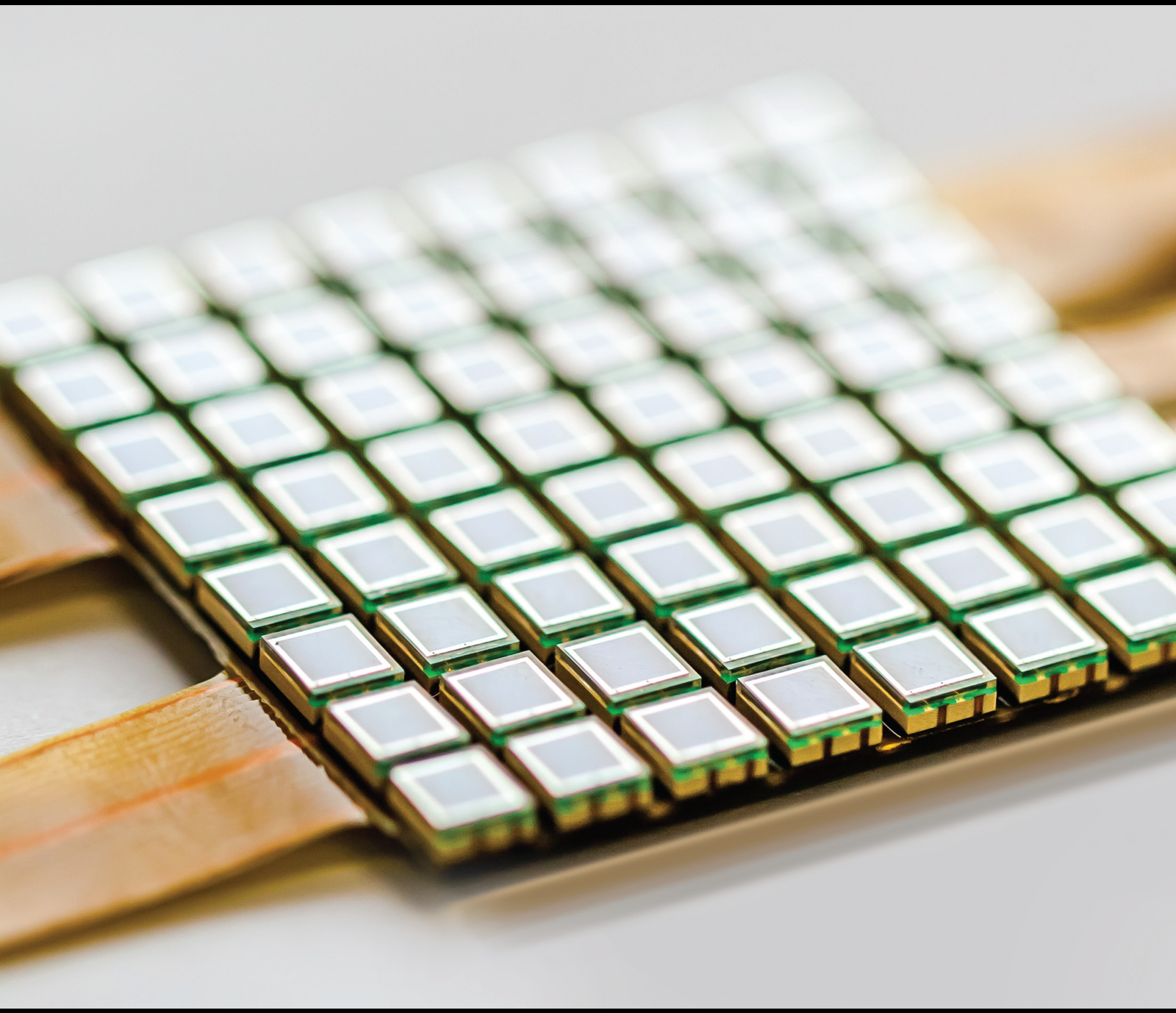
Journal of Sensors

Data Acquisition and Analysis for Smart Health Sensor Environments

Lead Guest Editor: Rocio Perez de Prado
Guest Editors: MOHIT MITTAL and Jaroslav Frnda



Data Acquisition and Analysis for Smart Health Sensor Environments

Journal of Sensors

Data Acquisition and Analysis for Smart Health Sensor Environments

Lead Guest Editor: Rocio Perez de Prado
Guest Editors: MOHIT MITTAL and Jaroslav Frnda

Contents

Hindawi
Journal of Sensors
Volume 2023, Article ID 9281230, 15 pages
https://doi.org/10.1155/2023/9281230

Research Article

Multiview Feature Fusion Attention Convolutional Recurrent Neural Networks for EEG-Based Emotion Recognition

Ruihao Xin,[1,2] Fengbo Miao,[1] Ping Cong,[1] Fan Zhang,[1] Yongxian Xin,[3] and Xin Feng[4,5]

[1]*College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, Jilin, China 130000*
[2]*College of Computer Science and Technology, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin, China 130012*
[3]*College of Business and Economics, Australian National University, ACT, Canberra, Australia 2601*
[4]*School of Science, Jilin Institute of Chemical Technology, Jilin, Jilin, China 130000*
[5]*State Key Laboratory of Inorganic Synthesis and Preparative Chemistry, College of Chemistry, Jilin University, Changchun, Jilin, China 130012*

Correspondence should be addressed to Xin Feng; fengxin@jlict.edu.cn

Received 4 August 2022; Revised 9 November 2022; Accepted 7 April 2023; Published 29 April 2023

Academic Editor: Mohit Mittal

Emotion recognition is essential for computers to understand human emotions. Traditional EEG emotion recognition methods have significant limitations. To improve the accuracy of EEG emotion recognition, we propose a multiview feature fusion attention convolutional recurrent neural network (multi-aCRNN) model. Multi-aCRNN combines CNN, GRU, and attention mechanisms to fuse features from multiple perspectives deeply. Specifically, multiscale CNN can unite elements in the frequency and spatial domains through the convolution of different scales. The role of the attention mechanism is to weigh the frequency domain and spatial domain information of different periods to find more valuable temporal perspectives. Finally, the implicit feature representation is learned from the time domain through the bidirectional GRU to achieve the profound fusion of features from multiple perspectives in the time domain, frequency domain, and spatial domain. At the same time, for the noise problem, we use label smoothing to reduce the influence of label noise to achieve a better emotion recognition classification effect. Finally, the model is validated on the EEG data of 32 subjects on a public dataset (DEAP) by fivefold cross-validation. Multi-aCRNN achieves an average classification accuracy of 96.43% and 96.30% in arousal and valence classification tasks, respectively. In conclusion, multi-aCRNN can better integrate EEG features from different angles and provide better classification results for emotion recognition.

1. Introduction

Emotion plays a significant role in our life, affecting human cognition and decision-making [1]. At the same time, it is also a relatively complex psychological state [2]. How to recognize emotions has become one of the issues in the industry [3]. At present, the mainstream methods of emotion recognition include two-dimensional valence and arousal coordinate system [4] and discrete assessment method [5]. In the two-dimensional valence and arousal coordinate system method, valence represents the positive or negative direction of the emotion, and arousal represents the intensity of emotion [6]. In discrete assessment, emotions are divided into multiple discrete categories. For example, Zheng and Lu classify emotions into positive, neutral, and negative categories [7], and Shanmugam and Padmanaban classify emotions into eight types: joy, trust, fear, surprise, sadness, disgust, anger, and expectation [8]. Emotion recognition is of great significance. It can help humans understand their own emotions, and it can also help computers better understand human emotions [9] so that computers can better serve humans.

With the development of computer science and information technology, human-machine interaction technology has attracted more attention [10]. As the cornerstone of human-machine interaction technology, emotion recognition has

inevitably attracted the attention of the academic community [11]. Generally speaking, emotion recognition methods can be divided into two categories: one is based on external signals of the human body [12], such as expression, posture, and voice; the other is based on the internal motions of the body [13], such as EEG, ECG, and EMG. Compared with external signals such as facial expression, posture, and voice, emotion recognition results based on internal cues such as EEG are more reliable because humans cannot control them intentionally [14].

In the traditional EEG emotion recognition method, first, screen out the hand-made features that are more relevant to the emotion recognition task [15], and then, input these emotional features into the machine learning model for classification. However, because deep learning does not require manual feature making and has better learning effect [16], researchers of EEG emotion recognition mostly use deep learning methods for research in recent years [17, 18]. Based on the characteristics of EEG signals, it can be extracted from the time domain, frequency domain, time-frequency domain, and nonlinear dynamical system [19]. Differential entropy (DE) is a representative nonlinear dynamic feature commonly used in EEG emotion recognition tasks [20]. The research of Garcia-Martinez et al. confirmed the effectiveness and robustness of the DE feature in EEG emotion recognition tasks [21]. Zhu and Zhong [22] classified DE features by using the 2DCNN-BiGRU network and achieved 87.89% and 88.69%, respectively, in the arousal and valence classification results of the DEAP dataset. However, a single convolution scale makes this method limited in spatial feature extraction, resulting in feature loss. In Yin et al.'s study [23], by using the ERDL model and combining the characteristics of the frequency domain and time domain, the classification accuracy on the DEAP dataset reached 90.45% and 90.6%, respectively. Shen et al. [24] proposed a four-dimensional convolutional recurrent neural network (4D-CRNN) to integrate the frequency domain, spatial domain, and time domain information of multichannel EEG signals to improve the accuracy of emotion recognition based on EEG. The accuracy of arousal and valence classification in the DEAP dataset reached 94.22% and 94.58%, respectively. However, these two methods ignore the differences between features and the impact of different features on classification results. In the process of feature fusion, it is easy to cause feature redundancy by not distinguishing different features. Cui et al. [25] proposed a DE-CNN-bi-LSTM network to remove DE features on different time slices in different frequency bands. After that, CNN and bi-LSTM were used to learn spatial and temporal information, so the classification accuracy on the DEAP dataset reached 94.86% and 94.02%, respectively. However, this method could not effectively deal with tag noise, which affected the classification results.

Aiming at the limitations of these methods, we propose a recursive network model based on multiview feature fusion. According to different spatial features extracted from convolutions of different scales, different spatial features of different periods are weighted and fused through the multihead attention mechanism to magnify the actual features and reduce the impact of invalid features, and label smoothing is used to reduce the impact of label noise. In conclusion, the main contributions of this study are as follows:

(1) To solve the noise in prediction labels of a single subject in EEG sentiment analysis, the method of label smoothing has dramatically reduced the influence of label noise on the classification accuracy of models to achieve a better effect of sentiment classification

(2) Multiscale convolution makes the extracted spatial features more comprehensive, and convolution is closely combined with bidirectional GRU (bi-GRU) so that the model can learn more comprehensive time-frequency features

(3) This paper proposes a multiview feature fusion attention convolutional recurrent neural network, which integrates the weight of frequency, space, and time-domain features. It effectively improves the classification accuracy of emotion recognition

In this paper, Section 1 is the introduction part, which introduces some basic concepts in the field of EEG emotion recognition, and briefly summarizes the previous research work. This paper made some breakthroughs based on prior studies. Section 2 is the method part, which mainly introduces key concepts in the data set and model. Section 3 is the experimental results and analysis, and the experimental process and results are introduced and analyzed in detail. Section 4 discusses, compares, and analyzes the existing paper results, reflecting the research significance and value of this paper. Section 5 is the conclusion, reviewing and summarizing this paper.

Table 1: DEAP dataset.

Array name	Array shape	Array contents
Data	40*40*8064(63*128)	Video/trial*channel*data
Labels	40*4	Video/trial*label

2. Methods

2.1. Dataset and Preprocessing. DEAP dataset [26] is a multichannel dataset collected by Koelstra et al., as shown in Table 1, who invited 32 subjects (including 16 males and 16 females) to watch 40 music videos used to study human emotion. The subjects invited for the experiment are in good physical condition and mental health and can generally respond to the stimulation of the video material. Each music video lasts for 1 minute, regarded as an experiment. For each experiment, the first 3 seconds (3 s) is the video conversion time, and the last 60 s is the music video play time, so the duration of each sample is 63 s, and the video conversion time of the first 3 s is the baseline time of the experiment. After playing the video, each subject had to score the video in valence, arousal, and other dimensions, ranging from 1 to 9. We selected five as the threshold and regarded the

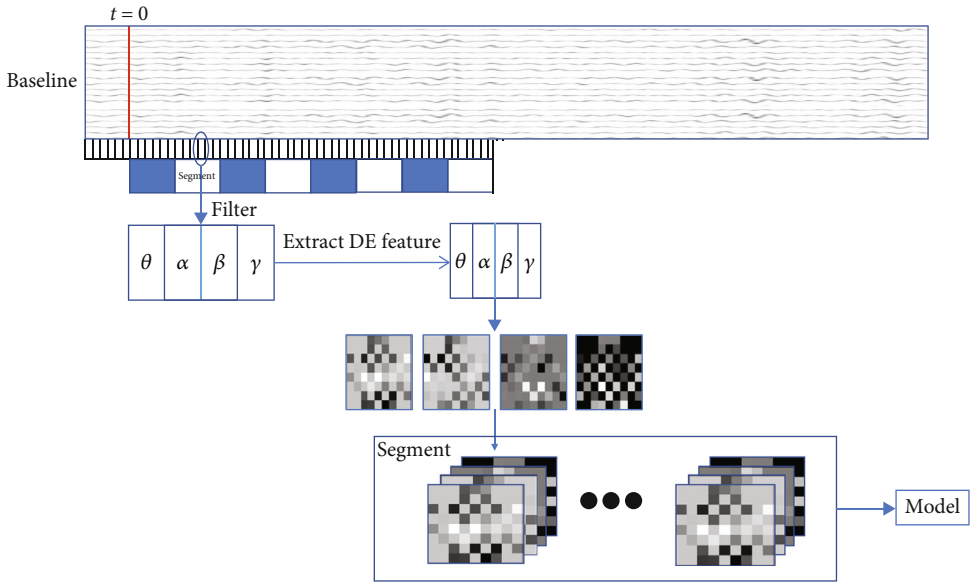


Figure 1: Data preprocessing flow chart.

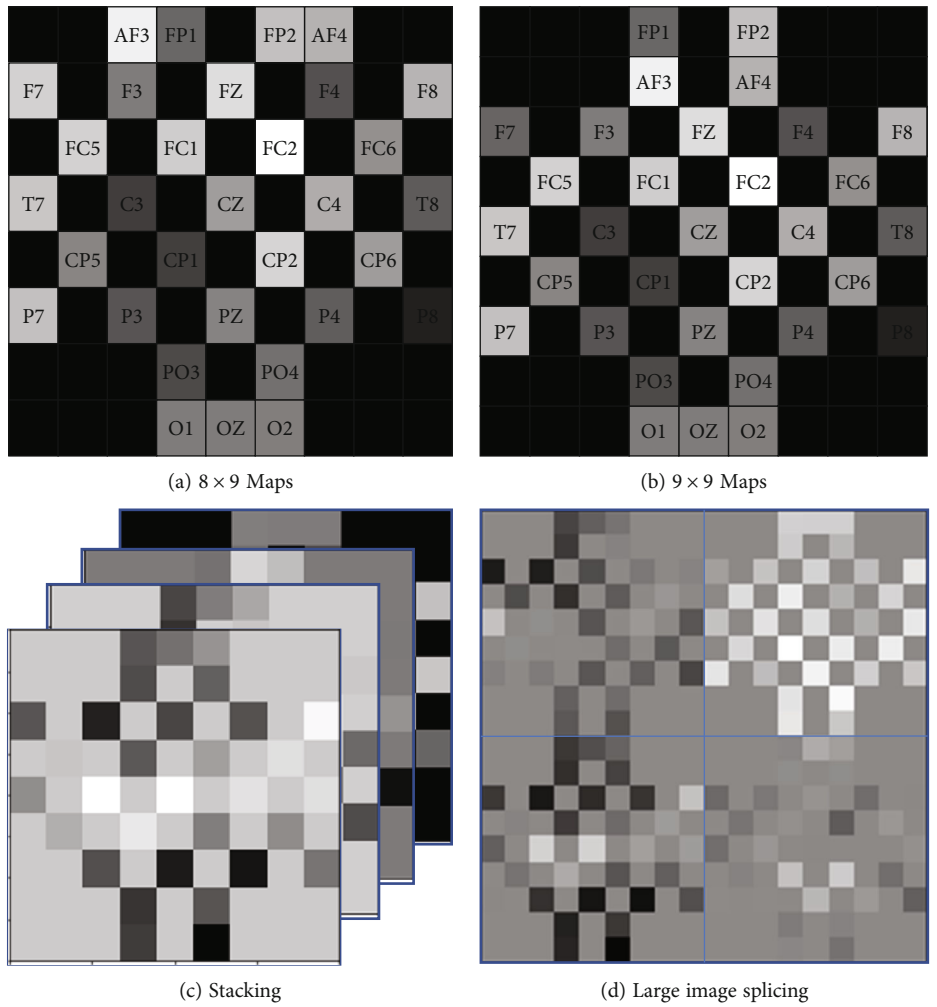
emotion recognition task of the DEAP dataset as two binary classification problems. EEG signals were sampled using a 32-channel electrode. The sampling frequency is 512 Hz, and electrodes were placed by the international standard lead 10-20. In addition, the experiment collected not only 32-channel EEG signals but also 8-channel ECG and EOG, a total of 40 physiological signal channels. Only the first 32 EEG signals were used in this paper. The website provides the preprocessed data by downsampling the frequency from 512 Hz to 128 Hz and removing noise such as ocular artifacts.

The 63 s EEG signal was collected for each trial. First, we cut the 63 s signal to 0.5 s. Then, each segmented EEG signal block was filtered to obtain four frequency bands of θ, α, β, and γ, and then, DE features were extracted from each frequency band. Since Yang et al. [27, 28] have proved that considering baseline signals can improve the classification effect of the model, we will carry out baseline correction according to the method in the paper. The DE features of each frequency band of the baseline signal were averaged. Then, the DE features of the stimulus signal were differentiated from the average of the corresponding frequency band to obtain the baseline-corrected DE features. The processed DE features were mapped into a two-dimensional map for each frequency band according to the electrode distribution. For each block, the 2D maps of the four frequency bands are concatenated to form a new feature matrix. Finally, the feature matrix of the block is combined according to the segment (1segment = 6 blocks) and sent into the model as a sample. The data preprocessing process is shown in Figure 1.

Electrodes can be converted into two kinds of 2D maps: one type is 8×9, as shown in Figure 2(a), and the other is a 9×9, as shown in Figure 2(b). Likewise, 2D maps of the four frequency bands can be jointed in two ways. One is stacked splicing, that is, to form a three-dimensional matrix. In this paper, 8×9 maps are spliced stacked, as shown in Figure 2(c). Or it can be assembled into a large picture. The picture jointed in this way is still a two-dimensional matrix. In this paper, 9×9 graphs are jointed in a large picture, as shown in Figure 2(d).

For a single subject, 60 seconds of stimulus signal data collected from 40 music videos watched is 40*60*2, divided into 800 samples for 4800 stimulus signals; each sample contains 6 time period information, and each period information contains 4 frequency bands. Taking the superposition result as an example, the characteristic of each frequency band is an 8*9 mapping matrix. There are a total of 32 subjects, and in the arousal classification, there are a total of 25600 samples, including 10860 low arousal samples and 14740 high arousal samples. Each sample contains 6 time period information. Each period information is a 4*8*9 feature matrix. In the valence classification, there are a total of 25600 samples, including 11,40 low-valence samples and 14160 high-valence samples, and each sample contains 6 time period information. Each period information is a 4*8*9 feature matrix. Because subjects respond differently to emotional stimuli, the number of positive and negative samples for (low/high) arousal and (low/high) valence will not be the same.

2.2. Spatial Feature Extraction Based on 2D-CNN. Convolutional neural networks are often used to process 2D data and usually consist of three parts: convolutional layers, pooling layers, and activation function layers [29]. The convolution layer performs the inner product operation on the input data through the convolution kernel. By setting the size and number of the convolution kernel, the model can extract different types of data features. At the same time, in the convolutional layer, the number of parameters that the model needs to be trained is reduced through "sparse connection" [30] and "weight sharing," thereby reducing the difficulty of training.



(a) 8 × 9 Maps

(b) 9 × 9 Maps

(c) Stacking

(d) Large image splicing

Figure 2: Electrodes conversion.

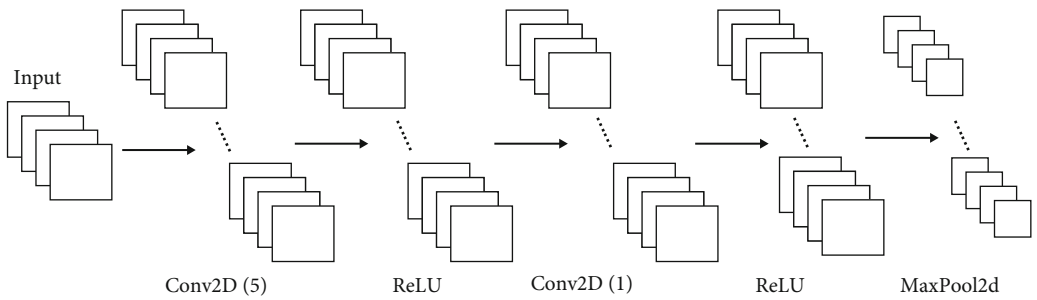


Figure 3: Part of the convolutional neural network model.

In addition, the pooling layer can further reduce the data provided by the model to the next layer of the network, reducing the difficulty of model training [31]. The activation function layer transforms the data to reduce the training difficulty and enhance the correlation between data. Part of the convolutional neural network model used in this paper is shown in Figure 3.

2.3. Time Series Feature Extraction Based on GRU. RNN (recurrent neural network) has certain advantages when dealing with time series data. When RNN processes the information at each moment, it can effectively preserve the original timing of the data, and the training parameters will not increase due to the increase in the sequence length. This paper uses an improved cyclic structure GRU (gated recurrent unit) model, as shown in Figure 4.

The GRU has a reset gate and an update gate. The reset gate determines the degree to which the input information will be combined with the previously memorized information. The update gate determines how much of

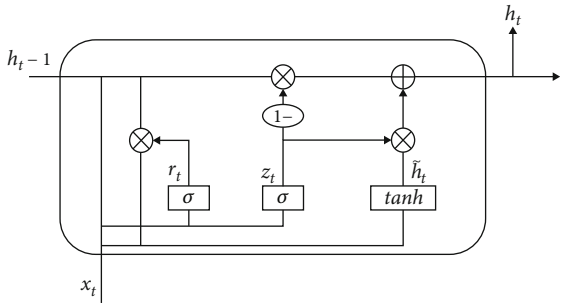


Figure 4: GRU (gated recurrent unit) model.

the previously memorized information can be retained to the current time step. The specific formula is as follows:

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_t - 1, x_t]), \\ r_t &= \sigma(W_r \cdot [h_t - 1, x_t]), \\ \tilde{h}_t &= \tan h(W \cdot [r_t * h_t - 1, x_t]), \\ h_t &= (1 - z_t) * h_t - 1 + z_t * \tilde{h}_t, \end{aligned} \tag{1}$$

where h_t is the hidden state at time t, x_t is the input at time t, r_t and z_t are the reset gate and update gate, respectively, $\tilde{h}_t$ is the candidate hidden state, σ is the sigmoid function, and $*$ is the Hadamard product.

In this paper, GRU is used to obtain the time series characteristics of data, and a comparative test is carried out for GRU and bi-GRU in Section 3.4.

2.4. Feature Fusion Based on Multihead Attention. Usually, scaled dot-product attention consists of three parts: Q (query), K (key), and V (value). The structure is shown in Figure 5(a). Assume that the dimensions of the input Q and K are d_k, and the dimension of V is d_v. Then, calculate the transposed multiplication of Q and K, divide by $\sqrt{d_k}$, pass the result through the Softmax function to get the weight, and multiply the weight by V to get the output matrix. The specific formula is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{\text{QK}^T}{\sqrt{d_k}}\right)V. \tag{2}$$

This paper uses multihead attention [32], and its structure is shown in Figure 5(b). Multihead attention can combine the information learned by different heads, which can be regarded as parallel processing of multiple scaled dot-product attention. Q, K, and V are first subjected to a linear transformation and then input to the scaled dot-product attention. Here is the scaled dot-product attention for head times and stacking the obtained results. The spliced result is then subjected to a linear transformation to obtain the value as the output of multihead attention.

2.5. AdamW and Label Smoothing

2.5.1. AdamW Optimization Algorithm. Adam optimization algorithm has been widely used in various deep learning models since its appearance, but experiments found that Adam has specific problems. Such as slow model convergence, nonconvergence, and other problems, various improved versions of Adam appeared. Different parameters in the Adam optimization algorithm adaptively learn at different learning rates. The formula is as follows:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) {g_t}^2, \end{aligned} \tag{3}$$

where g_t represents the gradient, the subscript t represents time, m_t is the first-order moment variable of the gradient, v_t is the second-order moment variable of the gradient, and β_1 and β_2 are the exponential decay rates (decay factors) of the moment estimation. When the values of m_t and v_t approach the 0 vectors, the result will be biased. This problem is solved by performing bias correction on m_t and v_t. The formulas for the bias correction value m_t and v_t of m_t' and v_t' are as follows:

$$\begin{aligned} m_t' &= \frac{m_t}{1 - \beta_1^t}, \\ v_t' &= \frac{v_t}{1 - \beta_2^t}. \end{aligned} \tag{4}$$

AdamW [33] adds a regular term to Adam's loss function and adds the result of the gradient of the regular term when calculating the gradient so that the gradient of the overall loss function is calculated when updating the model parameters, thereby updating the parameters. AdamW's loss function is

$$L = \text{loss} + \frac{1}{2}\|\theta\|^2. \tag{5}$$

Then, the formula for AdamW parameter update is

$$\theta_t = \theta_{t-1} - \eta\left(\frac{\alpha m_t'}{\sqrt{v_t'} + \xi} + \omega\theta_{t-1}\right), \tag{6}$$

where θ is a parameter in the model, η is the learning rate, α is 0.001, ξ is 10^{-8}, and ω is an actual number.

2.5.2. Label Smoothing. There are usually some noisy labels in machine learning samples, and these labels will have a certain impact on the prediction results. Label smoothing prevents the model from believing too much in the labels of the training samples by assuming that the labels may be wrong during training [34]. The formula looks like this:

$$y_i = \begin{cases} 1 - \varepsilon, & \text{if } i = \text{true}, \\ \dfrac{\varepsilon}{K - 1}, & \text{otherwise}. \end{cases} \tag{7}$$

Among them, ε is a defined hyperparameter, which generally takes a value of 0.1, K is the number of categories of

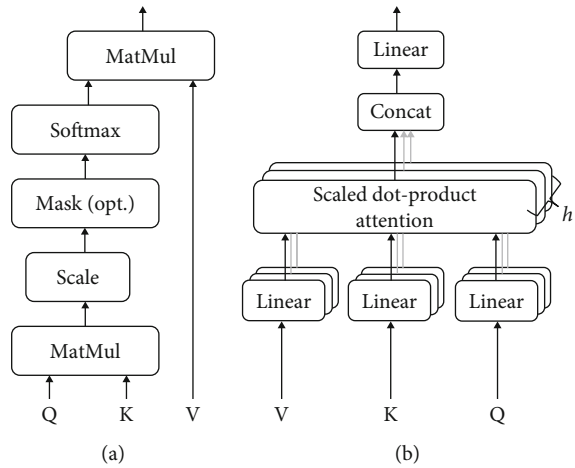


Figure 5: (a) Scaled dot-product attention. (b) Multihead attention.

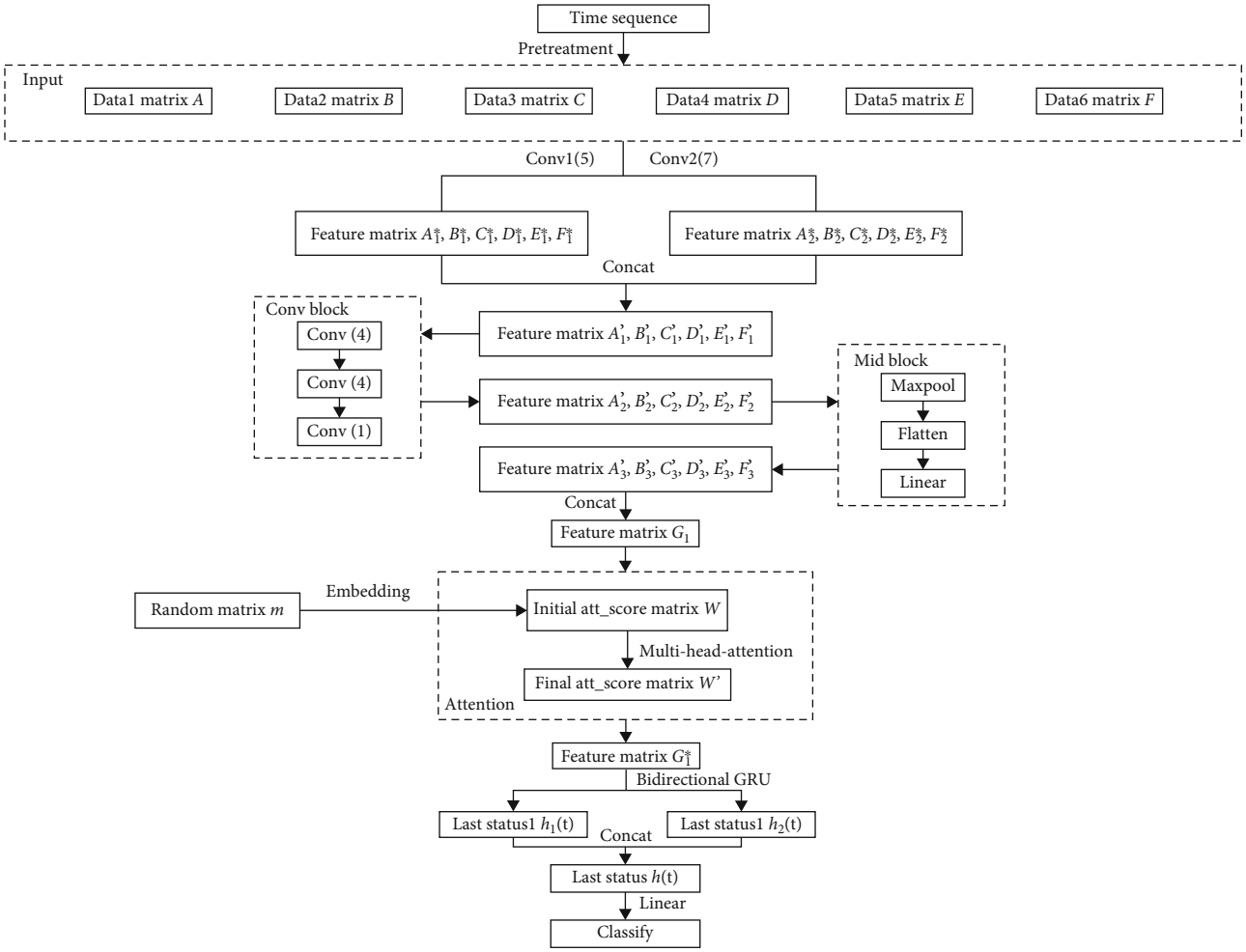


Figure 6: Model flow chart.

the classification problem, and y_i is the sample label. When training samples, it is usually not guaranteed that all sample labels do not contain noise interference. For example, if the label of a sample is wrong, then the sample may harm the training results during training. By letting the model know that the label of the sample is not necessarily correct, the trained model can better identify a small number of wrong samples.

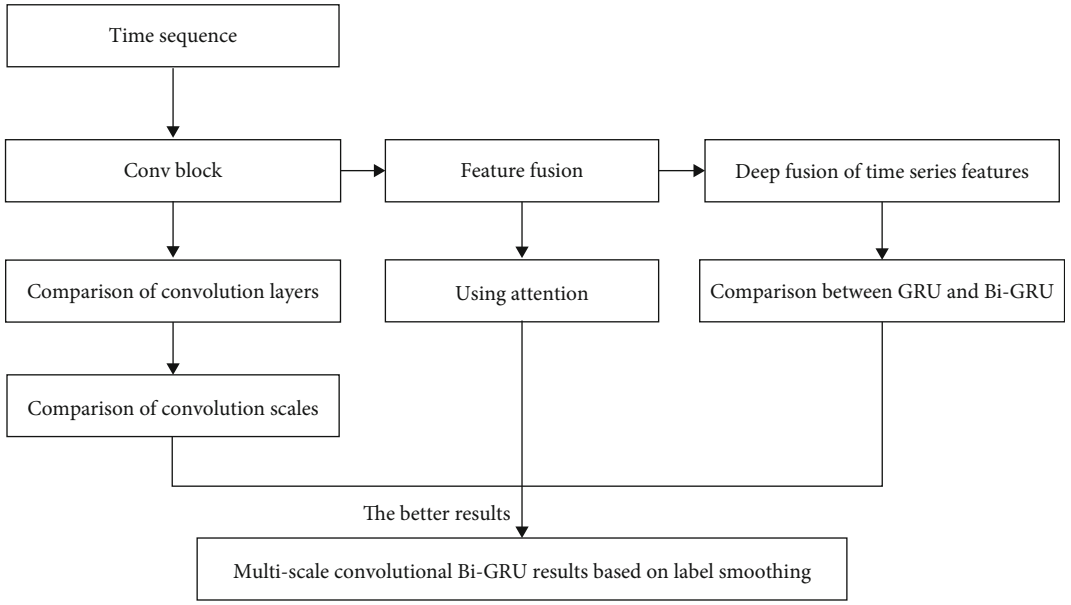


Figure 7: Overall experimental process.

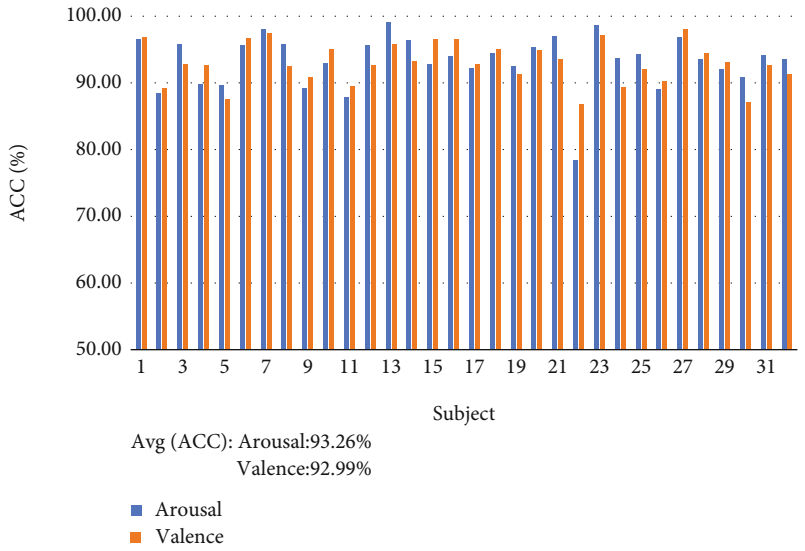


Figure 8: Two-layer convolution model results (the model is 5+1, there are two layers of the convolutional network, the size of one convolution kernel is 5, and the size of one convolution kernel is 1).

2.6. Multiview Feature Fusion Attention Convolutional Recurrent Neural Network Model. In this paper, after preprocessing the EEG time series data, the original sequence is divided into six data segments according to the period, thus effectively preserving the time series of the data. At the same time, to make the extracted features more comprehensive, multiscale convolution is used to extract spatial-domain features, and the extracted features are highly abstracted through convolution blocks. For the abstracted data, the extracted frequency-domain and spatial-domain features are weighted from the time series perspective through the attention mechanism, and the weighted data is classified through the bidirectional GRU model. The specific process is shown in Figure 6.

Perform spatial feature extraction to obtain feature matrix $A_1^*, B_1^*, C_1^*, D_1^*, E_1^*, F_1^*$ and feature matrix $A_2^*, B_2^*, C_2^*, D_2^*, E_2^*, F_2^*$. Concat the two spatial features to obtain matrix $A_1', B_1', C_1', D_1', E_1', F_1'$, and transfer the matrix to the Conv block. Abstract the spatial features through three-layer convolution in the Conv block to obtain feature matrix $A_2', B_2', C_2', D_2', E_2', F_2'$. The abstract matrix is subjected to maximum pooling, flatten and linear network layers to obtain feature matrix $A_3', B_3', C_3', D_3', E_3', F_3'$, and six matrices are Concat to obtain matrix G_1. At the same time, the random initialization matrix m is used as the initial weight matrix W of attention after passing through the embedding layer. After G_1 is input, the final weight matrix W' is obtained through multihead attention

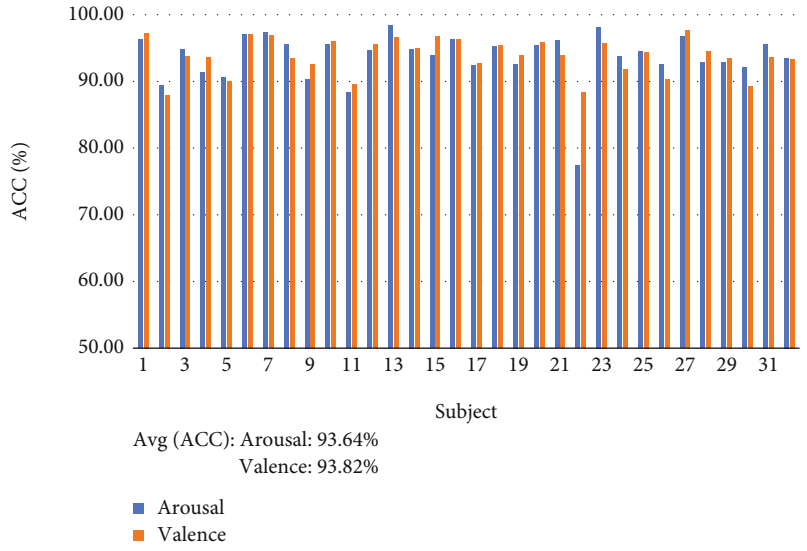


Figure 9: Deep convolution model results (the model is 5+4+4+1, and there are four layers of the convolutional network).

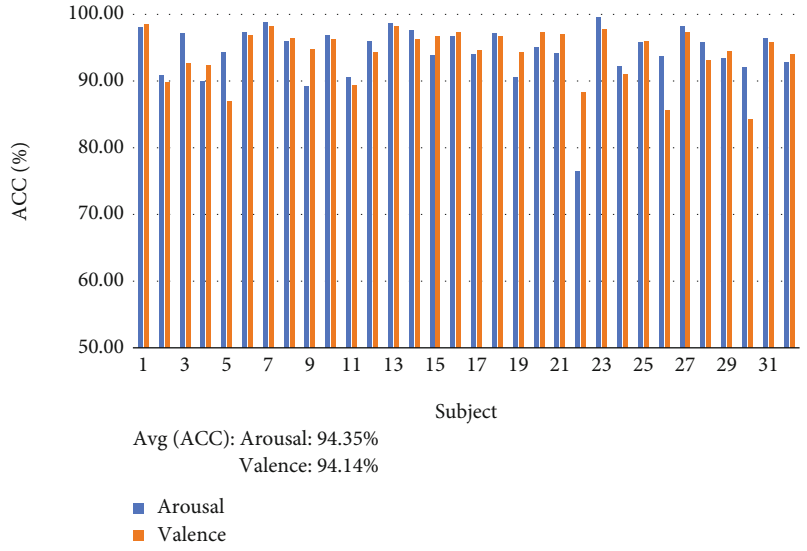


Figure 10: Results of deep convolution classification based on attention (the model is 5+4+4+1+Att, and attention is used in the model).

training, and the feature matrix G_1^* is obtained after weighting. The weighted matrix is passed into the bidirectional GRU model to extract the time series features of the data. Concat the states $h_1(t)$ and $h_2(t)$ at the last moment of forward and reverse to obtain the final output state matrix $h(t)$, and pass $h(t)$ through the linear network layer to achieve classification.

3. Experiment and Analysis

The batch size for training multi-aCRNN is 128, the dropout is 0.5, the maximum number of epochs is 500, the learning rate is $5 * 10^{-5}$, and the number of heads is 8. PyTorch implements the model, NVIDIA-SMI 460.67, CUDA Version: 11.2, python version 3.7.0, PyTorch version 1.11.0.

Table 2: Comparison of model classification results before and after adding attention.

Model	ACC (a)	Num (a)	ACC (v)	Num (v)
5+4+4+1	93.64%	3	93.82%	4
5+4+4+1+Att	94.35%	2	94.14%	6

Five cross-validations were used for each experiment in this paper. They were performed for the average classification accuracy (ACC) and the number of subjects (Num) with average classification accuracy below 90% for arousal (a) and valence (v) analysis. The overall experimental process is shown in Figure 7. The idea is to compare the number and scale of convolution layers, the use of attention, the

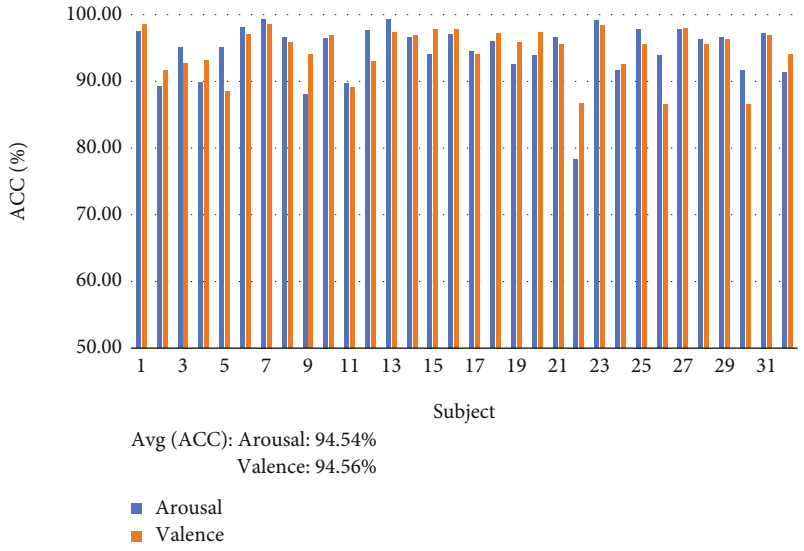


Figure 11: Attention-based deep convolution (the model is 7+4+4+1+Att).

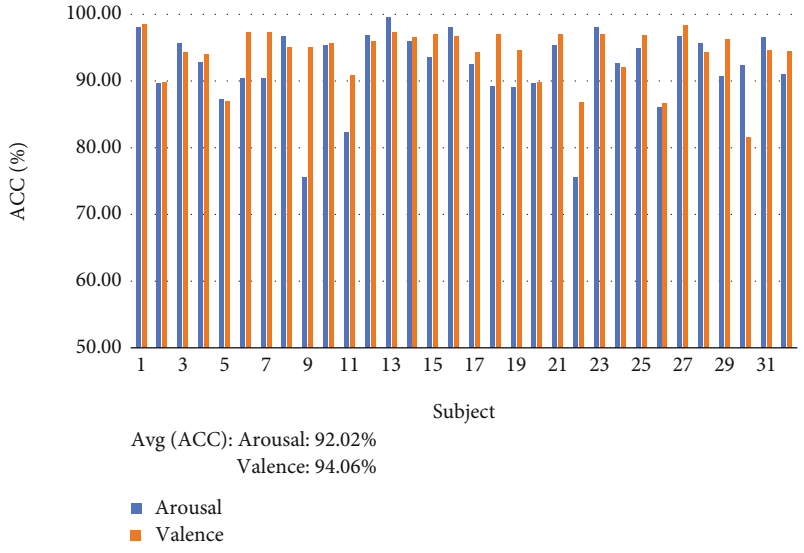


Figure 12: Attention-based deep convolution (the model is 7+5+5+1+Att).

comparison between GRU and bi-GRU, and the use of label smoothing.

