank  $\leftarrow \text{Rank} + 1$ , Flag  $\leftarrow 1$ 
(6) Else
(7) Match  $\leftarrow 0$ 
(8) end if
(9) Case 01: //update
(10) Rank  $\leftarrow 0$ , Match  $\leftarrow 0$ , Flag  $\leftarrow 0$ 
(11) Prefix  $\leftarrow \text{Input}_{prefix}$ , Hashcode  $\leftarrow \text{Rules}$ 
(12) Case 10: //rank
(13) Rankout  $\leftarrow \text{Rank}$ 
(14) Case 11: //set address
(15) Address  $\leftarrow \text{Input}_{Address}$ , Prefix  $\leftarrow \text{Input}_{prefix}$ 
(16) Hashcode  $\leftarrow \text{Rules}$ 
(17) Match  $\leftarrow 0$ , Flag  $\leftarrow 0$ 
(18) End Switch
(19) If flag == 1 then
(20) Flag  $\leftarrow 0$ , Match  $\leftarrow 0$ 
(21) End if

```

ALGORITHM 1: Implementation of the packet classifier micro-core.

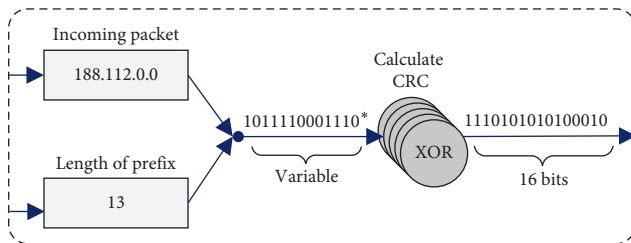


FIGURE 3: CRC Calculator module.

#### 4. Implementation and Evaluation

In this section, the results of implementation of the proposed micro-core architecture are discussed. This architecture is implemented on a XC6VLX75T chip from Virtex-6 FPGA family with a Xilinx ISE 14.7 simulator using VHDL language. Experiments are done on a system with the characteristics mentioned in Table 3.

The evaluation criteria in this experiment are throughput and memory. Throughput refers to the number of packets classified in a second. Assuming a minimum of 40 bytes for each packet [20], we use Gbps instead of Mpps for measuring the throughput. Memory usage is measured in bytes for each rule in the classifier.

We use Classbench tool to generate rules and experimental packets in our experiments. Classbench runs on Linux platform and is used for generating rulesets with desirable distributions in the model of the geometric space of rules. It generates rules and corresponding headers by using a set of input distribution parameters [23].

The highest throughput achieved so far belongs to Chang [16], which is 103.53 Gbps. However, it is 0.150 Gbps less than our throughput rate. Also, the memory usage for storing the classifier's rules in that architecture is four times greater than in our method. In fact, Chang's architecture resembles traditional TCAM-based architectures in terms of memory usage.

Table 4 compares the proposed micro-core architecture with the existing architectures. With a clock frequency of 170 MHz, the processing time of each micro-core is in the worst case 6.2 nanoseconds per packet and power consumption is 118 mW. With a dual-port memory which can process two packets simultaneously, the proposed micro-core is able to process 324 million packets of at least 40 bytes in a second which amounts to a throughput of more than 100 Gbps. In this simulation, our architecture used 137 Slice registers and 182 search tables.

In Table 4, the proposed micro-core architecture is compared with major recently proposed counterparts in terms of throughput and memory usage per rule. From among these architectures, Jiang and Prasanna [19] and Irfan et al. [17] require the least amount of memory, i.e., 17.4 and 18 bytes per rule, respectively. With a required memory of 14.5 bytes per rule, our proposed micro-core outperforms these two architectures. The major reason behind the low memory usage in our method is that, in contrast to the two mentioned architectures, our classifier does not rely on TCAM and mask for matching operation.

In a more fare approach, Table 5 compares the proposed design with other designs with regard to the performance per memory [19]:



TABLE 1: The function of input pins.

Name	Length (bit)	Description
Clk	1	Clock for all micro-cores
Prefix	6	Length of Prefix
Data	32	Width of input line for incoming packets
Rule	16	Maximum length of the hash code of rules
Address	16	Address of the selected processing core
En-address	1	Activation of updating and configuration operations
Select	2	Mode of processing in selected core
Rank-out	32	Width of the bus which is shared with all micro-cores and is used for updating
Match	1	A flag for signalling match/mismatch in a micro-core

TABLE 2: Modes provided by select pin.

Select	Modes	Corresponding lines of Algorithm 1
00	Classifying packets	3–7
01	Updating rules and property of micro-cores	9–11
10	Outputting rank of micro-cores on BUS	12–13
11	Initializing the selected micro-core	14–17

TABLE 3: System specification.

Specifications	Processor
Name	Intel Core i7-3720QM
Clock speed	2600 MHz
L3 cache	6 MB
Main memory	16 GB DDR3
Operation system	Windows 10 enterprise 18.03, 64 bit

TABLE 4: Comparison of the performance of the proposed method with different architectures.

Reference	Throughput (Gbit/s)	Frequency (MHz)	Memory (byte)	Seri	Chip
Pus and Korenek [18]	100	125	23.5	Virtex5	LX110T
Chang and Chen [16]	103.53	161.76	63.5	Virtex-6	XC5VFX200T
Fiessler et al. [24]	92.16	180	NA	Virtex-7	XC7VX690T
Orosz et al. [25]	100	312	NA	Virtex-6	XC6VHX255T
Zhou et al. [20]	147 mil	NA	NA	Virtex-7	XC7VX690T
Irfan et al. [17]	37.3	259	18	Virtex-6	XC6VLX760
Jiang and Prasanna [19]	100	167	17.4	Virtex-5	XC5VFX200T
Our design	<b>103.680</b>	<b>170</b>	<b>14.5</b>	<b>Virtex-6</b>	<b>XC6VLX75T</b>

TABLE 5: Comparison of performance per memory for various systems.

Approaches	Throughput (Gb/s)	Memory (byte)	Efficiency (throughput/memory)	Chip
Orosz et al. [25]	100	156	0.64	XC6VHX255T
<b>Our approach</b>	<b>101.7</b>	<b>14.5</b>	<b>7.01</b>	
Irfan et al. [17]	37.3	18	2.072	XC6VLX760
<b>Our approach</b>	<b>75.83</b>	<b>14.5</b>	<b>5.229</b>	
Ganegedara and Prasanna [21]	407	156	2.660	XC6VLX760
<b>Our approach</b>	<b>75.83</b>	<b>14.5</b>	<b>5.229</b>	
Qi et al. [26]	73.9	46.4	1.592	XC6VVSX475T
<b>Our approach</b>	<b>86.83</b>	<b>14.5</b>	<b>5.988</b>	
Pao and Lu [22]	108.8	18	6.04	XC6VLX75T
<b>Our approach</b>	<b>103.680</b>	<b>14.5</b>	<b>7.150</b>	

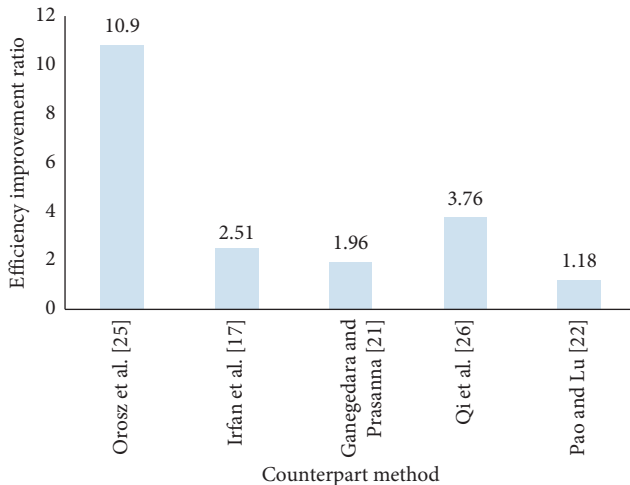


FIGURE 4: The ratio of efficiency improvement.

$$\text{efficiency} = \frac{\text{throughput (Gb/s)}}{\text{normalized memory (B/rule)}}. \quad (1)$$

For this purpose, the throughput as well as the memory consumption of the proposed architecture is measured on the chipsets that are used in the evaluation of the competitor designs. The memory usage per rule is always constant in the proposed design. Therefore, the performance per memory of the proposed design is the highest as compared to its major counterparts.

The ratio of superiority of the efficiency of the proposed method with respect to each counterpart is illustrated in Figure 4. Our comparisons suggest that the proposed architecture has considerably improved the throughput rate and memory usage of the Internet packet classification systems. The performance per memory of the proposed design is at least 18% and at most 990% better than the best and the worst designs, namely, Pao and Lu [22] and Orosz et al. [25].

## 5. Conclusion

In this paper, we proposed a new micro-core architecture for classification of Internet packets that is capable of being updated. The proposed architecture allows for adding or removing rules during processing and enjoys lower memory usage as well as higher throughput rate in comparison with other architectures. Our evaluations suggest that this micro-core can classify packets with a throughput of more than 103 Gbps, which amounts to about 324 Mpps. Another advantage of this architecture is that memory usage per rule is always constant. Therefore, the performance per memory of the proposed design is the highest as compared to its major counterparts. This achievement helps the proposed micro-core to avoid the problem of resource requirement in the longest prefix matching (LPM). A fruitful topic for future research would be to apply this inherent feature of the micro-core to the implementation of pipeline classifiers in which LPM is a great problem in pipeline processing.

## Data Availability

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

## Conflicts of Interest

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

## References

- [1] D. E. Taylor, "Survey and taxonomy of packet classification techniques," *ACM Computing Surveys (CSUR)*, vol. 37, no. 3, pp. 238–275, 2005.
- [2] N. T. Le, M. A. Hossain, A. Islam, D.-Y. Kim, Y.-J. Choi, and Y. M. Jang, "Survey of promising technologies for 5G networks," *Mobile Information Systems*, vol. 2016, Article ID 2676589, 25 pages, 2016.
- [3] R. M. A. Ujjan, Z. Pervez, K. Dahal, A. K. Bashir, R. Mumtaz, and J. González, "Towards sFlow and adaptive polling sampling for deep learning based DDoS detection in SDN," *Future Generation Computer Systems*, 2019.
- [4] M. Shakil, A. Fuad Yousif Mohammed, R. Arul, A. K. Bashir, and J. K. Choi, "A novel dynamic framework to detect DDoS in SDN using metaheuristic clustering," *Transactions on Emerging Telecommunications Technologies*, p. e3622, 2019.
- [5] A. R. Bevi, P. Shakthipriya, and S. Malarvizhi, "Design of software defined networking gateway for the internet-of-things," *Wireless Personal Communications*, vol. 107, pp. 1273–1287, 2019.
- [6] A. K. Bashir, R. Arul, S. Basheer, G. Raja, R. Jayaraman, and N. M. F. Qureshi, "An optimal multitier resource allocation of cloud RAN in 5G using machine learning," *Transactions on Emerging Telecommunications Technologies*, vol. 30, p. e3627, 2019.
- [7] P. Gupta and N. McKeown, "Packet classification on multiple fields," *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 4, pp. 147–160, 1999.
- [8] K. Kang and Y. S. Deng, "Scalable packet classification via GPU metaprogramming," in *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE)*, March 2011.
- [9] S. Shieh, F.-Y. Lee, and Y.-W. Lin, "Accelerating network security services with fast packet classification," *Computer Communications*, vol. 27, no. 16, pp. 1637–1646, 2004.
- [10] G. Varghese, "Chapter 12-packet classification," in *Network Algorithmics*, G. Varghese, Ed., pp. 270–301, Morgan Kaufmann, San Francisco, CA, USA, 2005.