3.1. Comparison of the Number of Convolutional Layers. Usually, in a convolutional neural network, the number of convolutional layers determines the degree of abstraction of features. Thereby, a more accurate prediction result can be obtained. Therefore, this section compares the two-layer and four-layer convolution results. The results are shown in Figures 8 and 9:

It can be seen from Figures 8 and 9 that the deep convolution is better than the two-layer convolution to a certain extent for the experimental results, especially in the valence result—the average accuracy of using deep convolution increases by 0.83%. At the same time, judging from the classification results of a single subject, when using two-layer convolution for classification, the arousal classification results of 7 subjects were lower than 90%. The valence classification results of 6 subjects were lower than 90%. In contrast, in the four-layer convolution, there are three subjects whose arousal classification results are lower than 90% and

Table 3: Comparison of classification results of different scale convolution kernel models.

Model	ACC (a)	Num (a)	ACC (v)	Num (v)
5+4+4+1+Att	94.35%	2	94.14%	6
7+4+4+1+Att	94.54%	5	94.56%	5
7+5+5+1+Att	92.02%	9	94.06%	6

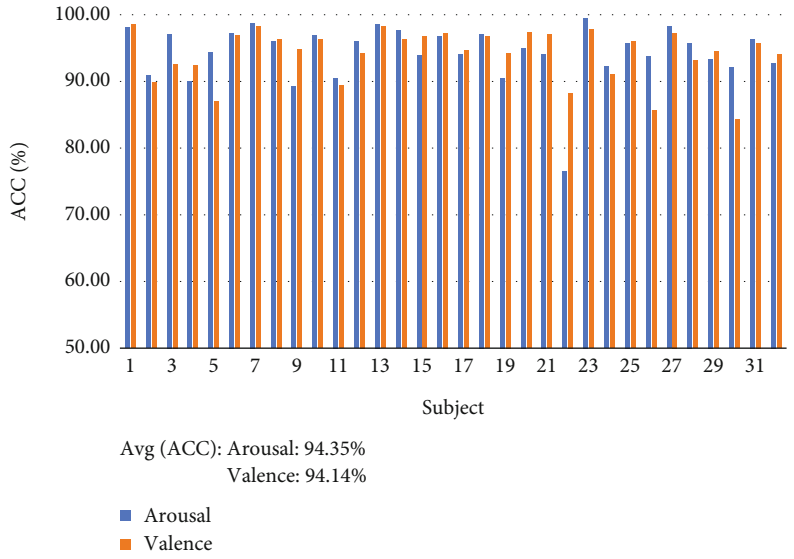


Figure 13: GRU model results (the model is 7+4+4+1+Att+GRU).

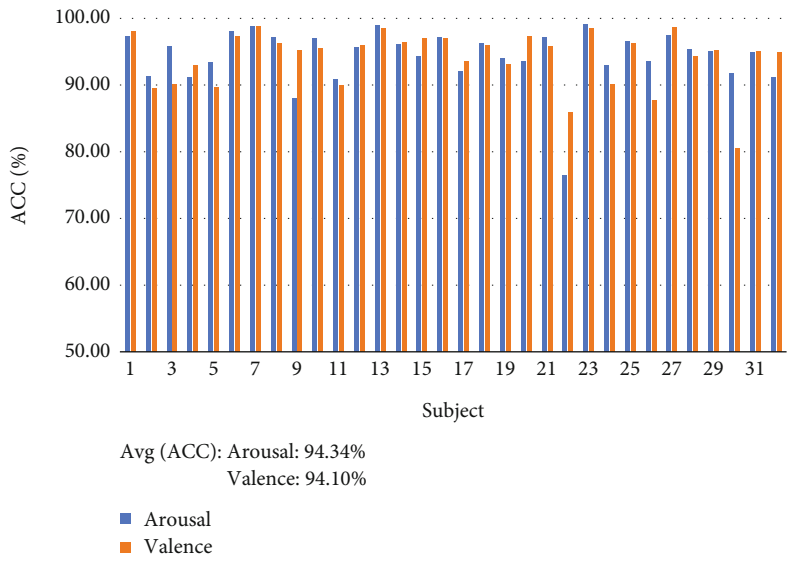


Figure 14: Bi-GRU model results (the model is 7+4+4+1+Att+bi-GRU).

four subjects whose valence classification results are lower than 90%, greatly reducing the degree of classification error.

3.2. Attention-Based Deep Convolution Classification Results. For the features extracted by the convolution layer for different periods, the traditional Concat method cannot sufficiently distinguish the effectiveness of these features. Aiming at this problem, we use the attention mechanism for feature fusion so that the frequency-domain and spatial-domain features of different periods can be distinguished by increasing the weight to achieve a better classification prediction effect. The results are shown in Figure 10.

Comparing Figures 10 and 9, after adding attention, the overall classification accuracy has been improved to a certain extent. It can be seen from Table 2 that the average accuracy of arousal classification increased by 0.71%, and the average accuracy of valence classification increased by 0.32%. For a single subject, using weighted feature fusion resulted in greater progress in the more difficult second subject to classify. At the same time, only two subjects had an accuracy rate below 90% for the arousal classification accuracy rate. Still, six subjects had an accuracy rate below 90% in the valence classification results. Since the weighted fusion is

Table 4: Comparison of classification results between GRU and bi-GRU models.

Model	ACC (a)	Num (a)	ACC (v)	Num (v)
GRU	94.35%	2	94.14%	6
Bi-GRU	94.34%	2	94.10%	5

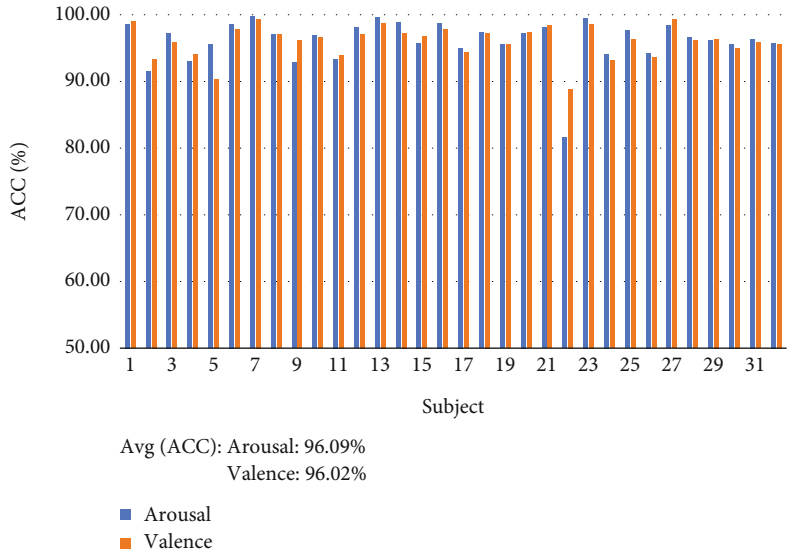


Figure 15: Single-scale convolutional bi-GRU results based on label smoothing (the model is 7+4+4+1+bi-GRU+lab).

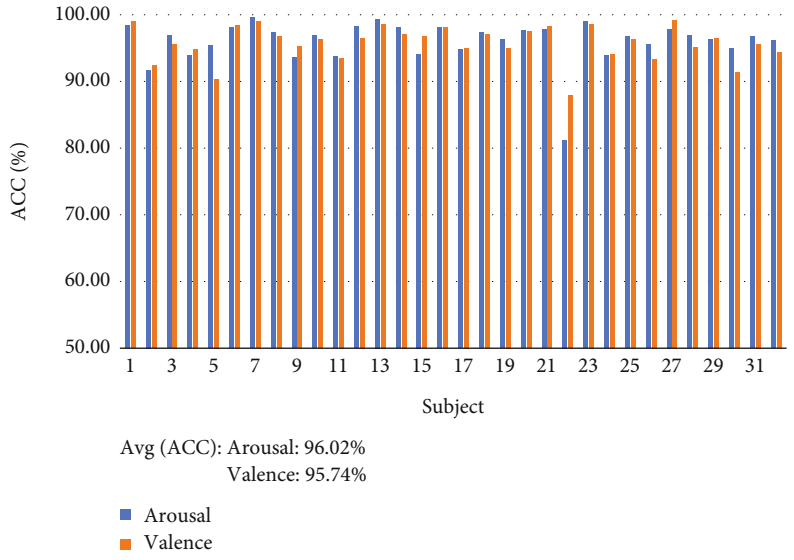


Figure 16: Multiscale convolutional GRU results based on label smoothing (the model is 5&7+4+4+1+GRU+lab).

carried out on the results of the convolution of six time periods, it has much to do with the convolution process. It is guessed that the locality of the initial convolution causes the attention of the model to be limited during learning, so the overall accuracy of valence classification has improved, and some single-subject results have declined.

3.3. A Comparative Study of Convolution at Different Scales. Since the convolutional layer is limited by the size of the convolution kernel when performing feature extraction, the size of the convolution kernel of the model is experimentally explored in this section. It can be seen from Figures 11 and 12 that different convolution kernel sizes have a significant impact on classification accuracy and classification stability. It can be seen from Table 3 that the first layer of convolution is the feature extraction of the original data, so by comparing the convolution kernels with the convolution kernel size of 5 and 7, it can be seen that the convolution kernel of 7 scales is used in the classification results better than five-scale convolution, and the accuracy of valence classification is improved by 0.42%. This is because large-scale convolution has a more extensive perception range when extracting features from the original data, which can significantly reduce the limitations of spatial feature extraction. However, this is only limited to the extraction of initial features. The effect of using a larger-scale convolution in the middle layer convolution is significantly reduced. This is because the middle layer convolution is a reabstraction of features. A larger-scale

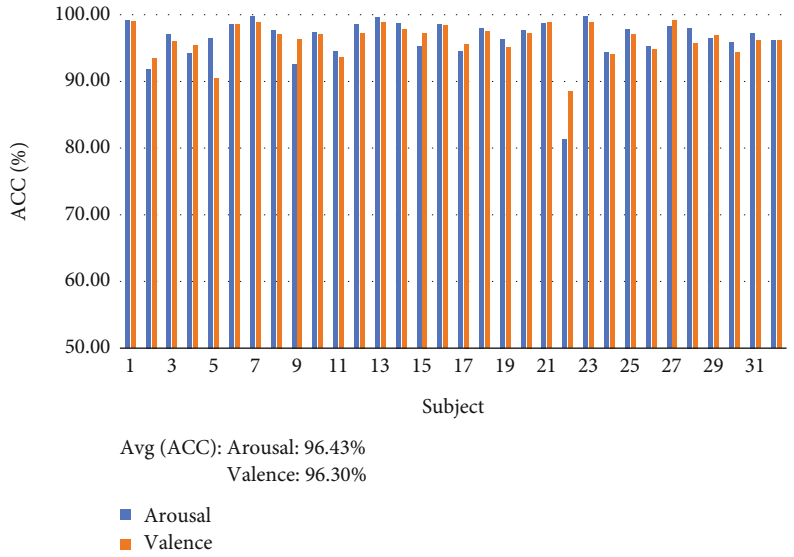


Figure 17: Multiscale convolutional bi-GRU results based on label smoothing (the model is 5&7+4+4+1+bi-GRU+lab).

Table 5: Comparison of multiscale and GRU and bi-GRU experimental results based on label smoothing.

Model	ACC (a)	Num (a)	ACC (v)	Num (v)
7+4+4+1+bi-GRU+lab	96.09%	1	96.02%	1
5&7+4+4+1+GRU+lab	96.02%	1	95.74%	1
5&7+4+4+1+bi-GRU+lab	96.43%	1	96.30%	1

convolution will affect the initial feature information and cover the part of the effective information in the initial features, resulting in a decrease in the classification results.

3.4. GRU and Bi-GRU Comparative Experiment. For time series, the time series features are extracted through a recurrent neural network to optimize the model. In this section, a comparative experiment is carried out on the GRU and bi-GRU networks, and the experimental results are shown in Figures 13 and 14. It can be seen from Table 4 that the classification accuracy results of GRU and bi-GRU are almost the same, and the classification results for a single subject are also relatively close. For this phenomenon, the use of GRU and bi-GRU will be further explored in Section 3.5.

3.5. Multiscale Fusion Model Based on Label Smoothing. In Section 3.4, GRU and bi-GRU are explored, but the experimental results are relatively close and cannot clearly show the pros and cons of the model. At the same time, we explore convolutional networks of different scales in Section 3.3 and find that larger initial convolution kernels are more effective for extracting spatial features but whether retaining the features of smaller-scale convolutions at the same time will promote emotion recognition to a certain extent. The multiscale fusion model based on label smoothing will be explored in this section. The results are shown in Figures 15–17 and Table 5.

After label smoothing, the model's accuracy has been significantly improved, and the results for a single subject are also more stable. In the single-scale Bi-GRU fusion model (7+4+4+1+Bi-GRU+lab), the accuracy of arousal and valence reached 96.09% and 96.02%, respectively, and there is only one subject with a classification accuracy below 90%. Since some subjects have certain errors in the experiment, it is difficult to improve the classification accuracy of some subjects. These noises will also affect the classification results of other subjects and even lead to the model's accuracy. The training results are getting worse and worse. It can be seen that AdamW and label smoothing can significantly promote the fitting of the model and the calibration of the network, which can dramatically reduce the impact of label noise.

Table 6: Comparison of experimental results between splicing and stacking preprocessing methods.

Methods	ACC (a)	Num (a)	ACC (v)	Num (v)
Splicing	94.96%	2	94.60%	4
Stacking	96.43%	1	96.30%	1

It can be seen from Table 5 that the training results of bi-GRU are significantly better than GRU. In the multiscale network, the results of bi-GRU (5&7+4+4+1+bi-GRU+lab) reach 96.43% and 96.30%, respectively, compared with 96.02% and 95.74% of the GRU network (5&7+4+4+1 +GRU+lab), the results of using bi-GRU are improved by 0.42% and 0.56%, respectively, and the overall optimization was achieved. It shows that in the EEG time series data,

Table 7: Comparison of results with other papers.

Nos.	Model	Information	ACC (a)	ACC (v)
1	2DCNN-BiGRU (2021) [22]	Spatial+temporal	87.89%	88.69%
2	ERDL (2020) [23]	Frequency+temporal	90.45%	90.6%
3	4D-CRNN (2020) [24]	Frequency+spatial+temporal	94.22%	94.58%
4	DE-CNN-BiLSTM (2022) [25]	Frequency+spatial+temporal	94.86%	94.02%
5	Multi-aCRNN (ours)	Frequency+spatial+temporal	96.43%	96.30%

the reverse time series information also has a certain effect, promoting the overall experimental results. At the same time, by comparing the experimental results of single-scale and multiscale, it can be seen that the results of using 5-scale and 7-scale convolution kernels at the same time to extract features from the original data are significantly better than the results of using 7-scale convolution kernels alone. Indicating the fusion of different scale features is more helpful for the model to learn more comprehensive and effective information to achieve better classification results.

3.6. Contrast of Splicing and Stacking Preprocessing. This section experimentally explores two different preprocessing methods, splicing and stacking. It can be seen from Table 6 that the stacking preprocessing method improves the arousal and valence classification results by 1.47% and 1.7%, respectively, compared with the splicing preprocessing method. And, for a single subject, the stacking results are more stable. At the same time, the splicing method has two subjects with an accuracy of less than 90% in arousal classification and four subjects with an accuracy of less than 90% in valence classification. So in this experiment, the stacking preprocessing method is used for the experiment.

4. Discussion

As shown in Table 7, when the model learns spatiotemporal information based on extracting frequency information, it tends to get a better experimental result. However, when the model does not distinguish the obtained information and trains all the information, it will affect the training effect of the model. We perform weight training on the extracted features through the attention mechanism, amplify effective information, reduce invalid information, and use label smoothing to reduce the impact of noise in the label on the final classification result. It can be seen by comparing the papers using the same feature information that multi-aCRNN (ours) outperforms the 4D-CRNN model by 2.21% and 1.72% on the arousal task and the valence task, respectively, and surpasses the DE-CNN-BiLSTM model by 1.57% and 2.28%. We can conclude that selectively training on frequency, spatial, and temporal information is more conducive to emotion recognition, and reducing label noise positively affects emotion recognition.

5. Conclusions

This paper proposes a multiview feature fusion attention convolutional recurrent neural network model for EEG sentiment analysis. This method extracts more comprehensive spatial feature information through multiscale convolution and combines the frequency-domain features and spatial-domain features of EEG data. The weight fusion is carried out from the time series perspective so that the model learns more accurate information for classification prediction. Through the comparison experiment between GRU and bi-GRU networks, the bi-GRU network is determined as the network layer for temporal feature extraction. At the same time, the noise in the label is smoothed, which effectively reduces the impact of label noise on the classification results, realizes emotion recognition in complex practical situations, and is verified on the DEAP dataset. The multi-aCRNN model achieved 96.43% and 96.30% on the arousal task and valence task, respectively. At the same time, this paper conducts an experimental comparison of stacking and splicing of motor conversion methods to understand the impact of different feature combination methods on sentiment analysis. These experiments can be the basis for further research on EEG characteristics and better experimental research on emotion analysis.

Although the experiments in this paper effectively fuse multiview features and obtain high classification accuracy, the model still has shortcomings in classification tasks. In future work, we will extend multi-aCRNN to a multiclass classifier to obtain more accurate emotional state localization and achieve more accurate and good classification results.

Data Availability

The experimental data in this paper comes from the public data set DEAP, and the data set source link is http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

XF and RH conceived the project, designed the experiments, and drafted the manuscript. RH, FB, PC, and ZF collected the data and conducted the experiments. RH, FB, and XF proofed and polished the manuscript and organized this project.

Acknowledgments

This work is supported by the Natural Science Foundation of Jilin Province (YDZJ202301ZYTS401 and YDZJ202301-ZYTS288), the Science and Technology Project of the Education Department of Jilin Province (JJKH20220245KJ and JJKH20220226SK), and the National Natural Science Foundation of China Joint Fund Project (U19A200496).

References

[1] K. Ajith, R. Menaka, and S. S. Kumar, "EEG based mental state analysis," *Journal of Physics Conference Series*, vol. 1911, no. 1, article 012014, 2021.

[2] A. H. Brooke and N. A. Harrison, "Neuroimaging and emotion," in *Stress: Concepts, Cognition, Emotion, and Behavior*, vol. 1, pp. 251–259, Elsevier, 2016.

[3] M. Iqbal, S. M. Ali, M. Abid, F. Majeed, and A. Ali, "Artificial neural network based emotion classification and recognition from speech," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, pp. 434–444, 2020.

[4] C. Wu, F. Wu, Y. Huang, S. Wu, and Z. Yuan, "THU_NGN at IJCNLP-2017 task 2: dimensional sentiment analysis for Chinese phrases with deep LSTM," in *Proceedings of the IJCNLP 2017, Shared Tasks*, pp. 47–52, Taipei, Taiwan, 2017.

[5] L. A. Moctezuma, T. Abe, and M. Molinas, "Two-dimensional CNN-based distinction of human emotions from EEG channels selected by multi-objective evolutionary algorithm," *Scientific Reports*, vol. 12, p. 1, 2022.

[6] M. Yen, Y. Huang, L. Yu, and Y. L. Chen, "A two-dimensional sentiment analysis of online public opinion and future financial performance of publicly listed companies," *Computational Economics*, vol. 59, no. 4, pp. 1677–1698, 2022.

[7] W. L. Zheng and B. L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[8] S. Shanmugam and I. Padmanaban, "Twitter emotion analysis for brand comparison using naive Bayes classifier," in *International Conference on Soft Computing and Its Engineering Applications*, pp. 199–211, Springer, 2021.

[9] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding emotions in text using deep learning and big data," *Computers in Human Behavior*, vol. 93, no. 4, pp. 309–317, 2019.

[10] D. Gorecky, M. Schmitt, M. Loskyll, and D. Zühlke, "Human-machine-interaction in the industry 4.0 era," in *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, pp. 289–294, Porto Alegre, Brazil, 2014.

[11] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis et al., "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[12] S. Wang, X. Wang, Z. Wang, and R. Xiao, "Emotion recognition based on static human posture features," in *International Conference on Computing, Control and Industrial Engineering*, pp. 529–539, Springer, Singapore, 2022.

[13] M. M. Rahman, A. K. Sarkar, M. A. Hossain et al., "Recognition of human emotions using EEG signals: a review," *Computers in Biology and Medicine*, vol. 136, no. 2, article 104696, 2021.

[14] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 81–84, San Diego, CA, USA, 2013.

[15] L. Jin, S. Gao, Z. Li, and J. Tang, "Hand-crafted features or machine learnt features? Together they improve RGB-D object recognition," in *2014 IEEE International Symposium on Multimedia*, pp. 311–319, Taichung, Taiwan, 2014.

[16] N. H. Nguyen, D. T. A. Nguyen, and J. Hu, "The application of machine learning and deep learning in sport: predicting NBA players performance and popularity," *Journal of Information and Telecommunication*, vol. 6, no. 2, pp. 217–235, 2022.

[17] N. G. Paterakis, E. Mocanu, M. Gibescu, B. Stappers, and W. V. Alst, "Deep learning versus traditional machine learning methods for aggregated energy demand prediction," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–6, Turin, Italy, 2017.

[18] Q. Chen, Q. Xie, Q. Yuan, H. Huang, and Y. Li, "Research on a real-time monitoring method for the wear state of a tool based on a convolutional bidirectional LSTM model," *Symmetry*, vol. 11, no. 10, p. 1233, 2019.

[19] Y. P. Lin, C. H. Wang, T. P. Jung et al., "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[20] W. Zheng, J. Zhu, Y. Peng, and B. Lu, "EEG-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, Chengdu, China, 2014.

[21] B. I. García-Martínez, A. Martínez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero, "A review on nonlinear methods using electroencephalographic recordings for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 801–820, 2021.

[22] Y. Zhu and Q. Zhong, "Differential entropy feature signal extraction based on activation mode and its recognition in convolutional gated recurrent unit network," *Frontiers in Physics*, vol. 8, no. 1, article 629620, 2021.

[23] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Applied Soft Computing*, vol. 100, no. 1, article 106954, 2021.

[24] F. Shen, G. Dai, G. Lin, J. Zhang, W. Kong, and H. Zeng, "EEG-based emotion recognition using 4D convolutional recurrent neural network," *Cognitive Neurodynamics*, vol. 14, no. 6, pp. 815–828, 2020.

[25] F. Cui, R. Wang, W. Ding, Y. Chen, and L. Huang, "A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition," *Mathematics*, vol. 10, no. 4, pp. 582–582, 2022.

[26] S. Koelstra, C. Mühl, M. Soleymani et al., "DEAP: a database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[27] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3D input for EEG-based emotion recognition," in *The 25th International Conference on Neural Information Processing*, p. 433, Springer, 2018.

[28] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, Rio de Janeiro, Brazil, 2018.

[29] H. Gao, S. Lin, Y. Yang, C. Li, and M. Yang, "Convolution neural network based on two-dimensional spectrum for hyperspectral image classification," *Journal of Sensors*, vol. 2018, Article ID 8602103, 13 pages, 2018.

[30] Y. Yan, Y. Mao, and B. Li, "SECOND: sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[31] M. Sun, Z. Song, X. Jiang, J. Pan, and Y. Pang, "Learning pooling for convolutional neural network," *Neurocomputing*, vol. 224, no. 8, pp. 96–104, 2017.

[32] A. Vaswani, N. M. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, vol. 5, no. 6-pp. 6000–6010, Curran Associates Inc, Long Beach, California, USA, 2017.

[33] J. P. Amorim, P. H. Abreu, M. Reyes, and J. Santos, "Interpretability vs. complexity: the friction in deep neural networks," in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, Glasgow, UK, 2020.

[34] J. Wang, P. Zhang, Q. He, Y. Li, and Y. Hu, "Revisiting label smoothing regularization with knowledge distillation," *Applied Sciences*, vol. 11, no. 10, pp. 4699–4699, 2021.

Hindawi
Journal of Sensors
Volume 2023, Article ID 3132863, 10 pages
https://doi.org/10.1155/2023/3132863

Research Article

Multisensor Intelligent Fall Perception Algorithm considering Precise Classification of Human Behavior Characteristics

Yafei Ding

College of Information Technology, Pingdingshan University, Pingdingshan 467000, China

Correspondence should be addressed to Yafei Ding; 2545@pdsu.edu.cn

Received 27 October 2022; Revised 16 December 2022; Accepted 5 April 2023; Published 21 April 2023

Academic Editor: Mohit Mittal

In order to improve the accuracy and efficiency of human motion perception, a multisensor intelligent fall perception algorithm considering the precise classification of human behavior characteristics is proposed. Multisensor devices (smart watches, smart phones) collect data such as acceleration and heart rate of the human body to obtain human behavior data. On the basis of human behavior data collection, the acceleration characteristics of a falling state are extracted, and the SVM method is used to classify human behavior characteristics. Cuckoo search is used to optimize the width of the SVM kernel and improve the accuracy of human behavior recognition. Finally, based on the behavior recognition results, the intelligent perception of human falling behavior is realized through the exercise preparation potential. The experimental results show that the perceptual accuracy of this method is high, which has reached 90%, and the perception efficiency is higher. The minimum perception time is only 0.56 s, which fully verifies the effectiveness of this method. It can be widely used in human-computer interaction, machine vision, and other fields.

1. Introduction

With the development of computing technology, the machine-centered computing model is changing to the human-centered computing model. Let people become a part of computing, promote the integration of the physical world and the information world, and realize high-level human-computer interaction is the future development direction [1, 2]. Accurate perception and understanding of human behavior are essential technical supports. As an important branch of image understanding, human behavior perception has a broad application prospect in computer vision fields such as video surveillance, human-computer interaction, and virtual reality. Considering the characteristics of human behavior, estimating, recognizing, and perceiving human behavior is a research hotspot and has been a difficulty in the field of computer vision in recent years [3, 4].

Yang et al. [5] proposed a method of human fall perception based on SECNN (squeeze and excitation cellular neural network), built a data acquisition environment, used a wireless radio frequency tomography network node as the communication foundation of the network, and used the network node to build a complete wireless sensor network communication system. The processing method of wireless sensor network data is "transmission, reception, and storage." According to the collected RF (radio frequency) signal strength value, effective links are extracted, denoising, wavelet transform, time-domain features, and wavelet domain features are extracted, and the SECNN model is trained using the extracted multidomain features. The XGBoost (Extreme Gradient Boosting) model is used to filter the multidomain features obtained by permutation and combination of the time-domain feature components and wavelet domain feature components to obtain the joint feature components with strong robustness, and the multidomain feature perception fingerprint database of the fall action is established. The SECNN can be trained by using the extracted multidomain features, thus realizing the perception of human fall. The experimental results show that this method is convenient and feasible and is suitable for most scenes. However, this algorithm currently only considers the single-person scenario in the detection environment, not the multiperson scenario. Hao et al. [6] proposed a highly robust human motion perception model based on WiFi signals. This model uses

antenna diversity to eliminate random phase shift and takes the Doppler frequency shift and FFT (fast Fourier transform) value caused by human motion in the frequency domain as identification features. Finally, the application performance of the model is verified by experiments. However, the generalization ability of this algorithm is weak and cannot recognize people's actions in a multiperson environment, the feature extraction speed is slow, and the robustness is poor. Zheng et al. [7] proposed a human perception algorithm based on WiFi channel state information, which was designed by using a neural network method to convert the amplitude characteristics of channel state information into a three-dimensional matrix structure so as to retain the spatial, temporal, and frequency correlation carried by single sample data to the greatest extent. Then, 2D convolution is used to extract features from the 3D matrix, and random discard neurons and batch normalization are used to reduce overfitting. Finally, multitask learning is used to realize the parallel perception of human gesture and identity. The experimental results show that this method has high recognition accuracy for 150 gestures. However, the algorithm does not combine with the method of migration learning, so it cannot obtain a high-performance training model, and the accuracy of identity recognition is low. Choi et al. [8] proposed a fall risk perception algorithm for patients in acute nursing hospitals. This study aims to develop a fall risk perception questionnaire for patients in acute nursing hospitals and determine its reliability and effectiveness. In order to prevent patients from falling during hospitalization, they must accurately perceive the risk of falling. Although it achieved good results, it failed to use tools with established reliability and effectiveness to regularly assess patients' fall risk awareness. Si et al. [9] proposed a 3D perceptual reconstruction algorithm of the retina based on a visual light field image. As an important part of the eye, the retina is very important to the human body. A large number of data were collected through a literature survey, and the related theories and algorithm processes of computer vision, light field imaging technology, and 3D reconstruction technology were introduced in detail, which laid a sufficient theoretical foundation for the 3D perceptual reconstruction of the retina. However, due to the uncertainty of the background image and the defects of current medical imaging materials, the quality of the reconstructed image is not very good. Rehman et al. [10] proposed the application of multimodal computer vision for suspicious activity identification based on the Internet of Things in smart city security. The proposed research model uses cross-combined human-computer interaction to implement an IoT-based architecture for efficient and real-time decision-making. Use a class dataset from the UCF crime dataset for activity identification. At the same time, the dataset extracted for suspicious object detection involves human-object interaction. This research is also applied to the detection and recognition of human activities on campus, for real-time suspicious activity detection and automatic alarm. Experimental results show that the proposed multimodal method achieves significant activity detection and recognition accuracy. However, the pattern of this algorithm is not customizable, flexible, and extensible, and it has certain limitations. Şengül et al. [11] proposed to apply smart watches to medical care based on the fall detection of deep learning. Fall is divided into falling from a chair and falling from a standing position. A mobile application is developed that can collect acceleration and gyroscope sensor data and transmit it to the cloud. The rolling update method is used to calculate 38 statistical data features and use them as the input of the classifier. Although it has achieved good results, the deep learning algorithm does not perform well on large and diverse datasets. Şengül et al. [12] proposed smart phone sensor data fusion for daily user activity classification. New mobile applications need to estimate user activity by using sensor data provided by smart wearable devices and provide context-awareness solutions for users living in an intelligent environment. This method is based on the matrix time series method for feature fusion as well as the improved, better-than-optimal fusion method and the random gradient descent algorithm and is used to build the optimal decision tree for classification. In order to estimate user activity, the statistical pattern recognition method is used, and the k-nearest neighbor and support vector machine classifiers are used. The classification accuracy rate of pedestrian motor traffic activities is 99.28%, but the algorithm has some limitations because more user activities are not classified and a large number of sensor-fused data are not processed in time.

After the above analysis, although the previous algorithms have achieved good results, there are still different problems. Therefore, this paper proposes a multisensor fall intelligent perception algorithm considering the fine classification of human behavior characteristics. Compared with the previous research algorithm, the innovation of the algorithm proposed in this paper is:

(1) The acceleration, heart rate, and other data of the human body are collected through multisensor equipment to obtain human behavior data

(2) On the basis of human behavior data collection, the acceleration features of the fall state are extracted, and the SVM method is used to precisely classify the human behavior features

(3) The cuckoo search algorithm is used to optimize the width of the SVM core to improve the accuracy of human behavior recognition

(4) Based on the results of behavior recognition, the intelligent perception of human fall behavior is realized through motion preparation potential

The research results of the algorithm are as follows:

(1) The perceptual accuracy of the method in this paper is high, which has reached 90%

(2) The perception efficiency is higher, and the minimum perception time is only 0.56 s, which fully verifies the effectiveness of the method. It can be widely used in human-computer interaction, machine vision, and other fields

2. Perception of Human Fall Behavior

The foundation of intelligent perception is to build an intelligent platform with strong data analysis capability. Its design does not require other redundant wearing links and only requires the collection of human behavior data through multisensors to achieve intelligent data analysis. In the intelligent perception platform, we can divide the areas where people are prone to fall at ordinary times, prevent falling areas in advance, and give early warning the first time of falling. This technology requires the accuracy of motion recognition. If normal actions are taken as falls, the effect of motion recognition will be affected. Therefore, in the process of action perception, this paper will maximize the ability of intelligent perception analysis to ensure the accuracy of human fall behavior perception.

2.1. Human Behavior Data Collection. At present, most human behavior recognition is based on video images. In order to ensure sufficient and high-quality video image data, traditional human behavior recognition methods based on video images often need to install a large number of cameras in a fixed area. Even so, when collecting images, they cannot achieve full-time continuous tracking of users. At the same time, scripted training of data collection personnel is required before data collection, which leads to a far cry from the real-life state and the inability to record the real behavior data of individuals. In addition, privacy protection is also a big challenge in the process of actually collecting image information. With the intelligent development of sensor equipment, accelerometer, magnetometer, heart rate meter, and other sensors are built into smart phones and smart bracelets, which can track and collect human behavior data for a long time and fully protect personal privacy.

The structure of human behavior data acquisition based on multisensor is shown in Figure 1.

According to Figure 1, firstly, data such as acceleration and heart rate of the human body are collected by multisensor devices (smart watches and smart phones), and the original data are scanned in a fixed time window to obtain samples as the input of the recognition model. Then, the behavior recognition model is used to identify specific categories and generate individual behavior datasets. In this paper, in the data analysis of intelligent perception, common actions such as walking, running, and jumping are added, and the actions such as walking in situ, tiptoeing, and backing are selected as references, and the data of the above six actions are collected as the behavior data of normal falls.

The statistics of human behavior intelligent perception parameters are shown in Table 1.

Table 1 shows a group of data collected and saved in a file in plain text. The X axis, Y axis, and Z axis of the data are the original sampled data of intelligent perception. Through the definition of intelligent perception parameters, the action state of the human body at this time can finally be obtained.

The human behavior data collection method based on multisensor devices allows users to complete data collection by wearing mobile phones and smart bracelets. The advantages of this data collection method are as follows:

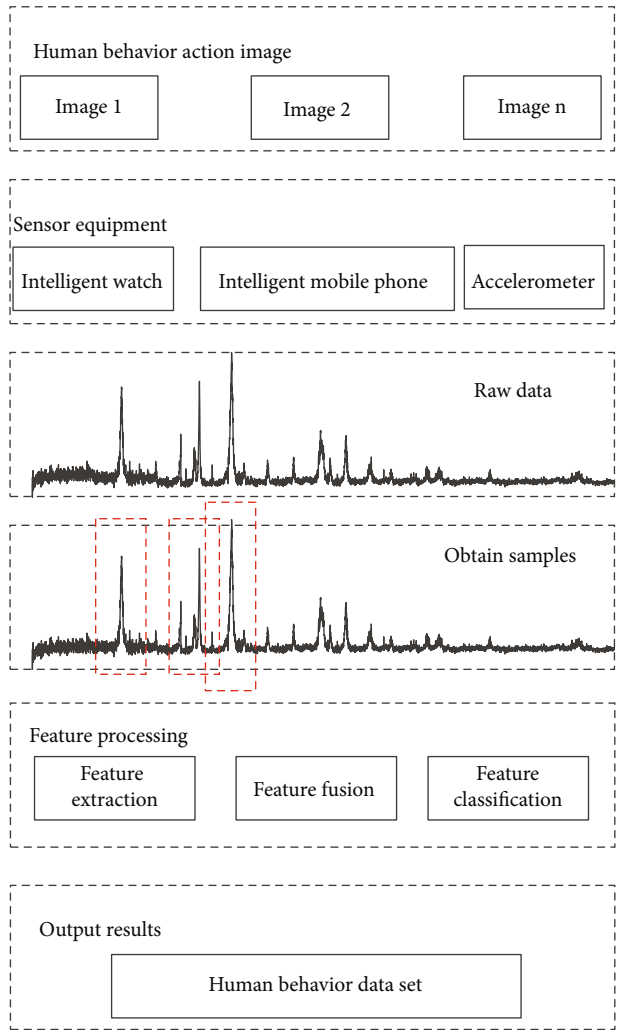


Figure 1: Structure diagram of human behavior data collection.

Table 1: Statistical table of human behavior intelligence perception parameters/met.

Action type	Intelligent sense parameters	X axis	Y axis	Z axis
Walk	273	396	219	640
Run	175	978	378	235
Beat	174	132	334	727
Walk in place	174	245	754	569
Tiptoe	198	880	285	203
Back off	205	741	520	870

(1) The data collection scenarios are diversified and are not limited by external conditions such as environment and weather

(2) The data is real, which can record a person's full-time motion data in a natural state

(3) The dataset is complete. Compared with video collection, data collection using multisensor devices

(smart phones, smart bracelets) is not subject to time, place, and human constraints

2.2. Feature Extraction of Falls. On the basis of human behavior data collection, the acceleration characteristics of the fall state are extracted, as shown in the following:

$$S(t) = \frac{s_x^2(t) + s_y^2(t) + s_z^2(t)}{2}. \quad (1)$$

In Formula (1), $S(t)$ is the state of human motion change; $s_x^2(t)$, $s_y^2(t)$, and $s_z^2(t)$ are the directional acceleration signal transmission status of X, Y and Z axes, respectively. Based on the big data analysis capability of intelligent perception, the human fall can be analyzed through the acceleration of $s_x^2(t)$, $s_y^2(t)$, and $s_z^2(t)$ in X, Y, and Z axes. The acceleration feature extraction method in this paper has a certain recognition effect, but its time-domain information still has a big loophole. Therefore, this paper will convert the time-domain features, and the conversion formula is as follows:

$$F_\eta(h) = \sum_{n=1}^{N} g(r)^{2\pi k}. \quad (2)$$

In Formula (2), F_η is the conversion coefficient; h is the robust feature in the time domain; N is time domain sequence; n is the time-domain sequence point; k is a constant. Because the conversion coefficients of the time domain characteristics are different, $0 \le h \le N$. In addition, when the time-domain sequence performs feature conversion, the sequence length also needs to be converted at the same time, and k needs to be guaranteed to be an integer power of 2. Under the same feature conversion frequency, this paper plans the time-domain sequence of acceleration uniformly, so that $F_\eta(h)$ and $c(n)$ are converted to $N = 0$, and the feature extraction results under the ideal state are as follows:

$$F_\eta(h) = \begin{cases} N & n \le \min F_\eta(h) \\ 0 & \text{other} \end{cases}. \quad (3)$$

In Formula (3), $c(n)$ is the ideal feature extraction state of the converted fall.

After the above feature extraction and conversion, six motion features, such as walking, running, jumping, walking in place, tiptoeing, and backing, are extracted to ensure the effect of motion perception.

2.3. Fine Classification of Human Behavior Characteristics. Based on the results of fall motion feature extraction, the human behavior features are precisely classified by the SVM (support vector machines) method. SVM has the characteristics of global optimization and good learning ability for small samples. It is a more suitable method for feature classification. The performance of SVM mainly depends on the kernel function. Due to different kernel functions, the performance of SVM will reflect different advantages and disadvantages. In the case of unknown data, the selection of the kernel function becomes a difficult problem. In the case of known data, the selection of an appropriate kernel function is the key problem to be solved. In order to construct an efficient SVM with good robustness, this paper proposes and uses an iterative method to improve the hyperkernel function that depends on training data, so as to reduce the learning risk and improve the convergence speed.

Support vector machines often use two types of kernel functions:

(1) Translation invariant kernel, in the form of:

$$W\left(y, y'\right) = f\left(\left\|y - y'\right\|^2\right). \quad (4)$$

In Formula (4), $f(\cdot)$ is an arbitrary function. Such kernel functions include the Gaussian kernel and polyhedral kernel.

(2) Rotation invariant kernel, in the form of:

$$W\left(y, y'\right) = u\left(\left\|y \cdot y'\right\|^2\right). \quad (5)$$

In Formula (5), $u(\cdot)$ is an arbitrary function, which includes the polynomial kernel and two-layer neural network kernel.

This paper constructs a hyperkernel function to overcome the shortcomings of a single kernel function:

$$W'\left(y, y'\right) = \alpha_1\left(\left\|y - y'\right\|^2\right) + \alpha_2\left(\left\|y \cdot y'\right\|^2\right). \quad (6)$$

In Formula (6), α_1 and α_2 are classifier parameters, $\alpha_1 + \alpha_2 = 1$. The first term uses the commonly used Gaussian kernel function, and the second term uses the second-order polynomial kernel function.

For a classification problem, suppose there is $\{y_i, z_i\}$, where $i = 1, 2, \cdots, j$ and y_i are the vector of the input space $A = E^n$, and z_i is the category index. SVM uses nonlinear mapping $\varepsilon(y)$ to map data y to high-dimensional feature space $F = E^n$ and uses hyperplane to separate the two types of problems, that is, to find a linear classification equation $f(y) = \mu \cdot \varepsilon(y) + b$ in the feature space. From a set point of view, this mapping defines the transition from input space A to feature space F. When feature space F is a Euclidean or Hilbert space, Riemannian metric can be defined:

$$p_{ij}(y) = W\left(y, y'\right) \times \left(\frac{\theta}{\theta y_i} + \frac{\theta}{\theta y_j}\right). \quad (7)$$

By increasing the spatial resolution of the interface in F, the distance between classes can be increased so as to improve the resolution performance of SVM and thus improve the classification accuracy of human behavior features.

2.4. Human Behavior Recognition. Although SVM can obtain the classification results of human behavior features when the number of training samples is small, the training time is long, and the running efficiency is low. RVM (relevance vector machine) integrates the advantages of SVM and neural networks. It not only overcomes the "overfitting phenomenon" caused by the small number of neural network training samples but also has the good classification ability of SVM, providing a new research tool for human behavior recognition. However, the kernel function parameters are still a problem to be solved. Therefore, in order to obtain an excellent human behavior perception effect, it is necessary to optimize the kernel width parameter ϕ.

2.4.1. Cuckoo Search Optimization Algorithm. The cuckoo algorithm is a swarm intelligence algorithm based on the transformation of the cuckoo's nest-seeking behavior in nature. Compared with the ant colony algorithm, annealing algorithm, and particle swarm algorithm under similar algorithms, it can effectively promote the clustering algorithm to quickly jump off the local optimal steady-state platform to achieve steady-state convergence. The cuckoo search optimization algorithm is proposed on the basis of the cuckoo search algorithm. It is a metaheuristic algorithm based on nest parasitism and flight search mechanism, which are characterized by low complexity and good global performance [13–17]. The algorithm first defines three assumptions, as follows:

(1) A cuckoo bird lays only one egg and is randomly deployed in the nest

(2) Some of the better nests will be reserved for the next generation, and only some of the worse nests will be updated

The probability of the host recognizing cuckoo eggs is $P_b \in [0, 1]$.

The path and location update operation of cuckoo's nest searching is as follows:

$$V_i^{s+1} = V_i^s + \delta L(\lambda). \tag{8}$$

In Formula (8), V_i^s and V_i^{s+1} represent the position vectors of the cuckoo nest of generation s and $s+1$; $L(\lambda)$ is the random search path of cuckoo flight; δ is a regulating factor.