- [11] J. Wee, J.-G. Choi, and W. Pak, "Wildcard fields-based partitioning for fast and scalable packet classification in vehicle-to-everything," *Sensors (Basel, Switzerland)*, vol. 19, no. 11, p. 2563, 2019.
- [12] Y. R. Qu, S. Zhou, and V. K. Prasanna, "High-performance architecture for dynamically updatable packet classification on FPGA," in *Proceedings of the Ninth ACM/IEEE Symposium Architectures for Networking and Communications Systems*, pp. 125–136, IEEE, San Francisco, CA, USA, October 2013.
- [13] Z. Ullah, M. K. Jaiswal, and R. C. C. Cheung, "Z-TCAM: an SRAM-based architecture for TCAM," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 1, pp. 402–406, 2014.

- [14] Z. Ullah, M. K. Jaiswal, and R. C. C. Cheung, "E-TCAM: an efficient SRAM-based architecture for TCAM," *Circuits, Systems, and Signal Processing*, vol. 33, no. 10, pp. 3123–3144, 2014.
- [15] M. Irfan, Z. Ullah, and R. C. C. Cheung, "Zi-CAM: a power and resource efficient binary content-addressable memory on FPGAs," *Electronics*, vol. 8, no. 5, p. 584, 2019.
- [16] Y.-K. Chang and H.-C. Chen, "Fast packet classification using recursive endpoint-cutting and bucket compression on FPGA," *The Computer Journal*, vol. 62, no. 2, pp. 198–214, 2018.
- [17] M. Irfan, Z. Ullah, and R. C. C. Cheung, "D-TCAM: a high-performance distributed ram based TCAM architecture on FPGAs," *IEEE Access*, vol. 7, pp. 96060–96069, 2019.
- [18] V. Puš and J. Korenek, "Fast and scalable packet classification using perfect hash functions," in *Proceedings of the ACM/SIGDA International Symposium on Field Programmable GateArrays*, ACM, pp. 229–236, Monterey, CA, USA, February 2009.
- [19] W. Jiang and V. K. Prasanna, "Field-split parallel architecture for high performance multi-match packet classification using FPGAs," in *Proceedings of the Twenty-first Annual Symposium on Parallelism in Algorithms and Architectures*, ACM, pp. 188–196, Calgary, Alberta, Canada, August 2009.
- [20] S. Zhou, Y. R. Qu, and V. K. Prasanna, "Large-scale packet classification on FPGA," in *Proceedings of the IEEE 26th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, pp. 226–233, IEEE, Toronto, Ontario July 2015.
- [21] T. Ganegedara and V. K. Prasanna, "StrideBV: single chip 400G+ packet classification," in *Proceedings of the IEEE 13th International Conference on High Performance Switching and Routing*, pp. 1–6, IEEE, Belgrade, Serbia, June 2012.
- [22] D. Pao and Z. Lu, "A multi-pipeline architecture for high-speed packet classification," *Computer Communications*, vol. 54, pp. 84–96, 2014.
- [23] D. E. Taylor and J. S. Turner, "Classbench: a packet classification benchmark," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 499–511, 2007.
- [24] A. Fiessler, C. Lorenz, S. Hager, B. Scheuermann, and A. W. Moore, "HyPaFilter+: enhanced hybrid packet filtering using hardware assisted classification and header space analysis," *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3655–3669, 2017.
- [25] P. Orosz, T. Tóthfalusi, and P. Varga, "C-GEP: adaptive network management with reconfigurable hardware," in *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pp. 954–959, IEEE, Toronto, Ontario, Canada, May 2015.
- [26] Y. Qi, J. Fong, W. Jiang, B. Xu, J. Li, and V. Prasanna, "Multi-dimensional packet classification on FPGA: 100 Gbps and beyond," in *Proceedings of the International Conference on Field-Programmable Technology*, pp. 241–248, IEEE, Beijing, China, December 2010.

## Review Article

# A Review of Deep Learning Security and Privacy Defensive Techniques

**Muhammad Imran Tariq** <sup>1</sup>, **Nisar Ahmed Memon**,<sup>2</sup> **Shakeel Ahmed**,<sup>2</sup> **Shahzadi Tayyaba**,<sup>3</sup> **Muhammad Tahir Mushtaq**,<sup>4</sup> **Natash Ali Mian**,<sup>5</sup> **Muhammad Imran**,<sup>6</sup> and **Muhammad W. Ashraf**<sup>6</sup>

<sup>1</sup>Department of Computer Science, Superior University, Lahore, Pakistan

<sup>2</sup>College of Computer Science and Information Technology (CCSIT), King Faisal University, Al-Ahsa, Saudi Arabia

<sup>3</sup>Department of Computer Engineering, The University of Lahore, Lahore, Pakistan

<sup>4</sup>School of Systems and Technology, The University of Management and Technology (UMT), Lahore, Pakistan

<sup>5</sup>School of Computer and Information Technology, Beaconhouse National University, Lahore, Pakistan

<sup>6</sup>Department of Physics (Electronics), Government College University, Lahore, Pakistan

Correspondence should be addressed to Muhammad Imran Tariq; [imrantariqbutt@yahoo.com](mailto:imrantariqbutt@yahoo.com)

Received 21 November 2019; Revised 24 January 2020; Accepted 12 February 2020; Published 7 April 2020

Guest Editor: Fawad Zaman

Copyright © 2020 Muhammad Imran Tariq 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 past years, Deep Learning presented an excellent performance in different areas like image recognition, pattern matching, and even in cybersecurity. The Deep Learning has numerous advantages including fast solving complex problems, huge automation, maximum application of unstructured data, ability to give high quality of results, reduction of high costs, no need for data labeling, and identification of complex interactions, but it also has limitations like opaqueness, computationally intensive, need for abundant data, and more complex algorithms. In our daily life, we used many applications that use Deep Learning models to make decisions based on predictions, and if Deep Learning models became the cause of misprediction due to internal/external malicious effects, it may create difficulties in our real life. Furthermore, the Deep Learning training models often have sensitive information of the users and those models should not be vulnerable and expose security and privacy. The algorithms of Deep Learning and machine learning are still vulnerable to different types of security threats and risks. Therefore, it is necessary to call the attention of the industry in respect of security threats and related countermeasures techniques for Deep Learning, which motivated the authors to perform a comprehensive survey of Deep Learning security and privacy security challenges and countermeasures in this paper. We also discussed the open challenges and current issues.

## 1. Introduction

Deep Learning is also called hierarchical learning and deep-structured learning, and it is comprised of supervised or unsupervised machine learning techniques. The idea of Deep Learning derived from the structure and functionality of the human brain and also the processing of signals through neurons in the human mind. Deep Learning is also taking the benefits of artificial neural networks, and it also consists of input, output, and many hidden layers. Each layer of Deep Learning relies upon the

nonlinear response based on the data provided through the input layer. For the last few years, the Deep Learning technique has been mostly and widely used in the signal processing of voice recognition, graphic recognition, discovery of the thing, and so numerous other areas, such as the discovery of the medicine for diseases and genomics [1]. Deep Learning developed a structure to deal with big data sets through a backpropagation algorithm to highlight in what way the device changes its core parameters that are being opted to calculate the representation in each rendering layer in the previous layer [2].

Despite their enormous size, successful Deep Neural Networks can make a very minor difference between training and test presentation. Traditional wisdom attributes the error of small circularization to the typical characteristics of the family or to the organizational techniques used during training [3].

The crucial problem of the DL is its encrypted data that flows from training and interface modules. The security and privacy issues are very important due to mostly adopted DL models in many applications as mentioned above. Further, actually Deep Learning prevailing in all models for training part relies upon a huge number of big data, sensitive, and confidential data of the user particularly training data. Keeping this in view, DL models must not disclose confidential and sensitive data. In this paper, systematic literature reviewed was conducted about the Deep Learning security threats, privacy threats regarding private data, and their corresponding developed defense techniques. The paper also included most secured techniques that use cryptographic primitives without the indulgence of the third party and the summary of the future challenges and opportunities.

*1.1. Application of Deep Learning.* Deep learning has introduced new ways to look at technologies. Artificial Intelligence (AIT) and its branches ML and Deep Learning have a lot of excitements. It is a reality that Deep Learning changed the ways of living and will also affect life in the near future. DL is grabbing market space day by day and we are sure, in coming five to ten years, the tools, techniques, and libraries of DL will include in every development toolkit.

Here, we will discuss the Deep Learning applications that captured the marked in 2019 and beyond.

*1.1.1. Self-Driving Car.* Many of the car manufacturing companies have built self-driving cars with the help of digital sensor systems. It is accomplished through training algorithms through the huge unstructured amount of data.

*1.1.2. DL in Healthcare.* Deep learning is also used to bring improvement in the field of Healthcare especially in breast cancer diagnostics and monitoring apps. It is also used to predict personalized medicine keeping in view the Biobank data. Deep learning completely reshaped the healthcare industry as well as life sciences. The key features of Deep Learning are advancing the future of health management.

*1.1.3. DL in Voice Search.* The most famous utilization of Deep Learning is voice recognition, searching, and activation. This facility is already available in every smartphone since 2011. Google and Apple are already offering these services, and now Microsoft Cortana has also launched a voice activation assistant.

*1.1.4. Automatic Machine Translation.* The google translator is the main example of the translation of one language into another language. The user entered the word, sentences,

paragraphs, and phrases of one language, and it easily converts to another language. Although this facility is available for a long time, DL is getting improvement in the results with the passage of time, and now machine translation is also translating images. Image to text conversion is an example of machine translation and is the innovation of Deep Learning.

*1.1.5. Automatic Handwriting Generation.* Deep Learning has also played a vital role in the automatic handwriting generation. The system automatically captures the movement of the pen and the letters to learn. The DL also facilitates the generation of new writing styles.

Also, there are numerous applications of the Deep Learning that cannot be covered in one paper, and the more applications of Deep Learning are as follows:

- (i) Image recolonization
- (ii) Face recolonization
- (iii) Automatic colorization
- (iv) Image captioning
- (v) Advertising
- (vi) Earthquake prediction
- (vii) Brain cancer detection
- (viii) Price forecasting
- (ix) Natural Language Processing
- (x) Gamming
- (xi) Cybersecurity

*1.2. Innovative Contributions of Deep Learning.* Deep learning has contributed to every field of science and brought innovative changes. Deep learning also uplifts every area of life by solving routine problems and also introduced new dimensions of research. The outstanding performance of Deep Learning is in the area of modern security systems. It is a very critical problem that today every small- and large-scale organization is facing; millions of new malware and virus threats are created, and large organizations like banks and government institutions are attacked by finding grey areas in the tools. Although many security solutions exist, security is an ongoing area in research. Deep learning presented new dimensions in the area of cybersecurity by detecting network attacks, removing malware, identifying vulnerabilities, and securing the system.

*1.3. Organization of Study.* Section 2 of the paper is related to background/literature review, Section 3 discusses Deep Learning private data frameworks, and Deep Learning treats and attacks are discussed in Section 4 of the paper, and defense techniques against security issues in Deep Learning briefly explained in Section 5 of the paper. The final conclusion of the paper is also discussed in Section 6 of the paper.

## 2. Background

*2.1. Deep Learning.* Deep learning permits high computational models that consist of multiple layers of processing to learn the depiction of data at multiple levels of abstraction layers. These techniques have vastly improved the state of the art in voice recognition, visual recognition, discovery of the object, and so many other areas, such as the discovery of the medicine for diseases and genomics. Deep learning artificial neural networks regularly contain additional trainable model parameters as compared with the number of samples in which they have been trained [4]. However, some of these models show a significantly lower circular error, that is, the difference between the training error and the test error. It is certainly easy to reach normal typical structures with little circulation [5]. What then distinguishes neural networks that generalize well from those that do not? A satisfactory answer to this question will not only help make neural networks more interpretable but can also lead to a more reliable and reliable architectural design. To answer this question, the theory of statistical learning proposed several different measures of complexity capable of controlling the error of generalization. These include the VC dimension, Rademacher complexity, and uniform stability. Also, when the number of parameters is large, the theory suggests that some type of regulation is needed to guarantee a small circular error. The regulation may be implicit as with the early suspension [6].