The relationship between random search path and time t follows the distribution of

$$L^s = t^{-\sigma}. \tag{9}$$

In Formula (9), σ is the power coefficient, and $1 \le \sigma \le 3$.

In the working process of traditional algorithms, most bird nests are randomly updated, which makes the important information in the area near the bird's nest not be fully utilized. Therefore, a selective elimination strategy is adopted, specifically:

$$V_{i,\text{nest}_1}^s = V_{i,\text{nest}_2}^s + m\left(V_{i,\text{nest}_2}^{s'} - V_i^{s'}\right). \tag{10}$$

In Formula (10), V_{i,nest_2}^s, $V_{i,\text{nest}_2}^{s'}$, and $V_i^{s'}$ are the nest positions at different times.

The traditional algorithm has a strong global search ability but a weak local search ability. Therefore, a fine search strategy is adopted to optimize the core width parameter ϕ, specifically:

$$V_{i,\text{nest}_1}^s = V_{i,\text{nest}_2}^s + \frac{V_{\text{best}}^s - V_{\text{worst}}^s}{\phi}. \tag{11}$$

In Formula (11), V_{best}^s and V_{worst}^s are the best and worst nest positions, respectively. If the cuckoo search optimization algorithm assumes that there are N adult cuckoo birds in the current population, the probability of their nests being found is P_a and usually set as 0.25; the initial spatial distribution of cuckoo birds is X_0, and the flight iteration distance and flight mode of the current population based on the moving step α and Levi distribution $\text{Levy}(\lambda)$ can be expressed as:

$$X^{t+1} = X_0 + \alpha \times \text{Levy}(\lambda) \times P_a \times V_{i,\text{nest}_1}^s. \tag{12}$$

In Formula (11), λ represents the data parameter in the Levy distribution $\text{Levy}(\lambda)$, and λ is usually set to 1.5, α is usually set to 1. The expression of $\text{Levy}(\lambda)$ is:

$$\text{Levy}(\lambda) = \frac{(1+\lambda) \times \sin(\pi \times \lambda/2)}{\lambda}. \tag{13}$$

The population position is iteratively updated through $\text{Levy}(\lambda)$, and the fitness function is recalculated. According to the results obtained, the data point set of the optimal solution is determined as $X_t = (x_1^t, x_2^t, \cdots, x_N^t)$, and the relationship between the preset value X_t and P_a is compared. If $X_t > P_a$, the population position is iteratively updated, and the above operations are repeated until the optimal solution or the preset iteration threshold is reached, otherwise, the current iteration result is retained as the optimal solution; thus, the research of the cuckoo search optimization algorithm is completed.

2.4.2. Steps of Human Behavior Recognition

(1) For the human motion features extracted in Section 2.2, they are normalized as follows:

$$\psi_{ij} = \frac{\psi_{ij} - \min(\psi_i)}{\max(\psi_i) - \min(\psi_i)}. \tag{14}$$

(2) Determine the value range of the Gaussian kernel function parameter ϕ of the correlation vector machine

Initialize the nest position vector, which includes a subset of human behavior features and ϕ.

(3) Calculate the fitness value of the nest position vector, as follows:

$$\kappa = \sum_{i=1,j=1}^{N} \rho_{ij} \times \omega. \quad (15)$$

In Formula (13), ρ_{ij} is the characteristic state; ω is the weight value.

(4) Some poor-quality nest position vectors are updated to generate new nest positions

(5) If the end condition of the algorithm is reached, the human behavior feature subset and ϕ are obtained according to the globally optimal nest position vector

(6) Build human behavior recognition model according to feature subset and ϕ [18, 19]

The workflow of the human behavior recognition model with the cuckoo search algorithm optimizing features and classifier parameters are shown in Figure 2.

2.5. Intelligent Perception of Human Fall Behavior. On the basis of human behavior recognition, the process of intelligent perception of human falling behavior is shown in Figure 3.

Firstly, the human behavior data collected by multisensor is used as the data source to preprocess the sensor signal, and the multisensor equipment is used to determine the starting time of human motion. Secondly, assuming that the starting time of human motion is zero, the sensor signal of -3.0–3.0 s is defined as an effective data segment to complete data segmentation and event alignment. Thirdly, the motion artifact component is removed from the segmented data.

After data processing is realized, an intelligent perception of human fall behavior is conducted based on motion readiness potential, and the steps are as follows:

First, according to the acquired characteristics of human behavior changes, the channel that conforms to the change rule of motor preparation potential is selected as the benchmark, and the falling edge of its motor preparation potential is taken as the matching template.

Secondly, the sliding step, threshold, and observation window are set for sliding matching. When the matching coefficient is greater than the set threshold, the motor readiness potential is considered to occur, and the time distance between the endpoint of the current observation window position and the starting time of the motion is calculated.

Finally, count the number of channels that can be matched; when it is greater than the threshold value of the number of channels set, it is considered that the human fall behavior occurred in this experiment, and determine the preperceived time of human fall [20–25].

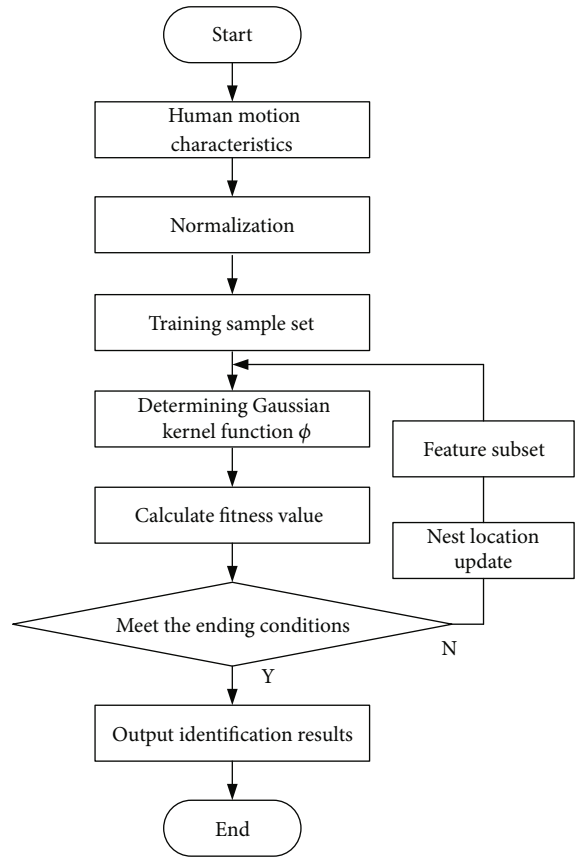


Figure 2: Flow chart of human behavior recognition.

3. Experimental Analysis

In order to verify the practicability and effectiveness of the multisensor intelligent fall perception algorithm considering the precise classification of human behavior characteristics, experimental research was carried out. In the experiment, a method of human fall perception based on SECNN and a highly robust human motion perception model based on WiFi signals were used as comparison methods and compared with this method.

3.1. Experimental Design and Arrangement. A total of 6 college students (labeled S1-S6) were selected in the experiment, including 5 males and 1 female, aged between 22 and 25. None of the 6 students had any history of sensorimotor impairment or any psychological disorder, and they signed an informed consent form with the subjects before the experiment started.

The experimental arrangement is as follows: each subject needs to complete two groups of lower limb autonomous movement experiments, including the left leg step and right leg step. In order to ensure the accuracy and integrity of the experiment, the subjects should always stand naturally during the experiment. After the start of the experiment, the subjects remained in a resting state for a certain time. Then, the subject can start the left or right leg's autonomous walking movement and keep it for about 1-2 seconds. The movement start time is completely controlled by the subject,

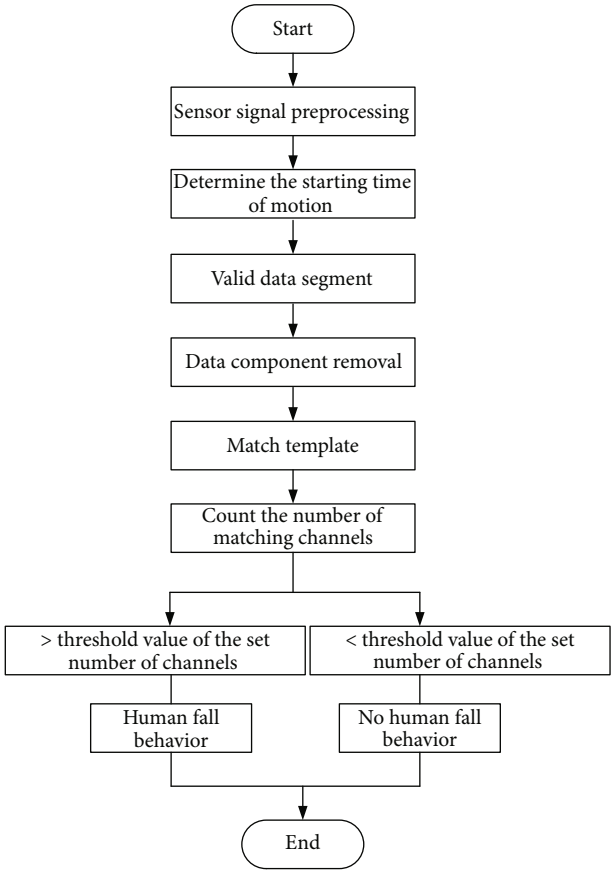


Figure 3: Flow chart of human fall behavior intelligent perception.

without any prompts or requirements. Thirdly, the subjects ran, jumped, tiptoed, and retreated. Finally, the subjects had a rest of about 10 s after completing the above actions and prepared for the next experiment. According to the above experimental arrangement, the subjects were tested many times, and the number of falls was counted. The process is shown in Figure 4.

3.2. Experimental Evaluation Index. In this experiment, three evaluation indexes—precision A, recall R, and F_1 value—are used. The precision is the proportion of the number of real positive cases in the detected positive cases; A is the most common evaluation index, and its calculation method is relatively simple, namely:

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}. \tag{16}$$

The recall rate is the proportion of all samples judged as positive examples in all positive examples. R is also called recall, and its calculation method is:

$$R = \frac{T_P}{T_P + F_P}. \tag{17}$$

The value F_1 is put forward on the above two evaluation indexes, which can evaluate the advantages and disadvantages of the perception algorithm. The F_1 value is the harmonic average of precision and recall. If only precision or recall is considered, it cannot be used as an indicator to evaluate a method. Therefore, the F_1 value is used to reconcile the two, which is compatible with precision and recall, and it is also one of the commonly used evaluation indexes. The calculation method is:

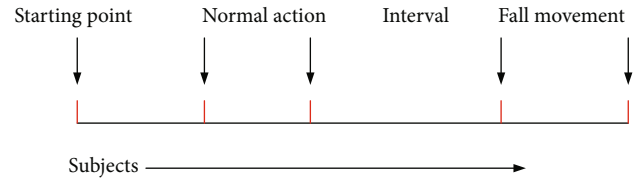


Figure 4: Schematic diagram of an experimental process.

$$F_1 = \frac{2 \times A \times R}{A + R}. \tag{18}$$

In the above formula, T_P is a true example, T_N is a true negative example, F_P is a false positive example, F_N is a false negative example, and the four parameters, respectively, indicate the number of positive samples, negative samples, negative samples, and negative samples.

3.3. Results and Analysis. Formulas (16)–(18) are used to calculate the intelligent perception results of human fall behavior based on a method of human fall perception based on SECNN proposed by Yang et al. [5], a highly robust human motion perception model based on WiFi signals, and this method proposed by Hao et al. [6], as shown in Figures 5–7.

According to Figures 5–7, compared with a method of human fall perception based on SECNN and a highly robust human motion perception model based on WiFi signals, this method has obvious advantages in three commonly used evaluation indexes, which fully show that this method has a better perception effect on human falling behavior. By analyzing the specific data, it can be seen that the A value of this method has achieved excellent results, and the accuracy has reached 90%. This method has also achieved good results in R value and F_1 value, which proves the advantages of this method. The above experimental results show that the motion perception effect of this method is better because the cuckoo search is used to optimize the kernel width parameter ϕ, which improves the correct rate of human behavior recognition.

In order to further verify the effectiveness of the method in this paper, a method of human fall perception based on SECNN, a highly robust human motion perception model based on WiFi signals, and the method in this paper are compared with the action perception efficiency as an experimental indicator. The results are shown in Table 2.

According to Table 2, the number of experiments increases gradually, and the perception time of the iterative actions of the three methods also increases gradually. In comparison, the perception time of this method is shorter, and its minimum value is only 0.56 s, which is 0.96 s and 1.04 s lower than that of a method of human fall perception

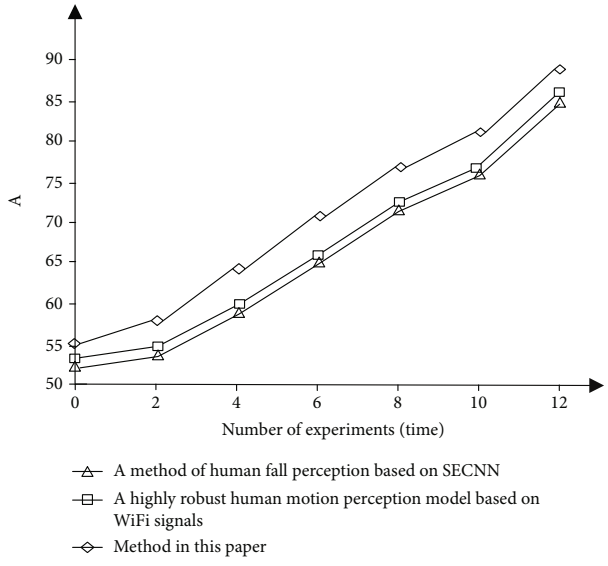


Figure 5: A value comparison results.

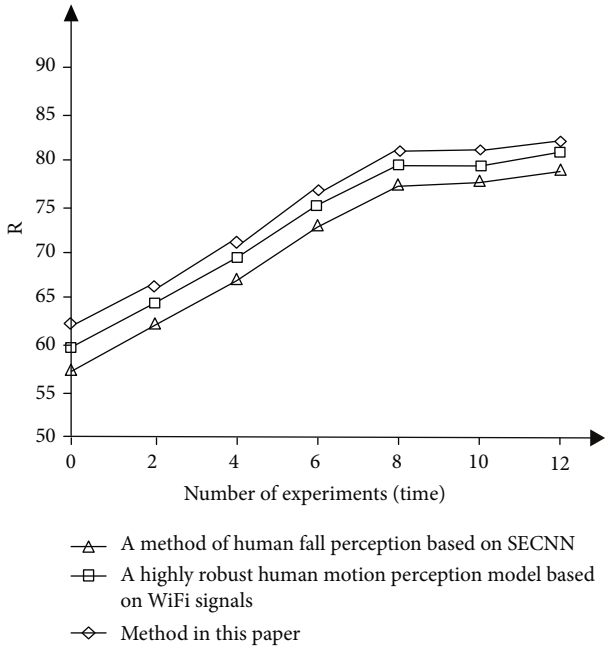


Figure 6: R value comparison results.

based on SECNN and a highly robust human motion perception model based on WiFi signals, respectively. The maximum perception time of this method is 1.13 s, which is 1.56 s and 1.42 s lower than that of a method of human fall perception based on SECNN and a highly robust human motion perception model based on WiFi signals, respectively. A method of human fall perception based on SECNN and a highly robust human motion perception model based on WiFi signals are 2.138 s and 2.115 s, respectively. The average perceived efficiency of this method is 0.82, which is 1.318 s and 1.295 s lower than the other two methods. To sum up, the method in this paper is more efficient in motion perception, which shows that this method can react to the human fall in a shorter time, which is conducive to avoiding

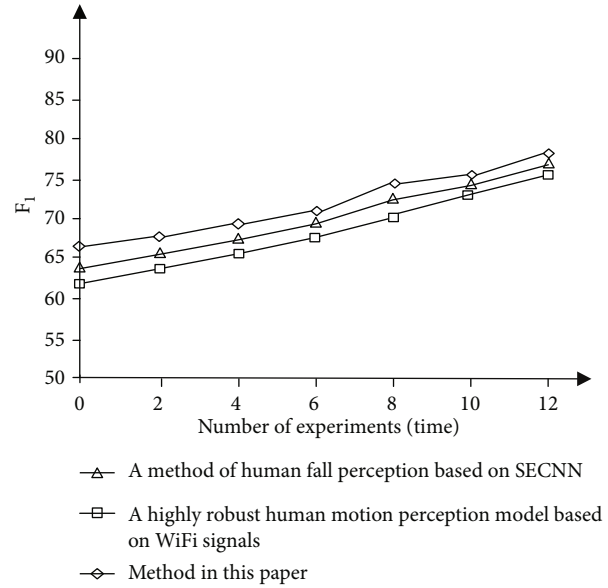


Figure 7: F_1 value comparison results.

Table 2: Comparison of perceived efficiency of falls.

Number of experiments/time	Fall motion perception time/s		
	A method of human fall perception based on SECNN	A highly robust human motion perception model based on WiFi signals	Method in this paper
1	1.52	1.60	0.56
2	1.87	1.89	0.69
3	2.01	2.03	0.72
4	2.23	2.17	0.80
5	2.51	2.45	1.02
6	2.69	2.55	1.13
Average value	2.138	2.115	0.82

the timely response of the human body and serious physical injury.

4. Conclusion

Many scholars have been constantly creating new technologies to realize more effective perception and analysis of human behavior, and many behavioral analysis technologies have been gradually applied to daily life. However, the existing methods still have the problem of low accuracy in motion perception. Therefore, to improve the accuracy and efficiency of human motion perception is the research purpose, this paper proposes a multisensor intelligent fall perception algorithm considering the precise classification of human behavior characteristics. Through research, the main conclusions are as follows:

(1) Multisensor devices (smart watches, smart phones) are used to collect human acceleration, heart rate,

and other data to obtain human behavior data and improve the comprehensiveness of the data

(2) Based on the results of human behavior data collection, the features of the fall state are extracted, and the SVM method is used to precisely classify human behavior features, which is conducive to improving the accuracy of motion perception

(3) Cuckoo search is used to optimize the width of the SVM kernel to improve the accuracy of human behavior recognition. Based on the behavior recognition results, an intelligent perception of human fall behavior is realized through motion preparation potential

(4) This method has high perception accuracy. Among them, the A value has achieved excellent results, the accuracy has reached 90%, and the perception efficiency is higher. The minimum value of perception time is only 0.56 s, which is 0.96 s and 1.04 s lower than the methods in reference, respectively, which fully verifies the effectiveness and application value of the method

Although the method in this paper has achieved some research results, the miscalculation rate of fall behavior is high, and the amount of calculation is too large, which can be further studied in the future.

Data Availability

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

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

References

[1] Y. Sha, T. Feng, X. Xiong, and T. Yang, "Designing online psychological consultation expert system using human-computer interaction," *Mobile Information Systems*, vol. 2021, Article ID 6458924, 12 pages, 2021.

[2] L. Hu, Q. Zeng, X. Wu, and Z. Lv, "Simulation of English classroom effectiveness based on human-computer interaction and facial identification," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 4, pp. 7025–7036, 2020.

[3] C. Dai, X. Liu, J. Lai, P. Li, and H. C. Chao, "Human behavior deep recognition architecture for smart city applications in the 5G environment," *IEEE Network*, vol. 33, no. 5, pp. 206–211, 2019.

[4] R. Liang, H. Zhi, and M. M. Kamruzzaman, "Methods of moving target detection and behavior recognition in intelligent vision monitoring," *Acta Microscopica*, vol. 28, no. 4, pp. 750–759, 2019.

[5] Z. Y. Yang, J. J. Wang, and L. Jin, "Human fall detection method based on SE-CNN," *Computer Engineering*, vol. 48, no. 6, pp. 270–277, 2022.

[6] Z. J. Hao, Z. Q. Qiao, X. C. Dang, D. Y. Zhang, and Y. Duan, "Wi-do: highly robust human motion perception model under WiFi signal," *Journal of Computer Research and Development*, vol. 59, no. 2, pp. 463–477, 2022.

[7] Q. Zheng, Z. M. Xu, and L. Q. Chen, "A human sensing algorithm based on WiFi channel state information," *Journal of Fuzhou University (Natural Science Edition)*, vol. 49, no. 1, pp. 20–25, 2021.

[8] J. Choi, S. M. Choi, J. S. Lee et al., "Development and validation of the fall risk perception questionnaire for patients in acute care hospitals," *Journal of Clinical Nursing*, vol. 30, no. 3-4, pp. 406–414, 2021.

[9] L. Si, G. Jiang, X. Hu, and B. Liu, "Retina 3D perception reconstruction algorithm based on visual light field image," *IEEE Access*, vol. 8, pp. 196804–196812, 2020.

[10] A. Rehman, T. Saba, M. Z. Khan, R. Damaševičius, and S. A. Bahaj, "Internet-of-things-based suspicious activity recognition using multimodalities of computer vision for smart city security," *Security and Communication Networks*, vol. 2022, Article ID 8383461, 12 pages, 2022.

[11] G. Şengül, M. Karakaya, S. Misra, O. O. Abayomi-Alli, and R. Damaševičius, "Deep learning based fall detection using smartwatches for healthcare applications," *Biomedical Signal Processing and Control*, vol. 71, article 103242, Part B, 2022.

[12] G. Şengül, E. Ozcelik, S. Misra, R. Damaševičius, and R. Maskeliūnas, "Fusion of smartphone sensor data for classification of daily user activities," *Multimedia Tools and Applications*, vol. 80, no. 24, pp. 33527–33546, 2021.

[13] J. Wang, D. Wu, Y. Gao et al., "Integral real-time locomotion mode recognition based on GA-CNN for lower limb exoskeleton," *Journal of Bionic Engineering*, vol. 19, no. 5, pp. 1359–1373, 2022.

[14] Z. Xiong, X. Weng, and Y. Wei, "SandplayAR: evaluation of psychometric game for people with generalized anxiety disorder," *The Arts in Psychotherapy*, vol. 80, article 101934, 2022.

[15] C. Huang, Z. Han, M. Li, X. Wang, and W. Zhao, "Sentiment evolution with interaction levels in blended learning environments: using learning analytics and epistemic network analysis," *Australasian Journal of Educational Technology*, vol. 37, no. 2, pp. 81–95, 2021.

[16] P. Chonggao, "Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2421–2431, 2021.

[17] X. Geng, "Research on athlete's action recognition based on acceleration sensor and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2229–2240, 2021.

[18] X. Zhang, D. Huang, H. Li, Y. Zhang, Y. Xia, and J. Liu, "Self-training maximum classifier discrepancy for EEG emotion recognition," *CAAI Transactions on Intelligence Technology*, 2023.

[19] S. Maity, A. Chakrabarti, and D. Bhattacharjee, "Robust human action recognition using AREI features and trajectory analysis from silhouette image sequence," *IETE Journal of Research*, vol. 65, no. 2, pp. 236–249, 2019.

[20] B. Q. He, W. Wei, B. Zhang, L. X. Gao, and Y. B. Song, "Improved deep convolutional neural network for human action recognition," *Application Research of Computers*, vol. 36, no. 3, pp. 945–949, 2019, 953.

[21] L. B. Jiang, X. L. Zhou, and L. Che, "Few-shot learning for human motion recognition based on carrier-free UWB radar," *Acta Electronica Sinica*, vol. 48, no. 3, pp. 602–615, 2020.

[22] J. Yu, L. Lu, Y. Chen, Y. Zhu, and L. Kong, "An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 337–351, 2021.

[23] X. Hou, L. Zhang, Y. Su et al., "A space crawling robotic biopaw (SCRBP) enabled by triboelectric sensors for surface identification," *Nano Energy*, vol. 105, article 108013, 2023.

[24] X. Zenggang, Z. Mingyang, Z. Xuemin et al., "Social similarity routing algorithm based on socially aware networks in the big data environment," *Journal of Signal Processing Systems*, vol. 94, no. 11, pp. 1253–1267, 2022.

[25] B. Q. Yin, Y. Deng, S. D. Wang, Z. C. Hu, B. Li, and L. Zuo, "Application of time-frequency generalized S transform and VL-MOBP neural network in human motion recognition," *Journal of Electronic Measurement and Instrumentation*, vol. 34, no. 11, pp. 1–9, 2020.

Hindawi
Journal of Sensors
Volume 2023, Article ID 1911385, 18 pages
https://doi.org/10.1155/2023/1911385

Research Article

IoT-Based Voice-Controlled Smart Homes with Source Separation Based on Deep Learning

Ghalib H. Alshammri

Department of Computer Science, Community College, King Saud University, Riyadh 11437, Saudi Arabia

Correspondence should be addressed to Ghalib H. Alshammri; galshammri@ksu.edu.sa

Received 23 August 2022; Revised 19 September 2022; Accepted 6 October 2022; Published 28 March 2023

Academic Editor: Jaroslav Frnda

The widespread availability of cutting-edge computer technologies has shed light on the relevance of artificial intelligence (AI) applications in almost all sectors of the economy. As a result of the incorporation of voice control processing into many Internet of Things (IoT) devices, many of these IoT devices may be operated using spoken commands. The environment that is controlled by speech may include several devices, each of which may be used for a separate activity; yet, all of the devices may collect and process the same command at the same time. This may be the case if the devices can communicate with one another. Because other devices may choose to ignore orders that are intended for particular devices if those devices are not equipped to deal with those orders, only the device that is designed to carry out the activity and process the command will be able to carry out the activity. This is because only the device that is designed to carry out the activity and process the command will be able to carry out the activity. On the other hand, when all of the voice-controlled devices capture the command through the microphone, there is a greater chance that it will mix with other sounds coming from a variety of sources. This is because the microphone is being used to capture the command from all of the voice-controlled devices. These noises may include those that are emanating from the television, music systems, and other sounds that are created by activities taking on inside the family, among other things. During the identification of instructions via processing, any blending of other sounds that are not the primary command is regarded as noise and has to be deleted. This is because any such blending is deemed to be noise. The direction of arrival (also known as DOA) of the sound waves is given primary consideration by this approach. This is done at the same time as the performance of the system, and the proposal for it are being evaluated. Based on the angle of arrival estimate, a specific room impulse response (RIR) from a collection of defined RIR is identified as a room acoustic characteristic, and source separation is carried out using the technique of independent component analysis (ICA). Following the completion of the analysis of the signals produced by the split command speech, the characteristics of the speech are retrieved from the signals. The Mel-frequency cepstral coefficients (MFCC) approach is used so that the operation of feature extraction may be carried out. This is the goal of the technique. Following that, a support vector machine classifier is used to the data in order to further split these characteristics into a large range of distinct groups. Comparisons are made between the performance of the SVM classifier and the performance of a large number of different classifiers, including decision trees, which are often used in applications that incorporate machine learning (DT). After analyzing its performance, the multiclass SVM classifier is found to have an accuracy of 91%, according to the conclusions of the study. Utilizing a classifier that is based on a probabilistic neural network, which is sometimes referred to as a PNN, is one way in which the accuracy of future classifications may be enhanced. This particular classifier is made up of three layers: one layer of gated recurrent units (GRU), one layer of long short-term memory (LSTM), and one layer that integrates the two of those different kinds of memory. This classification seems to have obtained an accuracy of 94.5 percent, which is higher than the classification accuracy attained by the multiclass SVM classifier.

1. Introduction

The vast majority of unpaid track royalties are mixtures of various sound assets, such as units or human voices. At this point in time, the bad effects and unfavorable effects of specific units are sometimes no longer accessible in distinct amounts. Even taking into consideration the fact that the units are recorded individually, the isolated side effects and side effects do exist, despite the fact that they are often not disseminated at this time [1]. This is the manner in which well-known track manifestations do their business the vast majority of the time. It is not an uncommon procedure for musicians to carry out their job while recording the bulk of the instrument pointers for a select number of orders, some of which include jazz, traditional track, or society. This practice is not rare since it is not an unusual activity. In this particular case, there is not an outstanding amount of disengaged currency owed [2]. In addition, as of today, the cash payment for a single gadget that was due for a variety of distinct vintage track manifestations has been missing. Despite this, getting closer to the precise equipment markers is vital for tracking cash due for a couple of usage seasons. It may be necessary, for example, to exclude the vocals of the original artist from the track that will be used for karaoke in order to obtain the desired effect. In order to make another track piece, a couple of skilled specialists are required to eliminate one instrument from the total amount and then remix it with the given quantity of cash that is due. Craftsmen have the ability to create their own coordinate tracks by withdrawing their tools and supplies from a track on which they still have room to improve their abilities [3]. When it comes to the market for consumer audio products, a combination will often have the standard number of channels included within it. In order to play back a recording on a system that has more channels, one may need to "up mix" the recording and replace the spatial area that was occupied by an instrument. This allows the recording to be played back on the system that has more channels. In addition, a personalized study into monetary commitments is carried out by way of an introduction to the pertinent monetary assets. Examples of models include replicating the singer persona, recreating the instruments or language of a song, and reporting the stanzas or the melodic score of a song [4].

1.1. Blind Source Separation. At the tail end of the 1980s, a strategy of managing solid indications known as the blind stock segment, which is sometimes abbreviated as BSS in certain circles, gained popularity. BSS was what was happening in present compositions in the fields of biological sciences, talk sign communication, image management, geography, and mining for insights from literary content [5]. This was accomplished via the utilization of certified sign management, phoney psyche connections, and record theory. The term "source separation" (SS), which is also known as "signal separation," was coined because the methodology focused on convalescing a gathering of difficult to comprehend supply alarms "(time series, pictures from a gathering of insights (for example expected alarms), which can be combinations of those supply alarms)." In this manner, the term "SS" was portrayed. The BSS arrangement is relevant to the circumstance, but the possible limitations and benefits of the concept are not discussed here. Figure 1 shows the blind source architecture.

Picture spotlight extraction, face confirmation, moving item ID, widespread picture watermarks, picture denoising, picture separation, and picture recovery are typical applications of BSS basically based totally on picture management [6]. The recovery of the support is the cognitive process that is being tested in this experiment. This reenacts an extra mounted literary material that was previously deleted and in some cases rewritten. Figure 1 illustrates the red, green, and blue channels in isolation from one another.

1.2. Machine Monitoring. Signal parcel could be used to protect against potential mechanical frustrations by isolating the acoustic component (in this particular scenario, the sound produced by a mechanical framework all through hurt) [7] from the weather conditions involving a combination of different assets that include various likely normal running parts. This would allow for the acoustic component to be protected from any potential mechanical issues.

1.3. Investigative Procedures in Medicine. It is common practice to use a wide variety of logical tools in order to investigate a particular constellation of adverse effects and adverse effects originating from the human body (as illustration EEG signals, and ECG signals). A signal parcel might be used to separate a sign of relevance that is connected with a selected significant cycle (or) lifts that is likely available. This separation could be accomplished by using a signal parcel. It is possible that this will prevent an erroneous diagnosis brought on by the influence of clatter inside the system [8].

This is a selected piece of programming that may be used globally in order to preserve melodic displays at 0 in just on certain instrument sounds and voices included inside the recordings. This might be specially employed to boost a chosen device, such as a percussion instrument that has to be somewhat more intense than expected [9].

The measurements provided by the Little Array are a very beneficial way of typifying the facts that have been obtained about DNA and protein explanations. Blind stock parcel in multicollector exhibiting depiction may probably be used to seclude the gene [10] significance, to accommodate as an option for delayed progressions, district of periodicity, grouping, and order of characteristics. Blind stock parcel in multicollector exhibits depiction.

The suggested VAD algorithm that is addressed in the proposed work is dedicated to building a deliberate VAD algorithm that is ideal for ASR systems. This is because simplicity and robustness are two of the primary issues that VAD algorithms face when used for ASR systems. Not only is this technique durable in a variety of different noisy settings [11], but it also has a low computing cost. The suggested method includes the incorporation of a noise estimate to characterize the various ambient noise signals. This estimation is then used to update the noise for each frame.

$$y(t) = H_{\cdot} x(t). \tag{1}$$

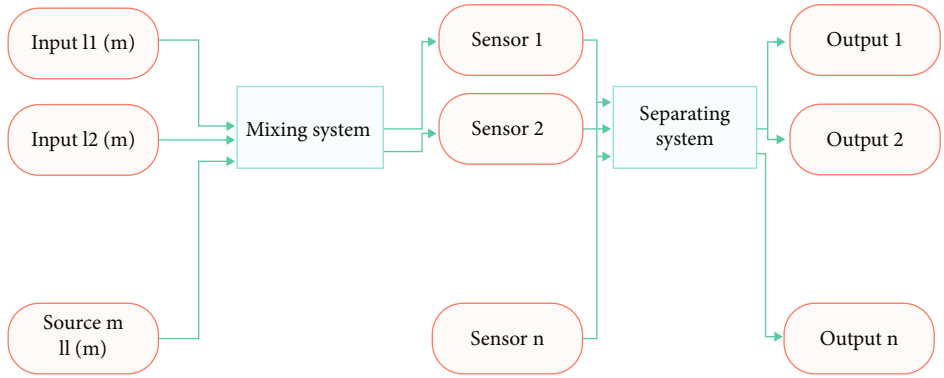


Figure 1: Blind source system architecture.

There have been many different VAD algorithms suggested up to this point. In order for these algorithms to be effective in combating the myriad of external sounds, a variety of discriminating characteristics have been implemented. The energy-based characteristics are the most sought-after of all of them due to the fact that they are both straightforward and efficient. In addition, it is challenging to build an effective combination algorithm that is appropriate for all situations. Traditional characteristics such as zero-crossing rate, pitch-based detection, and weak fricative detection are not used since they are not noise-immune. Other conventional features that are not utilized include weak fricative detection. These distinguishing traits have cheap calculation costs, which makes them excellent for automatic speech recognition because of their applicability.

Popular speech enhancement algorithms have been developed with the primary goal of improving the intelligibility and/or quality of the speech signal. When developing these algorithms, designers did not take into account the impact that their decisions may have on other speech processing systems.

The intelligibility of speech is not improved by the speech enhancement algorithms, not even the most advanced ones. The reason for this is that it does not possess a reliable estimate of the background noise spectrum, which is essential for the execution of the majority of algorithms. Accurate voice-activity detection techniques are necessary in order to accomplish this goal. It has been possible to make significant strides in the development of algorithms for noise estimation and voice activity detection in recent years. These algorithms are now able to constantly monitor the mean of the noise spectrum, at the very least. It is well known that algorithms for estimating noise work admirably in settings with a constant background noise level. As evidence of this, a little increase of 10% in intelligibility was noticed with speech processed in automotive contexts, but this was not the case with speech processed in other situations (e.g., babble). The constancy of the automobile noise, which made it possible to make reliable measurements of the noise, was thought to be the cause of the little improvement [12].

An accurate noise estimate may provide a contribution to increases in intelligibility, but it cannot, on its own, produce significant gains in intelligibility. This is because it is necessary to precisely monitor the spectrum of nonstationary noise. The fact that there is no discernible gain in intelligibility using the currently available speech enhancement algorithms is not just attributable to an inaccurate lack of estimations of the noise spectrum.

The single-channel speech augmentation scenario, in which only the distorted version of the original speech is available for recovery, is the one that poses the most challenge. When input signals are distorted as a result of background noise, the objective is to maximize the performance of voice communication systems while minimizing the influence of noise as much as possible.

A signal is said to be nonstationary if its power spectrum is constantly shifting over a number of distinct frequency bands. It may be difficult to determine if a portion of a nonstationary signal contains speech or noise for that portion of the signal. There are a great number of factors that contribute to the difficulties that often accompany voice recognition.

Natural speech is uninterrupted; in most cases, there are no gaps in between the individual words. As a result of this, among other things, it is more difficult to discern where the borders of the word are. Changes in global or local rates of speech, pronunciations of words within and among speakers, and phonemes in various contexts may all contribute to variations in natural speech. A large vocabulary can lead to misunderstandings. Variability in recorded speech may be attributed to factors such as the acoustics of the recording space, the characteristics of the channel and microphone [13], and the level of ambient noise. The updating of the estimate is only allowed during times when there is no speech when using traditional techniques of noise estimating, which are dependent on the detection of voice activity. In addition, VADs are notoriously difficult to tune, and their dependability is greatly compromised when dealing with speech components that are weak or when the input SNR is low. Therefore, the pace at which the noise estimate is updated is considered to be somewhat slow.

The estimates of the signal-to-noise ratio (SNR) and the noise powers are key difficulties in speech processing. The signal-to-noise ratio (SNR) is an important measure of the speech quality index that is used often in the process of data collection and classification. Additionally, speech improvement systems make use of the estimate of the local noise powers. Estimating the signal-to-noise ratio (SNR) and the local noise power of noisy speech may be very challenging in many situations since neither a clean reference signal nor speech activity is provided.

2. Survey of Related Work

In its most basic form, the procedure involves the processing of sound by use of an analog-to-digital converter (ADC), which converts analog sound signals into digital ones. In order to prepare the sound for subsequent processing, the technology removes noise and distortion. A sound signal is chopped up into smaller pieces at predetermined intervals and given a number. The algorithm that was created compares the produced phoneme to the framed phoneme that is utilized during the training phase. A number of different algorithms are used in order to carry out the pattern matching procedure for phonemes, words, and phrases. It is still a developing topic in terms of study areas [14], since the accomplishment of accuracy and performance is a tough problem. Word error rate, often known as WER, is the metric that is used to evaluate the performance of voice recognition systems. This metric detects words and creates mistakes that occur in the transcription representation. In recent years, there has been a rise in interest in "voice recognition technology," which is another name for the speech recognition system. However, throughout the last 65 years, this scientific field has seen significant change. SRS may be helpful in a variety of different real-time applications. The expansion of SRS will mostly take place in the following three areas: vocabulary size, speaker independence, and processing speed [15]. A method for recognizing single digits under the name "Audrey" was used. In the same year, Gunnar Fant created the source filter model for speech creation, which was the combination of vocal cords and an acoustic filter. This model was used to explain how speech is produced. Applications relating to voice proposed work and speech recognition found the source filter model to be helpful. Speech recognition technology known as "Shoebox" devices were first launched by IBM in 1962. These computers could understand 16 English words. At this point in time, Soviet researchers had constructed an analyzer for a vocabulary of 200 words and discovered the DTW (dynamic time warping) method. [16, 17] saw the beginning of the DARPA SUR (speech understanding research) project, which received funding from the Defense Advanced Research Projects Agency (DARPA) for a period of five years.

In 1987, a doll for kids named "Julie" was introduced, and it has the capability [18] of being taught by kids to reply to what they say. Dragon Dictate was the first consumer speech recognition tool for voice recognition call processing when it was originally made available by Dragon Company in the year 1990. In 1997, after making a few alterations, they came up with the idea for "Dragon NaturallySpeaking." In 2007, Google introduced its first product, GOOG-411, which was a directory that could be accessed through telephone. Siri, a virtual assistant developed by Apple and based on cloud computing, was made available to users in 2011 and responds to voice inquiries using a natural language user interface. Siri is able to answer questions or carry out tasks depending on the context. Google announced the availability of its voice search capabilities for iOS devices [19]. Since 2014 and continuing to the present day, the demand for virtual assistant services such as giving information and carrying out tasks has increased significantly. Amazon Alexa, Google Assistant, Amazon Echo, and Microsoft Cortana are just some of the most well-known examples of virtual assistants. Transcribing conversational telephony speech using several deep learning algorithms was one of the tasks that Microsoft researchers worked on in 2017 with the goal of achieving improved accuracy. Table 1 shows the existing methodology comparison.

The currently available voice recognition systems for Indian languages have reached a sufficient level of development and performance. But every system that is now in use has included an auditory and linguistic model. It was discovered that any system for English that is now available works solely on English characters and words. The researchers have not taken into account the most significant difficulties that pertain to digits, punctuation marks, and the end of utterance (EOU) [27]. It is far from exclusive that recurring components do not replace throughout the course of time while using DFT for the most significant part of the examination. As a consequence, the length of the window would not have an impact on the DFT outcomes, and banner houses hold as soon as from the beginning to the farthest limit of the window. Because it does not provide any measurements taken at the precise moment in time at which a repeat component takes place, a single DFT evaluation is often not sufficient for determining the significance of such markers [28]. STFT, also known as time-dependent Fourier transform (TDFT), is a notion that has been presented in order to govern such points. STFT is generally considered to be an important concept in conversation monitoring applications [29]. The goal of this project is to construct a model for voice recognition for the English language that takes into account English characters, numerals, and punctuation marks. To create a model that can identify the 'end of utterance' in speech, which is notoriously difficult to do. Furthermore, there is room for the development of the model described above, which has the potential to reduce the amount of time needed for the pattern finding process.

3. Methodology of Proposed Work

The following objectives will be met by the suggested works of art for supply primarily based on totally sound separation and request confirmation in IoT primarily based on fully sharp homegrown systems:

(1) To nurture new forms of important learning that are fundamentally founded on comprehensive supply segment computation for use in sound and request attestation

(2) To promote substantial learning primarily based on fully request test computation in voice-oversaw Internet of Things primarily based on utterly outstanding homemade situations

(3) To select how the suggested structure will be presented using a variety of sound combos and styles that will be taken into consideration

Table 1: Existing methodology comparison.

Ref.	Language	Techniques used	Category	Accuracy
[20]	Hindi	LVQ learning	Speech	95%
[21]	Hindi	ANN, RNN	Speech	94.63%
[22]	Assamese	CMU sphinx tool	Speech	93%
[23]	Assamese	HMM	Words	92.56%
[24]	Urdu	MFCC	Words	91.01%
[25]	Urdu	HTK	Speech	90.26%
[26]	Marathi	Sphinx 3 trainer	Speech	86.21%

(4) To dismantle the implementation of the suggested method regarding room acoustic homes

(5) To bring down the curtain on existing structures that are accessible for voice-controlled Internet of Things devices

3.1. Methodology. The improvements to the calculation of the supply sector's sound impact are included in the proposed changes. The sound dividers are mostly reliant on combination designs that count on the greatest possible massive component to fulfill the proposal ascribes of the space acoustics. It is predicted that the room inspiration response combination styles would be developed for the purpose of surveying the precise sound districts of the collector group area. Figure 2 shows the proposed work model.

3.1.1. Datasets. The sound barrier between the source and the receivers is determined by the direction of appearance (DOA) along the area of the receivers located within the room. The room's motivating response has an effect on the receiver's area as well as the sound it catches. Therefore, even when evaluating the framework's performance, such obstacles may be thought of. The suggested paintings have a variety of gestural elements, which are shown in discernment 1. The records and their respective warnings are first merged together, along with the concept of room motivation reaction. Because of the influence that the acoustic properties of the room have, the room pressure response is the most important factor that affects the sound symptoms and symptoms even as detecting and analyzing their individual roles in their individual sources. In spite of the fact that there are a number of different management hubs connected to the management, the original management is carried out according to the arrival direction (DOA). When compared to direct capture, room motivation response primarily based on fully mixing is dependable to develop nonstop acoustic results. This is because of the fact that it follows the guiding fundamental of amplifier display. In addition, the active mixing model, commotion effects, and sometimes the division of sound using the CBSS approach are detailed in this section.

3.1.2. Preprocessing Stage. The first thing that we do is provide the important rendition of convolutive mixes. A collection of N supply alerts, denoted by the notation $s(t) = (s1(t), \cdots, sN(t))$ are obtained from a set of M sensors at the discrete time interval t. The alerts that were sent out are suggested by the equation $x(t) = (x1(t), \cdots, xM(t))$. It is anticipated that the reassessment will be convolutive (or capably) combined in a few of relevant projects. The convolutive version provides the following connection between the mth contradictory message, the vital stock signals, and a pair of brought substance sensor disturbance $v_m s(t)$:

$$x_m(t) = \sum_{n=1}^{N} y \sum_{k=0}^{K-1} y\, a_{mnk} s_n(t-k) + v_m(t). \tag{2}$$

The going against message is a short sum of filtered assortments of each of the stock signs, and a_{mnk} addresses the related blending channel coefficients. The going against message may be thought of as a counterargument. In any case, for the sake of simplicity, we will assume that the coefficients will have a steady-to-combination variation. All things considered, those coefficients may also in like manner extrude throughout the span of time. In the statute, the channels are most likely of an unknowable length (which is most likely accomplished as IIR systems); despite this, when taken as a whole, it is more than adequate to anticipate that it will be K. It is possible that the convolutive adaptation will be formed as a result of the system structure:

$$x(t) = \sum_{k=0}^{K-1} y\, A_k s(t-k) + v(t), \tag{3}$$

where A_k is a M by N matrix that holds the kth filter coefficients and k is the index number of the matrix. The $M1$ noise vector is denoted by $v(t)$. It is possible to write the convolutive combination as follows in the z-domain:

$$X(z) = A(z)S(z) + V(z), \tag{4}$$

where $A(z)$ is a matrix that has FIR polynomials embedded inside each entry [22]. In special cases, there are certain unique applications of the convolutive mixture that may be simplified using equation (3). Assuming that all of the signals reach the sensors at the same time and are not filtered in any way, the convolutive mixture model may be simplified to

$$x(t) = As(t) + v(t). \tag{5}$$

The immediate or delay-less (straight) combination version is the name given to this variant. In this situation, $A = A0$ refers to a M by N grid that contains the mixing coefficients. In order to address the problem of the short aggregate, a number of computations were developed; for examples, see [15, 23]. In until further notice sources, assuming a resonance loose weather with engendering defers, the mixing version may be moved ahead to account for this.

$$x_m(t) = \sum_{n=1}^{N} a_{mn}\, s_n(t - k_{mn}) + v_m(t), \tag{6}$$

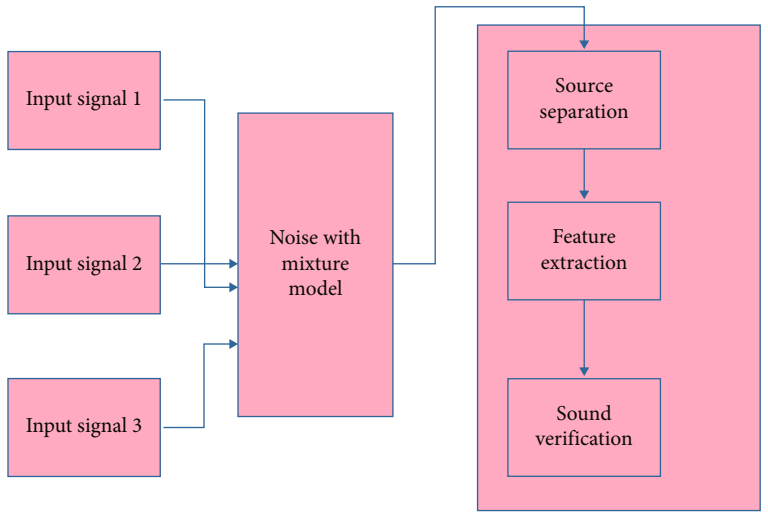


Figure 2: Block diagram of proposed work.

where k_{mn} denotes the time it takes for the signal to travel from source n to sensor m. In the process of deriving various algorithms, it is assumed that the convolutive model is noise-free, which means that there is no background noise in the model:

$$x(t) = \sum_{k=0}^{K-1} y\, A_k s(t-k). \tag{7}$$

In over and undersettled sources, frequently, it is miles expected that the quantity of sensors would draw near (or outperform) the quantity of reassessment, in which case, in an instant, strategies may work to reverse the quick blending. In any case, in the event that how much resets outperforms how many sensors the trouble is currently at this point not permanently set up, and surprisingly, underneath best realities at the blending system, methodologies cannot get well the resets to be composed withinside the repeat region quickly as fair-minded will increment for every repeat region:

$$X(\omega) = A(\omega)S(\omega) + V(\omega). \tag{8}$$

In most cases, the frequency transformation is calculated by using a discrete Fourier transform, often known as a DFT, within a time frame of length T, beginning at some point in time t:

$$X(\omega, t) = \mathrm{DFT}([x(t), \cdots, x(t+T-1)]), \tag{9}$$

and correspondingly for $S(!, t)$ and $V(1, t)$. Often a windowed discrete Fourier transform is used:

$$X(\omega, t) = \sum_{\tau=0}^{T-1} y\, w(\tau) x(t+\tau) e^{-j\omega\tau/T}, \tag{10}$$

where the window function $w(t)$ has been used because of the restricted temporal aperture in order to reduce the amount of band overlap. Because massive time-domain filters are commonly needed in acoustics, the fast Fourier transform (FFT) enables effective convolutions to be done in the discrete Fourier domain. This is crucial because convolutions may be done efficiently in the discrete Fourier domain. In frame blocking, the more generalized form of voice signal that is created by preprocessing and that will be used as input data will be called for. A series of speech frames is created from the input signal $s1(n)$, which represents the next occurrence of $s(n)$. A window will be formed out of the collection of speech frames. The steps involved in blocking and windowing are shown in Figure 3.

3.1.3. Feature Extraction. The decoding approach is used in order to locate the utterance feature vector that provides the greatest possible fit with the acoustic model. The process of decoding makes use of a dynamic programming method that is known as the Viterbi algorithm. The decoding step consists of modeling the acoustic data, the language data, and the pronunciation data. In continuous voice recognition, the system may identify an endless number of different sequences while trying to determine the optimum route of word sequence W for the input signal X. The Viterbi algorithm searches for the answer that is most plausible in light of the given speech. In deciphering looking for the word $W*$, it is possible to define it as in

$$W^* = \mathrm{argmax}_w(p(X|w)p(w)), \tag{11}$$

where $p(w)$ calculated from language model $p(X|w)$ can be calculated from available sequence of phonemes of words available in the dictionary in

$$p(X|w) = \mathrm{argmax}_s \pi\left(p\left(x|s_j\right)p\left(s_j\right)\right). \tag{12}$$

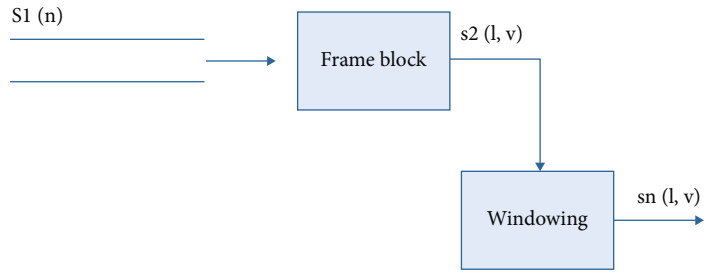


Figure 3: Preprocessing stage.

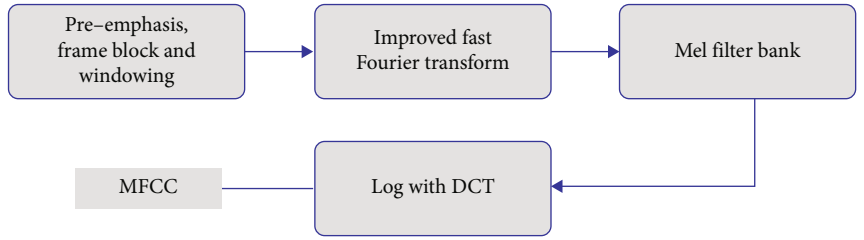


Figure 4: MFCC Architecture.

In MFCC computation, windowing process is performed by dividing speech signals into different frames. FFT algorithm is applied to calculate power spectrum for each frame. Filter bank processes the power spectrum using Mel-scale. Then, discrete cosine transform (DCT) is applied to speech signal to translate power spectrum to log domain to calculate MFCC coefficient as shown in Figure 4.

$$\text{mel}(f) = 2595x \log_{10}\left(1 + \frac{f}{700}\right), \tag{13}$$

where $\text{mel}(f)$ is the frequency and f represents frequency in (Hz). The MFCC calculation formula is

$$C_n = \sum_{k}^{n=1} (\log S\hat{}k) \cos\left[nk - \frac{1}{2}\frac{\pi}{k}\right]. \tag{14}$$

(i) The speech uttered by speaker in continuous form is entered as an input for the model which is shown in Figure 4

(ii) To differentiate voiced and unvoiced segment, threshold value is defined. If the uttered sound is less than the threshold value, it will be considered as unvoiced

(iii) For word detection process, from the voiced speech sequence of words is generated as $W(N)\bullet$. For each utterance of $W(i)$, feature vector is extracted

(iv) Pattern matching process is performed on $W(i)$, to match from the language model

(i) Word pattern can be words, digits, and punctuation mark. For word utterance, the best matched pattern gets it displayed in text format

(ii) Rule-based approach is implemented with the proposed model for digit and punctuation mark recognition

(iii) Rule-based approach stated that to append every digit pronunciation with "५:" to differentiate uttered digit with word sequences. Same approach can be used for punctuation mark representation that every punctuation mark pronunciation should end with "५:" to represent in symbol form and not in form of character sequence

(iv) If unknown word occurs, it matches with most relevant phoneme for the representation

Figure 5 shows the flowchart of the speech feature extraction model. The recommended model will perform an operation known as pattern matching on the uttered words. This operation is carried out on the spoken words. In the case that the said word does not locate a perfect match for itself, the algorithm will choose the most closely matched occurrence of that word. The model that has been suggested can take in data from a broad range of speakers with a diversity of accents and sound quality. For speech recognition to perform at its most efficient level, a significant quantity of training data is necessary. The gathering and preparation of a considerable quantity of data for training are carried out. There will be a total of eight different persons contributing their voices to the recording. These individuals will range in age, gender, and speaking style. The voice recognition model can accommodate a large number of users

Pseudocode of proposed methodology.
Determine the population size, the dimensions of the issue (NP, D), the crossover rate (c r), and the scaling factor (F).
Initialization: Establish a starting point for the population S i by setting its initial value to s $(1, i)t, s(2, i)t, \cdots, s(D, i)t$ with each individual being equally distributed in the range [s s"low," s"high,"]
Despite the fact that the requirements for termination are not fulfilled
In the population NP, for each unique goal vector, there is a total of NP
Using the following equation for mutation, choose three individuals at random from the population, and build a donor vector called *vit*.
$v(j, i)\hat{}t = s(j, p)\hat{}t + F\, i^{*}([\![\text{""}s]\!](j, r)\hat{}t + s(j, q)\hat{}t)$
Calculate the trial vector for the ith target vector using the following formula: $u(j, i)(t + 1)$:
$u(j, i)\hat{}t = \{\blacksquare(v(i, j)\hat{}t \,\&\text{"}\ \text{if"}\ \&\text{r i} \le \text{c r"}\ \text{or"}\ j = J\text{"}\ \text{rand"}\ @s(i, j)\hat{}\text{t\&\&"}\ \text{otherwise"})\dashv$
Apply the LSTM classifier as the fitness function f and then assess the *sit* and *uit* values as follows:
If $f(s\, i\hat{}t) \le f(u\, i\hat{}t)$ then $s\, i\hat{}(t + 1) = u\, i\hat{}t$
Else $s\, i\hat{}(t + 1) = s\, i\hat{}t$
Finish For
Finish While

Algorithm 1

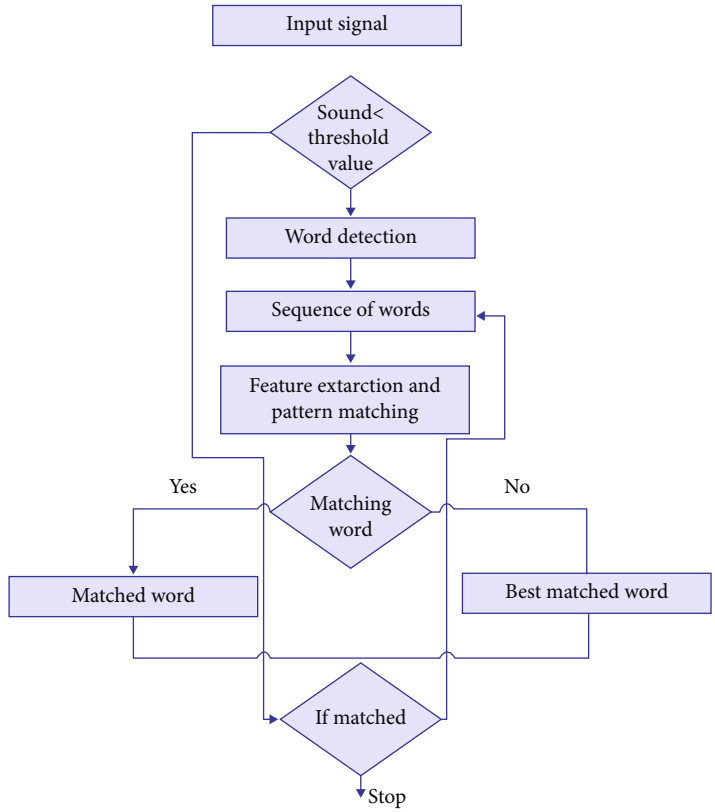


Figure 5: Flowchart of speech feature extraction.

simultaneously. The model has a threshold value that decides whether or not it will receive sound from a speaker. This value is determined by how loud the sound is. If the amount of voice in the sound is lower than the threshold value, then the sound will be classified as one that does not include any voice. The framework may be used in a different language by making certain fundamental adjustments to the way language is employed in its components. The model utilizes a method that is based on rules in order to properly represent numerals and punctuation marks. The framework may be used in a different language by making certain fundamental adjustments to the way language is employed in its components. The framework may be used in a different language by making certain fundamental adjustments to the way language is employed in its components. The graphical user interface that was built for the process of speech recognition is both user pleasant and includes all of the key interface functions. This section displays the textual

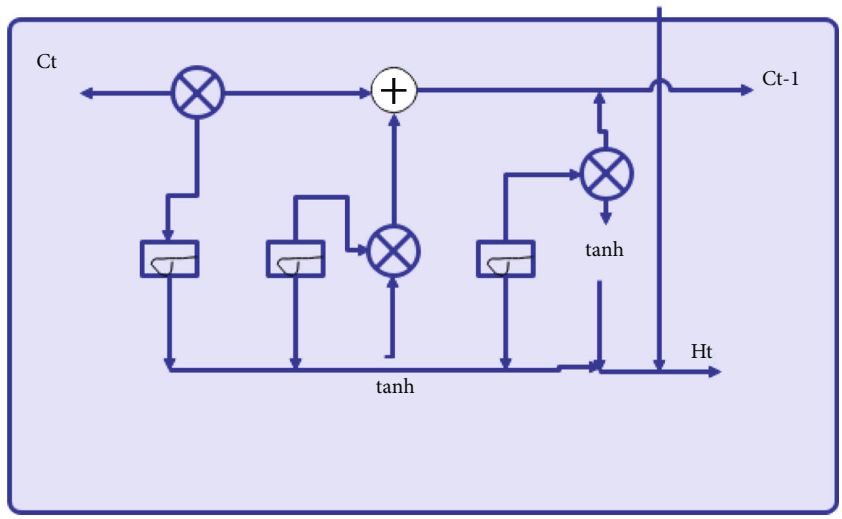


Figure 6: LSTM with GRU unit.

representation of the output that the recognizer generated. It is possible to generate a wide variety of output-oriented reports for the purpose of conducting further research and monitoring the recognition ratio and accuracy of the system. Even though the output is in the form of simple text, it is still possible to perform a number of text-based operations on it.

3.2. Classification. Candidate layer:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{15}$$

Input gate:

$$C\sim_t = \tan h(W_C \cdot [h_{t-1}, x_t] + b_c). \tag{16}$$

Output gate:

$$C_t = f_t * C_{t-1} + i_t * C\sim_t. \tag{17}$$

Hidden state:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o). \tag{18}$$

Memory state:

$$h_t = o_t * \tan h(C_t), \tag{19}$$

where W and b each represent a weight vector for the forget gate (f), the candidate (C), the input gate (I) and the output gate (o), respectively. The symbol "$*$" indicates element-by-element multiplication, while the "σ" symbol symbolizes the sigmoid function. The LSTM's internal structure is shown in Figure 6, which depicts the state of the system at time step t.

Even though LSTM solves the issue of vanishing gradients, gated recurrent unit (GRU), a generalized version of LSTM, was developed. GRU is a generalized variation of LSTM [4]. The GRU is a kind of unit that, like the LSTM unit, includes gating units that influence the flow of information inside the unit. However, unlike the LSTM unit, the GRU does not have distinct memory cells. The gated recurrent unit, or GRU, is responsible for calculating two gates referred to as the update gate and the reset gate. These gates are responsible for controlling the flow of information through each hidden unit. The following equations are used to determine the value of each hidden state at each time step t:

Update gate:

$$z_t = \sigma(W_z \cdot [h_{l-1}, x_t]). \tag{20}$$

Reset gate:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]). \tag{21}$$

New memory:

$$h\sim_t = \tan h(W \cdot [r_t * h_{t-1}, x_t]). \tag{22}$$

Final memory:

$$h_t = (1 - z_t) * h_{t-1} + z_t * h\sim_t, \tag{23}$$

where W denotes weight vector, * denotes element wise multiplication, and σ is the sigmoid function. Figure 6 demonstrates the internal structure of LSTM at time step t. Figure 6 shows the LSTM with GR unit.

LSTM calculates the hidden states by a set of equation as follows:

$$i = \sigma\left(x_t U^i + s_{t-1} W^i\right), \tag{24}$$

$$f = \sigma\left(x_t U^f + s_{t-1} W^f\right), \tag{25}$$

$$o = \sigma(x_t U^o + s_{t-1} W^o), \tag{26}$$

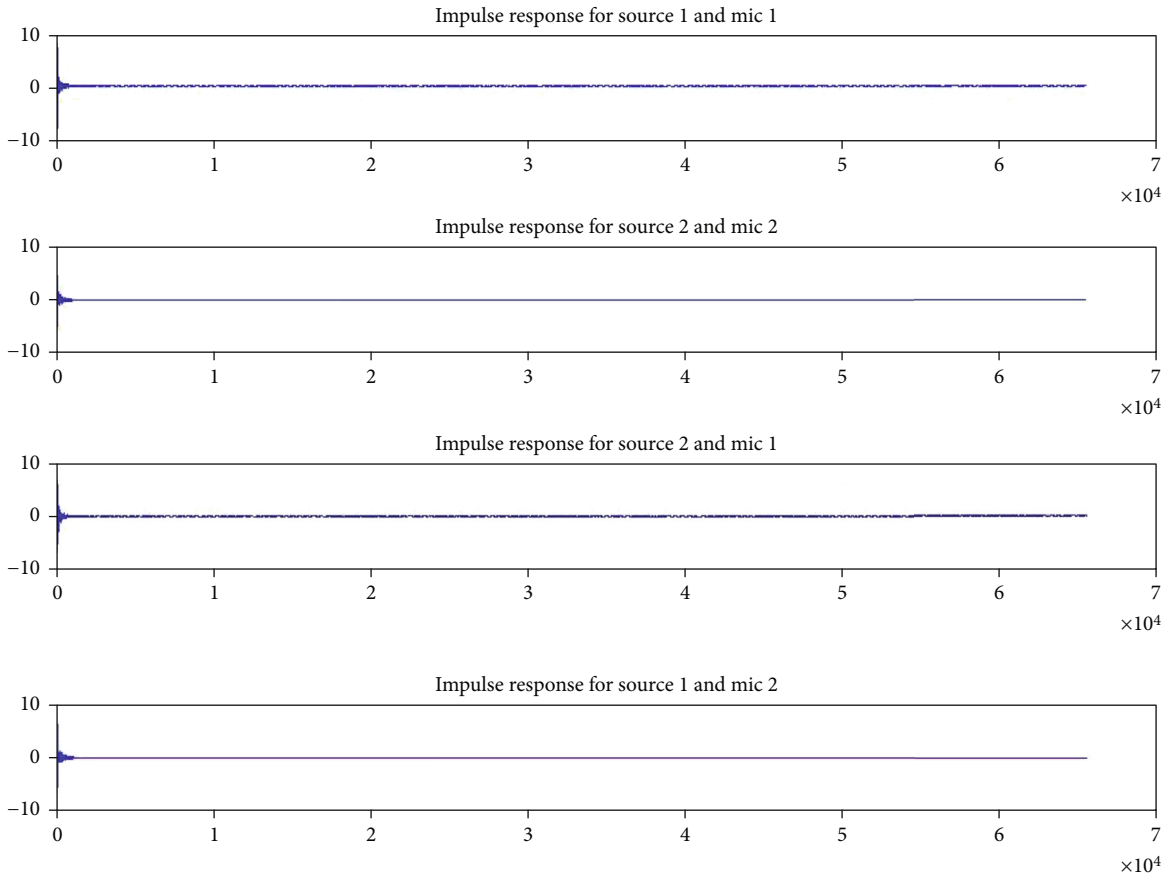


Figure 7: Room impulse response for node 1.

$$g = \tan h \tan h \left(x_t U^g + s_{t-1} W^g\right), \tag{27}$$

$$c_t = c_{t-1} \circ f + g \circ i. \tag{28}$$

Under these circumstances, the realities entrance, forget about entryway, provide doorway, and self-drawn-out independently are related with the letters I, f, and o, respectively, as seen in the previous sentence. Due to the fact that they provide the sense of being confused, the LSTM conditions might be discussed in more detail. I demonstrate how a large quantity of the spic and span measurements might possibly be supported by the memory mobileular. It is necessary for f to take responsibility for the insights, which need to be removed from mobile memory.

c.t. may be referred to on the grounds that it is the internal memory of the memory mobileular, which is the way a truckload element wise broadening of previous inward memory country through method of method for the brush to the side entrance and elementwise duplication of self-inauspicious country with enter entryway. At the end of the day, sh*t is related with the spine chiller country, which is now as of now not really chosen using the technique of approach for part proper duplication of the internal memory with the end-product entry. In addition, the end product that was not consumed will ultimately become part of the atmosphere as a result of the process outlined in Equation (27), which is comparable to Equation (28).

4. Results and Discussion

Vocabulary size A speech recognition system's vocabulary has an effect on its performance, which includes its accuracy and the amount of time it takes to complete recognition jobs. The needs of the system are what guide the development of the vocabulary in its entirety. One has the option of picking from one of four different groups, and this decision is based on how the system is being utilized.

The suggested system has undergone preliminary testing and error checking, both of which were carried out with the help of the Python programming language and its execution. The reaction that is produced as a result of the inspiration provided by the space is what establishes the genuine direction of the glance evaluation. Regarding the mouthpieces located at node 2 and the receivers located in center 1, the room inspiration response has been applied. As a consequence, for exploratory purposes and to account for the possibility of mistake, we have utilized combinations of the key inspiration response data. These centers observed for random causes in a room may have varying reactions to the power of the room. The plot of the room's inspiration vs. reaction may be seen in Figures 4 and 5, respectively.

(1) A vocabulary that is very restricted, consisting of little more than tens of words at most

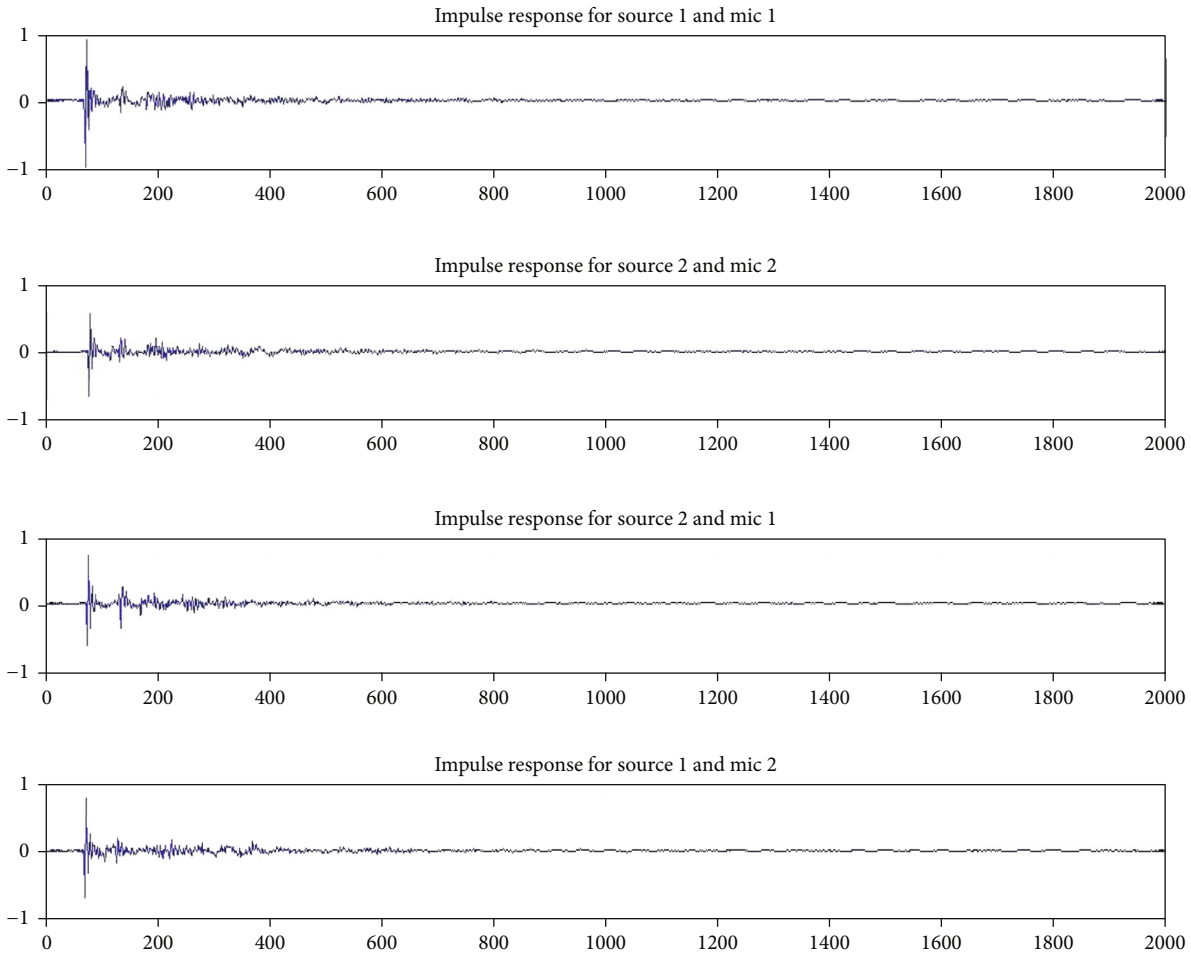


Figure 8: Room impulse response for node 2.

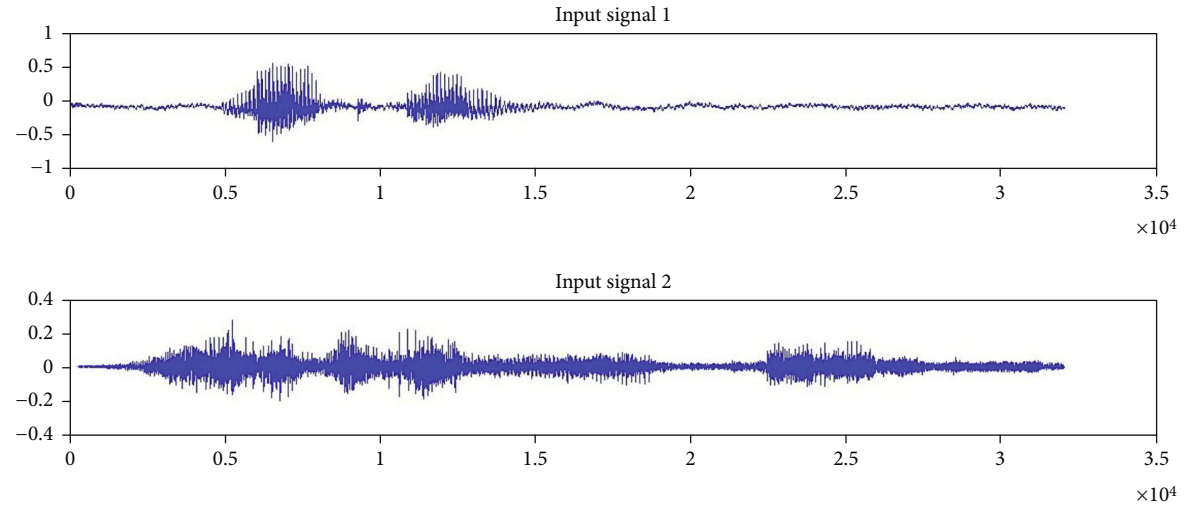


Figure 9: Input speech signals.

(2) A vocabulary of an intermediate level, comprised of hundreds of different terms

(3) A large vocabulary consisting of hundreds of different words

(4) A very extensive vocabulary that includes tens of thousands of words from a variety of sources Words that have the potential to be mixed up There are certain words in each language that, although having entirely distinct meanings, have the same sound.

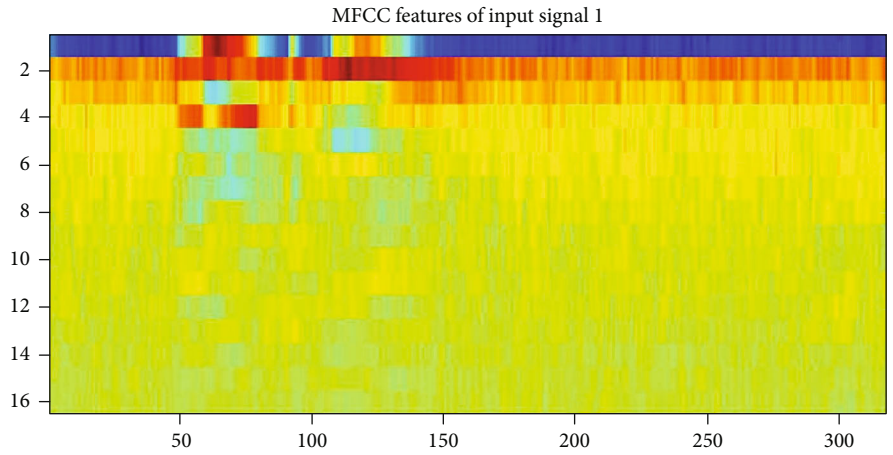


Figure 10: MFCC coefficient.

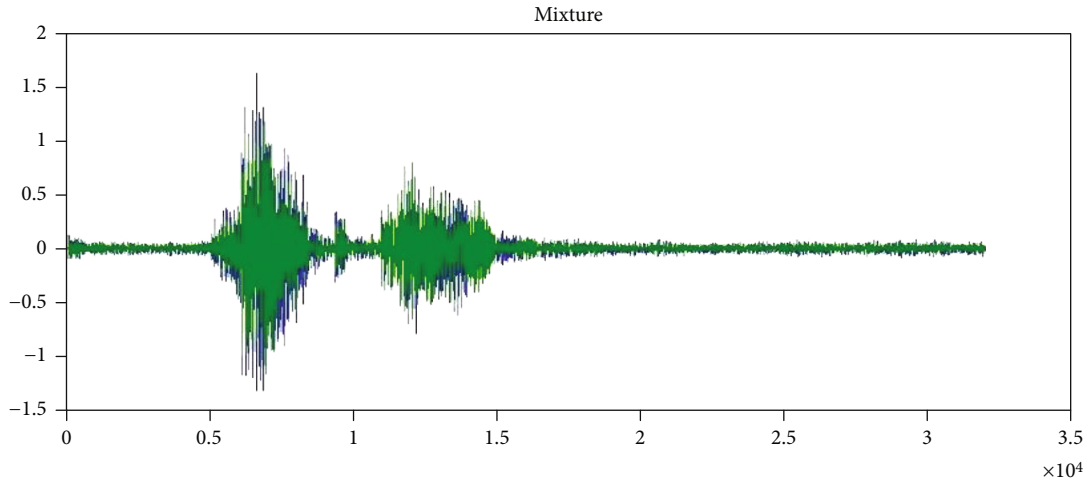


Figure 11: Mixed signal with respect to node 1 room impulse response.

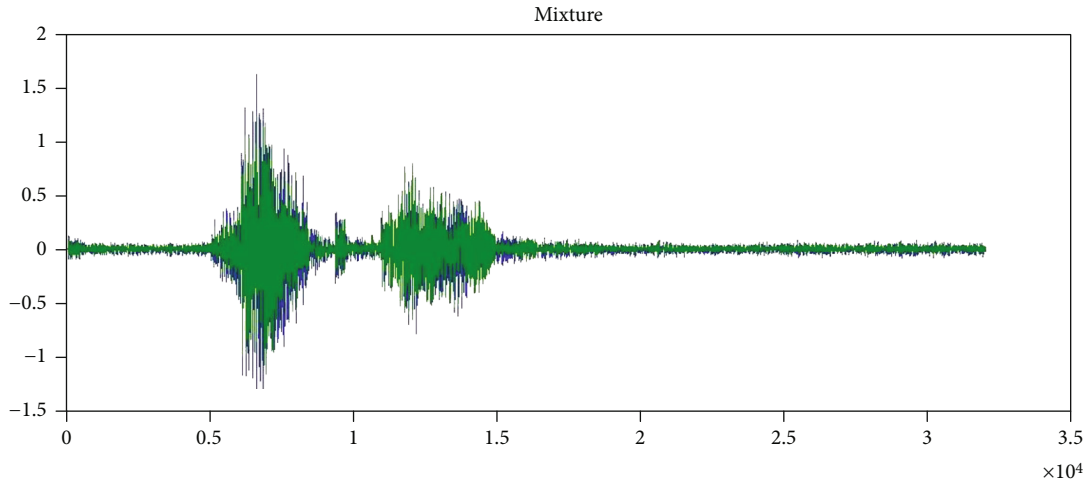


Figure 12: Mixed signal with respect to node 2 room impulse response.

These words might be misleading since they sound the same. The accuracy of the system suffers as a direct result of the confusion that is brought on by the usage of terminology of this sort. As an example, you may use ""and "", as well as ""and "", and so on. Speaker-dependent vs. speaker-independent: when

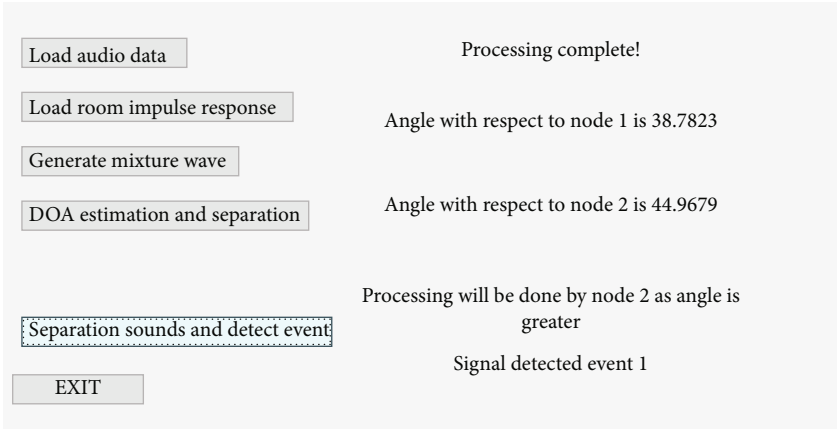


Figure 13: Result obtained for DOA and SVM-based event detection.

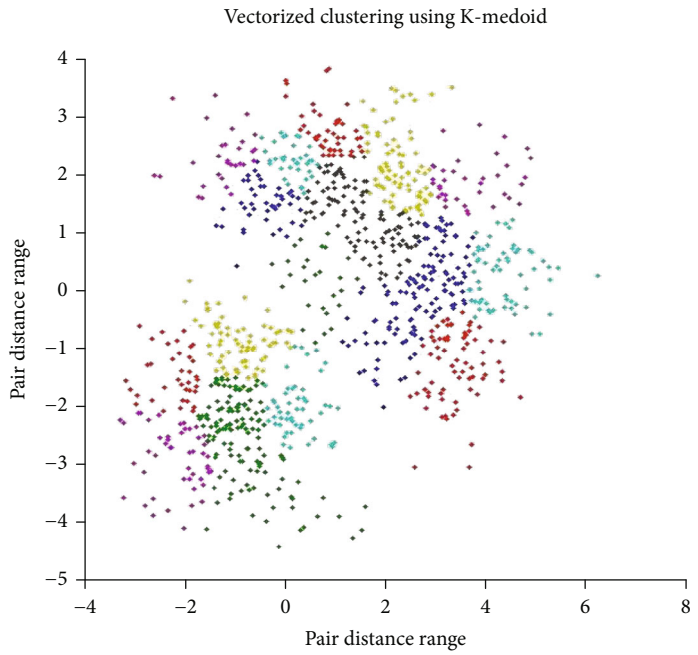


Figure 14: *K*-medoid clustering.

Table 2: Accuracy using machine learning methods.

Method	Accuracy	Specificity	Sensitivity
SVM	95.6%	0.91	0.89
KNN	93.2%	0.86	0.81
Naive Bayes (NB)	94.12%	0.89	0.82
Decision tree (DT)	94.05%	0.89	0.82

employing speaker-dependent systems, only certain speakers with constrained vocabularies are taken into consideration. A training phase must first be finished in order to save the speaker's utterances before the speech recognition system can be put into use. This phase must be finished before the system can be deployed. In comparison to the speaker-independent system, it has a greater degree of accuracy. A system that is independent of the speakers using it makes it possible for a large vocabulary to be utilized concurrently by several users. When compared to systems that are dependent on the speaker, the quantity of training data that is required here is far reduced. Isolated speech in addition to continuous and sporadic speech were both included. A better level of precision may be achieved by the use of individual words in the process of speech recognition when isolated speech is used. An example of discontinuous speech would be a string of words that are interspersed by silences in the course of a discussion. This kind of speech is easy to recognize since its boundaries are distinct. Continuous speech creates considerable difficulties in terms of real-time response as well as speech overlap, both of which have an impact on accuracy. These issues must be

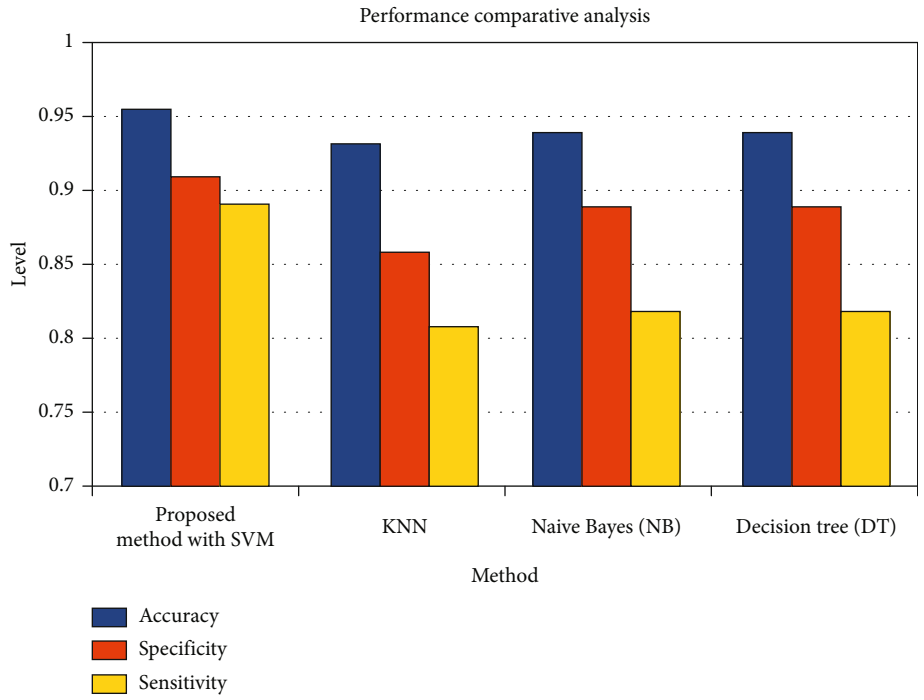


Figure 15: Graphical analysis of comparative methods.

Table 3: Accuracy using clustering and machine learning methods.

Method	Accuracy	Specificity	Sensitivity
SVM	96.80%	0.93	0.89
KNN	94.50%	0.89	0.82
Naive Bayes (NB)	94.12%	0.89	0.82
Decision tree (DT)	94.05%	0.89	0.82

Table 4: Deep learning-based model analysis.

Method	Accuracy	Specificity	Sensitivity
Model1	96.50%	0.91	0.9
Model2	96.80%	0.91	0.88
Model3	97.00%	0.89	0.89
Proposed model	98.50%	0.91	0.92

overcome. Taking notes on the circumstances, the circumstances under which the recording was made have an impact on the precision of the recognition. A loud setting, a broad channel bandwidth, the quality of the microphone, the speaker's dialect, and the speaker's speaking manner may all have a detrimental influence on the performance of a recognition system. Speech mistake: if the person speaking does not have a sufficient comprehension of the language or the topic at hand, then she or he will make mistakes in the pronunciation of specific words that are foreign to the person listening to them. The mental state of the speaker has an effect not only on the speaker's manner of speech but also on the data of their speech. Because of their worry, nervousness, and lack of self-confidence, the speaker uses a lot of unnecessary filler phrases and terminology that are hard to grasp, which brings the performance to a lower level

(i) Word Error Rate (WER)

The performance of speech recognition systems may be effectively tested using the standard measurement method, rate of errors in words (WER) [8]. The WER is an equation that indicates the system's degree of accuracy.

$$\mathrm{WER} = \frac{S + D + I}{N} * 100, \tag{29}$$

where S is the number of substitutions, D is the number of deletion, I is the number of insertion, and N is the number of words in the reference.

(ii) Real-Time Factor (RTF)

The real-time factor (RTF) is used to measure velocity parameter of speech recognition system. It is defined by the equation:

$$\mathrm{RTF} = \frac{P}{I}, \tag{30}$$

where P is the amount of time it takes to process input and I is the amount of time.