Machine learning technology operates many sides of current society like from online research to content filtering on social networks to recommendations on e-commerce sites and are increasingly present in consumer products such as cameras and smartphones. Machine learning systems are used to identify objects in pictures, convert voice into text, relate news items, publications or products with user interests, and identify relevant search results. Increasingly, all these applications are using Deep Learning [7].

According to [8], traditional machine learning techniques have not completed the ability to manipulate natural network data in its original shape. For decades, the establishment of a machine learning system requires precise engineering and substantial experience in the field to design a feature extractor that transforms raw information into an appropriate internal representation [9].

*2.2. Deep Neural Networks (DNNs).* This greater use of Deep Learning creates incentives for opponents to approach Deep Neural Networks (DNNs) to impose a poor classification of inputs. For example, Deep Learning applications use image workstations to differentiate themselves from inappropriate content, textures, and images to distinguish spam from nonintrusive mail [10]. An adversary capable of formulating erroneous inputs would benefit from the evasion of detection; even today, these attacks occur in classification systems other than Deep Learning. In the real world, consider a driverless car system that uses deep learning to identify traffic signals. If a change in the “stop” marks causes the

Deep Neural Networks to be incorrectly classified, the vehicle will not stop [11].

The neural network basically consists of 03 elements, one is called the input layer, which is basically the data that the user wants to analyze [12]. The second layer is actually hidden layers; it may consist of one node or maybe more than more nodes; the primary function of this node is to complete the computation in the light of the Deep Learning algorithm. The last layer is always the output layer, which calculates the result. Figure 1 illustrates the basic neural network, and Figure 2 illustrates the Deep Learning Neural Network.

For classification tasks, higher representation layers amplify important entry aspects of discrimination and suppress irrelevant differences. For example, the image comes in the form of an array of pixel values, and the features learned in the first rendering layer generally represent the presence or absence of edges in certain directions and locations in the image. The second layer usually discovers the motifs by detecting a certain arrangement of the edges, regardless of the small differences in the positions of the edges. The third layer can group shapes into larger groups that correspond to parts of familiar objects, and the following layers will discover the objects as groups of these parts.

The main feature of DL layers is that these layers are not designed by the human; actually, it has been learned from the data through a general-purpose learning procedure. Deep learning is making great progress in solving problems that have withstood the best efforts of the AI community for many years. It has proven to be very good at detecting complex structures in high-dimensional data and, therefore, is applicable to many fields of science, business, and government addition to multiply the registers in picture recognition and voice recognition; other machine learning methods have been overcome by actively predicting possible drug molecules, analyzing particle accelerator data, reconstructing cerebral circuits, and predicting the effects of mutations in noncoding DNA on gene expression and disease. Perhaps, most surprising thing is that Deep Learning has yielded very promising results for several tasks in the understanding of natural language, the classification of the particular topic, the analysis of morals, the answer to questions, and the translation of the language [13].

It is pertinent to add here that weaknesses in DL systems have recently been discovered in a big number of publications. It is very dangerous that these applications are based on a small understanding of security and privacy in DL systems [14].

Although many research studies have been published on attacks and the defense of the security and privacy of Deep Learning, they are still fragmented. Here, we review recent attempts to secure Artificial Intelligence and Private Data of Artificial Intelligence.

In order to meet the requirement for strong AI systems in information security and private data, we need to develop a take Secured Artificial Intelligence system. That secure Artificial Intelligence system should provide security

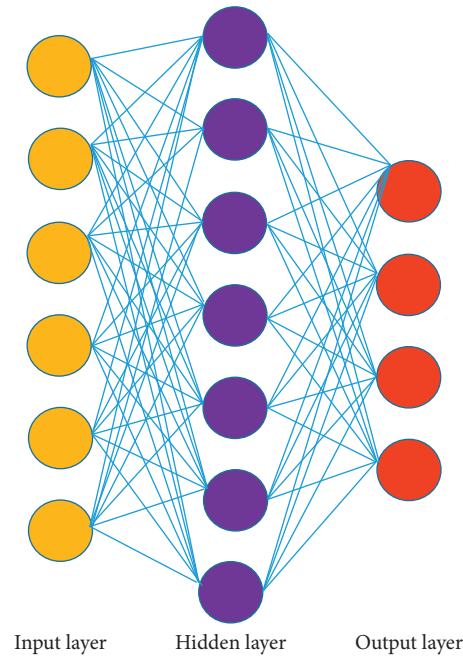


FIGURE 1: Basic neural network.

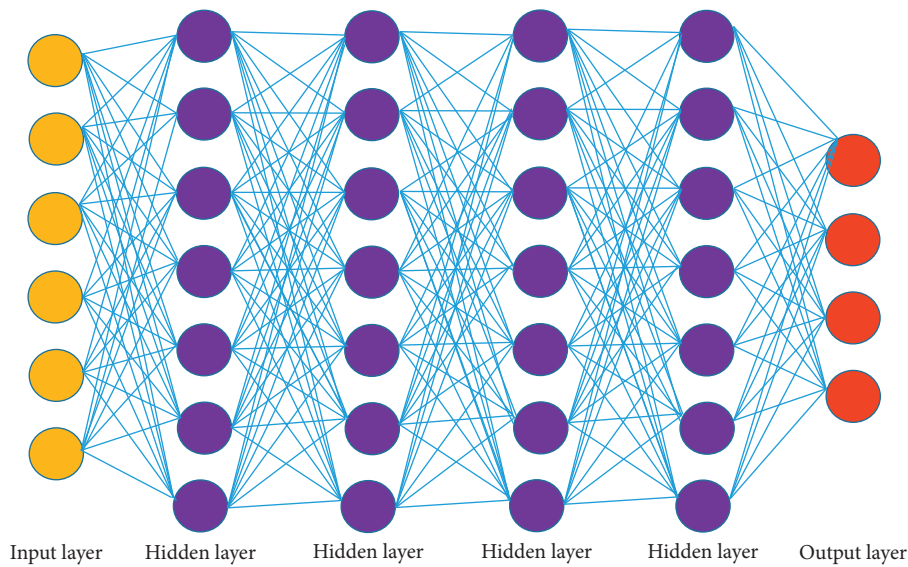


FIGURE 2: Deep Learning neural network.

guarantee, and Private Data Artificial Intelligence should maintain data privacy of the system [15].

The Secure Artificial Intelligence always focuses on attacks, threats, vulnerabilities, and accordingly defense of Artificial Intelligence systems, in respect of Deep Learning, which is a more effective model. The attacks on Deep Learning generate false predications by injecting wrong samples, such types of attacks are called white-box attacks, and it includes gradient-based techniques to compromise the system. In contrast, attacks from the black-box cause the suspect system to make fake predictions, without getting some information about the system. It has been observed

that almost every attack exploits the predictive confidence of the system without getting information about the structure and parameters of the system [16].

In order to develop defense against these attacks, various methods have been proposed such as adversarial training, generative adversarial network, statistical approach, and recurrent neural network.

The input data of the user contains sensitive data to the Deep Learning machines for recognition. The more secure option for the user is to install the Deep Learning model on its platform and execute it and obviously; it is not feasible for the user because the Deep Learning model always consists of

massive data and it processed them [17]. Every organization desires to keep their data confidential, and their competitors may not use it for their business purposes.

The upshot, the Deep Learning machine, should meet three main requirements while preserving privacy:

- (i) The data stored in the training model should not be disclosed to the cloud server
- (ii) The user request should not be disclosed to the cloud server
- (iii) The configurations of the cloud server should not be disclosed to the user

It is highly needed for the organizations using Deep Learning to establish privacy frameworks in which neither any intruder nor any attacker discloses information during the shared computation or modify it. In order to strengthen privacy computation in respect of Deep Learning, it is critically significant to plan new privacy-specific techniques that can minimize the complexity of secure function evaluation protocols [18].

The purpose of this research is to study the recent development of deep learning on private data and security issues attached to Deep Learning in different domains. Furthermore, we describe different types of Deep Learning possible security and privacy attacks along with different defense methods.

The core part of the Deep Neural Network is called Artificial Neuron. Artificial Neurons purely calculate the weighted amount of inputs and output, according to the following equation:

$$y = \sigma \sum_{i=1}^n w_i x_i, \quad (1)$$

where  $y$  is denoted as the output,  $x$  is for the input,  $\sigma$  is denoted as the activation function which is actually a nonlinear function, and  $w$  is called the weights. Artificial Neurons are basically used to develop construct layer (details are given in below figures), and if these layers are piled up, then it constructs DNN. The nonlinearity of the  $\sigma$  piles up the number of DNN layers that cultivates and allows the Deep Neural Networks to estimate the objective functions without any manmade feature selections.

**2.3. Artificial Intelligence in Deep Learning.** Figure 3 is a high-level group diagram of the learning process to develop a stereotype Deep Learning model. The performance of the DL model depends on the size of the existing available training data.

Nevertheless, training samples are typically gathered from the content of users stored on cloud machines that hold sensitive information, like photographs, video, sound, and location records. The privacy of the user is a major concern in Deep Learning during training and inference [19]. Internet service providing organizations are providing Deep Learning as a service where users can insert input to the cloud machines and obtain the result based on prediction.

**2.4. Architectures of DNNs.** The DNN model has different types of architectures that are briefly explained below.

**2.4.1. Feed-Forward Neural Network (FNN).** This is the fundamental and core building block of the Deep Neural Network. It consists of different types of the multiple layers, and these middle layers are completely connected with each other while the nodes within the layer are not linked to each other [20]. Figures 1 and 2 are examples of Feed-Forward Neural Network.

**2.4.2. Convolutional Neural Network.** This architecture is demonstrated in Figure 4. A CNN architecture consists of many convolutional and pooling layers. These layers use convolutional operations to compute and generate layerwise outcomes. The convolutional and pooling layer's operation permits the DNN network to get more knowledge about spatial. Hence, the CNN architecture shows exceptional results particularly on image applications [21, 22].

**2.4.3. Recurrent Neural Network.** It is extensively opted to process sequential information. As illustrated in Figure 5, the RNN calculates the output after updating the currently hidden units, past hidden units, and presently available input data [23]. The RNN also faces problems like gradient vanishing problem and long short-term memory. To solve these problems, the gated recurrent unit is used.

**2.4.4. Generative Adversarial Network.** This architecture of DNNs is basically comprised of two modules, one is called Discriminator (D) and the other is known as Generator (G). The Generator generates false data in the architecture while Discriminator is used in the architecture to inform whether the Generator's data are real or not? as illustrated in Figure 6. The Generator and Discriminator are usually used in DNNs, and it has many types of structures based upon the application of the network [24]. Generative Adversarial Networks are opted by many fields like image processing, voice recognizing, and domain adaptation.

**2.5. Deep Learning Privacy Preserving Techniques.** In the forthcoming section, the prevailing cryptographic primitives that are presently opted by the organizations for privacy preserving both for training and interface of the Deep Neural Networks (DNNs) are discussed.

**2.5.1. Homomorphic Encryption (HE).** Homomorphic Encryption (HE) is primitive encryption that allows a party to encrypt data and send it to another party that can then perform certain operations on the encrypted version of the data [25]. An encryption system that allows arbitrary calculations to be encoded on encrypted data without decryption or access to any symmetric cryptographic decryption key is known HE [26]. When the account ends, the encrypted version of the result is sent to the first party that can decrypt and get the result in plain text.