(iii) F-measure

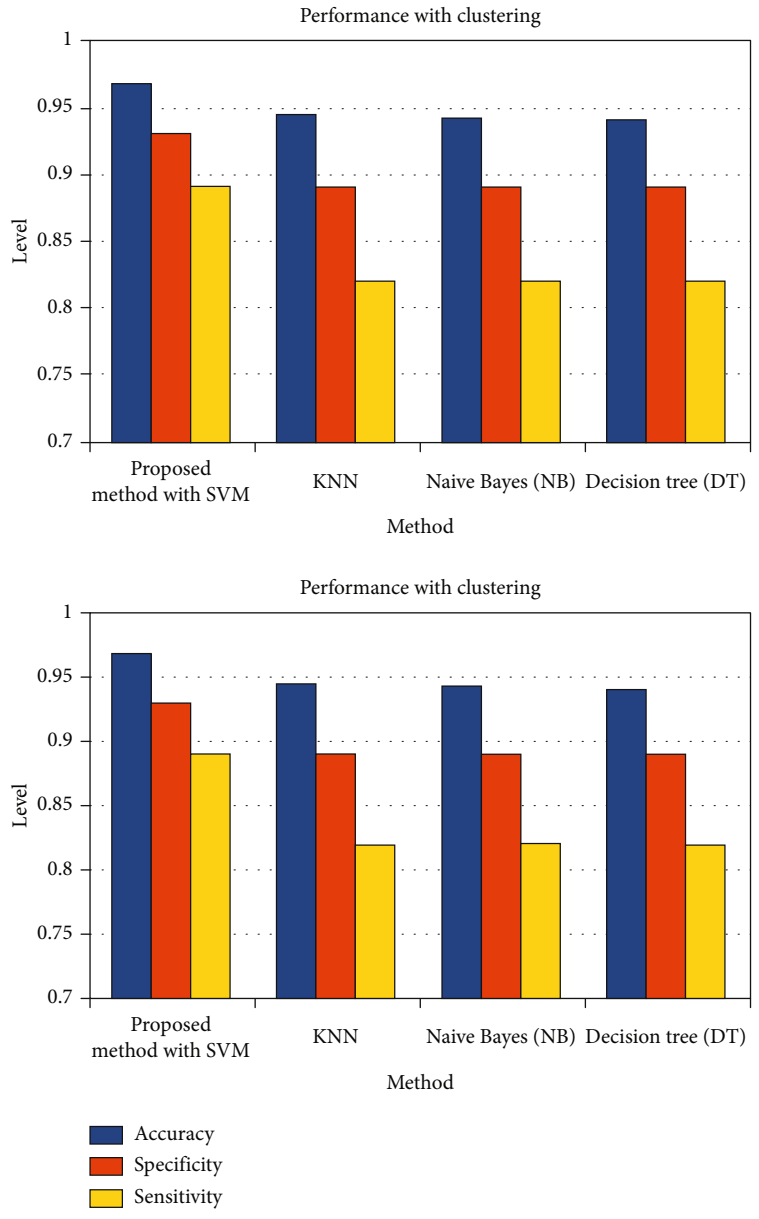


Figure 16: Graphical analysis of comparative methods.

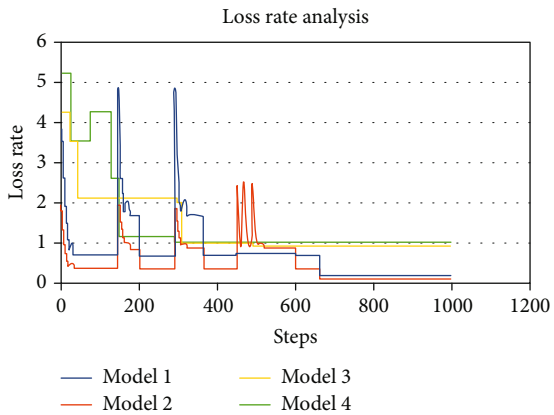


Figure 17: Loss rate analysis during training of DL models.

The F measure serves as a measure parameter in order to carry out performance analysis on the system. The F measure is often referred to as the $F1$ score or simply the F score. The weighted harmonic mean of the accuracy and recall of the test is what is used to define it.

The term "precision" refers to a measurement that compares the total number of words that were properly identified to the sum of the total numbers of correctly recognized words, substitutions, and insertions.

$$\text{Precision} = \frac{C}{(C+S+I)}. \tag{31}$$

Recall is a measure that depicts the total correctly recognized words to the sum of total numbers of correctly recognized words, substitutions, and deletions.

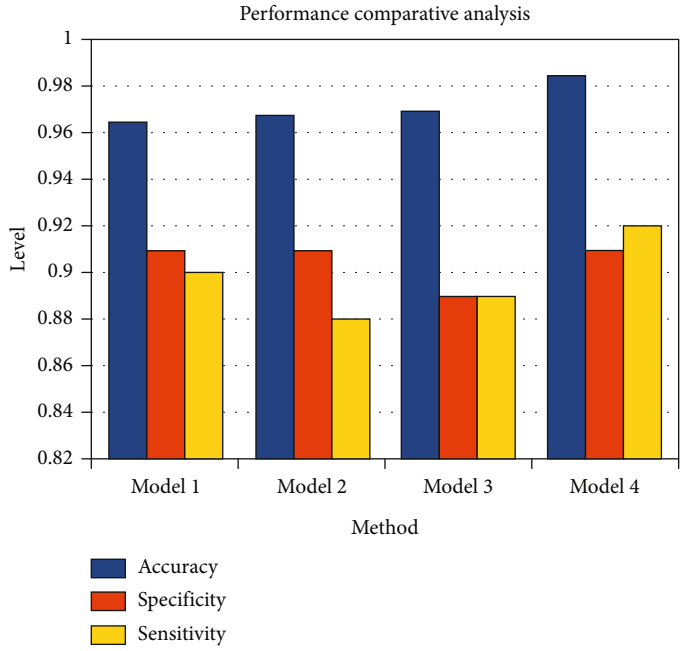


Figure 18: Comparative graph of performance parameters for deep learning models.

Table 5: Comparison of machine learning and deep learning methods.

Method	Accuracy	Specificity	Sensitivity
Model4 (proposed)	0.985	0.91	0.92
SVM	0.956	0.91	0.89

Table 6: Comparison of machine learning and deep learning methods.

Method	Accuracy	Specificity	Sensitivity
Model4	0.985	0.91	0.92
Clustering+model 4 (proposed)	0.99	0.95	0.92

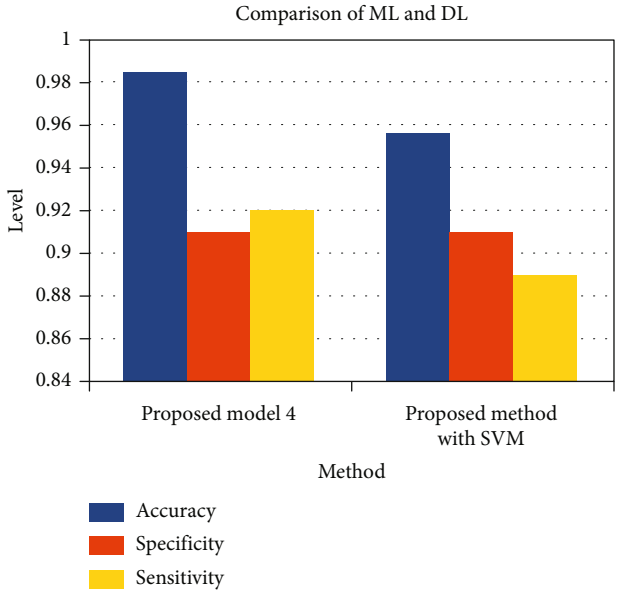


Figure 19: Comparison of ML and DL methods.

4.1. Signal Mixing and Feature Extraction. Using execution with the Python program, the preliminary testing and error checking for the proposed system has been completed. The true orientation of the glance assessment is determined by the response that is caused by the room's inspiration. Regarding the receivers of center 1 and the mouthpieces of node 2, the room inspiration response is implemented. These centers seen for random reasons in a room might have various reactions to the room's power, and as a result, for preliminary purposes and to account for potential errors, we have used mixtures of the essential inspiration reaction data. The room inspiration response plot is shown in Figures 7–9.

Figure 6 shows the input speech signals used for experimentation. The MFCC coefficient are presented in the spectrum graph as shown in Figure 7. The mixed signal is obtained based on room impulse response as shown in Figure 8.

The DOA estimation and SVM-based event detection is shown in Figure 9. As the angle with respect to node 2 is greater, further processing will be done by node 2. The sound separation using CBSS based on DOA is used to detect the estimate using MFCC coefficients. The resulting event is shown in the result (Figure 10). K-medoid clustering is applied on dataset features. The clustered data is plotted in a scatterplot as shown in Figure 11. Figure 12 shows the node 2 room impulse response, and Figure 13 shows the results obtained from the SVM classifier.

Figure 14 shows the K-means clustering scattered plot. Table 2 shows the accuracy comparison with the proposed and existing machine learning algorithms.

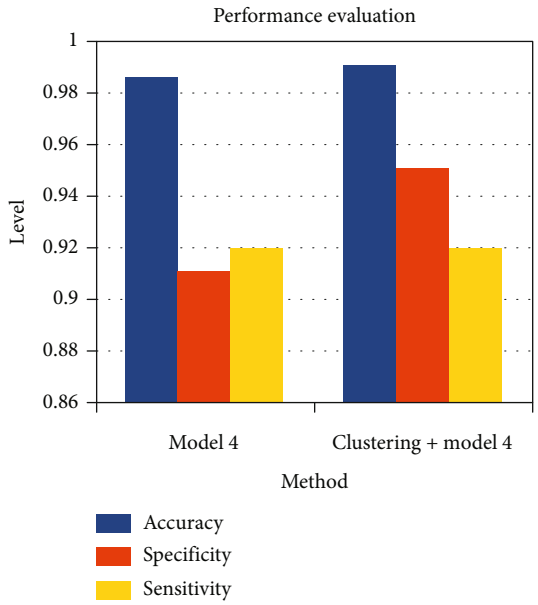


Figure 20: Comparison of clustering and without clustering methods.

Figure 15 shows the graphical analysis of the existing and proposed methodologies. Table 3 represents the performance metrics comparison.

K-nearest neighbor, naive Bayes (NB), and decision tree (DT) classifiers are considered for comparison. The values of accuracy, specificity, and sensitivity are obtained using formulae shown in Table 4. Figure 5 shows the graph of comparative analysis from which SVM results can be seen as better in terms of accuracy. The improved convergence for SVM model for different types of commands is the main reason for improved performance of the SVM model.

The training is done for 100 epochs with 10 steps per epoch using the dataset [12]. The loss rate for the training steps is tracked, and a loss analysis graph is plotted in Figure 16.

The results of the accuracy of classification for the identification of flawless instructions are studied in terms of accuracy, specificity, and sensitivity. Table 4 compares the scores of four different models, and Figure 17 depicts a graphical representation of the performance parameters. The accuracy, specificity, and sensitivity values are calculated using the formulas in Table 4.

Figure 18 shows the performance metrics of the deep learning models.

The performance obtained using machine learning and deep learning are compared as shown in Table 5 and graph in Figure 19. The comparative of use of clustering is also considered for comparison as shown in Table 6 and Figure 20. The performance of deep learning method is seen well.

5. Conclusion

Voice commands may be used to operate a wide range of devices that are connected to the internet of things network. It is possible to correctly identify orders coming from a single source in the absence of any other sounds that would be considered noise. When there are a large number of sound-producing sources in the surroundings around you, it is possible that the ambient noise will get muddled with the command voices. A number of different processing processes are required in order to separate speech instructions from a mixed sound stream. This study effort covers the separation and classification of command from mixed sound signal for the purpose of ensuring that the proper activity may be done through voice-controlled devices. The process consists of developing a database of mixed sound signals, which may contain a variety of verbal instructions as well as the noises of dogs, cats, or even television shows such as the news or other entertainment options. The evaluation of additional potential sources of noises can call for a large amount of the collection. The study work that was done demonstrates the noises that are usually regarded to originate from a variety of sources as well as their mixing in relation to the acoustic qualities of the space. The model that is able to accurately detect the command while simultaneously extracting the appropriate features. The initial step of the task consists of preparing a mixed sound dataset in such a way that the instructions that fall under a certain class may be appropriately identified. In the system that has been presented, the database that has been produced already has a number of instructions that have been gathered and segregated for the supervised training of the models. The second step of the work is comprised of the traditional blind source separation approach, which focuses its attention on the direction of arrival (DOA) of the sound signals. This is done while the system is being performed and proposed. Based on the angle of arrival estimate, a specific room impulse response (RIR) is regarded from a set of specified RIR as a room acoustic feature, and source separation is carried out using the independent component analysis (ICA) approach. The processed signals of the split command speech are then subjected to feature extraction. The Mel-frequency cepstral coefficient (MFCC) approach is used in order to carry out the process of feature extraction. Comparisons are made between the performance of the SVM classifier and that of other classifiers that are often used in machine learning applications, such as decision trees (DT). The effectiveness of the SVM classifier is evaluated, and the results reveal that it has an accuracy of 91%. Utilizing a classifier that is based on a recurrent neural network (RNN) helps to increase the accuracy of further categorization. This particular classifier is made up of three layers: one layer of gated recurrent units (GRU), one layer of long short-term memory (LSTM), and one layer of both. This classification seems to have an accuracy of 94.5 percent, which is higher than the SVM classifier.

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The author declare that he/she has no conflicts of interest to report regarding the present study.

References

[1] T. Ammari, J. Kaye, J. Y. Tsai, and F. Bentley, "Music, search, and IoT," *ACM Transactions on Computer-Human Interaction*, vol. 26, no. 3, pp. 1–28, 2019.

[2] D. Pal, C. Arpnikanondt, S. Funilkul, and W. Chutimaskul, "The adoption analysis of voice-based smart IoT products," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10852–10867, 2020.

[3] G. Muhammad, S. M. M. Rahman, A. Alelaiwi, and A. Alamri, "Smart health solution integrating IoT and cloud: a case study of voice pathology monitoring," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 69–73, 2017.

[4] P. Ni, Y. Li, G. Li, and V. Chang, "Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction," *Neural Computing and Applications*, vol. 32, no. 20, pp. 16149–16166, 2020.

[5] G. Alexakis, S. Panagiotakis, A. Fragkakis, E. Markakis, and K. Vassilakis, "Control of smart home operations using natural language processing, voice recognition and IoT technologies in a multi-tier architecture," *Designs*, vol. 3, no. 3, p. 32, 2019.

[6] A. Valera Román, D. Pato Martínez, Á. Lozano Murciego, D. M. Jiménez-Bravo, and J. F. de Paz, "Voice assistant application for avoiding sedentarism in elderly people based on IoT technologies," *Electronics*, vol. 10, no. 8, p. 980, 2021.

[7] A. F. Klaib, N. O. Alsrehin, W. Y. Melhem, and H. O. Bashtawi, "IoT smart home using eye tracking and voice interfaces for elderly and special needs people," *The Journal of Communication*, vol. 14, no. 7, pp. 614–621, 2019.

[8] S. Uma, R. Eswari, R. Bhuvanya, and G. S. Kumar, "IoT based voice/text controlled home appliances," *Procedia Computer Science*, vol. 165, pp. 232–238, 2019.

[9] M. Ali, "Developing applications for voice enabled IoT devices to improve classroom activities," in *In 2018 21st International Conference of Computer and Information Technology (ICCIT)*, pp. 1–4, IEEE, Dhaka, Bangladesh, (2018, December).

[10] H. Isyanto, A. S. Arifin, and M. Suryanegara, "Design and implementation of IoT-based smart home voice commands for disabled people using Google Assistant," in *In 2020 International Conference on Smart Technology and Applications (ICoSTA)*, pp. 1–6, IEEE, Surabaya, Indonesia, 2020.

[11] K. M. Malik, A. Javed, H. Malik, and A. Irtaza, "A light-weight replay detection framework for voice controlled IoT devices," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 982–996, 2020.

[12] Y. Meng, H. Zhu, J. Li, J. Li, and Y. Liu, "Liveness detection for voice user interface via wireless signals in IoT environment," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 6, pp. 1–3011, 2020.

[13] P. J. Rani, J. Bakthakumar, B. P. Kumaar, U. P. Kumaar, and S. Kumar, "Voice controlled home automation system using natural language processing (NLP) and internet of things (IoT)," in *In 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)*, pp. 368–373, IEEE, Chennai, India, 2017, March.

[14] CNET, "How to bring Alexa into every room of your home," 2017, https://www.cnet.com/how-to/how-to-install-alexa-in-every-room-of-your-home/.

[15] E. M. Mugler, M. C. Tate, K. Livescu, J. W. Templer, M. A. Goldrick, and M. W. Slutzky, "Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri," *Journal of Neuroscience*, vol. 38, no. 46, pp. 9803–9813, 2018.

[16] C. P. Browman and L. Goldstein, "Articulatory phonology: an overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[17] X. Lei, G. Tu, A. X. Liu, C. Li, and T. Xie, "The insecurity of home digital voice assistants - Amazon Alexa as a case study," 2017, Available: http://arxiv.org/abs/1712.03327.

[18] X. Lei, G. Tu, A. X. Liu, C. Li, and T. Xie, "The insecurity of home digital voice assistants - vulnerabilities, attacks and countermeasures," in *2018 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–9, Beijing, China, May 2018.

[19] J. Tan, C. Nguyen, and X. Wang, "SilentTalk: lip reading through ultrasonic sensing on mobile phones," in *IEEE INFOCOM 2017- IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, May 2017.

[20] Y. Meng, W. Zhang, H. Zhu, and X. S. Shen, "Securing consumer IoT in the smart home: architecture, challenges, and countermeasures," *IEEE Wireless Communications*, vol. 25, no. 6, pp. 53–59, 2018.

[21] W. Zhang, Y. Meng, Y. Liu, X. Zhang, Y. Zhang, and H. Zhu, "HoMonit: monitoring smart home apps from encrypted traffic," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 1074–1088, New York, NY, United States, 2018.

[22] Y. Zhang, R. Deng, D. Zheng, J. Li, P. Wu, and J. Cao, "Efficient and robust certificateless signature for data crowdsensing in cloudassisted industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 1–9, 2019.

[23] X. Yuan, Y. Chen, Y. Zhao et al., "CommanderSong: a systematic approach for practical adversarial voice recognition," in *Proceedings of 27th USENIX Security Symposium (USENIX Security)*, pp. 49–64, Baltimore, MD, 2018.

[24] Y. Chen, J. Sun, R. Zhang, and Y. Zhang, "Your song your way: rhythm-based two-factor authentication for multi-touch mobile devices," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, pp. 2686–2694, Hong Kong, China, April 2015.

[25] 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 IEEE Symposium on Security and Privacy (S & P)*, pp. 582–597, San Jose, CA, USA, 2016.

[26] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (ACM MobiCom)*, pp. 90–102, New York, NY, United States, 2015.

[27] Y. Liu, "WiVo: enhancing the security of voice control system via wireless signal in IoT environment," in *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 81–90, New York, NY, United States, 2018.

[28] P. Associates, "Top 10 consumer iot trends in 2017," 2017, http://www.parksassociates.com/whitepapers/top10-2017.

[29] Amazon, "Amazon alexa developer," 2019, https://developer.amazon.com/alexa.

Hindawi
Journal of Sensors
Volume 2023, Article ID 9210450, 12 pages
https://doi.org/10.1155/2023/9210450

Research Article

Field Validation of a Public Health Particulate Matter (PHPM) Low-Cost $PM_{2.5}$ Monitor and Commercial Sensors with Light Scattering Technology

Pornpun Sakunkoo,[1] Patcharintorn Kaewkun,[1] Watis Leelapatra,[2] Naowarat Maneenin,[3] Jetnapis Rayubkul,[4] Ratchaphon Suntivarakorn,[5] Chananya Jirapornkul,[3] and Sarawut Sangkham[6]

[1]*Department of Environmental Health Occupational Health and Safety, Faculty of Public Health, Khon Kaen University, Muang District, Khon Kaen 40002, Thailand*
[2]*Department of Computer Engineering, Faculty of Engineering, Khon Kaen University, Muang District, Khon Kaen 40002, Thailand*
[3]*Department of Epidemiology and Biostatistics, Faculty of Public Health, Khon Kaen University, Muang District, Khon Kaen 40002, Thailand*
[4]*Faculty of Public Health, Khon Kaen University, Muang District, Khon Kaen 40002, Thailand*
[5]*Department of Mechanical Engineering, Faculty of Engineering, Khon Kaen University, Muang District, Khon Kaen 40002, Thailand*
[6]*Department of Environmental Health, School of Public Health, University of Phayao, Muang District, 56000 Phayao, Thailand*

Correspondence should be addressed to Chananya Jirapornkul; chananya@kku.ac.th and Sarawut Sangkham; sarawut.sa@up.ac.th

Received 25 April 2022; Revised 23 September 2022; Accepted 26 November 2022; Published 23 February 2023

Academic Editor: Akhilesh Pathak

Low-cost $PM_{2.5}$ sensors are the key to creating high-resolution monitoring networks for localized reporting and increased public awareness. A low-cost $PM_{2.5}$ sensor using light-scattering technology Public Health Particulate Matter (PHPM) was developed and validated against measurements from a commercial device that reports real-time aerosol concentration and also collects gravimetric samples (DustTrak DRX Aerosol Monitor 8533). Linear regression of measurements from a controlled indoor site and three ambient air sites found that the novel $PM_{2.5}$ sensor correlated well with the measured $PM_{2.5}$ concentrations ($r \geq 0.90$) at the three ambient air locations with $PM_{2.5}$ concentrations ranging from 48 to 602 $\mu g/m^3$ measured at 30-minute intervals. The correlation was lower ($r = 0.71$) at the chamber, which used incense as a particulate matter source. The PHPM data of $PM_{2.5}$ concentrations was using multiply with 2.33. The hope is that the PHPM device can provide an additional tool for participatory in community and increased public awareness about $PM_{2.5}$ in Thailand. This PHPM device is more suitable for ambient air than the chamber. The PHPM device is inexpensive, portable, and can be charged with a power bank.

1. Introduction

Particulate matter air pollution with an aerodynamic diameter less than 2.5 microns ($PM_{2.5}$) is recognized as a global health threat that requires prioritized research [1]. While efforts to monitor ambient air quality have increased substantially across the globe, significant gaps remain, with many parts of the world having no regular $PM_{2.5}$ monitoring [2]. Localized measurements of $PM_{2.5}$ are important due to the high variability in ground level air pollution caused by complex interactions of topography [3], air patterns [4], and emission sources [5]. Two major barriers, however, to widespread implementation of $PM_{2.5}$ monitoring are the cost and technical requirements of high-accuracy equipment that relies on filtration and gravimetric mass determination [6]. Furthermore, gravimetric methods typically generate data with 24-hour measurement periods, while higher temporal resolution is preferable for real-time monitoring,

(a) (b)

Figure 1: Particulate matter monitoring devices used. (a) Low-cost particle scattering $PM_{2.5}$ sensor (PHPM) and (b) DustTrak DRX Aerosol Monitor 8533.

Table 1: Characteristics of SI, SN, and SV type dust measuring instruments used in research.

Characteristics	Instrument type SI	SN	SV
Approximate price (US dollars)	66.6	50	50
Packaging size (mm)	78 × 84 × 52	75 × 75 × 46	60 × 60 × 26
Product weight (g)	251	93	86
Estimated PM_x concentration	$PM_{2.5}$	$PM_{2.5}$	$PM_{2.5}$
Measurement of $PM_{2.5}$ concentration			
(i) Measuring range ($\mu g/m^3$)	0-999	—	0-500
(ii) Resolution ($\mu g/m^3$)	1	—	1
(iii) Accuracy ($\mu g/m^3$)	±15%, 10	—	±10%@100 ~ 500 $\mu g/m^{3/}$ ± 10 $\mu g/m^3$@0 ~ 100 $\mu g/m^3$
Measurement method	Laser scattering	Laser scattering	PLANTOWER laser particulate matter sensor
Display	—	—	—
Limitation	(i) Cannot be used in environments with high concentrations of $PM_{2.5}$ or greater than 999 $\mu g/m^3$ (ii) Cannot be used in an environment with too high temperatures (iii) Cannot be used in a humid environment	(i) Uses medical grade low current laser transmitter (ii) Uses temperature and humidity compensation algorithm which ensures the accuracy of product testing (iii) Uses a special air duct design by air intake every 30 seconds is 283 mL (iv) Use heat dissipation design which ensures that the laser transmitter 3000 h (v) Cannot be used in an environment with excessive temperature	(i) Cannot be used in environments with high concentrations of $PM_{2.5}$ or greater than 500 $\mu g/m^3$ (ii) Cannot be used in an environment with too high temperatures (iii) Cannot be used in an environment with too high temperatures (iv) Cannot be used in a humid environment

Table 2: Characteristics of SC, CKDPM, and PHPM type dust measuring instruments used in research.

Characteristics	Instrument type SC	CKDPM	PHPM
Approximate price (US dollars)	23.3	66.6	23.0
Packaging size (mm)	$60.3 \times 21.4 \times 84.3$	$48 \times 37 \times 12$	$15 \times 10 \times 5$
Product weight (g)	86	180	150
Estimated PM_x concentration	$PM_{2.5}$	$PM_{2.5}$, PM_{10}, PM_1	$PM_{2.5}$
Measurement of $PM_{2.5}$ concentration			
(i) Measuring range ($\mu g/m^3$)	—	0-500	0-500
(ii) Resolution ($\mu g/m^3$)	—	1	1
(iii) Accuracy ($\mu g/m^3$)	—	±10%@100 ~ 500 $\mu g/m^3$/± 10 $\mu g/m^3$@0 ~ 100 $\mu g/m^3$	±10%@100 ~ 500 $\mu g/m^3$/± 10 $\mu g/m^3$@0 ~ 100 $\mu g/m^3$
Measurement method	—	Laser scattering	Laser scattering
Display	LED digital display	LED digital display	LED digital display
Limitation	(i) Cannot be used in an environment with too high temperatures (ii) Cannot be used in a humid environment	(i) Cannot be used in environments with high concentrations of $PM_{2.5}$ or greater than 500 $\mu g/m^3$ (ii) Cannot be used in an environment with too high temperature (iii) Cannot be used in a humid environment	(i) Cannot be used in environments with high concentrations of $PM_{2.5}$ or greater than 500 $\mu g/m^3$ (ii) Cannot be used in an environment with too high temperature (iii) Cannot be used in a humid environment

Note: CKDPM type dust detector is a dust meter supported by the Chronic Kidney Disease Prevention and Slowing Project in the Northeast (CKDNET) sampler.

highly sensitive for the detection of glucose concentrations in human bodies [7], and biosensor has been proposed for detection of uric acid (UA) in human serum with a highly sensitive and selective optical fiber-based enzymatic [8]. In response, light scattering techniques (e.g., Chowdhury et al. [9]) have been developed that build the concept of optical particle counters and Mie's theory [10]. These light-scattering sensors are inexpensive and easy to operate.

Several low-cost particulate matter sensors have been developed and validated across the world [6, 9]. These sensors have become critical tools for providing localized air quality readings and raising awareness of particulate matter pollution [11]. Advances in computing power and power consumption have made low-cost sensing devices an important supplement to the granular network of regulatory stations established around the globe and help to engage local communities with air quality data [12]. Generally, these low-cost devices can quite accurately measure $PM_{2.5}$ concentrations under most conditions [13, 14]. However, the accuracy of such low-cost air sensors has been questioned, and the need for calibrating these devices is critical [15]. However, given the knowledge gaps, the aim of this study was to focus on estimating $PM_{2.5}$ concentrations using low-cost PHPM sensors in real-world situations. Therefore, calibrating and validating low-cost $PM_{2.5}$ sensors built on light-scattering technology is necessary before implementation in the community. This similarity metric can also guide the adoption of low-cost monitors and/or data selection in future field research as well as spatiotemporal [16]. While $PM_{2.5}$, which results from vehicle exhaust, biomass burning, and industrial emissions, has received substantial publicity and attention in Thailand [17], high-resolution monitoring of $PM_{2.5}$ is still limited in Thailand [18]. Widespread local data, however, is important. Attitudes, awareness, and risk perception affect the public's proenvironmental behaviors that influence $PM_{2.5}$ [19, 20].

Figure 2: For example, sampling locations among Faculty of Public Health area with operated device.

Thus, increasing awareness and publicity of $PM_{2.5}$ concentrations at the local level is important. Accordingly, the objectives of this research were to develop a simple, low-cost device for real-time reporting of $PM_{2.5}$ concentrations in the ambient air and to calibrate the device by determining the relationship between the device's readings and a standardized commercial aerosol sensor. This device is intended for widespread implementation at local health centers to

Figure 3: The sensor used to measure dust up to 2.5 microns in size according to this invention.

raise awareness about $PM_{2.5}$ at the local level, as access to real-world $PM_{2.5}$ data is essential for more effective and safe assessment of people's health risks and self-protection.

2. Materials and Methods

2.1. Instruments. A low-cost $PM_{2.5}$ sensing device Public Health Particulate Matter (PHPM) was developed for use in monitoring ambient air $PM_{2.5}$ concentrations at local health stations or other local agencies (Figure 1(a)). The PHPM uses the light scattering technique to measure particles suspended in the air. A red laser illuminates through a cavity with active ventilation provided by a fan. A photodiode measures the scattered light to determine the mass of particles according to Mie's theory. These readings were then converted to $PM_{2.5}$ concentrations by researcher. The estimates from the PHPM device were compared against a standardized commercial product, the DustTrak DRX Aerosol Monitor 8533 (TSI International, Shore-view, Minnesota, USA) (Figure 1(b)). This device can provide real-time aerosol concentration data, while simultaneously measuring both mass and size fraction. While also using a light-scattering laser photometer, it also collects gravimetric samples. These DustTrak monitors have been used to validate or calibrate several low-cost devices for both indoor and outdoor aerosol measurement [21, 22]. The estimates from the PHPM device were compared with the commercial sensors with light scattering technology which include SI, SN, SV, SC, and CKDPM. All commercial devices are shown in Tables 1 and 2.

2.2. Sampling Locations. The two instruments were set up together at four locations from December 2020 to February. Considering human activities in each sampling area to represent each site for reporting measurement results to cover the most residential areas. The first location (chamber) was inside a control room, in which incense was burned to provide a source of aerosols. The second location (University campus) was located on the campus of Khon Kaen University (16.4706°, 102.8255°) (Figure 2). The third location (urban center) was in downtown Khon Kaen, located near high-traffic roads and markets (16.4287°, 102.8354°). The fourth location (rural industrial) was located in a primarily agricultural area but in the vicinity of a sugar cane factors. Data from each instrument was logged every 30 minutes.

2.3. Sampling Method. The concentration of $PM_{2.5}$ samples were collected for 12 days, with data being recorded every

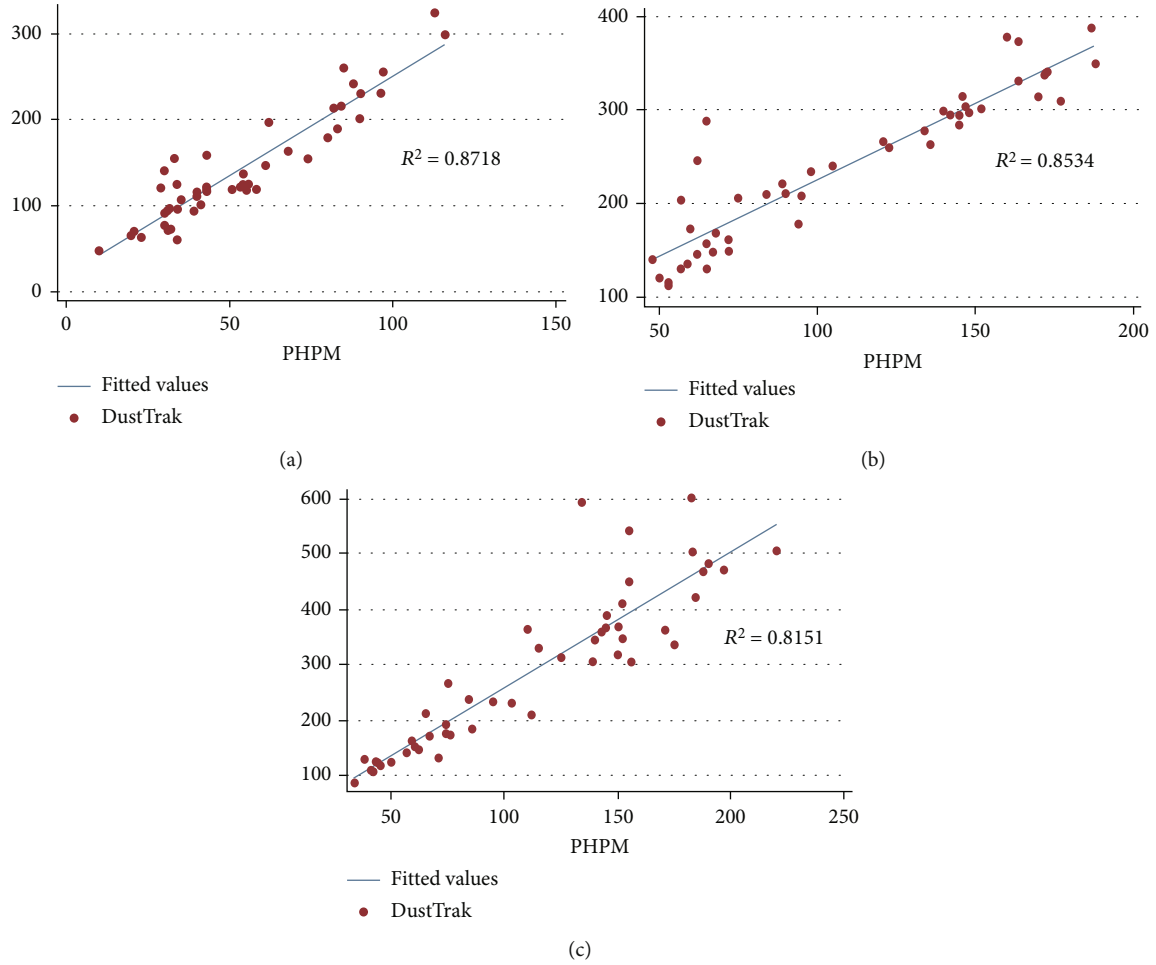


Figure 4: Observed $PM_{2.5}$ concentrations, as measured by the DustTrak Drx Aerosol Monitor 8533, on the y-axis compared to the estimate of $PM_{2.5}$ from the PHPM device, as measured at (a) a controlled indoor environment with incense (chamber); (b) the Khon Kaen University campus (University campus); and (c) at rural area near a sugar cane factory (rural industrial).

Table 3: Pearson's correlation coefficient, r, for each sampling location and summary statistics of measured $PM_{2.5}$ concentrations.

Site	Observed $PM_{2.5}$ in DustTrak 8533 ($\mu g/m^3$)†	Estimated $PM_{2.5}$ in PHPM ($\mu g/m^3$)†	Correlation, r
Chamber	241.8 (102-477)	139.3 (50-804)	0.71
University campus	143.2 (48-325)	53.6 (10-116)	0.93
Urban center	240.8 (112-388)	109.2 (48-188)	0.92
Rural industrial	292.6 (85-602)	113.9 (34-220)	0.90

†Average and range (in parentheses) are reported.

30 minutes from 8:00 a.m. to 8:00 a.m. the next day for 24 hours. Collect dust samples with a size of not more than 2.5 microns ($PM_{2.5}$) by installing a collector. Dust samples for both the High-Volume Air Sampler, 5 types of dust measuring instruments (commercially sold brands) and simple dust measuring instruments up to 2.5 μm that were developed from December 2020 to January 2021 for 24 hours from 8:00 a.m. to 8:00 a.m. of the next day for 24 days. Methods for collecting dust samples with a size not exceeding 2.5 μm ($PM_{2.5}$) by a High-Volume Air Sampler. Preparing filter paper, glass fiber filter is baked to remove moisture for 24 hours. After baking, put the filter paper in the dehumidifier cabinet for at least 24 hours. Then, put the filter paper in the filter head 1 sheet per time used sampling. When the sample collection is complete, the filter paper is placed in the dehumidifier for at least 24 hours and then weighed with a fine balance with 5 decimal places before and after the collected samples were to determine the weight of the dust particles on the filter paper. Method of collecting dust samples particulate matter up to 2.5 μm ($PM_{2.5}$) by DustTrak 8533, five types of dust measuring instruments (commercially sold brands), and a simple developed dust measuring device up to 2.5 μm. The method of collecting samples of these instruments is simple.

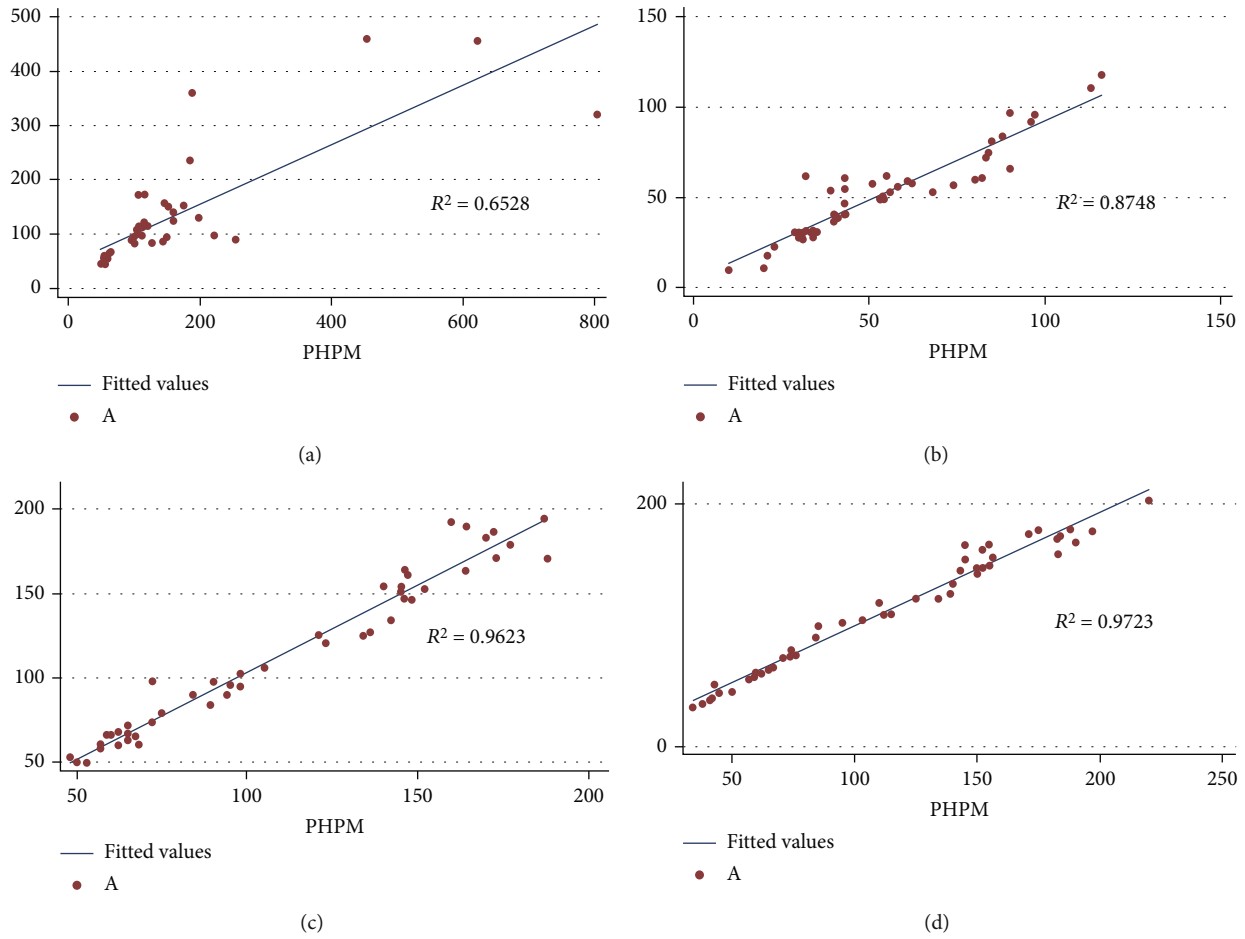


Figure 5: Observed $PM_{2.5}$ concentrations, as measured by the SI device, on the y-axis compared to the estimate of $PM_{2.5}$ from the PHPM device, as measured at (a) a controlled indoor environment with incense (chamber); (b) the Khon Kaen University campus (University Campus); (c) in downtown Khon Kaen (urban center); and (d) at rural area near a sugar cane factory (rural industrial).

The machine will suck air into the system and analyze the data and display results immediately on the display screen, which will take samples every 30 minutes from 8:00 a.m. to 8:00 a.m. of the following day from December 2020 to January 2021 and averaged by using standard 24 h. Analyzing fine particulate matter $PM_{2.5}$ by gravimetric method by weighing the dust filter paper with a size of not more than 2.5 μm ($PM_{2.5}$) before and after, samples were taken, and the weight difference was determined to calculate the concentration of fine particulate matter using the formula as follows:

$$\text{PM2.5}\left(\mu g/m^3\right) = \frac{\left(W_f - W_i\right) \times 10^6}{\text{Vstd}}, \quad (1)$$

where W_f = filter paper weight after sampling (g)

W_i = presampling paper weight (g)

Vstd = volume of air at standard conditions (cubic meter unit)

10^6 = convert grams to micrograms

Source: Office of Air Quality and Noise Management, Pollution Control Department, Thailand.