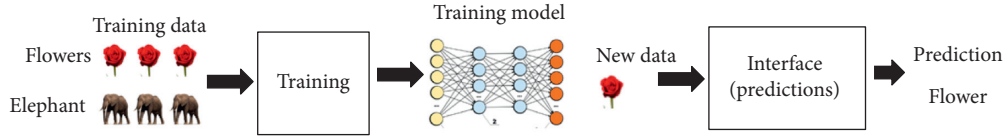


FIGURE 3: Training and interface in Deep Learning.

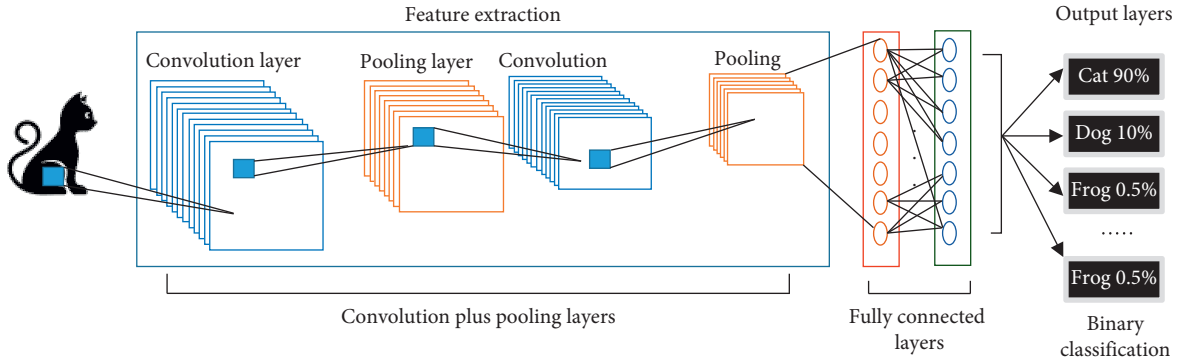


FIGURE 4: Structure of convolutional neural network.

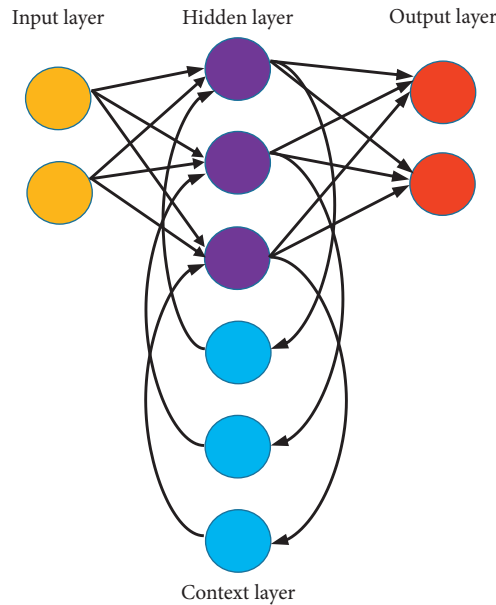


FIGURE 5: Structure of recurrent neural network.

Homomorphic encryption methods can be partially divided into completely Homomorphic Encryption and partially Homomorphic Encryption [27]. For example, the Paillier encryption system only supports adding to the two-digit encrypted version, which is partially Homomorphic Encryption. In contrast, a fully symmetric encryption system supports arbitrary functional logic. The Homomorphic Encryption scheme (Enc) follows the following equation:

$$\text{Enc}(a) \Delta \text{Enc}(b) = \text{Enc}(a * b). \quad (2)$$

where  $\text{Enc}: X \rightarrow Y$  is a Homomorphic Encryption scheme wherein  $X$  is used for a set of messages and  $Y$  is used for

ciphertext. Furthermore,  $a$  and  $b$  are messages in  $X$  and  $\Delta, *$  are linear operations. At the beginning when Homomorphic Encryption used partial scheme and with the passage of time, researchers developed a full Homomorphic Encryption scheme which allowed complete computation on any type of data.

2.5.2. *Garbled Circuits (GCs)*. Yao’s garbled circuit method provides a general mechanism for building a secure two parties  $x$  and  $y$ , respectively, to develop an arbitrary Boolean function  $f(x, y)$  without disclosing information regarding

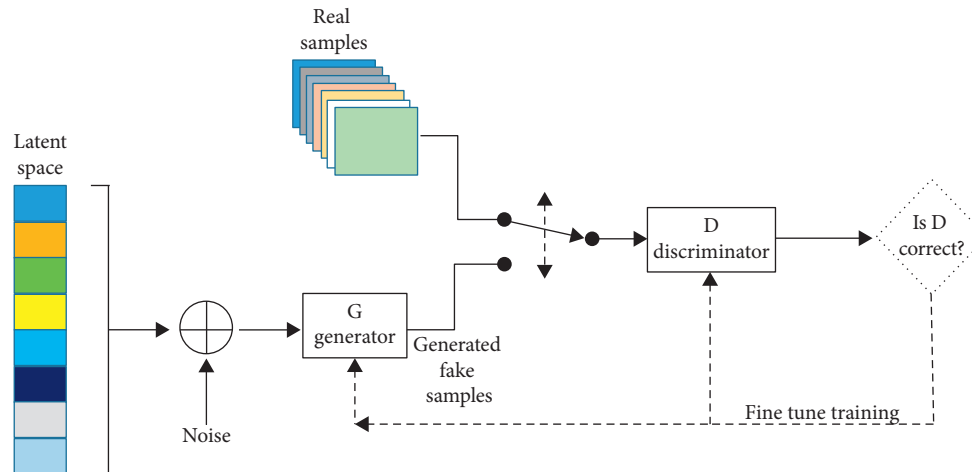


FIGURE 6: Generative adversarial network.

inputs irrespective of output of the function. The basic idea behind this algorithm is that one party will prepare the encrypted version of the circuit by computing  $f$  and the second party will obviously compute the output of the garbled circuit without knowing any value and information of the first circuit [28]. For example, in the 1st step, the first party will assign random keys to each wire of the circuit. The mentioned circuit has gates, and the first party shall encrypt output keys of the gates by using the associated input key and generate a garbled table [29]. The first party will then send the developed tables to the second party along with the associated input keys. On the other hand, the second party will get generated garbled tables and input keys. The second party then decrypts each gate that was encrypted by the first party until they find the output keys of the circuit [30]. The first party after decryption of the circuit will map the output keys to generate the plain text of the circuit.

**2.5.3. Goldreich Micali Wigderson (GMW).** It is also a generic secure function evaluation protocol, and it was developed in the year 1987 with the idea to evaluate the circuit through wire values by using secure linear secret sharing. This is like the Garbled Circuit protocol; this also requires the function that designates as a Boolean circuit [31, 32]. However, unlike Garbled Circuits, two users are required to cooperate for each AND gate. Thus, all AND gates are handled independently and in parallel, and the linear complexity is used in respect of the circuit. This technique is only used in short-level communication.

**2.5.4. Differential Privacy (DP).** DP is a metric that determines how much information about one entry in a database is exposed when a query is made to the database [33]. To preserve the privacy of database entries, carefully selected noise is added to the database so that the statistical properties of the database are retained while each data point is changed due to added noise [34]. Equally, DP can be considered as a way to reduce the dependency between the query result and individual data points in the database, thus

reducing the leakage of information. It ensures that the attacker cannot infer any high confidence information from the databases or forms that have been released [35].

**2.5.5. Share Secret (SS).** It is a way to distribute the secret to two or more parties where each share does not give any information/data about the secret, but the secret can be reconstructed from the posts. One of the utmost famous Share Secret variants is Share Secret additive. In this case, the secret is shared by taking random samples and creating the last post so that collecting all the shares gets the secret value [36]. The secret of the algorithm can be reconstructed by inserting all the shares.

### 3. Deep Learning Private Data Frameworks

In this section, we will briefly describe the most efficient private data security frameworks for Deep Learning. All the given below frameworks are highly protected in the light of the Honest-but-Curious (HbC) adversary model. All parties adhering with this protocol are supposed to follow the protocol's instruction, but it is also observed that parties might infer more information. The said protocol is very secured as it stops the malicious attacks and also stops parties to deviate from protocol norms.

**3.1. Shokri and Shmatikov.** The authors suggested a method for maintaining privacy based on Differential Privacy (DP) for Deep Learning when the data are laid with different parties. In this situation, each party locally installs its own version of the neural network and selectively participates in some parameters updated with other parts. The authors proposed that the algorithm should be run on different machines in parallel, and then the results of the separate machines shall be aggregated to generate the final result. In order to protect the private data of the users, the Differential Privacy algorithm shall be applied when the parameters are shared instead of sharing the initial values. As a result, an

exchange is introduced between the precision of the trained neural network and the specificity of the data.

**3.2. SecureML.** It is a system to learn to maintain privacy in general and neural networks in particular. The system is based on the HE, GC, and SS protocols. Data owners secretly share their data with servers that do not comply with the rules and that train the particular neural network [37]. SecureML uses a more efficient custom activation feature to train a neural network using secure account protocols [31]. At the end of the account, the managed model is shared privately between the servers. In addition to training, SecureML also provides a conclusion to maintain privacy.

**3.3. Google.** A secure collection protocol was introduced for high-dimensional operators maintained by premium users. These protocols can be used in a unified education in which users maintain their databases and forms [38]. The core server recognizes the intelligent intelligence model by securely assembling the user's learning updates. The method is based on the covert exchange of the code and is powerful against users who exit the protocol at any time [39].

**3.4. CryptoNets.** CryptoNets, by applying ML to the problem regarding medical, educational, financial, or other kinds of confidential data, requires not only accurate forecasts but also careful cares to keep them safe and secure [40]. CryptoNets is basically developed by the Microsoft Research group, by introducing levelled Homomorphic Encryption (LHC). Due to nonlinear activation functions that cannot be achieved using LHE, the authors proposed that the activation functions are approached using polynomials of multiple degrees [41]. Therefore, the neural network must be retrained in plain text with the same activation function to maintain good prediction accuracy. Another disadvantage of this approach is that there is a certain limit on the number of serial multipliers imposed by LHE that makes the solution prohibitive. In addition, CryptoNets has an exchange of privacy/utility to achieve a higher level of privacy, and accuracy must be reduced within the same computing capabilities.

**3.5. MiniONN.** The authors observed that there are still privacy-preserving risks, and clients are still facing disclosure of sensitive information threats [42]. The MiniONN introduced the method for transmuting the existing DNN to the newly developed Oblivious Neural Network that addresses the privacy-preserving risks. It offers that the server does not know about the input of the client-side and the client also does not know about the model [42]. The performance of the MiniONN is better than CryptoNets and SecureML. It influences additive Homomorphic Encryption, Garbled Circuits, and secret sharing and also supports activation functions viz-a-viz pooling for CNN. It also has two main stages:

- (i) An offline phase that supports additive Homomorphic Encryption that is not dependent on input
- (ii) An online phase consists of GC and SS; nonlinear layers use GC and SS for processing

**3.6. Chameleon.** This protocol consists of mix frameworks regarding privacy preservation. This framework gets the benefits of the existing work of GMW protocol for in-depth analysis of the activation function and other Garbled Circuits for complicated activation functions and pooling layers. Chameleon uses secret sharing for arithmetic and addition functions. It has offline and online phases like in MiniONN [41]. The offline computation provided more fast computation for prediction instead of the online phase. Like SecureML, the Chameleon also requires two noncolluding machines, and unlike SecureML, it does not allow the involvement of the third party during the online phase. The Chameleon is more efficient as compared with all other discussed techniques.

**3.7. DeepSecure.** It is one of the modern frameworks based on the Garbled Circuit protocol. Since garbled circuit is a generic function evaluation protocol, the framework supports all nonlinear activation functions. DeepSecure offers the idea of decreasing the size of the data and the network before the implementation of the Garbled Circuits, thus compressing the account and connecting up to two things in size [43]. The preprocessing phase is independent of the basic encryption protocol and can be adopted by any other backend engine for its inference. DeepSecure also supports secure outsourcing of the account to a secondary server when the client has restricted resources.

## 4. Deep Learning Threats and Attacks

Deep learning faces various types of threats and attacks, and all famous threats and attacks are listed below.

**4.1. Security Attack Taxonomy.** Ji et al. [44] proposed classification of security threats for Deep Learning in 3 different angles, which influence classifieds, security breaches, and privacy of attacks.

In the view of impact, security risks and threats of Deep Learning are characterized into two categories.

**4.1.1. Causative Attack.** In the causative attack used to decrease the performance and reliability of the training processes, the machine learning algorithm provided incorrect training data after modification in the labels of the samples that are not covered under the decision limit. Many researchers performed causative attacks on the images and revealed that it expressively decreases the performance of the training phase.

This means that the opponents have the ability to change the input of training data, which becomes the cause of changes in the parameters of the learning models during

recycling, resulting in a substantial reduction in the presentation of jobs in succeeding taxonomy tasks.

*4.1.2. Exploratory Attack.* The exploratory attacks basically do not influence on a training dataset. The key objective of the exploratory attacks is to get knowledge with respect to the learning algorithm as much as it can about the basic system. Model invasion attack, model extraction, and membership inference are the examples of the exploratory attacks.

In a security break viewpoint, threats to Deep Learning may be characterized into 3 groups:

(1) *Integrity Attack.* The integrity attack occurs and then the Deep Learning models failed to trace the negative cases when categorizing harmful samples. The output of the system will clearly show that the integrity of the learning machine has been compromised. Suppose, we used spam filter to stop unwanted/harm messages, and if the attacker sends a message that has unwanted/harm words then, the filter does not get it. The integrity attack is tested through exploratory testing.

(2) *Availability Attack.* The availability attack is the opposite of an integrity attack in which the Deep Learning models filtered out the legitimate cases during the categorization of the unwanted/harmful samples. The output of the system will clearly show that the availability of the learning machine has been compromised and it is no more available and hacked. The DoS attack is one of the examples of availability wherein legitimate cases failed to cross the filters and ultimately the system becomes compromised.

(3) *Privacy Violation Attack.* In the privacy violation attack, the attacker becomes successful to get the sensitive/confidential information of the system from both training and learning models. In terms of attack privacy, security threats for Deep Learning have further 02 categories.

*4.1.3. Targeted Attack.* It is highly dangerous, and it directly decreases the performance of the classifier in a single specific sample or set of one of the samples.

*4.1.4. Indiscriminate Attack.* An indiscriminate attack is the subtype of the poisoning attack. The attacker's key goals are to increase the general classification error. Further, the indiscriminate attack always chooses a random value from the training sample. It randomly fails the classifier.

*4.2. Deep Learning Attack Types.* Although Deep Learning becomes successful to get draw the attention of the industry its security and privacy challenges, unfortunately, it could not get full attention as it should have. Here, we discuss the attack surface of the machine learning and discuss the weaknesses in the implementation of Deep Learning.

During the research, numerous types of attacks targeting DL applications and containing DoS attacks, evasion attacks,

and organic termination attacks are revealed. Though all these attacks are different in its nature and in terms of their offensive objectives, the attacker's attack sources in Deep Learning applications are essentially from the following three angles.

*4.2.1. Deep Learning Attack Surface Type-I.* Deep learning application after trained mostly works on input data of the user for its classification. The attacker planned a malformed input attack on the input files or sometimes the network [24]. This type of attack applies to image recognition application which uses files on input and also applied to the applications that use sensors and cameras on the input. Due to the input type of the application, this risk can be reduced to implement risk mitigation techniques but the risk cannot be eliminated.

*4.2.2. Deep Learning Attack Surface Type-II.* This surface attack is also called a poisoning attack. The earlier surface type attack is due to the contaminated input data type of the application. This type of attack is not dependent on the application flaws or software breaches. However, defects in applications can become the reason of data poisoning easier. Suppose we observed variation in the procedure of analyzing the image in the frame and in common desktop applications. This variation allows the contamination of confidential data without being observed by the people who monitor the training process.

*4.2.3. Deep Learning Attack Surface Type-III.* It is a great chance of an attack on the Deep Learning applications if the developer will opt the model developed by the experts. Even though many programmers plan and create models from the beginning, many templates of the models exist for programmers who do not sufficient knowledge of machine learning. In this scenario, the attacker has also access to the template of the models. Like attacks of data poisoning, an attacker can easily attack all those applications and can get access to the private data that uses external models without any barrier. However, implementation flaws, such as a security vulnerability in the form analysis code, help attackers hide damaged models.

The readers should keep in mind that there are many types of attack surfaces and differ from each other, and it depends on the particular application, but above these 03 types of attack, surfaces cover most of the attack area. The comparison of attacking techniques against Deep Learning is given in Table 1.

*4.3. Types of Threats.* During the literature review, the authors studied many types of threats that affect the functionality of Deep Learning, and these threats targets different stages of Deep Learning. Here, in this paper, we are going to present the threat caused by the malformed input with the assumption that Deep Learning applications are taking input from files or networks.

TABLE 1: Comparison of attacking techniques against Deep Learning.

Attacking technique	Advantages	Disadvantages	Countermeasure technique
Causative attack [45, 46]	Influence on training data and exploits misclassifications	Time consuming Not fit for large dataset	[45, 47–49]
Exploratory attack [47]	Changes the discriminant results Misclassifies positive sample	Resource consuming	[50–52]
Integrity attack [53]	False negative passes through the system	Easily detected	[54–56]
Availability attack [57]	False positive results in blocking records	Time and resource consuming	[58–60]
Privacy violation attack [61]	Easily exploit the training dataset	Its performance is not reliable as it based on iterations	[62–64]
Targeted attack [65]	Misclassified to any arbitrary class	It does not provide assurance about the generated samples	[66–68]
Indiscriminate attack [69]	Good trade-off Highly efficient	Perturbation is high	[70, 71]

*4.3.1. Deep Learning Threat Type-I.* The most common weaknesses in Deep Learning frameworks are program errors that which cause software crashes, an infinite loop, or full memory depletion. The immediate threat of these errors is the denial of service attacks for applications running at the top of the window [72].

*4.3.2. Deep Learning Threat Type-II.* Deep Neural Networks are vulnerable to attacks at the time of its testing [45–48]. For example, in image recognition, an attacker may insert little noise to test a sample so that the error is classified as a DNN [73]. An example of a noise test is called an adversarial example. The noise is usually so small for a human. The benign is the alternate name of the adversarial example.

Evasion attacks are one of the Deep Learning attacks that restrict sensitive security and protection applications, like vehicles that drive on their own. Examples of self-driving adversaries can make unwanted decisions [74–78]. For example, one of the basic capabilities of autonomous cars is to automatically identify stop signals and traffic lights of the road.

Let us say, the adversary generates an adversarial stop, which means that the adversarial adds many imperceptible points to the stop, so that the vehicle that is driving alone is not recognized as a stop. As a result, vehicles that drive on their own will not stop at the stop sign and may collide with other vehicles, which could lead to serious traffic accidents.

There are many memory corruption-related bug in Deep Learning framework which may be a cause of wrong output. The evasion can be achieved through exploiting bugs in the Deep Learning framework by overwriting classification and control flow. In order to develop an effective defense against evasion attack, Goodfellow et al. [79] proposed adversarial training and adversarial example by introducing training of a DNN through augmenting training dataset. In order to train a DNN, the system generates training adversarial example through evasion attacks. The learner understands both the original training example and relating adversarial examples.

The adversarial training is weak as compared with adversarial examples that cannot be seen during training. Papernot et al. [80] developed a decontamination based

technique to train Deep Neural Networks and Carlini and Wagner [81] revealed that their generated attacks have maximum success for Deep Neural Networks trained with concentration. Furthermore, Carlini and Wagner [81] determined that all measures must be assessed against the taxonomy of evasion attacks.

*4.3.3. Deep Learning Threat Type-III.* The software bugs of the systems that hosted Deep Learning applications on its operating system can be hijacked due to remote compromise and application bugs [44, 82]. This mostly happens when the system is connected with the cloud system and the Deep Learning applications are also running on that cloud-based system. All the input to the Deep Learning system is received through the network.

## 5. Defense Techniques against Security Issues in Deep Learning

During the literature review, many defense techniques against security concerns of Deep Learning were found, and we categorized these techniques into two major categories known as evasion and poisoning. Further, there are many evasion attack mitigation techniques, but in this chapter, only well-known and effective types are explained herein. Whereas, in a similar fashion, the defense techniques against the poisoning attack proposed by the researcher are also given in Section 5.1. These defense techniques cannot 100% overcome the attacks, but these techniques can improve the prediction of the results.

*5.1. Defense against Evasion Attacks.* The most effective method of defense against evasion attack is to augment the adversarial examples and detect adversarial examples, adversarial training, and defensive distillation.

*5.1.1. Detecting Adversarial Examples.* The researchers [81, 83, 84] proposed different techniques to detect adversarial examples in the input and to create different benign and adversarial examples. As we mentioned earlier, the target of the attacker is to add more noise to formulate

effective adversarial examples. According to [83], it is not easy to detect such adaptive attacks, and some detection techniques effectively work while some ineffective. The main problem in the detection of adversarial examples is that it is unclear, and it is very hard to manage the testing example that is used to predict the adversarial example. Therefore, the expert should label the test examples manually. We give the above example of an automated/self-driving car which automatically takes decisions; it is not possible for the human to mark the label manually to detect adversarial example [75, 85–90].

Meng and Chen [84] proposed an approach to verify adversarial examples through testing examples and also the template of the testing example. According to the authors, if during verification of the adversarial example, it is proved through testing examples, then there is no need to label the classifier; otherwise, in the case of not predicted, the testing examples are required to be reformed through the reformer by removing unwanted noise from the testing example. After the completion of this task, the classifier shall label the example of testing to the Deep Neural Network and will consider it a genuine testing example. The experiments of MagNet show that it successfully presented defense against the evasion attacks.

*5.1.2. Adversarial Training.* Goodfellow et al. [79] presented a technique to train a Deep Neural Network through expanding dataset of training along with several adversarial examples and named it as adversarial training. In order to handle the evasion attack, the author proposed training benign examples against each training adversarial example. The learner of the system will use the backpropagation algorithm to get the knowledge of the Deep Neural Network through the original benign example and the attack adversarial example. The following authors also proposed the variants of the adversarial training. The authors used robust optimization techniques to solve min-max optimization problems. The core issue in the adversarial training is accuracies in the benign example.

*5.1.3. Defensive Distillation.* Sethi et al. [50] projected a method dependent on distillation for Deep Neural Network Training. The Deep Neural Network is trained first using a typical method. For each training example, Deep Neural Network produces a set of confidence levels. Confidence levels are treated as a soft mark for the training example. Due to software labels and training examples, Deep Neural Network weights are retrained. The named  $T$  parameter is used for the distillation temperature in the soft top layer during both training sessions to control confidence levels. In addition, noise is added to good example when hostile examples generated are slightly higher in distilled Deep Neural Network than in non-Deep Neural Network.

*5.2. Defense against Poisoning Attack.* The framework suggested in [91] takes the method of eliminating extreme values that fall outside the relevant group. In the binary

grouping, they seek to discover the midpoints of the positive and negative categories. Then, the authors eliminate the points that are not near to the relevant focal point. To get information about these points, they use the defense field that eliminates points outside the radius of the ball, and a slab defense ignores points away from the line in a complementary manner.