2.4. Data Analysis. Pearson's correlation coefficient, r, was computed to measure the association between the $PM_{2.5}$ estimates from the PHPM and measurements from the DustTrak 8533 for each sampling location independently. Linear regression was used to calibrate the readings of the PHPM to the DustTrak 8533 readings. All data analysis was completed in STATA version 15 (copyright of Khon Kaen University). Simple linear regression was used to describe the relationship between the two variables and to make predictions. By constructing an equation showing a linear relationship between two variables from the linear equation as follows: $yi = a + bxi$ or $yi = \beta_0 + \beta 1xi$, where y value is the concentration of $PM_{2.5}$ obtained from the $PM_{2.5}$ measuring device. A simple developed and x is the reference DustTrak 8530, where a or β_0 is the y-intercept constant and b or $\beta 1$ is the slope or coefficient, where r^2 is 0 to 1, and approaching 1, the regression equation can explain a lot of the change. The data were analyzed from the sampling results of dust concentrations up to 2.5 microns by statistical methods, using Pearson correlation coefficient between the two variables. The estimation equation for Pearson correlation coefficient (r_{xy}) is given as follows:

$$r = \frac{\text{COV}(X, Y)}{S_x S_y}, \quad (2)$$

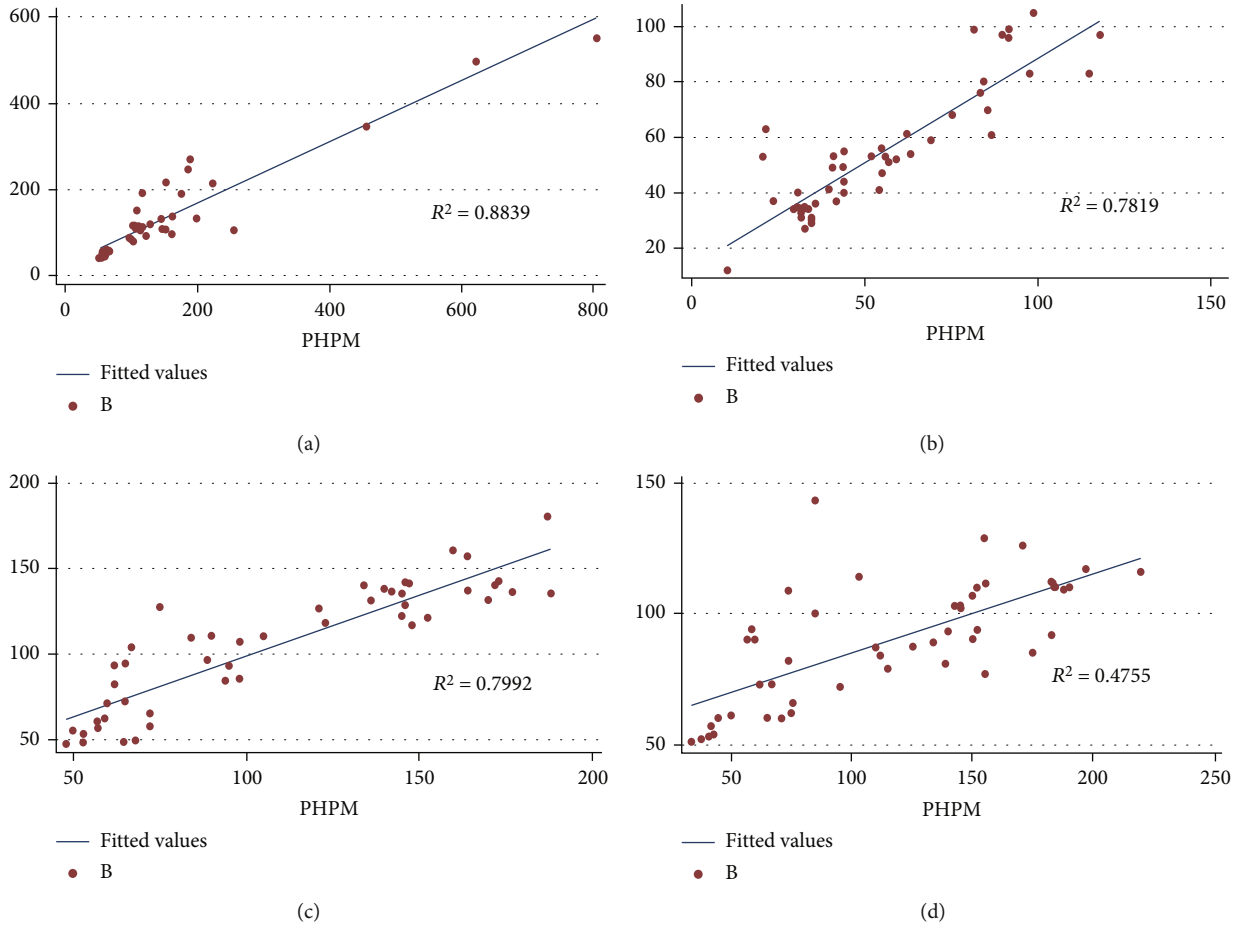


Figure 6: Observed $PM_{2.5}$ concentrations, as measured by the SN device, on the y-axis compared to the estimate of $PM_{2.5}$ from the PHPM device, as measured at (a) a controlled indoor environment with incense (chamber); (b) the Khon Kaen University campus (University campus); (c) in downtown Khon Kaen (urban center); and (d) at rural area near a sugar cane factory (rural industrial).

with

$$\mathrm{COV}(X, Y) = \sum_{i=1}^{n} \left(\frac{(xi - \bar{x})(yi - \bar{y})}{(n-1)} \right), \tag{3}$$

where $\mathrm{COV}(X, Y)$ is the sample covariance between two random variables X and Y that are normally distributed with means $\bar{x}$ and $\bar{y}$ standard deviations S_x and S_y, respectively.

3. Results

The graph shows the high correlation of the data with $r = 0.93$ and $R^2 = 0.87$, meaning that the resulting regression can easily account for the distribution of the results measured with a dust measuring instrument up to 2.5 μm, 87 percent. Concentration prediction obtained from a simple dust measuring device up to 2.5 microns in size using a simple regression equation was as follows: $y = 18.58 + 2.33X$. It was found that dust concentrations of no more than 2.5 microns were found, obtained from a simple dust measuring device that measures no more than 2.5 μm with an increase of 1 $\mu g/m^3$. As a result, the dust concentration of up to 2.5 μm obtained by the DustTrak 8533 was increased by 2.33 $\mu g/m^3$, where y is the predicted value of the particulate concentration equation of 2.5 μm obtained from simple dust measuring instruments up to 2.5 μm which are given as values from DustTrak 8533, and x is the concentrations of dusts up to 2.5 μm obtained from dust detectors. Size is not more than 2.5 μm easily. A PHPM instrument for monitoring $PM_{2.5}$ concentration is shown in Figure 3.

3.1. Observed and Estimated $PM_{2.5}$ Concentrations. Overall, a positive linear relationship was observed between the PHPM estimates and the $PM_{2.5}$ data from the DustTrak 8533 (Figure 4). The highest $PM_{2.5}$ concentrations were observed in the chamber (Figure 4(a)). Ambient air $PM_{2.5}$ concentrations at the three outdoor sites were similar. However, the $PM_{2.5}$ was lowest at the University campus site (Figure 4(b)) and highest at rural area near a sugar cane factory (rural industrial) (Figure 4(c)). Overall, the estimates from the PHPM were lower than the data from the DustTrak 8533 (Table 3).

3.2. Correlation of PHPM Data to DustTrak 8533 Data. Considering only the outdoor ambient air sites, the correlation between PHPM and DustTrak 8533 was high, with a Pearson's correlation coefficient values ($r = 0.90$) (Table 3). The

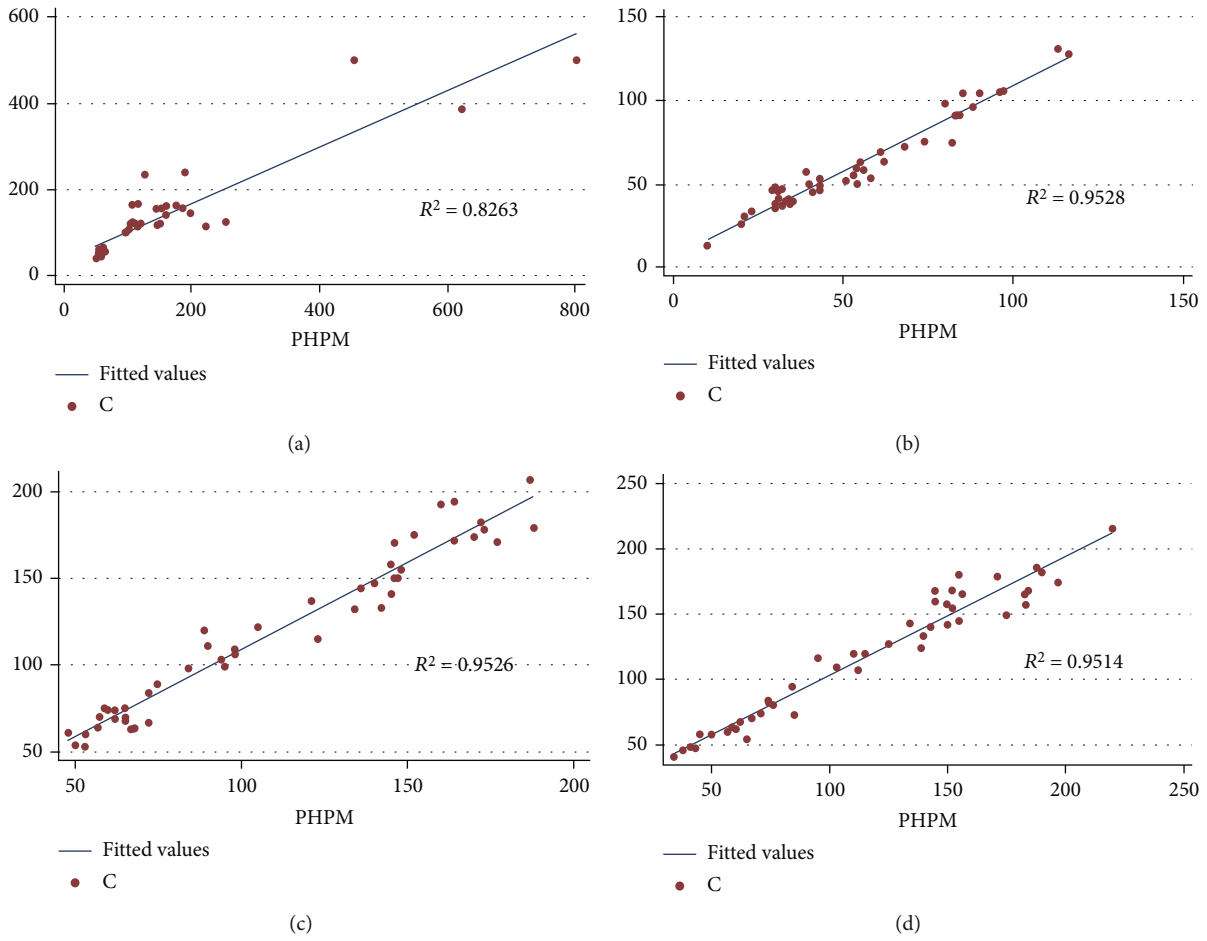


Figure 7: Observed $PM_{2.5}$ concentrations, as measured by the SV device, on the y-axis compared to the estimate of $PM_{2.5}$ from the PHPM device, as measured at (a) a controlled indoor environment with incense (chamber); (b) the Khon Kaen University campus (University campus); (c) in downtown Khon Kaen (urban center); and (d) at rural area near a sugar cane factory (rural industrial).

chamber data, however, show nonlinear relationship and lower correlation. In general, the values reported by the PHPM system were lower than the DustTrak 8533 estimates, indicating the PHPM device was underestimating $PM_{2.5}$. Nevertheless, the PHPM system performed well with high correlation with $PM_{2.5}$ concentrations in its targeted environment of ambient air. Linear regression determined the relationship of $PM_{2.5}$ from the PHPM device to be

$$\text{PM2.5}\left(\mu g/m^3\right) = 18.6 + 2.33 \times \text{PHPM} \quad (4)$$

The data suggests that this linear relationship can be used to correct the PHPM data for 30-minute $PM_{2.5}$ concentrations using a correction factor of 2.33 and is applicable for use in ambient air with $PM_{2.5}$ concentrations in the range of 50 to 600 $\mu g/m^3$. The results of the study of the relationship of the data in the study method are shown in Figures 5–9.

4. Discussion

This research found that the correlation between the low-cost light-scattering sensor and the higher accuracy DustTrak device was relatively high ($R^2 > 0.80$) across all observed $PM_{2.5}$ concentrations in the three outdoor ambient air test sites. However, the lowest observed $PM_{2.5}$ concentration was 48 $\mu g/m^3$ according to the DustTrak device. Testing of multiple commercial low-cost $PM_{2.5}$ sensors has demonstrated that the correlation between such low-cost devices and the DustTrak device can have poor correlation at concentrations less than 50 $\mu g/m^3$ [23]. Therefore, it is possible that the PHPM device may be inaccurate at concentrations less than 50 $\mu g/m^3$. However, the targeted application of the device is to increase awareness of $PM_{2.5}$ concentrations during $PM_{2.5}$ pollution events (generally higher than 50 $\mu g/m^3$). Therefore, inaccuracy at this range does not substantially impact the intended use.

A limitation of the device is that substantial variations in conditions may affect the correction factor determined from this research. While air quality data from low-cost sensors is extremely valuable for both research and public health usage, it should be carefully evaluated, especially under unusual atmospheric or genesis conditions. In addition, the sensors will automatically alarm when the dust concentration is detected over the standard. In the dust season, people were alarmed, and they felt panic; thus, we decided to turn off the alarm. The sensors need to be used with Internet

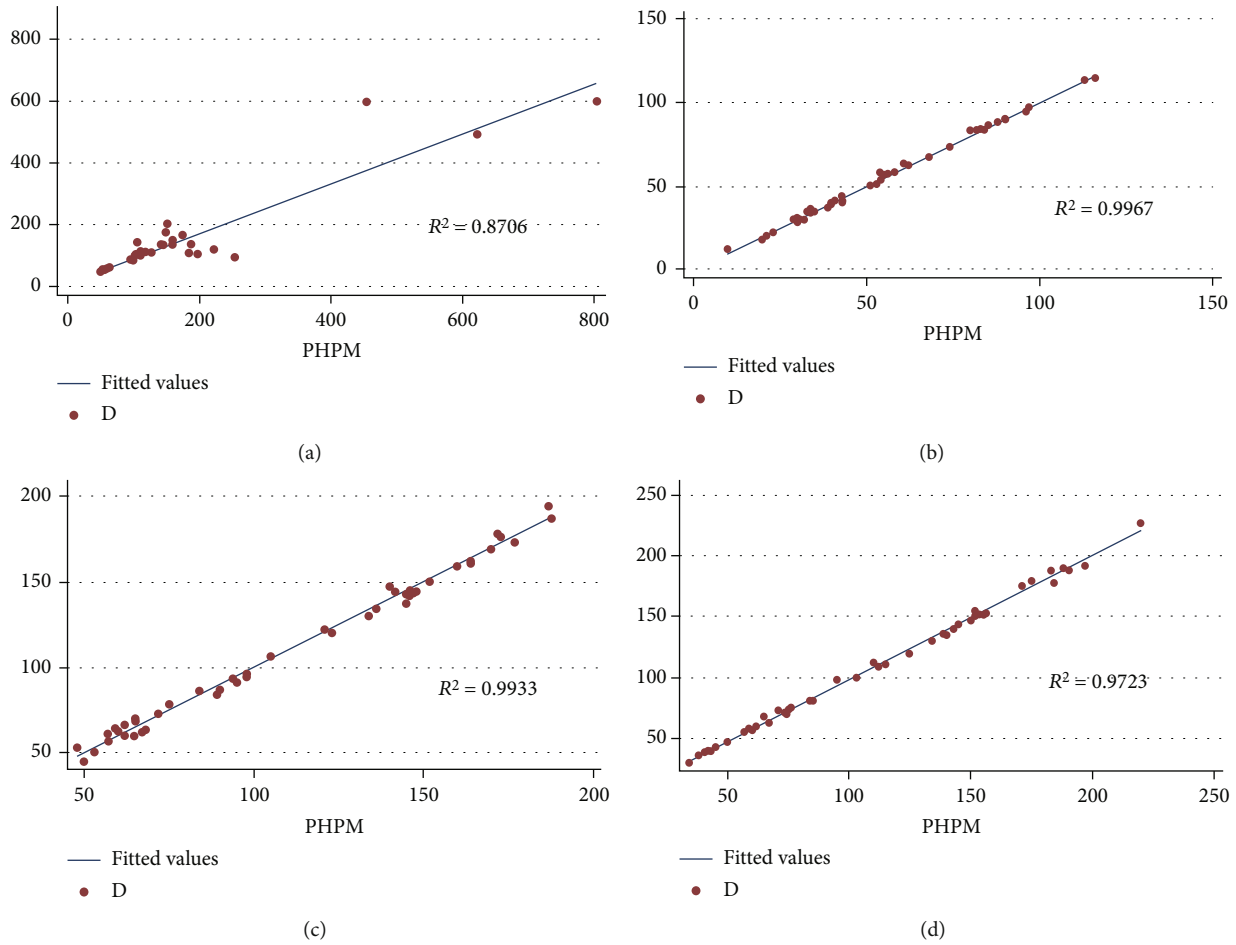


Figure 8: Observed $PM_{2.5}$ concentrations, as measured by the SC device, on the y-axis compared to the estimate of $PM_{2.5}$ from the PHPM device, as measured at (a) a controlled indoor environment with incense (chamber); (b) the Khon Kaen University campus (University campus); (c) in downtown Khon Kaen (urban center); and (d) at rural area near a sugar cane factory (rural industrial).

signal and also need a cloud storage for the application in order to link between dust quantity and health; anyway in some area, there is no Internet available, so the receivers needs to be installed and also a cloud storage. First, the performance of low-cost $PM_{2.5}$ sensors, such as optical particle counters, can be impacted by weather conditions. High relative humidity has been shown to cause optical particle counters to overestimate $PM_{2.5}$ concentrations [24]. Under high humidity conditions, hygroscopic growth of particles and reduced molecular mass of water result in an overestimation of particle mass [25]. Secondly, the varying size and composition of particles can also substantially affect the performance of low-cost $PM_{2.5}$ sensors [26, 27]. The effect of differing particle composition was observed in this study, too. While the sensors performed well ($r > 0.90$) in the ambient outdoor environments, the correlation was low in the chamber test using incense ($r = 0.71$). For this study, this PHPM device is more suitable for ambient air than the chamber. Thirdly, variations in the composition of atmospheric dust or meteorological conditions may also require adjustments in the computation methods for $PM_{2.5}$ estimation or the use of short time periods (e.g., one hour) [24]. Thus, caution should be applied in using the PHPM sensor outside of the tested conditions. If changes in the composition or size of the $PM_{2.5}$ particles, relative humidity, or desired temporal resolution occur, additional validation will be needed.

Another potential solution to the need for revised correction factors under varying conditions is to incorporate these confounding factors into the regression. While linear regression was used in this study, the accuracy of the low-cost device could be improved with multivariate regression that accounts for effects of confounding factors such as relative humidity or site-specific composition of the particulate matter [6] or nonlinear regression to account for nonlinear associations [28]. The widespread adoption of a low-cost $PM_{2.5}$ sensor will help increase public awareness of current $PM_{2.5}$ concentrations. The goal of increased awareness is to encourage coping activities, such as wearing a mask, [20] and pollution mitigation actions, such as using public transport [29]. Low-cost sensors can also facilitate an environment of citizen science, in which nonscientists actively participate in generating and sharing air quality data [30]. Such low-cost sensors have been used in open-source platforms to create a network of thousands of $PM_{2.5}$ monitoring

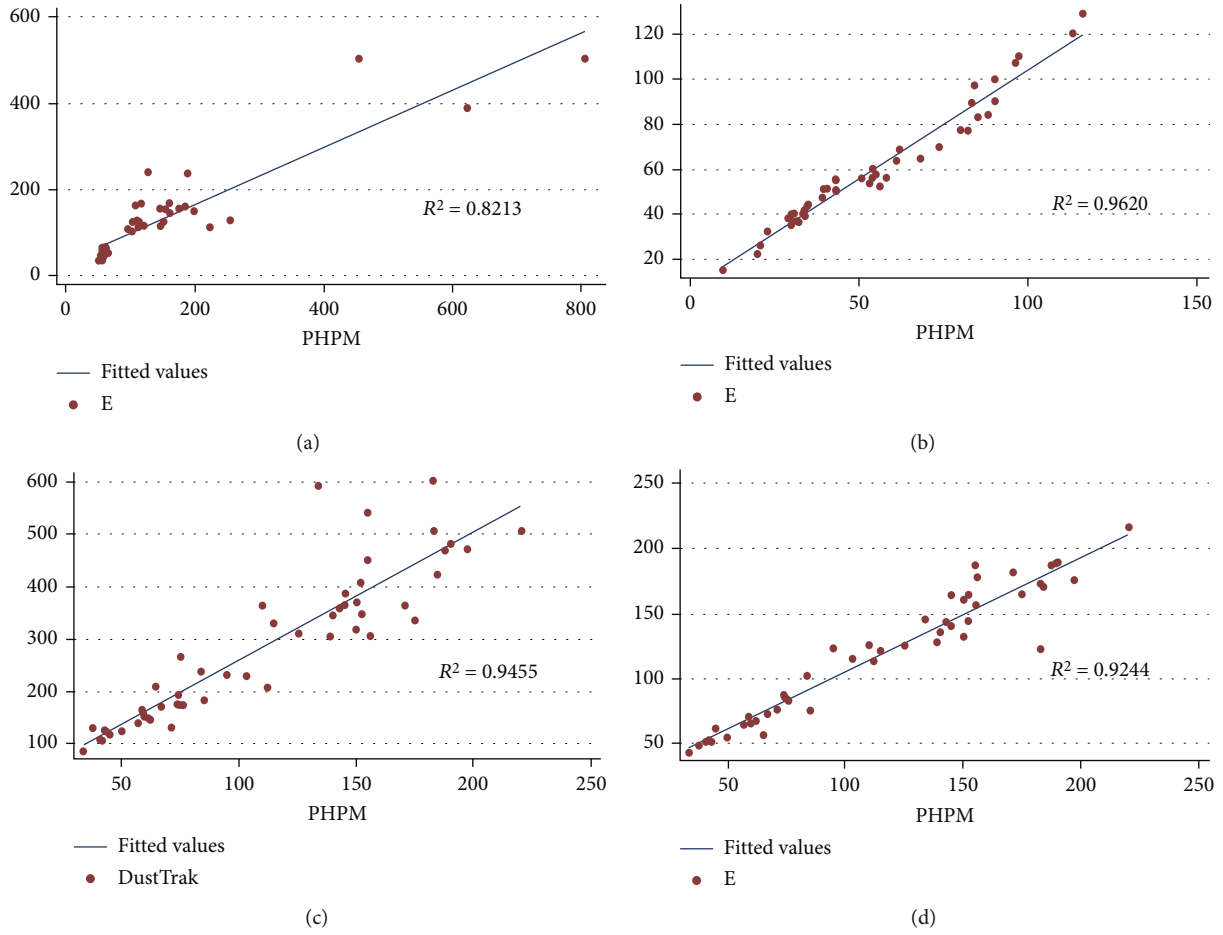


Figure 9: Observed $PM_{2.5}$ concentrations, as measured by the CKDPM device, on the y-axis compared to the estimate of $PM_{2.5}$ from the PHPM device, as measured at (a) a controlled indoor environment with incense (chamber); (b) the Khon Kaen University campus (University campus); (c) in downtown Khon Kaen (urban center); and (d) at rural area near a sugar cane factory (rural industrial).

devices [31], in networks of community air monitoring stations [32], and to measure other pollutants besides just $PM_{2.5}$ [33]. In addition, depending on the sensing capabilities of the developed and maintenance sensor probe, working on the method is to be specific and accurate for detection [7, 8].

5. Conclusion

A novel low-cost $PM_{2.5}$ sensor using light-scattering technology, PHPM, was developed and validated under both the chamber and field ambient air conditions. The PHPM has a large LED display screen, to be easy to see when installing in public areas where people can be aware of the amount of dust in the area where they live. In addition, the tool can be used outdoors due to its heat-resistant metal body. The instrument is working; it will suck the air through the air intake in the sensor. There will be a laser light source acting as part of the beam to strike particles of various sizes that enters the machine by a dust measuring device that does not exceed 2.5 microns in size.

Validation against a DRX Aerosol Monitor 8533 found that the average concentration of $PM_{2.5}$ for the PHPM device was using multiply with 2.33, according to linear regression. Under ambient air conditions with $PM_{2.5} > 50\,\mu g/m^3$, the PHPM device correlated well with the DustTrak device ($r > 0.90$). However, the correlation was weaker under indoor ambient air conditions with incense as a particle source. The hope is that the PHPM device can provide an additional tool for participatory citizen science and increased public awareness about $PM_{2.5}$ in Thailand. The PHPM device is inexpensive, portable, and can be charged with a power bank. Additionally, a simple measuring device for dust of less than 2.5 μm with a unique characteristic body made of heat-resistant metal can be used outdoors. The front has a large digital display of the amount of dust particles up to 2.5 μm. In the front part, there is a dust level display, no more than 2.5 μm, red, yellow, and green lights on the top, and there is a ventilation hole. The internal sensor used to read the amount of dust, the size of not more than 2.5 μm, is PMS 7003. There are components, which are composed of a heat-resistant metal body, an invention that has never been done before. (1) The front has a digital display of the amount of dust particles up to 2.5 μm. (2) In the front part, there is the level display of particulate matter up to 2.5 μm in red, yellow, and green lights is a new invention (3) with dust

levels in the range of ≥50 $\mu g/m^3$ red indicator light. When the dust concentration is higher than the standard value, it affects health, and there will be an alert sound immediately. The level of dust is in the range of 0-25 $\mu g/m^3$ green indicator light. When the dust is concentrated, there is no health effect, and the level of dust is in the range of 26-49 $\mu g/m^3$ yellow indicator light, when the dust concentration is starting to affect health. The sensor is a processor that controls the status light. A heat-resistant metal measuring instrument is built into the unit with a PMS 7003 sensor. The sensor has an air intake port attached to the PMS 7003 sensor unit. The top of the measuring instrument has several air intake holes to allow the sensor to absorb a lot of air. This is a new invention. (4) It allows the sensor to read the amount of dust particles up to 2.5 μm. (5) The sensor's characteristics are a one-time response time of less than 1 second, integrated response time of less than or equal to 10 s charging 5 V, less than or equal to 100 mA working environment -10~+60°C, humidity 0~99% measuring range 0-500 $\mu g/m^3$, and resolution 1 $\mu g/m^3$ meter accuracy ±10% @100 ~ 500 ± 10 $\mu g/m^3$, @0~100 $\mu g/m^3$. The limitation is that it cannot be used in environments with a dust concentration of 2.5 μm or more 500 $\mu g/m^3$ and cannot be used in environments with excessive humidity and temperature.

Data Availability

The data used during the current study are available from the authors upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

P.S., P.K., and S.S., are responsible for the conceptualization; P.S., P.K., W.L., and R.S. for methodology; W.L. for software; P.S., P.K., W.L., R.S., N.M., and J.R. for validation; P.K., W.L., and R.S. for formal analysis; W.L., R.S., and C.J. for resources; P.S. and S.S. for writing—original draft preparation; P.S. and S.S. for writing—review and editing; P.K. for visualization; and P.S. for funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We express our sincere appreciation to the Occupational Health Safety and Environmental Epidemiology Group (OHSEE) at the Faculty of Public Health, Khon Kaen University for data collection and operator. This research was funded by research program funding from Khon Kaen University (grant number: RP64006).

References

[1] P. J. Landrigan, R. Fuller, H. Hu et al., "Pollution and global health–an agenda for prevention," *Environmental Health Perspectives*, vol. 126, no. 8, article 084501, 2018.

[2] R. V. Martin, M. Brauer, A. van Donkelaar, G. Shaddick, U. Narain, and S. Dey, "No one knows which city has the highest concentration of fine particulate matter," *Atmospheric Environment: X*, vol. 3, article 100040, 2019.

[3] S. Zhao, D. Yin, Y. Yu, S. Kang, D. Qin, and L. Dong, "$PM_{2.5}$ and O_3 pollution during 2015-2019 over 367 Chinese cities: spatiotemporal variations, meteorological and topographical impacts," *Environmental Pollution*, vol. 264, article 114694, 2020.

[4] C. Ratti, S. Di Sabatino, and R. Britter, "Urban texture analysis with image processing techniques: winds and dispersion," *Theoretical and Applied Climatology*, vol. 84, no. 1-3, pp. 77–90, 2006.

[5] H. Xu, J. Cao, J. C. Chow et al., "Inter-annual variability of wintertime $PM_{2.5}$ chemical composition in Xi'an, China: evidences of changing source emissions," *Science of the Total Environment*, vol. 545-546, pp. 546–555, 2016.

[6] B. Alfano, L. Barretta, A. Del Giudice et al., "A review of low-cost particulate matter sensors from the developers' perspectives," *Sensors*, vol. 20, no. 23, p. 6819, 2020.

[7] Q. Yang, G. Zhu, L. Singh et al., "Highly sensitive and selective sensor probe using glucose oxidase/gold nanoparticles/graphene oxide functionalized tapered optical fiber structure for detection of glucose," *Optik*, vol. 208, article 164536, 2020.

[8] S. Kumar, R. Singh, G. Zhu et al., "Development of uric acid biosensor using gold nanoparticles and graphene oxide functionalized micro-ball fiber sensor probe," *Bioscience*, vol. 19, no. 2, pp. 173–182, 2020.

[9] Z. Chowdhury, R. D. Edwards, M. Johnson et al., "An inexpensive light-scattering particle monitor: field validation," *Journal of Environmental Monitoring*, vol. 9, no. 10, pp. 1099–1106, 2007.

[10] C. M. Sorensen, J. Gebhart, T. J. O'Hern, and D. J. Rader, "Optical measurement techniques: fundamentals and applications," in *Aerosol Measurement: Principles, Techniques, and Applications; Pramod Kulkarni*, P. A. Baron and K. Willeke, Eds., pp. 269–312, John Wiley & Sons, Hoboken, NJ, USA, 2011.

[11] P. Kumar, L. Morawska, C. Martani et al., "The rise of low-cost sensing for managing air pollution in cities," *Environment International*, vol. 75, pp. 199–205, 2015.

[12] L. Morawska, P. K. Thai, X. Liu et al., "Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone?," *Environment International*, vol. 116, pp. 286–299, 2018.

[13] H. Mei, P. Han, Y. Wang et al., "Field evaluation of low-cost particulate matter sensors in Beijing," *Sensors*, vol. 20, no. 16, p. 4381, 2020.

[14] F. Reisen, J. Cooper, J. C. Powell, C. Roulston, and A. J. Wheeler, "Performance and deployment of low-cost particle sensor units to monitor biomass burning events and their application in an educational initiative," *Sensors*, vol. 21, no. 21, p. 7206, 2021.

[15] N. Castell, F. R. Dauge, P. Schneider et al., "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?," *Environment International*, vol. 99, pp. 293–302, 2017.

[16] J. Bi, N. Carmona, M. N. Blanco et al., "Publicly available low-cost sensor measurements for PM2.5 exposure modeling: guidance for monitor deployment and data selection," *Environment International*, vol. 158, article 106897, 2022.

[17] C. ChooChuay, S. Pongpiachan, D. Tipmanee et al., "Impacts of PM2.5 sources on variations in particulate chemical compounds in ambient air of Bangkok, Thailand," *Atmospheric Pollution Research*, vol. 11, no. 9, pp. 1657–1667, 2020.

[18] N. R. Fold, M. R. Allison, B. C. Wood et al., "An assessment of annual mortality attributable to ambient PM2.5 in Bangkok, Thailand," *International Journal of Environmental Research and Public Health*, vol. 17, no. 19, p. 7298, 2020.

[19] H. Shi, S. Wang, and D. Zhao, "Exploring urban resident's vehicular PM2.5 reduction behavior intention: an application of the extended theory of planned behavior," *Journal of Cleaner Production*, vol. 147, pp. 603–613, 2017.

[20] Z. Xu, J. Li, J. Shan, and W. Zhang, "Extending the theory of planned behavior to understand residents' coping behaviors for reducing the health risks posed by haze pollution," *Environment, Development and Sustainability*, vol. 23, no. 2, pp. 2122–2142, 2021.

[21] D. Liu, Q. Zhang, J. Jiang, and D. R. Chen, "Performance calibration of low-cost and portable particular matter (PM) sensors," *Journal of Aerosol Science*, vol. 112, pp. 1–10, 2017.

[22] A. Curto, D. Donaire-Gonzalez, J. Barrera-Gómez et al., "Performance of low-cost monitors to assess household air pollution," *Environmental Research*, vol. 163, pp. 53–63, 2018.

[23] R. Jayaratne, X. Liu, K.-H. Ahn et al., "Low-cost PM2.5 sensors: an assessment of their suitability for various applications," *Aerosol and Air Quality Research*, vol. 20, no. 3, pp. 520–532, 2020.

[24] E. Brattich, A. Bracci, A. Zappi et al., "How to get the best from low-cost particulate matter sensors: guidelines and practical recommendations," *Sensors*, vol. 20, no. 11, p. 3073, 2020.

[25] G. Bagtasa, N. Takeuchi, S. Fukagawa, H. Kuze, and S. Naito, "Correction in aerosol mass concentration measurements with humidity difference between ambient and instrumental conditions," *Atmospheric Environment*, vol. 41, no. 8, pp. 1616–1626, 2007.

[26] K. K. Johnson, M. H. Bergin, A. G. Russell, and G. S. W. Hagler, "Field test of several low-cost particulate matter sensors in high and low concentration urban environments," *Aerosol and Air Quality Research*, vol. 18, no. 3, pp. 565–578, 2018.

[27] T. Sayahi, D. Kaufman, T. Becnel et al., "Development of a calibration chamber to evaluate the performance of low-cost particulate matter sensors," *Environmental Pollution*, vol. 255, Part 1, article 113131, 2019.

[28] M. Alvarado, F. Gonzalez, A. Fletcher, and A. Doshi, "Towards the development of a low cost airborne sensing system to monitor dust particles after blasting at open-pit mine sites," *Sensors*, vol. 15, no. 8, pp. 19667–19687, 2015.

[29] M. Greaves, L. D. Zibarras, and C. Stride, "Using the theory of planned behavior to explore environmental behavioral intentions in the workplace," *Journal of Environmental Psychology*, vol. 34, pp. 109–120, 2013.

[30] M. Jerrett, D. Donaire-Gonzalez, O. Popoola et al., "Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science," *Environmental Research*, vol. 158, pp. 286–294, 2017.

[31] L. Chen, Y. Ho, H. Lee et al., "An open framework for participatory PM2.5 monitoring in smart cities," *Access*, vol. 5, pp. 14441–14454, 2017.

[32] P. English, H. Amato, E. Bejarano et al., "Performance of a low-cost sensor community air monitoring network in Imperial County, CA," *Sensors*, vol. 20, no. 11, p. 3031, 2020.

[33] P. Arroyo, J. Gómez-Suárez, J. I. Suárez, and J. Lozano, "Low-cost air quality measurement system based on electrochemical and PM sensors with cloud connection," *Sensors*, vol. 21, no. 18, p. 6228, 2021.

Hindawi
Journal of Sensors
Volume 2023, Article ID 9213689, 10 pages
https://doi.org/10.1155/2023/9213689

Research Article

Rural Folk Dance Movement Recognition Based on an Improved MCM-SVM Model in Wireless Sensing Environment

Li Xie

College of Music and Dance, Yulin Normal University, Yulin, 537000 Guangxi, China

Correspondence should be addressed to Li Xie; 20040643@ylu.edu.cn

Received 21 September 2022; Revised 19 October 2022; Accepted 28 January 2023; Published 16 February 2023

Academic Editor: Rocio Perez de Prado

In order to improve the accuracy and timeliness of folk dance movement recognition, this paper proposes an improved MCM-SVM recognition model to recognize the lower limb human motion of ethnic dance in rural areas based on sensors. In order to recognize these actions, the SVM algorithm is used to identify the current action, and the MCM is used to optimize the recognition result. The experimental results show that the proposed improved model achieves higher recognition rate compared to the SVM algorithm for the recognition of different dance moves. The average recognition rate exceeds 93%, and the average recognition time is about 0.6 ms, which verifies the effectiveness of the proposed model. The proposed model will provide guidance and practicality for the design and construction of future dance movement recognition systems.

1. Introduction

The application of modern information technology in the protection of traditional cultural resources and cultural inheritance makes the spread of national art and culture present a dynamic and technological sense. Ethnic dance has a distinctive feature in China and is the main form of art in rural areas of China. However, with the passage of time, many ethnic dance moves are on the verge of being lost. Therefore, the recording, protection, and inheritance of these intangible cultural heritage have become an important research topic.

Motion capture is an accurate three-dimensional human motion recording method, which can be used in ethnic dance to comprehensively record and protect dynamic arts such as ethnic dance [1]. Human action recognition is a hot research field that has emerged in recent years. With the development of pattern recognition and artificial intelligence, more and more universities, research institutes, and companies have invested in research in this field. As an important branch in the field of pattern recognition and artificial intelligence, human action recognition has important significance and far-reaching development prospects in human-computer interaction, medical health, sports analysis, intelligent monitoring, and even homeland security. Divided from the research content, the human body action recognition can be divided into two aspects: gesture recognition and dynamic process recognition [2]. Among them, gesture recognition is mainly to classify and distinguish static objects, which can be subdivided into hand shape recognition and posture recognition. Dynamic process recognition is the recognition of dynamic objects, which can be specifically subdivided into gesture recognition, facial expression recognition, gait recognition, motion recognition, etc.

The recognition of lower limb human motion of ethnic dance belongs to dynamic process recognition, which mainly includes the recognition of motion information such as lower limb human motion mode, spatial position, lower limb joint angle, and angular acceleration. At present, in the process of human action recognition, the accuracy and timeliness of recognition need to be further improved. The low recognition accuracy causes the movement of the exoskeleton of the lower limbs to be inconsistent with the movement that the wearer wants to complete, and the poor recognition timeliness causes the device to always lag behind the movement of the human body [3]. Human motion recognition based on wearable sensors is one of the important research directions. It mainly analyzes and processes human

body motion information to identify the human body's motion state.

Lower limb human motion recognition is mainly used in motion analysis, medical health, and bionic robots, but the research on its application in the field of dance is relatively scarce. This research will take the inheritance of ethnic dance in rural areas of China as the background to explore the sensor-based recognition method of ethnic dance lower limb human motion to fill the research gap in this field. In order to recognize these actions, the support vector machine (SVM) algorithm is used to identify the current action, and the Markov chain model (MCM) is used to optimize the recognition result.

2. Literature Review

Psychologist Johansson was the first to study human movements. He obtained the movement process of the human body by binding high-brightness reflective sequins to key parts of the human body [4]. This shows that people can distinguish the type of human body movement by observing the movement time series of key parts of the human body. This experiment opened the prelude to the study of human movement and had an important impact on future research.

With the development of computers and sensors, the research on human motion recognition has opened a diversified, high-precision research model. In order to achieve more accurate motion recognition, Peng et al. added physiological signals to the motion capture data, using the characteristics of the physiological signals directly as the input of the classifier, and the sensor signals were extracted from the topic through the topic model. After the distribution is input into the classifier, this method has achieved a good recognition effect on complex behaviors [5]. Human movements are highly complex and diverse in styles. Wang et al. focused on the difference in motion between the elderly and young people, extracted the common time-domain and frequency-domain features in the motion capture signal, and completed the establishment of the recognition model through commonly used machine learning algorithms [6]. Gibson et al. introduced and evaluated a system for remote health monitoring, which can realize fall detection and diagnosis based on acceleration sensors [7]. Since there is currently little research work on combined classifiers, one of the contributions of the article is to use a combination of multiple classifiers with different attributes to improve the performance of the system and improve the single classifier system through majority voting.

The threshold-based fall detection method uses whether the acceleration peak, valley, or other characteristic value reaches a predetermined threshold to determine whether a fall has occurred. They can detect when a fall occurs; however, the false alarm rate is an important issue for this type of algorithm. Thakkar and Pareek discussed various ML and DL technologies of human behavior recognition (HAR) from 2011 to 2019 and gave the advantages and disadvantages of action representation methods, dimensionality reduction methods, and action analysis methods [8]. Dhiman and Vishwakarma proposed a human action recognition framework with invariable depth of view, which integrates two important action cues: motion and shape time dynamics (STD) [9]. Ludl et al. introduced a modular simulation framework that provides training and verification algorithms in various human-centered scenarios. Laboratory experiments show that based on motion capture data and 3D avatars, you can train with only simulation data. A recurrent neural network achieves almost perfect results in classifying human actions on real data [10]. Du and Mukaidani discussed a two-stream structure human action recognition method based on a linear dynamic system, proposed a dual-stream deep feature extraction framework based on a preprocessed convolutional neural network, and verified the effectiveness of the method [11]. Sedmidubsky and Zezula proposed an evaluation procedure for 3D human action recognition to determine the best combination in a very effective way [12].

3. Methodology

3.1. Data Collection Scheme of Lower Limb Human Motion of Ethnic Dance. The movement of the lower limbs of the human body is completed by the driving force generated by the contraction of skeletal muscles to make the joints move [13]. The movement of the lower limbs of the human body mainly drives the movement of the ankle, knee, and hip joints through muscle contraction, thereby driving the bone movement and finally completing the corresponding gait movement. The form of motion of the joint is mainly rotation, and in the three motion modes of the ankle and hip joints, bending/extension has a larger range of motion, while the knee joint can only perform bending/extension motion. Joint movement under a specific action will also drive the bone to produce corresponding acceleration [14]. Therefore, these joint angle information and acceleration information of specific parts can be used to analyze the motion of the human body.

In order to recognize actions, the information of these actions should first be converted into electrical signals that can be processed by a computer. Physical interaction signals are used to analyze the lower limb human motion of ethnic dance. Try to use as few sensors as possible to accurately identify human movement intentions. Reducing the number of sensors can not only improve the comfort of the wearer but also reduce the amount of calculation to shorten the calculation time [15]. Due to the symmetry of the human body, only the movement of one leg was analyzed in the study, which helps to extend to both legs.

In the experiment, angle encoders were chosen to be placed at the ankle, knee, and hip joints to detect the angle signals of the joints in the sagittal plane. The Angtron-RE-38 series rotary angle encoder is used to detect the joint angle signal, the range is 0-360°, and the measurement accuracy is less than 0.1°. The MODEL4630 acceleration sensor is used to measure the sagittal acceleration on the underside of the thigh, with a range of ±10 g and a nonlinearity (%FSO) of ±0.1. The x-axis and y-axis of the acceleration sensor are in the sagittal plane, the x-axis is perpendicular to the thigh bone forward, and the y-axis is upward along the thigh. In

order to determine the start and end time of the gait action, the experiment placed a pressure sensor on the sole of the back foot to segment the movement data of the entire gait cycle. The effective area of the FlexiForce A401 thin-film pressure sensor is 25 mm, the measuring range is 45 kg, the measurement error is ±3%, and the response time is less than 5 ms. The sensor meets the requirements of this experiment and is more comfortable and convenient to wear and at the same time reduces the impact on the normal movement of the human body.

The sensor is connected to the capture card. Transfer the collected data to the computer for analysis and processing.

3.2. Signal Noise Reduction Method. The original signal has noise, especially the acceleration signal noise is more serious, and it appears as a more serious glitch phenomenon on the signal graph [16]. These noises will have a great impact on the feature extraction and the final dance action recognition accuracy. Therefore, these noises must be filtered out as much as possible, while trying not to destroy the original motion information of the signal.

The human body acceleration signal value fluctuates within a certain range and is subject to strong random interference. In this article, moving average filtering is used to filter signal noise.

The following is the principle of the moving average filtering method: first, the continuous sampling data is regarded as a queue with a fixed length of N. After a new measurement, the first data of the above queue is removed, the remaining $N-1$ data are moved forward in turn, and the new sampled data is inserted as the tail of the new queue. Finally, arithmetic operations are performed on this new queue, and the result is the final result of this measurement.

The original signal data is recorded as

$$X = \langle x_1, x_2, x_3, \cdots x_t \rangle. \tag{1}$$

The data processed by the moving average filtering method is

$$Y = \langle y_1, y_2, y_3, \cdots y_t \rangle. \tag{2}$$

The corresponding relationship between y_t and x_t is

$$y_t = \frac{1}{2k+1} \sum_{j=-k}^{k} x_{t+j}, \qquad N = 2k+1. \tag{3}$$

N is the number of adjacent sampling points used by the average value.

3.3. Signal Segmentation Processing. The human motion signal collected by the sensor is data composed in time sequence, which belongs to a typical time sequence. In order to use these time series data to quickly identify the motion of the human body, it is first necessary to segment it to obtain a data segment containing the corresponding motion. The result of data segmentation has a greater impact on the accuracy and real-time performance of the action recognition.

According to the experimental characteristics of this article, the important point segmentation method is selected. Segmenting the human body motion data segment contains only one complete gait cycle. The data fragment not only contains a complete gait movement information but also does not exceed the redundant data after the movement, so it is very suitable for real-time analysis of dance movements.

3.4. Action Feature Extraction Method. The human body motion information obtained by the sensor contains too much data. It is very time-consuming and inefficient to directly use these signal data to analyze the human body motion. In order to process these sensor data efficiently, the method of extracting signal characteristics is generally used to solve this problem, and the extracted characteristics are used to replace the signal itself to achieve the purpose of simplifying the signal. From another perspective, feature extraction is also a method of compression and dimensionality reduction of the original signal. While extracting signal features to reduce the dimensionality of the time series, the important information contained in the signal itself must be kept as much as possible. In the process of behavior expression, the posture of the object breeds spatial information, and the motion information is reflected in time and space. Therefore, the temporal dynamic information is very important for behavior expression. Therefore, the temporal feature extraction method is used in this article for the feature extraction of human motion information.