Sun et al. [57] selected to rename the data points that are external values instead of deleting them. Attack flipping label is a distinct item for data poisoning that permits an attacker/hacker to control the appointment of a trifling number of training points. The author further describes a mechanism that studies points beyond the limits of the resolution to be harmful and reclassifies them. The procedure resets the label of every case.

Paudice et al. [92] also propose a protection mechanism to alleviate the intensity of poisoning attacks through remote sensing. The label tries to have the utmost influence on the protector with an inadequate number of poison points. The external detection process computes the external result of every  $x$  in the original data set. Further, there are many and different methods to calculate the external result.

It is stated that the impact functions are used to trail the predictions of the model and find the best persuasive data points that are accountable for the given forecast. It shows that the approximation of functions is still able to provide important materials that are nontransferable and nondiscriminatory models where the theory collapses [93]. The authors also assert that by using impact functions; the protector can verify the priority data only by the degree of impact. This method is superior to the previous methods to determine the greatest loss of training to eliminate contaminated samples.

The authors of this paper, to convince of the researchers, compared the advantages and disadvantages of existing countermeasure methods of Deep Learning, as presented in Table 2.

Various Deep Learning security attacks and corresponding countermeasures have drawn the attention of the industry and researchers. Table 3 presents comparative results and qualitative analysis of attacks and corresponding defensive techniques.

## 6. Observations and Recommendations

Deep Learning is providing new techniques to solve security problems. It introduced significant improvements over stereotype techniques and classical ML algorithms. Table 4 is a list of Deep Learning papers related to Deep Learning that we reviewed during the literature review. This table consists of methods used to solve the problems and citations of each paper. The authors reviewed 41 papers in this survey; the majority of the researchers conducted their study on malware detection and intrusion detection. During the survey, we also noticed some new areas of health security and vehicle security wherein Deep Learning techniques can be applied. Autoencoder technique is the most favorite one for the researchers to detect malware; thereafter, the Recurrent Neural Networks (RNNs) are also used for the same purpose

TABLE 2: Comparison of countermeasure techniques of Deep Learning.

Countermeasure methods	Advantages	Disadvantages
Adversarial training [94]	Very easy to understand and implement Scalable and have the ability to handle the complex dataset	It depends upon the sample size in the training phase
Defense distillation [80]	Sample and have the defense ability	Difficult to converge and high complexity
Ensemble method [95]	Model-independent, good generalization	Do not rebut the training data and computation overhead
Differential Privacy [96]	Preserves the privacy of training and learning data Low overhead, low complexity	It also affects legitimate data and model-independent
Homomorphic Encryption [97]	Maintains security and privacy of data and simple	It increases the data size and extensive computation overhead

TABLE 3: Comparison of attacking and defensive techniques in Deep Learning.

Attack/defense	Technique	Training/testing	Taxonomy
Attack	Adversarial label flips	Training	Confidentiality, integrity, and reliability
Attack	Enchanting	Training	Integrity and reliability Exploratory attack Exploratory attack
Attack	Obfuscation	Training	Targeted attack Integrity and reliability
Attack	Poisoning	Training	Confidentiality, integrity, and reliability Causative attack Indiscriminate attack
Attack	Impersonate	Training	Exploratory attack Integrity and reliability
Defense	Adversarial training	Training	Creates a fool-proof system, improves the safety and security of the system, and defeats security attacks
Defense	Defense distillation	Training	It ensures the integrity, availability, reliability, and authenticity. Smooth classifier
Defense	Ensemble method	Training	Detects anomalies in the network Boosts data mining and intrusion detection
Defense	Differential Privacy	Training and testing	It protects the privacy of the data
Defense	Homomorphic Encryption	Training and testing	It protects the privacy of the data to ensure confidentiality

TABLE 4: Survey of Deep Learning approaches, methods, and security applications.

DL method	Citation	No of times cited (as of 17.01.2020)	Security application
Autoencoder	Hardy et al. [98]	59	Malware detection
Autoencoder	Rhode et al. [99]	50	Malware detection
Autoencoder	Kalash et al. [100]	35	Malware classification
Autoencoder	Wang and Yiu [101]	17	Malware classification
Autoencoder	Chalopathy and Chawla [102]	61	Anomaly detection
Autoencoder	Chen and Ye [103]	7	Adversarial malware attacks
Autoencoder	Maniath et al. [104]	10	Ransomware detection
Autoencoder	Zakaria [105]	-	Ransomware detection
Autoencoder	Demetrio [106]	10	Adversarial malware binaries
Autoencoder	James and Aimone [107]	18	File Type Identification
Autoencoder	Wang [108]	132	Traffic Identification
Autoencoder	Fadlullah et al. [109]	238	Network traffic control systems
Autoencoder	Aminanto et al. [110]	46	Wi-Fi impersonation detection
Autoencoder	Aceto et al. [111]	34	Mobile encrypted traffic
Autoencoder	Mi et al. [112]	15	Spam identification
Autoencoder	Shi et al. [113]	57	User authentication
Autoencoder	Catak and Yazi [114]	—	Malware classification
CNN	Gibert [115]	33	Malware classification

TABLE 4: Continued.

DL method	Citation	No of times cited (as of 17.01.2020)	Security application
CNN	Cha et al. [116]	517	Crack damage detection
CNN	Murata and Yamanishi [117]	01	Download attack
CNN	Vinayakumar et al. [118]	58	Network intrusion detection
CNN	Wang et al. [119]	109	Malware traffic classification
RNN	Yin et al. [120]	225	Intrusion detection
CNN RNN	Maleh [121]	-	Malware classification
CNN RNN	Kolosnjaji et al. [122]	179	Malware detection
CNN RNN	Tobiyama et al. [123]	98	Malware detection
CNN RNN	Yu et al. [124]	33	Intrusion detection
CNN (dynamic)	Hill and Bellekens [125]	03	Malware detection
DNN	M.-J. Kang and J.-W. Kang [126]	224	Intrusion detection
DNN	Potluri and Diedrich [127]	67	Intrusion detection
DNN	Dahl et al. [128]	283	Malware classification
DNN	Sebastián et al. [129]	146	Massive malware labeling
DNN RNN	Mi et al. [130]	15	Insider threat
DNN RNN	Mi et al. [131]	03	Spam detection
GAN	Anderson et al. [132]	67	Intrusion detection
GAN	Yu et al. [133]	23	Character detection
GAN	Zhauniarovich et al. [134]	20	Malicious domains detection
RBM	Alrawashdeh and Purdy [135]	63	Intrusion detection
RBM	Yuan et al. [136]	202	Malware detection
RBM	Wang et al. [137]	216	Defect prediction
RBM	Chen et al. [138]	67	Detecting android malware

as well as to detect information security threats. Restricted Boltzmann Machines (RBMs) are also used for the same purpose, but we cannot find much study using this technique for security purposes. Different authors combined autoencoders and RNN techniques to train the unlabelled data. RBM is a popular technique due to its easy implementation and simplicity.

After studying the above techniques, it is very difficult for the authors to exactly define the performance of the techniques due to different datasets and metrics. It is pertinent to add here that the performance of these techniques/methods varies across security areas. The information security domain has a vast range of data collected through different sources to apply Deep Learning tests. The researches/studies could not be completed and generate accurate results because a large volume of datasets is not publically available. The majority of the dataset sources are small and old. To develop a security solution through the meaningful method, it is necessary to test the method on large, updated, and reliable datasets. The results of the methods should be compared with each other through real-time scenarios.

## 7. Conclusion

Deep learning has now become part of our daily lives, and when new technology invested, definitely security and privacy issues arise. In recent years, extensive research was carried out on the security and privacy preserving issues and its counter frameworks for Deep Learning and Deep Neural Network's training and interface modules. Therefore, security and privacy become very critical and important issues as in the other technologies that cannot be overlooked.

During the literature review, we found two basic types of security attacks: evasion and poisoning. We also presented

the effective countermeasures of these two types of attacks. We explained both security and private attacks, frameworks, and countermeasure techniques.

These frameworks have cryptographic primitives and numerous characteristics. It should be noted that private interference frameworks have no complete capability to provide DNNs security and privacy. We outline the details of different types of security attacks on Deep Learning. There are many types of attacks that are invested to exploit the Deep Learning results so that model information may be extracted or get the knowledge about the training data like model inversion, model extraction, and membership inference. The said attacks steal training data and generate expected results. The private training section of Deep Learning has more computation overhead as compared with the interface. Therefore, more concentration and research are required in this direction to develop a more efficient solution for the privacy preservation of the data while maintaining models.

Privacy risks always persist due to various characteristics of the Deep Neural Networks which is actually relying upon a huge amount of input training data. In this chapter, we also discussed possible privacy threats on sensitive and confidential Deep Learning model's data. Various studies have been conducted on privacy preserving attacks by using Deep Learning.

For future work, it is essential for the researchers to deeply investigate different cryptographic primitive's solutions for DNNs. A mixed protocol technique can reduce the computation overhead on the security and privacy preserving solutions. Furthermore, customization of the privacy and security protocols for DNNs is also an interesting and open research area to develop a viable solution. The authors are also intended to perform their research in the application



of Deep Learning especially in the area of astrophysics, plasma physics, atomic physics, thermodynamics, electromagnetic, machines, nanotechnology, fluid mechanics, electro hydrodynamics, signal processing, power, energy, bioinformatics, economy, and finance.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this article.

## References

- [1] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, 2018.
- [2] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.
- [3] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [5] M. A. Nielsen, *Neural Networks and Deep Learning*, Vol. 25, Determination press, San Francisco, CA, USA, 2015.
- [6] R. Miikkulainen, "Evolving deep neural networks," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, pp. 293–312, Elsevier, Amsterdam, Netherlands, 2019.
- [7] Y.-G. Jiang, Z. Wu, J. Wang, X. Xue, and S.-F. Chang, "Exploiting feature and class relationships in video categorization with regularized deep neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 352–364, 2018.
- [8] V. Kotu and B. Deshpande, "Deep learning," in *Data Science*, pp. 307–342, Elsevier, Amsterdam, Netherlands, 2019.
- [9] S. Gollapudi, "Deep learning for computer vision," in *Learn Computer Vision Using OpenCV*, pp. 51–69, Apress, Berkeley, CA, USA, 2019.
- [10] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8614–8618, Vancouver, Canada, May 2013.
- [11] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4580–4584, South Brisbane, Australia, April 2015.
- [12] T. P. Huster, C. J. Chiang, R. Chadha, and A. Swami, "Towards the development of robust deep neural networks in adversarial settings," in *Proceedings of the 2018 IEEE Military Communications Conference (MILCOM)*, pp. 419–424, Los Angeles, CA, USA, October 2018.
- [13] U. Shaham, A. Cloninger, and R. R. Coifman, "Provable approximation properties for deep neural networks," *Applied and Computational Harmonic Analysis*, vol. 44, no. 3, pp. 537–557, 2018.
- [14] P. Mohamed Shakeel, S. Baskar, V. R. Sarma Dhulipala, S. Mishra, and M. M. Jaber, "Maintaining security and privacy in health care system using learning based deep-Q-networks," *Journal of Medical Systems*, vol. 42, no. 10, p. 186, 2018.
- [15] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security—CCS'15*, pp. 1310–1321, Denver, Colorado, USA, 2015.
- [16] R. Vinayakumar, K. P. Soman, and P. Poornachandran, "Detecting malicious domain names using deep learning approaches at scale," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1355–1367, 2018.
- [17] B. Feng, Q. Fu, M. Dong, D. Guo, and Q. Li, "Multistage and elastic spam detection in mobile social networks through deep learning," *IEEE Network*, vol. 32, no. 4, pp. 15–21, 2018.
- [18] J. Li, L. Sun, Q. Yan, Z. Li, W. Srisa-an, and H. Ye, "Significant permission identification for machine-learning-based android malware detection," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3216–3225, 2018.
- [19] W. W. Stead, "Clinical implications and challenges of artificial intelligence and deep learning," *JAMA*, vol. 320, no. 11, pp. 1107–1108, 2018.
- [20] D. Cohen, J. Foley, H. Zamani, J. Allan, and W. B. Croft, "Universal approximation functions for fast learning to rank," in *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval—SIGIR'18*, pp. 1017–1020, New York, NY, USA, 2018.
- [21] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856, Long Beach, CA, USA, 2018.
- [22] A. Hassan, M. Kamran, A. Illahi, and R. M. A. Zahoor, "Design of cascade artificial neural networks optimized with the memetic computing paradigm for solving the nonlinear Bratu system," *The European Physical Journal Plus*, vol. 134, no. 3, p. 122, 2019.
- [23] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing Chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.
- [24] Q. Yang, P. Yan, Y. Zhang et al., "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.
- [25] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–35, 2018.
- [26] D. Boneh, "Threshold cryptosystems from threshold fully homomorphic encryption," in *Advances in Cryptology—CRYPTO 2018*, pp. 565–596, Springer, Berlin, Germany, 2018.
- [27] S. Halevi, Y. Polyakov, and V. Shoup, "An improved RNS variant of the BFV homomorphic encryption scheme," in *Topics in Cryptology—CT-RSA 2019*, pp. 83–105, Springer, Berlin, Germany, 2019.
- [28] Q. Yang, G. Peng, P. Gasti et al., "MEG: memory and energy efficient garbled circuit evaluation on smartphones," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 913–922, 2019.
- [29] A. Saleem, A. Khan, F. Shahid, M. Masoom Alam, and M. K. Khan, "Recent advancements in garbled computing: how far have we come towards achieving secure, efficient and reusable garbled circuits," *Journal of Network and Computer Applications*, vol. 108, pp. 1–19, 2018.

- [30] A. Dupin, D. Pointcheval, and C. Bidan, "On the leakage of corrupted garbled circuits," in *Proceedings of the Provable Security*, pp. 3–21, Jeju, South Korea, October 2018.
- [31] S. Sharma and K. Chen, "Privacy-preserving boosting with random linear classifiers," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2294–2296, New York, NY, USA, 2018.
- [32] H. Ahmad, L. Wang, H. Hong et al., "Primitives towards verifiable computation: a survey," *Frontiers of Computer Science*, vol. 123, pp. 451–478, 2018.
- [33] J. Wang, J. Zhang, W. Bao, X. Zhu, B. Cao, and P. S. Yu, "Not just privacy," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2407–2416, New York, NY, USA, 2018.
- [34] N. C. Abay, Y. Zhou, M. Kantarcioglu, B. Thuraisingham, and L. Sweeney, "Privacy preserving synthetic data release using deep learning," in *Machine Learning and Knowledge Discovery in Databases*, pp. 510–526, Springer, Berlin, Germany, 2019.
- [35] N. Hynes, D. Dao, D. Yan, R. Cheng, and D. Song, "A demonstration of sterling," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 2086–2089, 2018.
- [36] L. T. Phong and T. T. Phuong, "Privacy-preserving deep learning via weight transmission," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 11, pp. 3003–3015, 2019.
- [37] P. Mohassel and Y. Zhang, "SecureML: a system for scalable privacy-preserving machine learning," in *Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP)*, pp. 19–38, San Jose, CA, USA, 2017.
- [38] G. Lin, N. Sun, S. Nepal, J. Zhang, Y. Xiang, and H. Hassan, "Statistical twitter spam detection demystified: performance, stability and scalability," *IEEE Access*, vol. 5, pp. 11142–11154, 2017.
- [39] K. Bonawitz, V. Ivanov, B. Kreuter et al., "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, New York, NY, USA, 2017.
- [40] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the GAN," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security—CCS'17*, pp. 603–618, New York, NY, USA, 2017.
- [41] M. S. Riazi, C. Weinert, O. Tkachenko et al., "Chameleon," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security—ASIACCS'18*, pp. 707–721, New York, NY, USA, 2018.
- [42] J. Liu, M. Juuti, Y. Lu, and N. Asokan, "Oblivious neural network predictions via MiniONN transformations," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security—CCS'17*, pp. 619–631, New York, NY, USA, 2017.
- [43] B. D. Rouhani, M. S. Riazi, and F. Koushanfar, "Deepsecure," in *Proceedings of the 55th Annual Design Automation Conference—DAC'18*, pp. 2:1–2:6, New York, NY, USA, 2018.
- [44] Y. Ji, X. Zhang, S. Ji, X. Luo, and T. Wang, "Model-reuse attacks on deep learning systems," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pp. 349–363, New York, NY, USA, 2018.
- [45] C. Burkard and B. Lagesse, "Analysis of causative attacks against SVMs learning from data streams," in *Proceedings of the 3rd ACM on International Workshop on Security And Privacy Analytics—IWSPA'17*, pp. 31–36, Scottsdale, AZ, USA, March 2017.
- [46] Y. Li and D. S. Yeung, "A causative attack against semi-supervised learning," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 196–203, Qingdao, China, July 2014.
- [47] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar, "Adversarial machine learning," in *Proceedings of the 4th ACM workshop on Security and artificial intelligence*, pp. 43–58, Chicago, IL, USA, October 2011.
- [48] Y. Shi and Y. E. Sagduyu, "Evasion and causative attacks with adversarial deep learning," in *Proceedings of the MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM)*, pp. 243–248, Baltimore, MD, USA, October 2017.
- [49] B. Miller, "Adversarial active learning," in *Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop*, pp. 3–14, Angers,, February 2014.
- [50] T. S. Sethi, M. Kantardzic, and J. W. Ryu, "'Security theater': on the vulnerability of classifiers to exploratory attacks," in *Proceedings of the Pacific-Asia Workshop on Intelligence and Security Informatics*, pp. 49–63, Jeju Island, South Korea, May 2017.
- [51] T. S. Sethi and M. Kantardzic, "Data driven exploratory attacks on black box classifiers in adversarial domains," *Neurocomputing*, vol. 289, pp. 129–143, 2018.
- [52] L. Halawi, R. McCarthy, and N. Muoghalu, "Student approaches to learning: an exploratory study," *Issues in Information Systems*, vol. 10, no. 1, p. 13, 2009.
- [53] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar, "Can machine learning be secure?" in *Proceedings of the 2006 ACM Symposium on Information, computer and communications security*, pp. 16–25, Taipei Taiwan, March 2006.
- [54] S. Ntalampiras, "Automatic identification of integrity attacks in cyber-physical systems," *Expert Systems with Applications*, vol. 58, pp. 164–173, 2016.
- [55] B. Biggio and F. Roli, "Wild patterns: ten years after the rise of adversarial machine learning," *Pattern Recognition*, vol. 84, pp. 317–331, 2018.
- [56] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 309, pp. 2805–2824, June 2019.
- [57] L. Sun, J. Wang, P. S. Yu, and B. Li, "Adversarial attack and defense on graph data: a survey," 2018, <https://arxiv.org/abs/1812.10528>.
- [58] A. Erba, "Real-time evasion attacks with physical constraints on deep learning-based anomaly detectors in industrial control systems," 2019, <https://arxiv.org/abs/1907.07487>.
- [59] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li, "Manipulating machine learning: poisoning attacks and countermeasures for regression learning," in *Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP)*, pp. 19–35, San Francisco, CA, USA, May 2018.
- [60] M. Sun, "Data poisoning attack against unsupervised node embedding methods," 2018, <https://arxiv.org/abs/1810.12881>.
- [61] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, "A survey on security threats and defensive techniques of machine learning: a data driven view," *IEEE Access*, vol. 6, pp. 12103–12117, 2018.
- [62] W. Li, Y. Wang, H. Li, and X. Li, "Leveraging Memory PUFs and PIM-based encryption to secure edge deep learning systems," in *Proceedings of the 2019 IEEE 37th VLSI Test Symposium (VTS)*, pp. 1–6, Monterey, CA, USA, April 2019.
- [63] A. Siddiqi, "Adversarial security attacks and perturbations on machine learning and deep learning methods," 2019, <https://arxiv.org/abs/1907.07291>.

- [64] Z. E. Mrabet, N. Kaabouch, H. E. Ghazi, and H. E. Ghazi, "Cyber-security in smart grid: survey and challenges," *Computers & Electrical Engineering*, vol. 67, pp. 469–482, 2018.
- [65] K. Eykholt, "Robust physical-world attacks on deep learning models," 2017, <https://arxiv.org/abs/1707.08945>.
- [66] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, pp. 828–841, 2019.
- [67] J. Rauber, W. Brendel, and M. Bethge, "Foolbox v0. 8.0: a python toolbox to benchmark the robustness of machine learning models," vol. 5, 2017, <https://arxiv.org/abs/1707.04131>.
- [68] S. Shen, S. Tople, and P. Saxena, "A uror: defending against poisoning attacks in collaborative deep learning systems," in *Proceedings of the 32nd Annual Conference on Computer Security Applications*, pp. 508–519, New York, NY, USA, December 2016.
- [69] M. Barreno, B. Nelson, A. D. Joseph, and J. D. Tygar, *The Security of Machine Learning*, p. 26, Springer, Berlin, Germany.
- [70] S. L. Wang, K. Shafi, C. Lokan, and H. A. Abbass, "Robustness of neural ensembles against targeted and random Adversarial Learning," in *Proceedings of the International Conference on Fuzzy Systems*, pp. 1–8, Barcelona, Spain, July 2010.
- [71] J. Peng and P. P. K. Chan, "Revised Naive Bayes classifier for combating the focus attack in spam filtering," in *Proceedings of the 2013 International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 610–614, Tianjin, China, July 2013.
- [72] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proceedings of the 2016 IEEE European Symposium on Security and Privacy (EuroS P)*, pp. 372–387, London, UK, April 2016.
- [73] C. Szegedy, "Intriguing properties of neural networks," 2013, <https://arxiv.org/abs/1312.6199>.
- [74] S. A. Butt, M. I. Tariq, T. Jamal, A. Ali, J. L. Diaz Martinez, and E. De-La-Hoz-Franco, "Predictive variables for agile development merging cloud computing services," *IEEE Access*, vol. 7, pp. 99273–99282, 2019.
- [75] M. I. Tariq, "Towards information security metrics framework for cloud computing," *International Journal of Cloud Computing and Services Science*, vol. 1, no. 4, p. 209, 2012.
- [76] M. I. Tariq, *Providing Assurance to Cloud Computing through ISO 27001 Certification: How Much Cloud is Secured after Implementing Information Security Standards*, CreateSpace, Scotts Valley, CA, USA, 2015.
- [77] M. I. Tariq, "Analysis of the effectiveness of cloud control matrix for hybrid cloud computing," *International Journal of Future Generation Communication and Networking*, vol. 11, no. 4, pp. 1–10, 2018.
- [78] M. I. Tariq, "Agent based information security framework for hybrid cloud computing," *KSII Transactions on Internet & Information Systems*, vol. 13, no. 1, 2019.
- [79] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, <https://arxiv.org/abs/1412.6572>.
- [80] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, "Distillation as a defense to adversarial perturbations against deep neural networks," in *Proceedings of the 2016 IEEE Symposium on Security and Privacy (SP)*, pp. 582–597, San Jose, CA, USA, May 2016.
- [81] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP)*, pp. 39–57, San Jose, CA, USA, May 2017.