4. The Lower Limb Human Motion Recognition Model of Ethnic Dance

4.1. Data Collection Experimental Results of Lower Limb Human Motion of Ethnic Dance. In order to obtain objective and reasonable data, a total of 6 folk dance performers from rural areas were recruited in this collection process, aged between 29 and 49, and none of their lower limbs had any disease. Among them, 3 men are between 170 and 185 cm in height; 3 women are between 155 and 172 cm in height.

Before the start of the experiment, the tested persons were instructed, and their gait speed and other parameters were standardized to ensure that the movement data was collected under normal gait actions. The tested persons were all familiar with the procedures and precautions of the entire testing process. After fixing the sensor to the person to be detected, the movement information of the flexion, extension, abduction, external rotation, and internal rotation in the ethnic dance in rural areas is detected.

During the experiment, take appropriate rest according to the state of the tested person to prevent fatigue from adversely affecting the exercise data. Obviously, abnormal signals were removed according to the law of signal images, 120 combined grid data were obtained for each tester in each gait, and a total of 3,600 sets of data were obtained for 5 gaits. Each group of data contains 3 pieces of angle signal data, 2 pieces of acceleration sensor data, and 1 piece of pressure sensor data. In the experiment, the acquisition frequency of the sensor is set to

50 Hz, the data processing platform is Lenovo PC, the main frequency is 2.5 GHz, the memory is 8 GB, and the software platform is MATLAB R2016b.

Take the knee joint angle signal as an example. Figure 1 shows the knee joint angle signal during the lower limb human motion of these five kinds of ethnic dance. It can be seen intuitively from the figure that the joint angle signals have obvious differences under these 5 different actions, so these signals can be used to identify the actions of the human body's lower limbs.

4.2. Experimental Results of Signal Noise Reduction. The real-time nature of the signal is very important. The principle of selecting the N value is to reduce the N value as much as possible on the basis of achieving the noise reduction effect. In order to find the appropriate value of N, this paper uses different values of N for comparative experiments. Taking the acceleration in the y-axis direction of the thigh as an example, the filtering effect of the signal is obtained through experiments. The study found that when $N = 3$, the processed signal has poor smoothness, and when $N = 7$, the processed signal makes the original characteristics of the signal inconspicuous. After comprehensively considering the signal characteristics such as smoothness and peaks, it is found that the filtering effect of the moving average method is better when $N = 5$. Therefore, the value of N is set as 5 in this paper.

4.3. Feature Extraction and Optimization of Lower Limb Human Motion Signal of Ethnic Dance

4.3.1. Feature Standardization. We strive to achieve a higher accuracy of action recognition with a small amount of calculation. When analyzing sensor data, select time-domain features that have better real-time performance and are easy to analyze and process. The specific time-domain features selected include mean, median, variance, skewness, maximum, and minimum.

Since the dimensions of the signals collected by different types of sensors are different, there are many types of features extracted from the same type of sensor signals. This leads to large numerical differences between the extracted features. If these features are directly used for classification and recognition, the role of features with a larger value will be amplified and become the main factor, and the role of features with a smaller value will be reduced and become a secondary factor. This makes those features with good classification performance but small values fail to exert their excellent performance, which will cause greater interference to the overall classification and recognition performance. In order to solve this problem, different types of features need to be standardized to a certain range. In this article, each type of feature is standardized by feature category. The formula is

$$x_k^* = \frac{(x_k - \mu)}{\sqrt{\sigma}}. \tag{4}$$

μ is the average value of this type of feature, and $\sqrt{\sigma}$ is the standard deviation.

4.3.2. Feature Dimensionality Reduction. Multiple sensors are used to detect dance moves in this article. Except for the pressure sensor, there are 5 signals in total. Six features are extracted for each sensor signal, and there are 30 features in total. There are still too many such features, and the information expressed between different features is likely to overlap. The higher feature dimension not only results in a substantial increase in the amount of calculation but also is not conducive to reducing the redundant information of the feature, thereby reducing the classification performance. In order to further reduce the feature dimensions to obtain features that are more suitable for classification, it is necessary to perform dimensionality reduction processing on these feature vectors.

There are two main ways of feature dimensionality reduction. One is feature screening. The key issue is how to determine the importance of features and how to select them. The other is the combination optimization of features. The original features are reorganized through transformation to produce new features, that is to say, the new features are a function of the original features. This article uses the latter to optimize features.

Principle component analysis (PCA) is used to optimize the combination of feature vectors to achieve the purpose of dimensionality reduction processing. PCA is a widely used and efficient dimensionality reduction method. It uses the covariance matrix according to the principle of maximizing variance to convert multiple original features into a small number of representative and better integrated features, so as to reduce the dimensionality of the feature vector. It is generally ensured that these reconstructed comprehensive features can reflect more than 85% of the information contained in the original features and that the various features are independent of each other to avoid overlap of information.

Write the standardized features obtained from the training set data in the form of (3-4) matrix Z.

$$Z = \left(z_{ij}\right). \tag{5}$$

Take each signal feature as a column, and take the number of samples collected as a row. The number of rows is the number of samples, and the number of columns is 30.

Input the original feature data to calculate

$$\sum_{i=1}^{p} \lambda_i = 30. \tag{6}$$

Arrange the calculated eigenvalues λ from large to small, and the first five eigenvalues are 9.0624, 6.6358, 6.0065, 3.1528, and 1.0865. According to the principle of selection of principal components, this article selects the first five principal components, and the total contribution rate can be obtained:

$$\frac{\sum_{i=1}^{5} \lambda_i}{\sum_{i=1}^{30} \lambda_i} = 88.06\%. \tag{7}$$

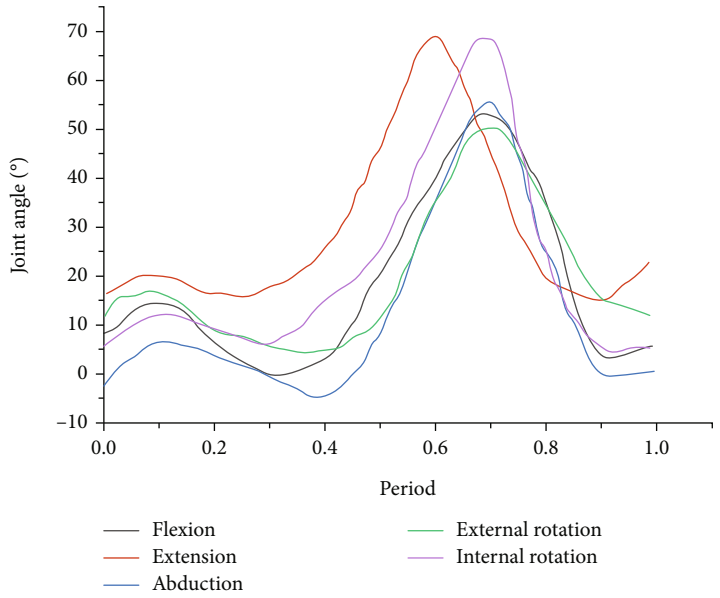


Figure 1: Knee joint angle signals under five lower limb human motions of ethnic dance.

After inputting the collected data, the eigenvector corresponding to the eigenvalue $\{\lambda 1, \cdots, \lambda 5\}$ is calculated as shown in Table 1. Among them, the feature column vector α_i corresponds to the feature value λ_i, each column corresponds to the coefficient of the respective principal component, and each row value corresponds to the weight of the original feature of the corresponding sensor signal in the principal component. Substituting the values of the eigenvectors in Table 1 into formula (8), the corresponding 5 principal components $F1, \cdots, F5$ can be obtained.

$$F_k = a_{1k}x_1 + a_{2k}x_2 + \cdots + a_{pk}x_p. \tag{8}$$

The contribution rate of each principal component is 30.38%, 22.19%, 20.08%, 10.54%, and 3.67%, and the contribution rate histogram is shown in Figure 2. The PCA method is used to compress 30 features into 5 comprehensive features. Moreover, the compressed 5 kinds of comprehensive features contain 86.86% of the information of the original features.

In order to determine the classification performance of the acquired features, this paper conducts a comparative experiment on the features before and after optimization. The features proposed from the same training group and verification group signals will also use the SVM algorithm (using radial basis kernel function) to train and recognize the action model. The results shown in Table 2 are obtained. The results show that the feature extraction scheme used in this paper has obvious advantages in the accuracy and efficiency of recognition.

The overall process of national dance movement recognition data processing is shown in Figure 3.

4.4. The Lower Limb Human Motion Recognition Model of Ethnic Dance

4.4.1. SVM Model. SVM is a very powerful model that perform well on a variety of datasets. SVM allows the decision boundary to be complex, even if the data has only a few features. It performs well on both low-dimensional data and high-dimensional data, meeting the needs of the design of this paper. Therefore, the SVM model was selected to classify and recognize the five movements of the lower limb human motion of ethnic dance. SVM is a two-classifier. Suppose there are N samples, x_i is the sample point, and the category label is

$$y_i = \{-1, +1\}. \tag{9}$$

For positive samples,

$$w^T x_i + b \geq 0. \tag{10}$$

For negative samples,

$$w^T x_i + b < 0. \tag{11}$$

So, in summary,

$$y_i\left(w^T x + b\right) \geq 0. \tag{12}$$

But in actual use, most samples are not so obvious, and there may be a small amount of confusion between the two types of samples. In order to solve this problem, the researchers introduced the concepts of slack variables and penalty factors to deal with samples that violated the inequality (13), thereby obtaining the following optimization problem:

$$\min\left(\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i\right), \tag{13}$$

Table 1: The first 5 feature vectors of signal features.

$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$
-0.14	0.19	-0.26	0.05	-0.03
-0.23	0.17	-0.16	0.01	0.02
-0.21	0.19	-0.16	0.13	0
0.29	-0.02	-0.13	0.16	-0.08
0.25	-0.02	-0.11	-0.33	0
-0.22	0.18	-0.14	0.13	0
0.33	0.07	-0.03	-0.17	-0.03
0.18	0.32	0.16	0.03	-0.07
0.28	0.25	-0.01	-0.02	-0.1
0.27	-0.09	-0.08	-0.28	-0.01
0.17	-0.23	0.18	-0.19	0.19
0.24	0.27	-0.02	0.17	-0.07
-0.04	0.24	0.33	0.05	-0.02
0.27	0.17	-0.15	-0.04	0.2
0.17	0.33	0.14	0.04	0.19
-0.19	0.05	0.31	0.06	-0.14
-0.1	0.02	-0.16	-0.25	-0.47
0.18	0.3	0.19	0.12	0.25
0.14	-0.13	-0.01	0.39	-0.18
0.01	0.3	-0.01	-0.09	0.42
-0.08	-0.1	-0.08	-0.11	0.12
0.24	-0.08	-0.11	0.25	-0.27
0.14	0.17	-0.13	0.11	-0.24
0.25	0.17	-0.14	-0.13	-0.06
0.17	-0.19	0.21	0.28	0.08
-0.07	0.11	0.29	-0.21	-0.24
0.13	0.1	0.37	-0.03	-0.22
0.22	-0.13	-0.01	0.4	0.09
0.06	-0.27	0.13	-0.16	0.07
0.09	0.17	0.37	0.02	-0.19

$$\text{s.t.} y_i\left(w^T x_i + b\right) \geq 1 - \xi_i. \tag{14}$$

ξ_i is a slack variable, and a small amount of misclassification is allowed, $\xi_i \geq 0, i = 1, 2, \cdots, N$. C is the penalty factor, which is a manually set parameter greater than 0, used to punish samples that violate the inequality, so as to avoid sample misclassification caused by slack variables. By introducing slack variables and penalty factors, the application value of SVM is further improved.

There are 10 SVM subclassifiers. The training and use process of the SVM recognition model is shown in Figure 4.

The MATLAB platform is used to build the SVM motion recognition model. To use SVM, the kernel function must be selected first, but there is still no unified standard for the selection rules of kernel function. In order to select the appropriate kernel function, this paper uses the same training group and validation group data to compare and analyze the performance of the kernel function in the SVM model. The experimental results are shown in Table 3. The results show that, for the action recognition model in this paper, the radial basis kernel function has the highest recognition rate, and at the same time, it is better in recognition efficiency. Therefore, this paper finally chooses the radial basis kernel function in the SVM model.

After determining the structure and kernel function of the SVM recognition model, the database containing the training set and the verification set is input into the SVM model for model training and optimization. First, the 5-dimensional feature vector extracted from the training dataset is input into the SVM model for training, so as to optimize the parameters in the model to improve the classification performance. After completing the training of the model, use the validation data to test the recognition accuracy of the SVM model. The main algorithm steps are as follows:

(1) Import feature database

(2) Select the radial basis kernel function

(3) Find the optimal c and g parameters

(4) Use training data to train the recognition model to determine various parameters

(5) Test the SVM recognition model

(6) Draw a classification effect diagram

After the above steps, the training task of the SVM action recognition model is finally completed, and then, the test group motion data is input to the model for final recognition effect verification. Experiments show that the model has an average recognition rate of 85.6% for the five actions of flexion, extension, abduction, external rotation, and internal rotation (codenamed 1~5). The recognition rate of each action is shown in Table 4.

4.4.2. MCM. There will be some misrecognition in the recognition of human lower limb movements, especially for similar movements. There are specific rules for the transition between the movements of the lower limbs of the human body. We use this law to improve the accuracy of human lower limb human motion recognition. The Markov chain model (MCM) is used to solve this problem.

Set 5 gait movements: flexion, extension, abduction, external rotation, and internal rotation. The code names are 1~5. Remember the corresponding probability of each action at the last moment:

$$P^T(0) = (p_1, p_2, p_3, p_4, p_5). \tag{15}$$

The corresponding probability of each action at the current moment is

$$P^T(1) = (p_1(1), p_2(1), p_3(1), p_4(1), p_5(1)). \tag{16}$$

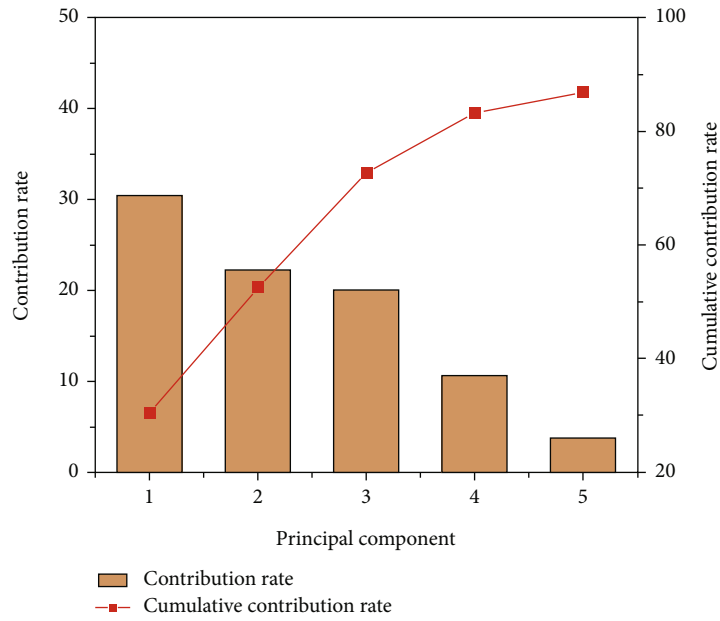


Figure 2: Contribution rate of principal component of signal feature.

Table 2: Signal feature performance comparison.

Feature extraction scheme	Number of features	Recognition time (ms)	Recognition accuracy rate (%)
Comparison plan	9	1.05	82.3
Original feature	30	2.49	82.1
Features after optimization	5	0.48	84.6

The transition probability between gait actions is

$$P = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}. \quad (17)$$

Therefore, the predicted probability of each action at the current moment is

$$P^T(1) = P^T(0)P. \quad (18)$$

The accuracy of the current gait action probabilities predicted by MCM using the previous action probabilities is poor and cannot be directly used as the result of action recognition. This article mainly uses it to optimize the recognition results of SVM and eliminate some obvious misrecognitions, thereby improving the recognition accuracy.

4.5. Model Optimization. Multiply the processed votes corresponding to the current actions obtained by SVM by the probabilities of the current actions predicted by MCM to correct the number of votes obtained by SVM. The action corresponding to the maximum value of the final result is regarded as the current action. The number of votes for each gait action obtained and processed by the SVM recognition model is

$$A = \{a_1, a_2, a_3, a_4, a_5\}. \quad (19)$$

The probability of each current action predicted by MCM is

$$B = \{b_1, b_2, b_3, b_4, b_5\}. \quad (20)$$

The value after fusion is

$$C = \{c_1, c_2, c_3, c_4, c_5\}. \quad (21)$$

The elements are

$$c_i = a_i b_i. \quad (22)$$

In order to improve the accuracy of action recognition as much as possible, a condition for optimizing the SVM model using the MCM mechanism is set. The rules are as follows: if the number of votes recognized by the SVM as a certain action is equal to 4 (the maximum number of votes), the result of the action is directly output, and MCM is no longer used for processing; If the number of votes identified as an action is less than 4, the last identification is made. If the number of votes for a certain action is 4, the recognition result of the SVM model is optimized using MCM. In this way, it can optimize the situation where the SVM

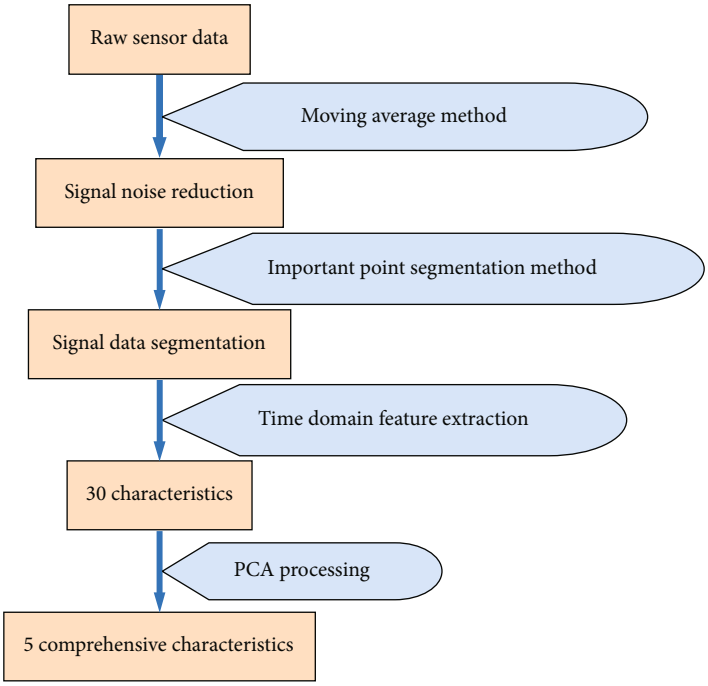


Figure 3: Data processing flow of the sensor.

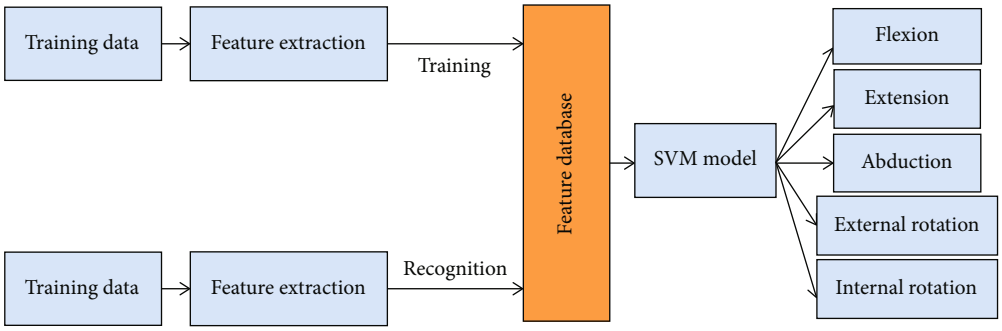


Figure 4: Action recognition model framework based on SVM.

Table 3: Performance comparison of kernel functions.

Kind of kernel function	Average recognition time (ms)	Average recognition rate (%)
Radial basis kernel function	0.57	85.02
Sigmoid kernel function	0.64	82.35
Linear kernel function	0.7	70.94
Polynomial kernel function	0.51	65.43

recognition accuracy rate is low, and it will not affect the recognition result when the SVM recognition accuracy rate is high and finally achieve the purpose of improving the recognition accuracy rate of the lower limb human motion of the human body. The optimized action recognition process is shown in Figure 5.

After determining the action recognition optimization plan, the same experimental data is used to test the effect of MCM on SVM optimization. Similarly, the 5-dimensional feature vector is classified and recognized using the SVM model, and the recognition results of the SVM model are optimized using MCM. The average recognition accuracy is improved to 93.4%. The accuracy comparison of these two recognition schemes is shown in Table 5. From the comparison of the histogram, it can be seen directly that the recognition accuracy of the action recognition model optimized by MCM has been greatly improved.

5. Result and Discussion

This paper presents a new sensor-based research on the recognition method of ethnic dance lower limb human motion in rural areas. On the basis of extracting the temporal features of ethnic dance lower limb human motion, the SVM algorithm is used to identify the current action, and the MCM is used to optimize the recognition result. Three angle

Table 4: Action recognition accuracy rate of SVM model.

Dance motion	Recognition rate (%)
Flexion	90.5
Extension	84.9
Abduction	82.8
External rotation	86.3
Internal rotation	83.5

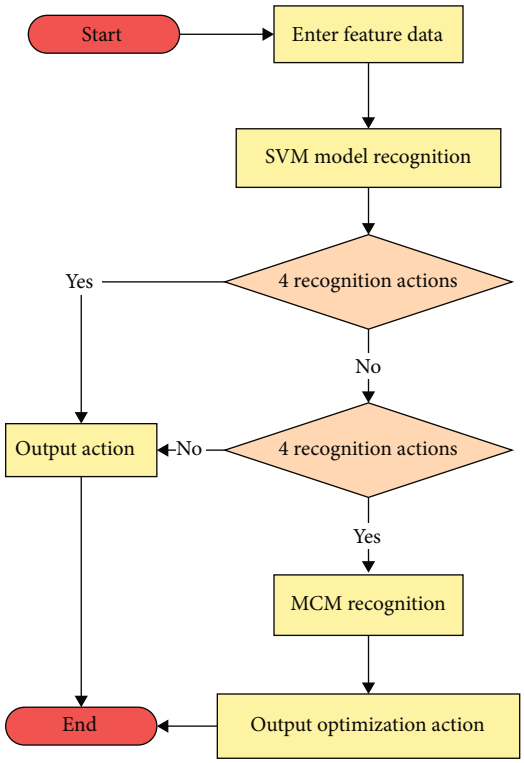


Figure 5: The overall process of recognition.

Table 5: Recognition results after fusion of SVM and MCM.

	Recognition rate (%)	
Dance motion	SVM	SVM-MCM
Flexion	90.0	93.6
Extension	85.2	94.2
Abduction	82.1	94.9
External rotation	86.3	92.9
Internal rotation	82.7	91.5

sensors are used to detect the angle signals of the ankle, knee, and hip joints. An acceleration sensor is placed on the underside of the thigh, and a pressure sensor is placed on the sole of the hind foot. The moving average filtering method with a small amount of calculation is used to filter the signal noise. The time-domain feature extraction method is used to extract each signal feature. The principal component analysis method is adopted to reduce the dimension of the feature vector. The SVM model was used to identify the five lower limb human motions of ethnic dance, with an average recognition accuracy rate of 85.6%. MCM theory is used to optimize the recognition results of the SVM model. Experimental results prove that the optimized action recognition model has a higher accuracy rate, with an average recognition accuracy rate of 93.4%.

This paper is an exploratory research and proposes an ethnic dance lower limb human motion recognition scheme in rural areas. In addition, it is important that the number of sensors can be appropriately increased without excessively increasing the amount of calculation to improve the accuracy of action recognition.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The author declares no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Guangxi University Humanities and Social Science Key Research Base "Research on Art and Culture Construction of New Rural Communities in Guangxi—Taking Southeast Guangxi as an Example." The project number is 2020YJJD0009.

References

[1] G. D. Maclean, *An Examination of the Characteristics of Short Term International Midwifery Consultants [M]*, University of Surrey (United Kingdom), 1998.

[2] H. V. A. Chenarlog and F. Razzazi, "Multi-stream 3D CNN structure for human action recognition trained by limited data," *IET Computer Vision*, vol. 13, no. 3, pp. 338–344, 2019.

[3] H. Kim, Y. J. Shin, and J. Kim, "Design and locomotion control of a hydraulic lower extremity exoskeleton for mobility augmentation," *Mechatronics*, vol. 46, no. 46, pp. 32–45, 2017.

[4] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[5] L. Peng, L. Chen, X. Wu, H. Guo, and G. Chen, "Hierarchical complex activity representation and recognition using topic model and classifier level fusion," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1369–1379, 2017.

[6] D. Wang, A. H. Tan, and D. Zhang, "Non-intrusive robust human activity recognition for diverse age groups," in *IEEE/Wic/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 368–375, Singapore, 2015.

[7] R. M. Gibson, A. Amira, N. Ramzan, P. Casaseca-de-la-Higuera, and Z. Pervez, "Multiple comparator classifier framework for accelerometer-based fall detection and diagnostic," *Applied Soft Computing*, vol. 39, no. C, pp. 94–103, 2016.

[8] A. Thakkar and P. Pareek, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," *Artificial Intelligence Review*, vol. 25, pp. 2259–2322, 2021.

[9] C. Dhiman and D. K. Vishwakarma, "View-invariant deep architecture for human action recognition using two-stream motion and shape temporal dynamics," *IEEE Transactions on Image Processing*, vol. 29, pp. 3835–3844, 2020.

[10] D. Ludl, T. Gulde, and C. Curio, "Enhancing data-driven algorithms for human pose estimation and action recognition through simulation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 3990–3999, 2020.

[11] Z. Du and H. Mukaidani, "Linear dynamical systems approach for human action recognition with dual-stream deep features," *Applied Intelligence*, vol. 3, pp. 1–19, 2022.

[12] J. Sedmidubsky and P. Zezula, "Efficient combination of classifiers for 3D action recognition," *Multimedia Systems*, vol. 27, no. 5, pp. 941–952, 2021.

[13] M. Nazarahari, A. Noamani, N. Ahmadian, and H. Rouhani, "Sensor-to-body calibration procedure for clinical motion analysis of lower limb using magnetic and inertial measurement units," *Journal of Biomechanics*, vol. 85, no. 27, pp. 224–229, 2019.

[14] D. R. Hume, A. Navacchia, A. A. Ali, and K. B. Shelburne, "The interaction of muscle moment arm, knee laxity, and torque in a multi-scale musculoskeletal model of the lower limb," *Journal of Biomechanics*, vol. 76, pp. 173–180, 2018.

[15] N. Dey, A. S. Ashour, F. Shi, S. J. Fong, and R. S. Sherratt, "Developing residential wireless sensor networks for ECG healthcare monitoring," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 442–449, 2017.

[16] P. Tamburini, M. C. Bisi, and R. Stagni, "Frequency evaluation of gait trunk acceleration signal: a longitudinal study," *Gait & Posture*, vol. 57, pp. 8-9, 2017.

Hindawi
Journal of Sensors
Volume 2022, Article ID 3590973, 17 pages
https://doi.org/10.1155/2022/3590973

Research Article

Detection and Classification of ADHD from EEG Signals Using Tunable Q-Factor Wavelet Transform

R. Catherine Joy,[1] S. Thomas George,[2] A. Albert Rajan,[3] M. S. P. Subathra,[4] N. J. Sairamya,[5] J. Prasanna,[2] Mazin Abed Mohammed,[6] Alaa S. Al-Waisy,[7] Mustafa Musa Jaber,[8,9] and Mohammed Nasser Al-Andoli[10]

[1]*Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore 641114, India*
[2]*Department of Biomedical Engineering, Karunya Institute of Technology and Sciences, Coimbatore 641114, India*
[3]*Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences, Coimbatore 641114, India*
[4]*Department of Robotics Engineering, Karunya Institute of Technology and Sciences, Coimbatore 641114, India*
[5]*Department of Electrical and Computer Engineering, Université du Québec à Trois-Rivières, 3351 Bd des Forges, Trois-Rivières, QC, Canada G8Z 4M3*
[6]*College of Computer Science and Information Technology, University of Anbar, 31001 Ramadi, Anbar, Iraq*
[7]*Computer Technologies Engineering Department, Information Technology College, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq*
[8]*Department of Computer Science, Dijlah University College, Baghdad, Iraq*
[9]*Department of Computer Science, Al-Turath University College, Baghdad, Iraq*
[10]*Computer Science & Information Systems Department, Faculty of Science, Sa'adah University, Sa'adah, Yemen*

Correspondence should be addressed to Mohammed Nasser Al-Andoli; mnalandoli@saada-uni.edu.ye

Received 12 July 2022; Accepted 2 September 2022; Published 15 September 2022

Academic Editor: Mohit Mittal

The automatic identification of Attention Deficit Hyperactivity Disorder (ADHD) is essential for developing ADHD diagnosis tools that assist healthcare professionals. Recently, there has been a lot of interest in ADHD detection from EEG signals because it seemed to be a rapid method for identifying and treating this disorder. This paper proposes a technique for detecting ADHD from EEG signals with the nonlinear features extracted using tunable Q-wavelet transform (TQWT). The 16 channels of EEG signal data are decomposed into the optimal amount of time-frequency sub-bands using the TQWT filter banks. The unique feature vectors are evaluated using Katz and Higuchi nonlinear fractal dimension methods at each decomposed levels. An Artificial Neural Network classifier with a 10-fold cross-validation method is found to be an effective classifier for discriminating ADHD and normal subjects. Different performance metrics reveal that the proposed technique could effectively classify the ADHD and normal subjects with the highest accuracy. The statistical analysis showed that the Katz and Higuchi nonlinear feature estimation methods provide potential features that can be classified with high accuracy, sensitivity, and specificity and is suitable for automatic detection of ADHD. The proposed system is capable of accurately distinguishing between ADHD and non-ADHD subjects with a maximum accuracy of 100%.

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is one of the prevailing neuropsychiatric disorders among children, and it frequently persists into adulthood [1, 2]. The word-wide study shows that, about 5-12% of prevalence of ADHD is observed among school-going children, and more manifestation is experienced among male children [3–5]. This

disorder has the subtypes such as predominantly inattentive, predominantly hyperactive-impulsive, and the combined type with the primary symptoms of inattention, impulsivity, and hyperactivity [6–8]. Early detection and identification of this disorder and treating in an early stage will be extremely beneficial to children, parents, and especially community health. Currently, clinical interviews, observations, and ratings from multiple sources such as parents and teachers are used to examine and diagnose ADHD [9–11]. The traditional clinical evaluation procedures are time consuming and are subject to ambiguity. Therefore, there is a great need for objective clinical diagnostic methods from the biological signals that reflect the behaviors of ADHD and its subtypes.

Electroencephalography (EEG) is the record of the electrical activities of a human brain, which can reveal a great deal about physiology and pathology. EEG signals have been employed in the diagnosis of several neurological illnesses by extracting unique features and classifying them with different classifiers in automated detection systems. Neurophysiological disorders such as alcoholism [12], dementia [13, 14], epileptic seizure [15], schizophrenia [16, 17], Parkinson's disease [18, 19], and depressive disorder [20, 21] are some of the areas where EEG signals are employed in automatic detection. The EEG signals of ADHD children are different from that normal child in terms of complex randomness, amplitude, and frequency. Researchers have employed several feature extraction techniques and classifiers to analyse EEG signals in the identification of ADHD [22–25]. Researchers have experimented several machine learning algorithms and nonlinear feature extraction approaches such as entropy estimators and classifiers such as support vector machine (SVM), multilayer perceptron, and k-nearest neighbor (KNN) [26–29] to detect ADHD using EEG data. These techniques suffer from higher computational complexity and lower classification accuracy.

A quantifiable brain reaction that happens as a direct result of a sensory, cognitive, or motor event is known as an event-related potential (ERP). Mueller et al. used ERP features to analyse 75 ADHD and 75 normal children and have classified ADHD and normal subjects with a classification accuracy of 91% [30]. Different authors worked on the ERP and extracted probable features that can help detect ADHD, and classification is done with a multilayer neural network and categorized with an accuracy close to 96.7%. [27, 31–33]. While some authors investigated the application of complex deep learning algorithms to diagnose ADHD from EEG signals, others explored more efficient machine learning methodologies. For most of the ADHD detection, the authors extracted nonlinear features and classified with standard classifiers such as support vector machine (SVM), multilayer perceptron, and KNN [26, 28, 29, 34]. A deep convolutional neural networks and deep learning networks were experimented to diagnose ADHD in adults and children [35–37].

Literature reveals that wavelet transform techniques have higher computational efficiency, and they have the added benefit of being able to distinguish tiny details in a signal. Ahmadlou and Adeli [38] employed a wavelet-synchronization pattern recognition methodology with RBF neural network classifier, to detect ADHD with a maximum accuracy of 95.6%. Sadatnezhad et al. [31] used fractal dimension, AR model, and EEG band power to diagnose ADHD children and achieved a maximum classification of 86.4%. Allahverdy et al. [39] analysed EEG data with the nonlinear features extracted using fractal dimension methods and distinguished ADHD subjects with a classification accuracy of 86%. Ahmadlou and Adeli et al. employed the synchronization likelihood (SL) and fuzzy synchronization likelihood (FSL) frameworks to assess functional connectivity and achieving classification accuracies of 87.5% and 95.6%, respectively, for a synchronization pattern in the theta and delta frequency bands [38, 40]. These wavelets transform methods could not achieve the maximum classification accuracy as they could not dynamically adjust the Q value and had lower reconstruction capabilities. To classify ADHD versus normal subjects, these studies largely used artificial intelligence approaches. Based on the literature, different authors expressed that classification accuracy has to be further improved and the computational complexity has to be reduced. Moreover, challenges prevail in identifying the better feature extraction technique and applying the best classifier algorithm for achieving maximum classification accuracy in ADHD diagnostic methods.

To address these difficulties, the authors have experimented an efficient algorithm using tunable Q-wavelet transform (TQWT) with Katz and Higuchi fractional dimension method, that is lighter in computational complexity and with an ANN classifier that provides a maximum classification accuracy. Children with ADHD are identified by the experts and their EEG signals recorded under eyes-open and eyes-closed states are used for this analysis. Potential features are derived from Katz and Higuchi fractal dimensions, which are estimated from the segmented EEG signal subbands. The features extracted through the fractional dimension techniques are classified using the ANN classifier which is a proven effective classifier. The results show that the suggested method is effective in classifying ADHD and normal subjects EEG signals. The main contributions list can be summarized as follows:

(i) The combination of TQWT with Katz and Higuchi fractional dimension method is proven to be an efficient feature extraction method for ADHD detection

(ii) The potential features extracted with the Katz and Higuchi fractional dimension techniques with an ANN classifier brought out with a maximum classification accuracy of 100%

(iii) As the proposed system is lighter in computation with maximum classification accuracy, it can be a resourceful technique for clinical detection of ADHD from the EEG signals

This work is composed of four major sections. Section 2 presents the Materials and Methods, which include data acquisition and feature extraction methods. Section 3 proposes the Results and Discussion, which includes the

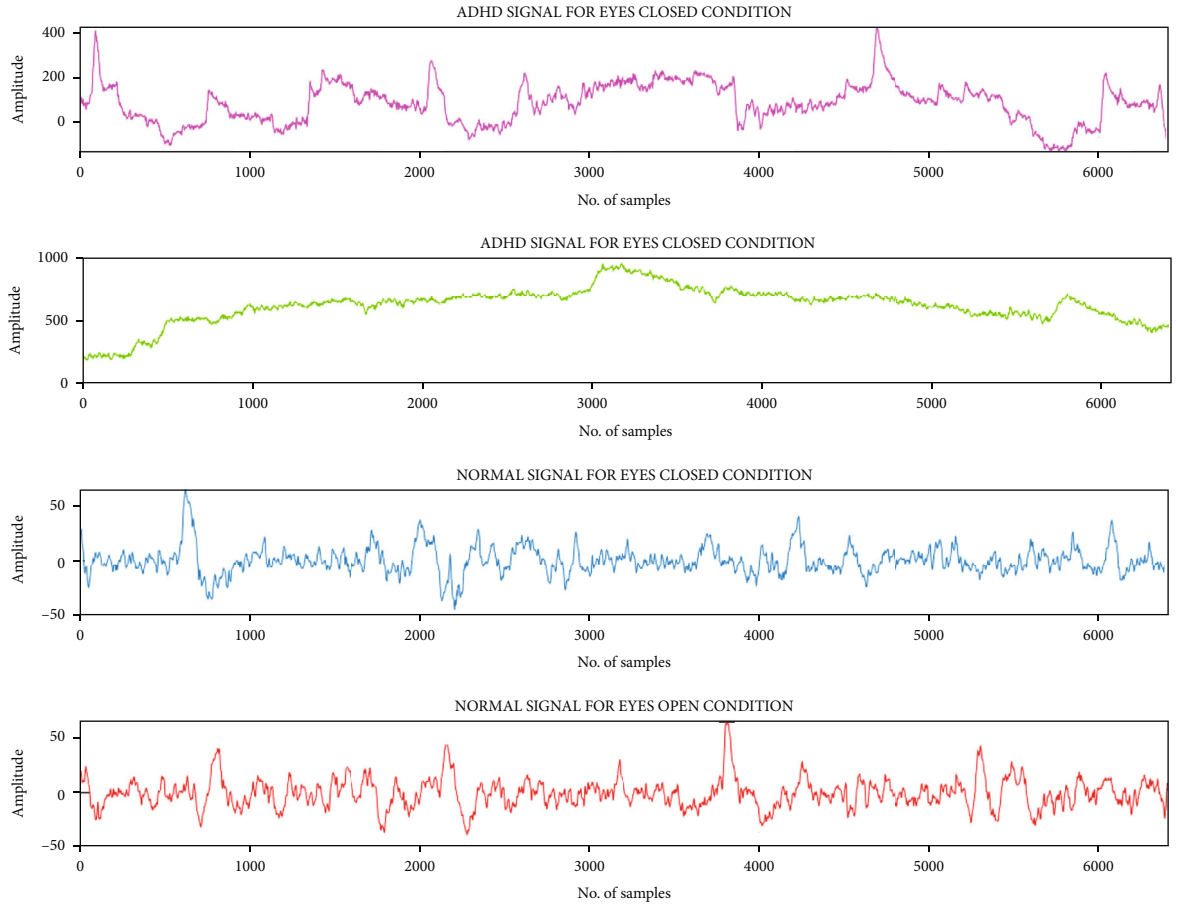


Figure 1: Single-channel EEG signals for ADHD and normal subjects in eyes-closed and eyes-open condition.

analysis of various classifier algorithms, performance metrics, and discussion on comparing similar works. Finally, the Conclusion and future works are presented in Section 4.

2. Materials and Methods

2.1. Data Acquisition. The EEG data set for the proposed approach is created with 5 subjects of ADHD and 5 subjects of normal, in each category under eyes-closed and eyes-open resting state. The EEG signals were recorded from the children age group 7 to 12, after getting parental consent and the children's consent [25]. Using the unipolar setup, the EEG signals of the individuals are measured using the 10–20 electrode placement system. Individual scalps are carefully prepared with a contact impedance of less than 5KΩ for EEG signal measurement. With a sample rate of 256 Hz, each EEG signal consisting of 6400 sampling points was recorded using a 16-electrode unipolar montage. Each signal was captured for 300 seconds with a 24-bit resolution. Signals which acquired are divided into 25 seconds in the eyes-closed resting state and eyes-open resting state in both ADHD and normal subjects. The EEG signals of ADHD and normal subjects are presented in Figure 1.

The acquired signals are preprocessed using MATLAB (MatLabR2018a) to determine the needed range of signals from each channel. Visual inspection and computerized review are carried out with the aid of specialists and a range of signals that are not acceptable for analysis and further processing are removed. A bandpass filter with cutoff frequencies of 1 Hz and 60 Hz is used to reduce signal noise and to eliminate other artefacts and noise during eye blinking. The power frequency noise is suppressed using a 50 Hz notch filter. Experiments are conducted on a laptop with 4 GB of RAM, a 3.2 GHz CPU, and intel core processor. MATLAB 2018a is used to execute the simulations, and the statistical data are recorded for analysing various performance measures. The functional blocks of the proposed methodology for classifying ADHD and normal subjects using the TQWT algorithm is presented in Figure 2.

In this proposed methodology, the EEG signals of both ADHD and normal categories are decomposed into 15 levels of subbands by using TQWT technique. The unique fractal dimension features such as katz and higuchi features are extracted from the decomposed subbands reflecting the ADHD and normal behavior. These features are fed as input to the different classifiers such as linear discriminant, logistic regression, support vector machine, artificial neural networks, and ensemble techniques are experimented to compare the performance of each classifier. The best classifier algorithm with higher classification accuracy is evaluated for choosing the better combination to perform the feature

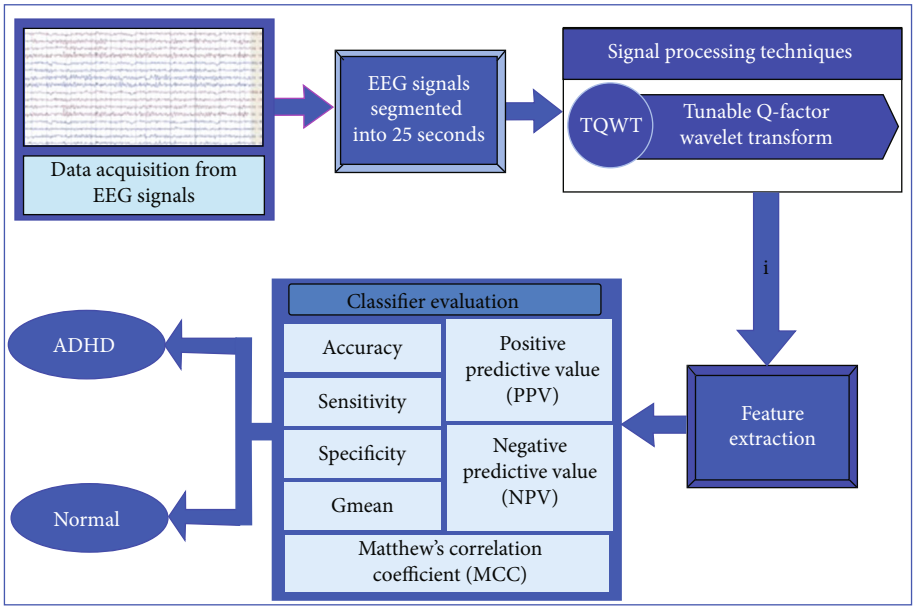


Figure 2: Functional blocks of the proposed methodology.

extraction and classification. The performance of the proposed methodology is verified with different performance metrics for ensuring its best performances.