- [82] J. Saxe, R. Harang, C. Wild, and H. Sanders, "A deep learning approach to fast, format-agnostic detection of malicious web content," in *Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW)*, pp. 8–14, San Francisco, CA, USA, May 2018.
- [83] N. Carlini and D. Wagner, "Adversarial examples are not easily detected," in *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security—AISec'17*, pp. 3–14, New York, NY, USA, 2017.
- [84] D. Meng and H. Chen, "MagNet: a two-pronged defense against adversarial examples," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security—CCS'17*, pp. 135–147, New York, NY, USA, 2017.
- [85] M. I. Tariq, S. Tayyaba, H. Rasheed, and M. W. Ashraf, "Factors influencing the cloud computing adoption in higher education institutions of Punjab, Pakistan," in *Proceedings of the 2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*, pp. 179–184, Islamabad, Pakistan, March 2017.
- [86] M. I. Tariq, D. Haq, and J. Iqbal, "SLA based information security metric for cloud computing from COBIT 4.1 framework," *International Journal of Computer Networks and Communications Security*, vol. 1, no. 3, pp. 95–101, 2013.
- [87] M. I. Tariq, S. Tayyaba, M. U. Hashmi, M. W. Ashraf, and N. A. Mian, "Agent based information security threat management framework for hybrid cloud computing," *International Journal of Computer Science and Network Security*, vol. 17, no. 12, p. 57, 2017.
- [88] M. I. Tariq, S. Tayyaba, M. W. Ashraf, and V. E. Balas, "8—deep learning techniques for optimizing medical big data," in *Deep Learning Techniques for Biomedical and Health Informatics*, B. Agarwal, V. E. Balas, L. C. Jain, R. C. Poonia, and Manisha, Eds., pp. 187–211, Academic Press, Cambridge, MA, USA, 2020.
- [89] M. I. Tariq, S. Tayyaba, M. W. Ashraf, and H. Rasheed, "Risk based NIST effectiveness analysis for cloud security," *Bahria University Journal of Information & Communication Technologies (BUJICT)*, vol. 10, no. Special Is, 2017.
- [90] M. I. Tariq, S. Tayyaba, M. W. Ashraf, H. Rasheed, and F. Khan, "Analysis of NIST SP 800-53 rev. 3 controls effectiveness for cloud computing," in *Proceedings of the 1st National Conference on Emerging Trends and Innovations in Computing & Technology*, pp. 88–92, Karachi, Pakistan, 2016.
- [91] J. Steinhardt, P. W. Koh, and P. Liang, "Certified defenses for data poisoning attacks," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 3520–3532, Long Beach, CA, USA, December 2017.
- [92] A. Paudice, L. Muñoz-González, A. György, and E. C. Lupu, "Detection of adversarial training examples in poisoning attacks through anomaly detection," 2018, <https://arxiv.org/abs/1802.03041>.
- [93] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," 2017, <https://arxiv.org/abs/1703.04730>.
- [94] F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel, "Ensemble adversarial training: attacks and defenses," 2018, <https://arxiv.org/abs/1705.07204>.
- [95] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amaratunga, "Ensemble deep learning for regression and time series forecasting," in *Proceedings of the 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)*, pp. 1–6, Orlando, FL, USA, December 2014.

- [96] 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—CCS’16*, pp. 308–318, Vienna, Austria, 2016.
- [97] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, “Privacy-preserving deep learning via additively homomorphic encryption,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, 2018.
- [98] W. Hardy, L. Chen, S. Hou, Y. Ye, and X. Li, “DL4MD: a deep learning framework for intelligent malware detection,” in *Proceedings of the International Conference on Data Mining (DMIN)*, p. 61, Las Vegas, NE, USA, July 2016.
- [99] M. Rhode, P. Burnap, and K. Jones, “Early-stage malware prediction using recurrent neural networks,” *Computers & Security*, vol. 77, pp. 578–594, 2018.
- [100] M. Kalash, M. Rochan, N. Mohammed, N. D. Bruce, Y. Wang, and F. Iqbal, “Malware classification with deep convolutional neural networks,” in *Proceedings of the 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pp. 1–5, Paris, France, February 2018.
- [101] X. Wang and S. Yiu, “A multi-task learning model for malware classification with useful file access pattern from API call sequence,” 2016, <https://arxiv.org/abs/1610.05945>.
- [102] R. Chalapathy and S. Chawla, “Deep learning for anomaly detection: a survey,” 2019, <https://arxiv.org/abs/1901.03407>.
- [103] L. Chen and Y. Ye, “SecMD: make machine learning more secure against adversarial malware attacks,” in *Proceedings of the Australasian Joint Conference on Artificial Intelligence*, pp. 76–89, Melbourne, Australia, August 2017.
- [104] S. Maniath, A. Ashok, P. Poornachandran, V. Sujadevi, A. P. Sankar, and S. Jan, “Deep learning LSTM based ransomware detection,” in *Proceedings of the 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE)*, pp. 442–446, Noida, India, October 2017.
- [105] W. Z. Zakaria, M. F. Abdollah, and A. F. Mohd Ariffin, “On Ransomware Detection,” in *Proceedings of the Seventh International Conference on Informatics and Applications (ICIA2018)*, pp. 12–17, Takamatsu, Japan, November 2018.
- [106] L. Demetrio, B. Biggio, G. Lagorio, F. Roli, and A. Armando, “Explaining vulnerabilities of deep learning to adversarial malware binaries,” 2019, <https://arxiv.org/abs/1901.03583>.
- [107] C. D. James and J. B. Aimone, *A Signal Processing Approach for Cyber Data Classification with Deep Neural Networks*, Sandia National Lab.(SNL-NM), Albuquerque, NM, USA, 2015.
- [108] Z. Wang, *The Applications of Deep Learning on Traffic Identification*, Vol. 24, TechRepublic, Louisville, KY, USA, 2015.
- [109] Z. M. Fadlullah, F. Tang, B. Mao et al., “State-of-the-Art deep learning: evolving machine intelligence toward tomorrow’s intelligent network traffic control systems,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2432–2455, 2017.
- [110] M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, “Deep abstraction and weighted feature selection for Wi-Fi impersonation detection,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 621–636, 2017.
- [111] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, “Mobile encrypted traffic classification using deep learning,” in *Proceedings of the 2018 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 1–8, Vienna, Austria, June 2018.
- [112] G. Mi, Y. Gao, and Y. Tan, “Apply stacked auto-encoder to spam detection,” in *Proceedings of the International Conference in Swarm Intelligence*, pp. 3–15, Beijing, China, June 2015.
- [113] C. Shi, J. Liu, H. Liu, and Y. Chen, “Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT,” in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, p. 5, Chennai, India, July 2017.
- [114] F. O. Catak and A. F. Yazı, “A benchmark API call dataset for windows PE malware classification,” 2019, <https://arxiv.org/abs/1905.01999>.
- [115] D. Gibert, *Convolutional Neural Networks for Malware Classification*, University Rovira i Virgili, Tarragona, Spain, 2016.
- [116] Y.-J. Cha, W. Choi, and O. Büyükoztürk, “Deep learning-based crack damage detection using convolutional neural networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.
- [117] M. Murata and K. Yamanishi, *Detecting Drive-By Download Attacks from Proxy Log Information Using Convolutional Neural Network*, Osaka University, Osaka, Japan, 2017.
- [118] R. Vinayakumar, K. P. Soman, and P. Poornachandran, “Applying convolutional neural network for network intrusion detection,” in *Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1222–1228, Manipal, India, September 2017.
- [119] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, “Malware traffic classification using convolutional neural network for representation learning,” in *Proceedings of the 2017 International Conference on Information Networking (ICOIN)*, pp. 712–717, Da Nang, Vietnam, January 2017.
- [120] C. Yin, Y. Zhu, J. Fei, and X. He, “A deep learning approach for intrusion detection using recurrent neural networks,” *IEEE Access*, vol. 5, pp. 21954–21961, 2017.
- [121] Y. Maleh, “Malware classification and analysis using convolutional and recurrent neural network,” in *Handbook of Research on Deep Learning Innovations and Trends*, pp. 233–255, IGI Global, Harrisburg, PA, USA, 2019.
- [122] B. Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, “Deep learning for classification of malware system call sequences,” in *Proceedings of the AI 2016: Advances in Artificial Intelligence*, pp. 137–149, Cham, Switzerland, December 2016.
- [123] S. Tobiyama, Y. Yamaguchi, H. Shimada, T. Ikuse, and T. Yagi, “Malware detection with deep neural network using process behavior,” in *Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 577–582, Atlanta, GA, USA, June 2016.
- [124] Y. Yu, J. Long, and Z. Cai, “Network intrusion detection through stacking dilated convolutional autoencoders,” *Security and Communication Networks*, vol. 2017, Article ID 4184196, 10 pages, 2017.
- [125] G. D. Hill and X. J. A. Bellekens, “Deep learning based cryptographic primitive classification,” 2017, <https://arxiv.org/abs/1709.08385>.
- [126] M.-J. Kang and J.-W. Kang, “Intrusion detection system using deep neural network for in-vehicle network security,” *PLoS One*, vol. 11, no. 6, Article ID e0155781, 2016.
- [127] S. Potluri and C. Diedrich, “Accelerated deep neural networks for enhanced Intrusion Detection System,” in *Proceedings of the 2016 IEEE 21st International Conference on*

- Emerging Technologies and Factory Automation (ETF A)*, pp. 1–8, Berlin, Germany, September 2016.
- [128] G. E. Dahl, J. W. Stokes, L. Deng, and D. Yu, “Large-scale malware classification using random projections and neural networks,” in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3422–3426, Vancouver, Canada, May 2013.
- [129] M. Sebastián, R. Rivera, P. Kotzias, and J. Caballero, “AVclass: a tool for massive malware labeling,” in *Research in Attacks, Intrusions, and Defenses*, pp. 230–253, Springer, Cham, Switzerland, 2016.
- [130] G. Mi, Y. Gao, and Y. Tan, “Apply stacked auto-encoder to spam detection,” in *Advances in Swarm and Computational Intelligence*, pp. 3–15, Springer, Cham, Switzerland, 2015.
- [131] G. Mi, Y. Gao, and Y. Tan, “Term space partition based ensemble feature construction for spam detection,” in *Data Mining and Big Data*, pp. 205–216, Springer, Cham, Switzerland, 2016.
- [132] H. S. Anderson, J. Woodbridge, and B. Filar, “DeepDGA: adversarially-tuned domain generation and detection,” in *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security—ALSec’16*, pp. 13–21, Vienna, Austria, 2016.
- [133] B. Yu, J. Pan, J. Hu, A. Nascimento, and M. De Cock, “Character level based detection of DGA domain names,” in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Rio de Janeiro, Brazil, 2018.
- [134] Y. Zhauniarovich, I. Khalil, T. Yu, and M. Dacier, “A survey on malicious domains detection through DNS data analysis,” *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–36, 2018.
- [135] K. Alrawashdeh and C. Purdy, “Toward an online anomaly intrusion detection system based on deep learning,” in *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 195–200, Anaheim, CA, USA, December 2016.
- [136] Z. Yuan, Y. Lu, Z. Wang, and Y. Xue, “Droid-sec: deep learning in android malware detection,” in *Proceedings of the 2014 ACM conference on SIGCOMM—SIGCOMM’14*, pp. 371–372, Chicago, IL, USA, 2014.
- [137] S. Wang, T. Liu, and L. Tan, “Automatically learning semantic features for defect prediction,” in *Proceedings of the 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)—ICSE’16*, pp. 297–308, Austin TX, USA, May 2016.
- [138] S. Chen, M. Xue, Z. Tang, L. Xu, and H. Zhu, “StormDroid: a streaming-based machine learning-based system for detecting android malware,” in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security—ASIA CCS’16*, pp. 377–388, Xi’an, China, 2016.