2.2. Tunable Q-Factor Wavelet Transform (TQWT). The ratio of the centre frequency to the bandwidth of the filters employed in the transform is known as the Q-factor of a wavelet transform. TQWT has the property of fast decomposition and perfect reconstruction which makes it suitable for application in many biomedical signal processing problems. TQWT's efficient decomposition and perfect reconstruction properties make it well-suited to a wide range of biological signal processing applications [41]. The TQWT is a wavelet transform that is analogous to the rational-dilation wavelet transform and has been used to investigate EEG signals [12, 42, 43]. The TQWT provides perfect reconstruction of the signal and the energy of the signal is divided into subbands by the TQWT coefficients. It is done with a discrete wavelet transform that performs a double-channel multirate filter bank with low and high-pass filters. With an adjustable Q-factor and a powerful transform for oscillatory signal analysis, this approach is suited for the discrete-time signal analysis [42]. TQWT's fundamental parameters are its Q-factor (Q), redundancy (r), and the number of levels of decomposition (j) which allow it to analyse signals with a diversity of oscillatory characteristics [44]. TQWT filters are straightforward to reconstruct and implement since they are built up of nonrational transfer functions utilising a Fast Fourier Transform (FFT) with an adjustable Q-factor. In our proposed work, the Q-value factors are modified between 1 and 10, and the classification results are analysed to find the best Q-factor value. The TQWT approach applied to this work is depicted in Figure 3.

Figure 3 shows the different stages of decomposition applied with the TQWT analysis and synthesis filter banks [45]. An EEG signal is divided into j levels by iteratively applying two-channel filter banks on the signal. At each level of decomposition, the input signal $x[n]$ with sampling frequency f_s is decomposed into a high-pass subband signal $x_1[n]$ and a low-pass subband signal $x_0[n]$, with sampling frequencies of αf_s and βf_s, respectively. Here α and β refer scaling factors for the filter banks. Selesnick [42] elaborated a comprehensive description of scaling parameters and proposed that the scaling parameters must fulfil the following criteria to limit redundancy while ensuring perfect reconstruction.

$$\begin{aligned} 0 < \beta \le 1, \\ 0 < \alpha < 1, \\ \alpha + \beta = 1. \end{aligned} \tag{1}$$

The characteristic equation of TQWT can be represented as follows:

$$H_0(\omega) = \begin{cases} 1, |\omega| \le (1-\beta)\,\pi \\ \theta\left(\dfrac{\omega + (\beta-1)\pi}{\alpha+\beta-1}\right), (1-\beta)\pi \le |\omega| < \propto \pi, \\ 0, \alpha\pi \le |\omega| \le \pi \end{cases} \tag{2}$$

$$H_1(\omega) = \begin{cases} 0, |\omega| \le (1-\beta)\,\pi \\ \theta\left(\dfrac{\alpha\pi - \omega}{\alpha+\beta-1}\right), (1-\beta)\pi \le |\omega| < \alpha\pi, \\ 1, \alpha\pi \le |\omega| \le \pi \end{cases} \tag{3}$$

$$\theta(\omega) = \frac{1}{2}\,(1 + \cos\omega)\sqrt{2 - \cos\omega} \text{ for } |\omega| \le \pi. \tag{4}$$

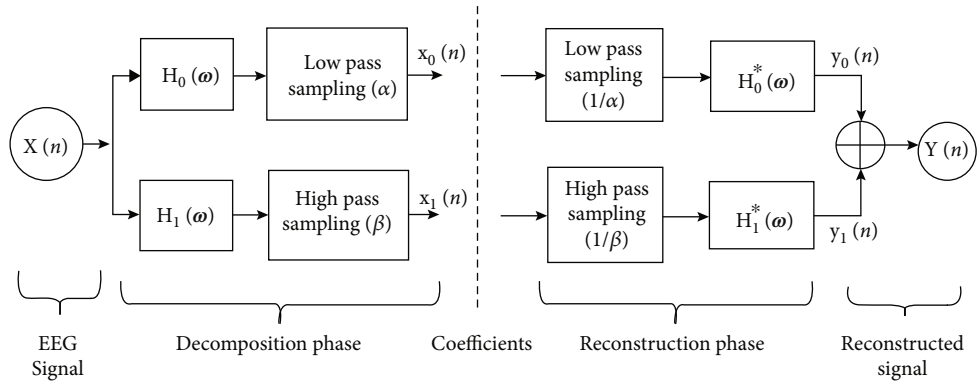


Figure 3: Flow diagram of TQWT analysis and synthesis filter banks.

The bands of $H_0(\omega)$ and $H_1(\omega)$ are constructed using transition function $\theta(\omega)$, which is derived from the Daubechies filter with two vanishing moments. To accomplish the perfect reconstruction criteria, the low-pass filter $H_0(\omega)$ and high-pass filter $H_1(\omega)$ can be assessed using the relation given in equation (2).

$$|H_0(\omega)|^2 + |H_1(\omega)|^2 = 1. \tag{5}$$

The decomposed EEG signals can be reconstructed with the use of a synthesis filter. The relationship between TQWT input parameters and scaling parameters α and β is related by;

$$r = \frac{\beta}{1-\alpha}, \tag{6}$$

$$Q = \frac{2-\beta}{\beta}. \tag{7}$$

The criterion of dominant frequency is used to select the appropriate value of decomposition levels (j). According to the dominant frequency criterion, the number of decomposition levels is kept in such a way that the decomposed subbands have the greatest correlation with substantial EEG frequency ranges.

In this paper, the following feature vectors are evaluated and analysed:

(a) *Q-factor*. The value of Q in TQWT determines the oscillatory behaviour of the signals. EEG signals, in particular, are highly oscillatory in nature and have a high Q-value. The theoretical definition of the Q-factor is expressed as $Q = (2-\beta)/\beta$ and $\alpha = 1 - (\beta/r)$. Based on the values of Q and r, the values of α and β are computed. The value of the Q-factor can be chosen based on the input signal behaviour because it reflects the oscillatory behaviour of the wavelet. If the proposed Q-value is compatible with the input signal's features, it can accurately extract useful information from the EEG signal.

(b) *The maximum number of levels* $j_{\max}$. The scaling parameters α and β, the number of samples (N) in the input signal are used to calculate $j_{\max}$, The maximum levels of decomposition,

$$j_{\max} = \frac{\log(\beta N/8)}{\log(1/\alpha)} \tag{8}$$

(c) *Oversampling rate/redundancy parameter (r)*. The resonance is controlled by the redundancy factor r, which allows the wavelet to be focused in time without affecting its shape. The oversampling rate is defined as r in this case $r = \beta/(1-\alpha)$. When analysing biological signals, the specific number $r = 3$ has been previously recommended [46]. As a result, throughout this research, the redundancy parameter r is chosen as 3

The wavelet transforms technique shall be applied to signals with little or no oscillatory characteristic with a low Q-factor. Most wavelet transforms, except for the continuous wavelet transform, are unable to adjust their Q-factor. This difficulty is solved by TQWT, which allows the Q-factor to be regulated. Moreover, TQWT has been widely employed to investigate a variety of physiological signals [47–49]. Due to the rational transfer functions, the filters are computationally efficient and hence provide direct representation in the frequency domain. As TQWT is a powerful tool for analysing oscillatory physiological signals with lesser computational complexity, the authors felt to apply this technique for the proposed work.

2.3. Feature Extraction. The EEG signals are complex and highly nonlinear in nature. Because of the nonlinear and intricate behaviour, nonlinear methods are appropriate tools for analysing brain dynamics and behaviours from the EEG signals. The Higuchi and Katz fractional dimension based feature extraction methods are more predominantly used for EEG signal analysis [50, 51]. In our work, the Higuchi and Katz nonlinear feature extraction techniques are used to identify the potential features that can help discriminate ADHD and normal subjects.

2.3.1. Higuchi Fractal Dimension. A fractal dimension is a tool for determining nonperiodic and irregular time series.

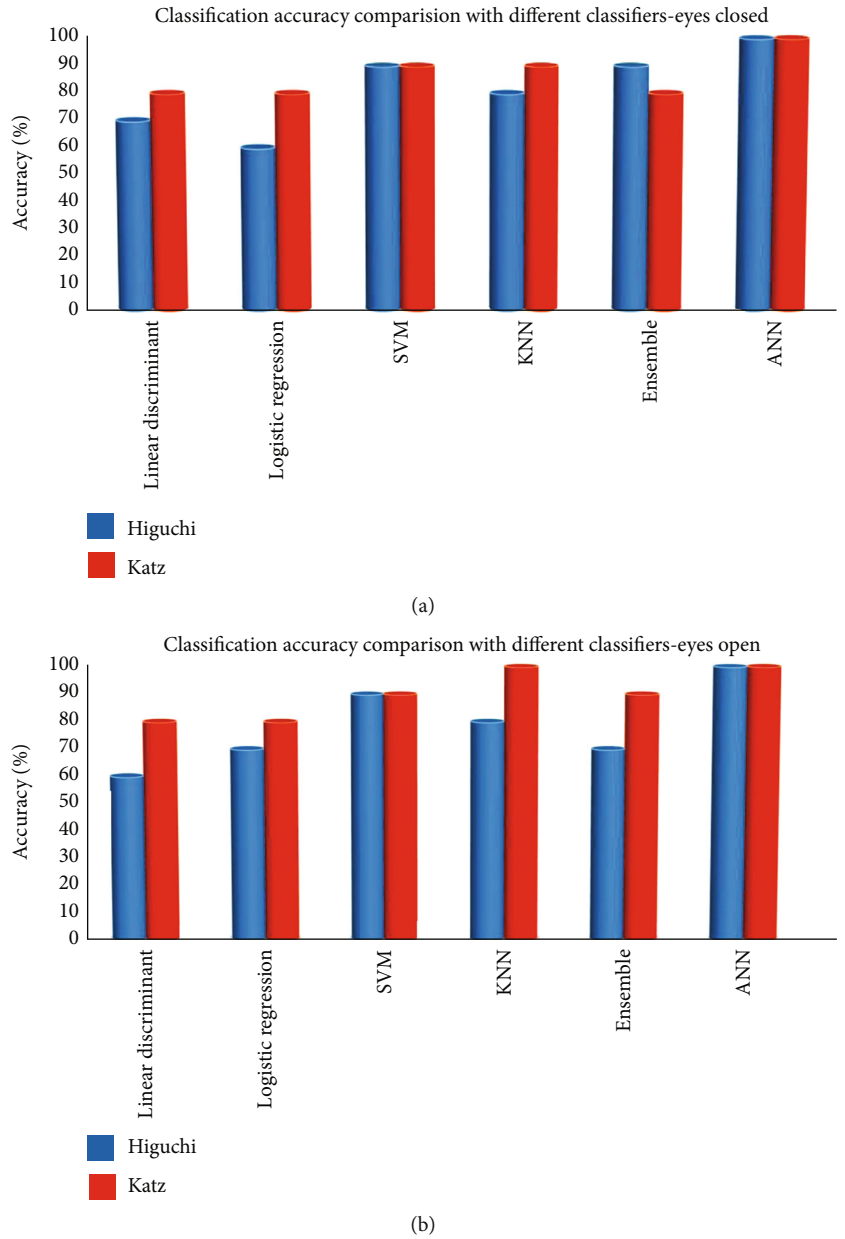


Figure 4: Classification accuracy comparison for different classifiers (a) eyes-closed condition (b) eyes-open condition.

The Higuchi fractional dimension has its high accuracy and efficiency in determining fractal dimensions based on curve length measurements. The time series of the EEG signal is segmented into k number of samples, and the mean length of the series/curve is measured using the segment of k samples [52, 53]. The FD Higuchi estimation can be obtained by following four steps for a finite set of time series, i.e., $S(\tilde{n})$; $\{\tilde{n} = 1, 2, \cdots N\}$; N is the number of points on the curve:

Stage 1. Generate k number of new time series, for values of k ranging from 1 to $k_{\max}$, calculate S_m^k from given time series data.

$$S_m^k = \left\{ S(m), S(m+k), S(m+2k), \cdots, S\left(m + \text{int}\left(\frac{N-m}{k}\right).k\right)\right\}. \tag{9}$$

In this, the discrete time interval between sample points is represented by k, and the initial time value is represented by $m\,(m = 1, 2, 3, \cdots\ldots, k)$.

Stage 2. The length $L_m(k)$ is calculated for each of the constructed time series S_m^k.

$$L_m(k) = \left[\left(\sum_{i=1}^{\text{int }(N-m/k)} |S(m+i.k) - S(m+(i-1).k|).\frac{N-1}{\text{int }(N-m/k).k}\right)\right].\text{k}^{-1}. \tag{10}$$

Here the curve length is given by $L_m(k)$ and $(N-1) * (\text{int }(N-m/k).\text{k})^{-1}$ is the normalization factor.

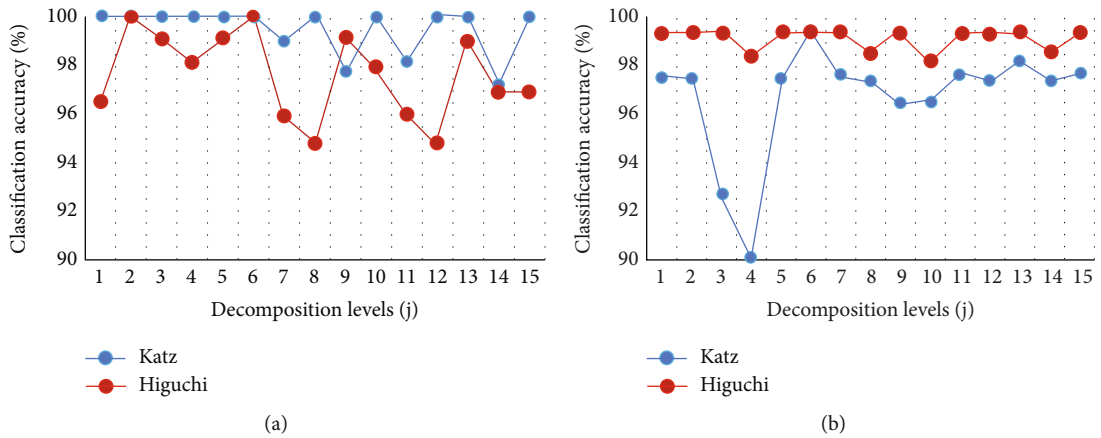


FIGURE 5: Classification accuracy for different decomposition levels for EEG signals under (a) eyes-closed state (b) eyes-open state.

Stage 3. The average length $L_{avg}(k)$ of the curve is calculated by the following equation for each interval of k:

$$L_{\text{avg}}(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k) \tag{11}$$

For all values of k ranging from 1 to $k_{\max}$, average length calculation is repeated.

Stage 4. The length of the total average curve, $L_{avg}(k)$ is proportional to k^{-D}, where D is the Higuchi fractal dimension (HFD). The slope of the least-squares linear best fit is the estimation of the fractal dimension in the curve of $\ln(L_{avg}(k))$ versus $\ln(1/k)$ in the curve [52]. The parameter time interval k determined the HFD value. In this analysis, we employed k in a certain range of values, which resulted in a specific HFD value.

2.3.2. Katz's Fractal Dimension. In the Katz fractional dimension method [54], the ratio of the curve's total length to the line for the maximum Euclidean distance from the starting point line. In general, the planar curve's fractal dimension, FD_{katz}, is determined by the given equation,

$$\text{FD}_{\text{katz}} = \frac{\log(L)}{\log(d)}, \tag{12}$$

where L denotes the overall length of the curve, or the sum of distances between successive points, and d denotes the diameter, which is calculated as the distance between the first and farthest point on the sequence and given as,

$$L = \sum_{i=1}^{N} \|w_{i+1} - w_i\|, \tag{13}$$

$$d = \max \|w_i - w_1\|. \tag{14}$$

Here, the Euclidean distance is denoted by "$\|.\|$". The fractal dimension (FD) compares the number of units that make up a curve to the smallest number of units required to generate a structure with the same spatial area. The measurement units used to compute FDs have an impact on the results. The average step or average distance between successive points, "a" is created as a general unit or yardstick in Katz's approach [50, 55]. According to Katz's approach, the fractal dimension (FD_{katz}), is expressed as:

$$\text{FD}_{\text{katz}} = \frac{\log(L/a)}{\log(d/a)} = \frac{\log(N)}{\log(d/L) + \log(N)} \tag{15}$$

where, Katz proposed normalize L is the length of the middle stage and d is the average distance between successive points $a = L/N$, where N is the number of steps in the curve.

3. Results and Discussions

The tunable Q-factor wavelet transform approach is used to extract unique features from all 16 channels of EEG signals. EEG data from ADHD and normal subjects are decomposed into multiple levels, and Higuchi and Katz's fractal dimensional features are obtained. The total extracted features for each fractal measures are 112 since six level wavelet decompositions have been performed using TQWT. Hence, a total of 7 wavelet coefficients with respect to 16 EEG channels ($16^*7 = 112$) the size of the features that have been taken. For improved classification accuracy, the optimal selection of quality factor (Q) and decomposition levels (j) is investigated. The redundancy (r) value is fixed to 3 in order to perform better [42, 51]. The TQWT technique is applied on the EEG signals of both ADHD and normal subjects under eyes-closed and eyes-open states for extracting different subbands for different Q and j values. Initially, keeping the Q as 1 and the features are extracted for all the 15 decomposition levels. The unique features extracted with Higuchi and Katz fractional dimension decomposition methods are fed into different classifiers for investigating the efficiency in terms of classification accuracy. Classifiers such as linear discriminant, logistic regression, support vector machine, k-nearest neighbour, ensemble and artificial neural networks are experimented. In all of the experimental conditions, decomposition level 6 achieved maximum classification accuracy. After setting the j value to 6, the Q value is changed from 1 to 10, and the Katz and Higuchi features are computed

for each of the 6 + 1 (7) subbands including one low pass subband are considered. The total number of samples examined in this study is 6400, with 25 second EEG data sets collected. The classification accuracy of each classifier for the EEG signals under the eyes-open state and eyes-closed state are demonstrated in Figure 4.

On observing the classification accuracies of different classifiers, the ANN classifier has exhibited the highest classification accuracy among all the classifiers in both eyes-closed and eyes-open states EEG analysis. Because the ANN classifier outperformed all other classifiers in terms of classification accuracy, only the ANN classifier is used for further investigation. The results are analysed under eyes-open state and eyes-closed for both ADHD and normal subjects. The potential features extracted from all the 15 decomposed levels of EEG signal under eyes-closed and eyes-open states. Figure 5 demonstrates the classification accuracy of both feature extraction methods under eyes-closed and eyes-open states, respectively, for different decomposition levels. It is evident that the potential features extracted from the Katz have maximum classification accuracy in a greater number of decomposition levels than the Higuchi under eyes-closed state. In the meantime, the Higuchi fractional dimension could provide more features that can reflect ADHD in more decomposition levels under the eyes-open state.

In order to determine the best Q and j values, a series of experiments are carried out. While keeping the quality factor $Q = 1$ and redundancy factor $r = 3$ as constants and extracting the unique features from Katz and Higuchi fractal dimensions for all the 15 decomposition levels, it exhibits that the classification accuracy reaches its maximum significant in the 6^{th} level in both the feature extraction methods under eyes-closed and eyes-open states. After choosing $j = 6$ as the decomposition level, the quality factor is changed from 1 to 10, and the characteristics extracted from all 7 subbands $(j + 1)$ for each value of Q are compared. Unique features extracted using Higuchi fractal dimension and Katz's fractal dimension techniques are distinct for ADHD and normal subjects EEG signals. An artificial neural network classifier with a 10-fold cross-validation method is used to validate the classified results. The classification accuracy at different Q values while keeping the decomposition levels 6 as constant is shown in Table 1(b).

The feature extraction techniques used are Katz and Higuchi for both eyes-closed and eyes-open condition. For each technique, 112 features are extracted and given to the ANN classifier. The classification accuracy obtained are shown in the Tables 1–3 for different levels 3, 6, and 8 with Q-factor varying from 1 to 10. Among these, level 6 is giving the best accuracy. The features extracted through Katz fractional dimension have a higher potential to discriminate the ADHD and normal subjects under an eyes-closed state. The EEG signals with eyes-closed states are more significant with higher classification accuracy than the eyes-open state in ADHD diagnosis using Katz fractional dimension estimation method. The classification accuracy became maximum at the decomposition level 6 consistently at $Q = 1$ for both feature extraction methods indicating that the filter banks are perfectly tuned to the optimal classification accuracy at the 6^{th} decomposition level.

Table 1: (a) Classification accuracy for different Q values for a fixed decomposition level $j = 3$, (b) Classification accuracy for different Q values for a fixed decomposition level $j = 6$, (c) Classification accuracy for different Q values for a fixed decomposition level $j = 8$

(a)

Level = 3	Eyes-closed		Eyes-open	
Q	Katz	Higuchi	Katz	Higuchi
1	99.17	91.67	99.00	99.00
2	98.75	95.92	87.42	98.75
3	95.92	93.17	91.67	91.67
4	95.92	69.33	84.08	87.42
5	99.17	70.00	88.17	95.92
6	99.17	82.50	84.08	99.17
7	99.17	84.08	91.67	99.17
8	99.17	88.83	84.08	95.92
9	99.17	87.42	88.17	98.75
10	99.17	88.17	98.75	95.92

(b)

Level = 6	Eyes-closed		Eyes-open	
Q	Katz	Higuchi	Katz	Higuchi
1	100.00	94.25	100.00	100.00
2	99.17	98.75	88.17	99.00
3	98.17	95.92	92.92	92.67
4	98.17	70.00	87.83	89.00
5	100.00	69.33	91.67	97.17
6	100.00	84.08	86.50	100.00
7	100.00	82.50	93.17	100.00
8	100.00	87.42	86.58	97.17
9	100.00	88.83	92.00	99.17
10	100.00	78.25	93.83	97.75

(c)

Level = 8	Eyes-closed		Eyes-open	
Q	Katz	Higuchi	Katz	Higuchi
1	99.17	91.67	99.17	99.17
2	98.75	95.92	84.08	98.75
3	98.75	91.67	90.67	90.17
4	98.75	69.33	86.50	88.17
5	99.00	68.17	90.17	95.92
6	99.00	82.50	86.50	99.17
7	99.00	84.50	91.67	99.17
8	99.00	86.50	86.50	98.75
9	99.00	84.08	89.67	98.75
10	99.00	69.33	91.67	95.92

Table 2: Performance metrics for Katz and Higuchi feature extraction methods (a) eyes-closed state (b) eyes-open state.

(a)

Level = 6	Accuracy		Sensitivity		NPV		MCC		F1-score		G-mean	
Q	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi
1	100	94.25	100	89.17	100	91.3	100	90.16	100	93.33	100	93.95
2	99.2	98.75	98.33	97.5	98.75	98.3	98.54	97.89	99	98.33	99.08	98.54
3	98.2	95.92	95.83	91.67	97.5	94.2	96.6	92.82	97.33	94.67	97.62	95.24
4	98.2	70	95.83	66.67	97.5	65	96.6	41.37	97.33	69.52	97.62	65.13
5	100	69.33	100	58.33	100	66.7	100	43.46	100	65.17	100	67.23
6	100	84.08	100	76.67	100	80.4	100	70.52	100	82	100	83.39
7	100	82.5	100	75.83	100	80.4	100	67.56	100	80.17	100	81.7
8	100	87.42	100	80.83	100	84.6	100	76.4	100	85.67	100	86.75
9	100	88.83	100	86.67	100	90.8	100	78.97	100	87.24	100	87.16
10	100	78.25	100	73.33	100	77.1	100	58.79	100	76.5	100	76.99

(b)

Level = 6	Accuracy		Sensitivity		NPV		MCC		F1-score		G-mean	
Q	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi	Katz	Higuchi
1	100	100	100	100	100	100	100	100	100	100	100	100
2	88.2	99	76.67	97.5	83.33	98.75	79.8	98.06	85	98.33	86.63	98.54
3	92.9	92.67	87.5	89.17	89.58	92.92	88.46	87.1	91.83	91	92.75	91.39
4	87.8	89	75.83	82.5	83	87.17	79.12	81	84.17	86.79	85.98	87.44
5	91.7	97.17	84.17	95.83	88.33	97.5	86.08	94.66	89.5	96.62	90.74	96.15
6	86.5	100	72.5	100	80.83	100	76.34	100	82.33	100	84.24	100
7	93.2	100	86.67	100	90.42	100	88.37	100	91.17	100	92.21	100
8	86.6	97.17	73.33	100	81.33	100	77	95.2	82.5	97.29	84.51	97.25
9	92	99.17	84.17	100	88.33	100	86.11	98.54	90	99.29	91.02	99.08
10	93.8	97.75	88.33	100	90.83	100	89.51	96.22	92.67	98	93.4	97.62

3.1. Performance Metrics. The classifier's performance was measured using accuracy, sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV), F1-score, G-mean, and Matthew's correlation coefficient (MCC). The mathematical background of each performance metric is given as follows:

The percentage of true positives (TP) and true negatives (TN) over the total number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) individuals was used to calculate the classification performance for accuracy.

$$\text{Accuracy}\,(\%) = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} * 100 \tag{16}$$

Sensitivity is calculated by dividing the number of true positive (TP) cases by the number of genuine positive cases, i.e., the total of true positive (TP) and false negative (FN) cases.

$$\text{Sensitivity}\,(\%) = \frac{\text{TP}}{(\text{TP} + \text{FN})} * 100 \tag{17}$$

Specificity refers to the number of true negative (TN) cases found among all actual negative cases, i.e., the total of true negative and false positive (FP) cases.

$$\text{Specificity}\,(\%) = \frac{\text{TN}}{(\text{TN} + \text{FP})} * 100 \tag{18}$$

The ratio of true positives to the number of positive brain maps is known as the positive predictive value (PPV).

$$\text{Positive Predictive Value}(\%) = \frac{\text{TP}}{(\text{TP} + \text{FP})} * 100 \tag{19}$$

The ratio of true negative to the number of negative brain maps defines the negative predictive value (NPV).

$$\text{Negative Predictive Value}(\%) = \frac{\text{TN}}{(\text{TN} + \text{FN})} * 100 \tag{20}$$

Table 3: Summary of comparison for automated detection of ADHD with state-of-the-art techniques.

S.no	Authors	Year	Participants	Age group	Feature extraction methodology	Classifiers	Classification accuracy (%)
1	Ghaderyan et al. [56]	2022	14 ADHD and 19 healthy children	6 to 11 years	Dynamic frequency wrapping	Sparse nonnegative least-square coding	99.17%
2	Tor et al. [34]	2021	45 ADHD, 62 (ADHD+CD) and 16 CD	6 to 12 years	Nonlinear features	KNN classifier	97.88%
3	Ahmadi et al. [35]	2021	13 ADHD-C subtype, 12 ADHD-I subtype, 14 control	6 to 11 years old	Spatial and frequency band features	Convolutional neural network	99.46%
4	R et al. [25]	2021	5 ADHD and 5 normal subjects with eyes-open and eyes-closed state	7 to 12 years	Permutation entropy, Sure entropy, log energy entropy, fuzzy entropy	ANN classifier	99.82%
5	Ekhlasi et al. [36]	2021	61 ADHD and 60 healthy children	9.62 is the mean age of ADHD, 9.85 is the mean age of control group	Directed phase transfer entropy	ANN classifier	89.1%
6	Moghaddari et al. [37]	2020	31 ADHD and 30 healthy children	7 to 12 years	Frequency band separation	Deep CNN	98.48%%
7	Rezaeezadeh et al. [57]	2020	12 ADHD and 12 normal subjects with eyes-closed condition	7–12 years old	Nonlinear and linear univariate features	SVM, KNN, PNN	99.58%
8	Dubreuil-Vall et al. [58]	2020	20 healthy and 20 ADHD patients		Features extracted from neural network	CNN, RNN, SNN	88%
9	Kaur et al.[59]	2020	47 ADHD and 50 control during the eyes-open, eyes-closed, and continuous performance test (CPT) condition	20.3 is the mean age of ADHD, 20.6 is the mean age of normal	Phase space reconstruction, statistical features	SVM, KNN, neural dynamic classifier, enhanced probabilistic neural network, and naive-Bayes classifier	93.3%
10	Altınkaynak et al. [60]	2020	23 ADHD and 23 healthy controls	7–12 years	Morphological, wavelets, and nonlinear based features	Multilayer perceptron, Naïve Bayes, support vector machines, k-nearest neighbor, adaptive boosting, logistic regression and random forest	91.3%
11	Chang et al. [61]	2019	100 ADHD, 44 Normal subjects	10.9 ± 2.4, 10.6 ± 1.9 mean age of ADHD, 11.3 ± 2.2 mean age of normal	Time points, channels input features, saliency maps	Deep EEGNet	83%
12	Boroujeni et al. [29]	2019	50 ADHD and 26 normal cases	4 to 15 years old	Lyapunov exponent, fractal dimension, correlation dimension and sample, fuzzy and approximate entropies	SVM classifier	96.05%

Table 3: Continued.

S.no	Authors	Year	Participants	Age group	Feature extraction methodology	Classifiers	Classification accuracy (%)
13	Chang et al. [61]	2019	30 ADHD and 30 control subjects	Mean age in the ADHD group was 8 years 5 months ±1 year 11 months, control group was 8 years 5 months ±1 year 8 months	Univariate and multivariate features	SVM classifier	Area under the curve = 87.78%, sensitivity = 80.0%, specificity = 80.0%
14	Bashiri et al. [62]	2018	95 ADHD subjects	7 to 18 years old	QEEG features, integrated visual and auditory continuous performance test	ANN classifier	To determine the existence or absence of attention and response control in ADHD subjects
15	Khoshnoud et al. [23]	2018	12 ADHD and 12 normal children	7–12 years old	Frequency band powers, nonlinear features	SVM classifier	83.33%
16	Karimu et al. [28]	2018	20 ADHD and 20 normal children	7–10 years.	Mixture of expert fuzzy models	Continuous wavelet transform (CWT) and standalone classifier	98.01%
17	Chow et al. [63]	2018	30 ADHD and 30 controls.	Mean age for control group 7 years and 10 months ±2 years and 2 months, mean age for ADHD 8 years and 1 month ±2 years	ApEn and TBR	Logistic regression	81.7%
18	Mohammadi et al. [27]	2016	30 ADHD and 30 healthy children	9.62 ± 1.75 years for ADHD, 9.85 ± 1.77 years for healthy children	Fractal dimension (FD), approximate entropy and Lyapunov exponent	Multilayer perceptron	93.65%
19	Khoshnoud et al. [64]	2016	12 ADHD and 10 control groups	8-13 years old	Largest Lyapunov exponent and approximate entropy	Probabilistic neural network classifier	87.5%
20	Helgadóttir et al. [65]	2015	310 ADHD and 351 controls	5.8 to 14 years	Spectral features extraction	Multivariate diagnostic classifier	76%
21	Tenev et al. [8]	2014	67 ADHD, 50 controls	18 to 50 years of age	Forward selection for feature extraction	SVM classifier	82.3%
22	González et al. [66]	2013	22 ADHD and 21 healthy controls	4–15 years	Multivariate linear and nonlinear interdependence measures	Logistic regression	86.7%
23	Nazhvani et al. [67]	2013	12 healthy ones, 12 with ADHD and 12 with BMD	10 to 22 years old	Wavelet denoising and synchronous averaging features	KNN classifier	92.85%
24	Abibullaev et al. [68]	2012	7 ADHD and 3 normal groups	7–12 years	Shanon's entropy, mutual information measures	SVM classifier	97%
25	Ahmadlou and Adeli [40]	2011	12 ADHD and 12 control participants	8–13 years old	Fuzzy synchronization likelihood	Leave one out cross-validation method	87.50%

Table 3: Continued.

S.no	Authors	Year	Participants	Age group	Feature extraction methodology	Classifiers	Classification accuracy (%)
26	Ahmadlou and Adeli [38]	2010	47 ADHD and 7 control individuals with eyes-closed	7-12 years old	Wavelet decomposition and synchronization likelihood method.	RBF neural network classifier	95.6%
27	Proposed work	2022	5 ADHD and 5 non-ADHD subjects with eyes-open and eyes-closed state	6-12 years old	Tunable Q-factor wavelet transform	Linear discriminant, logistic regression, support vector machine, artificial neural networks, and ensemble	100%

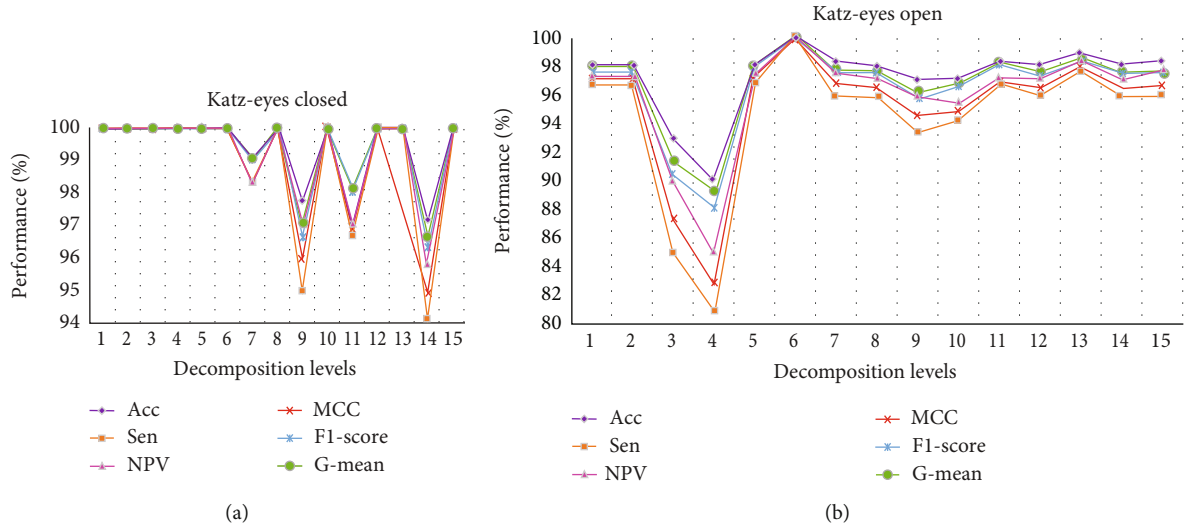


Figure 6: Performance analysis of Katz fractional dimension with different decomposition levels (a) eyes-closed state (b) eyes-open states.

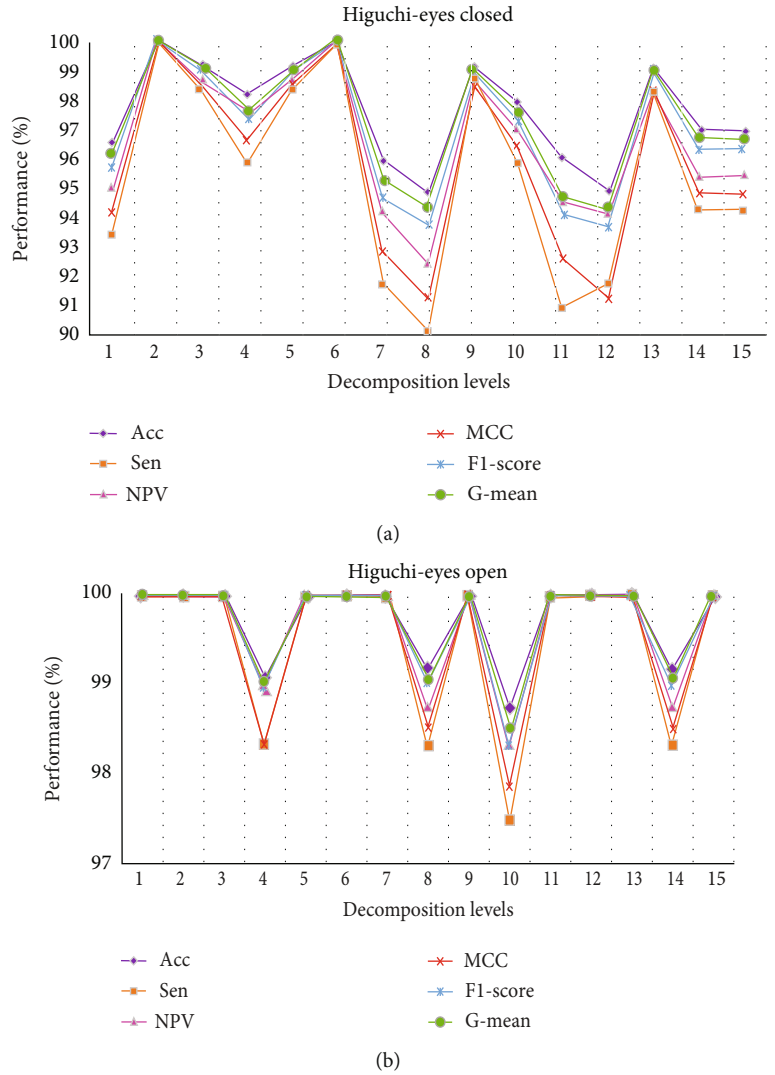


Figure 7: Performance analysis of Higuchi fractional dimension with different decomposition levels (a) eyes-closed state (b) eyes-open states.

The harmonic and geometric measurements of sensitivity and specificity are the F1-score and G-mean, respectively.

$$\text{F1} - \text{Score}\,(\%) = \frac{(2 * \text{Sensiitivity} * \text{PPV})}{(\text{Sensiitivity} + \text{PPV})} * 100 \quad (21)$$

$$\text{G}_{\text{mean}}\,(\%) = \sqrt{\text{Sensitivity} * \text{Specificity}} * 100 \quad (22)$$

Matthew's correlation coefficient is a balanced metric that determines both true and false positives and negatives, even if the classes are of different sizes [55]. To determine a result, the MCC takes into account the test's true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), and a significant prediction is one. Matthew's correlation coefficient (MCC) is calculated as:

$$\text{MCC} = \frac{(\text{TP} * \text{TN}) - (\text{FP} * \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FN})}} * 100 \quad (23)$$

The classification process for ADHD and normal subjects under eyes-closed and eyes-open conditions with different decomposition levels j varying from 1 to 15 is performed. The above performance metrics are evaluated and plotted against the decomposition levels which is shown in Figures 6 and 7. While comparing the performances in the Katz fractional dimension method, the eyes-closed state shown in Figure 6(a) reflects that the classification accuracy, sensitivity, and all other performance matrices are consistently stable at 100% in the lower decomposition level up to level 6 and the variation observed in higher levels of decomposition. This indicates that significant features are extracted in the eyes-closed state and it has a higher ability of discrimination at lower decomposition levels. In the meantime, the EEG signals with eyes-open states shown in Figure 6(b) indicate that the performance metrices has a lesser value at lower decomposition levels, and improve at higher decomposition levels. It reflects that the potential features to discriminate the ADHD are more significant and the ability of the classifier is increased in higher decomposition levels.

Table 2 presents the statistical details of different performance matrices while keeping the decomposition level 6 as constant and varying the tuning factor Q. The filter banks are tuned for its best performance at decomposition level 6, and observed that the Katz feature extraction method showed higher significance in classifying the ADHD under eyes-closed state with higher classification accuracy and sensitivity. Meanwhile, the Higuchi feature extraction method showed higher performance with higher classification accuracy and sensitivity under eyes-open state.

The classification accuracy of the proposed work is compared with the similar works of various authors and observed an improvement in performance. Table 3 shows the performance comparison specifically in terms of classification accuracy. The proposed methodology exhibits the maximum performance with 100% classification accuracy.

It is clear, classify, and diagnose attention deficit hyperactive disorder from the EEG signals has been used many techniques and applications such as ANN, Fog computing, Internet of Medical Thing, and other methods [69–74].

The merit of the proposed methodology is that it has given the highest classification accuracy and this is a robust system as it is validated with 10 fold cross-validation for different classifiers. The limitation of the present work is that it needs preprocessed artifact-free EEG signals, that add an additional signal processing stage to the implementation of the proposed method for real-time diagnostic purposes. Also, the data set shall be extended to fine-tune the system performances in analysing the effectiveness of the proposed system.

4. Conclusion

The present work is an effective methodology to classify and diagnose attention deficit hyperactive disorder from the EEG signals. The EEG signals of both ADHD and normal subjects recorded under eyes-closed and eyes-open states are preprocessed and the tunable Q-wavelet transform is applied to decompose into different subbands. The Katz and Higuchi fractional dimension feature extraction techniques are applied to extract the features for the effective classification of ADHD and normal subjects with the possible maximum accuracy. The Q-value and the decomposition levels are optimally tuned to extract the potential features that can bring out the maximum classification accuracy. Different classifiers have experimented and the artificial neural network classifier with a 10-fold cross-validation method is found to be an effective classifier with a maximum classification accuracy of 100%. According to the findings, the level of decomposition and the Q-factor parameter has a significant impact on feature extraction performance. The classification accuracy varies dramatically with different Q-factor values, with decomposition level j of 6 being the most appropriate. Moreover, the Katz fractional dimension algorithm shows better results in eyes-closed states and the Higuchi FD algorithm demonstrates better results under eyes-open states. Different performance metrics are used to measure the effectiveness of the classifier algorithm that has justified the observed results. The features extracted through Katz and Higuchi from EEG signal under the eyes-closed state in lower decomposition levels have higher significance in discriminating ADHD from the normal subjects. Also, the features extracted from the eyes-open state in higher decomposition levels have higher significance in estimating ADHD with higher classification accuracy. With the eyes-closed EEG signals, the Katz feature extraction method showed greater significance with higher sensitivity in diagnosing ADHD. Meanwhile, the Higuchi feature extraction method showed higher performance with higher sensitivity under eyes-open state signals. As the proposed system has given the highest classification accuracy with higher sensitivity, this shall be used in the clinical diagnosis of ADHD.

Data Availability

The data sets are not publicly available. 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 is no conflict of interest.

References

[1] K. Konrad and S. B. Eickhoff, "Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder," *Human Brain Mapping*, vol. 31, no. 6, pp. 904–916, 2010.

[2] E. Cormier, "Attention seficit/hyperactivity disorder: a review and update," *Journal of Pediatric Nursing*, vol. 23, no. 5, pp. 345–357, 2008.

[3] M. Smith, "Hyperactive around the world? The history of ADHD in global perspective," *Social History of Medicine*, vol. 30, no. 4, pp. 767–787, 2017.

[4] J. A. Venkata and A. S. Panicker, "Prevalence of attention deficit hyperactivity disorder in primary school children," *Indian Journal of Psychiatry*, vol. 55, no. 4, pp. 338–342, 2013.

[5] J. Monge, C. Gómez, J. Poza, A. Fernández, J. Quintero, and R. Hornero, "MEG analysis of neural dynamics in attention-deficit/hyperactivity disorder with fuzzy entropy," *Medical Engineering & Physics*, vol. 37, no. 4, pp. 416–423, 2015.

[6] A. R. Clarke, R. J. Barry, R. Mccarthy, and M. Selikowitz, "EEG-defined subtypes of children with attention-deficit / hyperactivity disorder," vol. 112, pp. 2098–2105, 2001, http://www.elsevier.com/locate/clinph.

[7] M. Arns, H. Heinrich, and U. Strehl, "Evaluation of neurofeedback in ADHD: the long and winding road," *Biological Psychology*, vol. 95, no. 1, pp. 108–115, 2014.

[8] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller, and G. Candrian, "Machine learning approach for classification of ADHD adults," *International Journal of Psychophysiology*, vol. 93, no. 1, pp. 162–166, 2014.

[9] L. Bergeron, J. P. Valla, J. J. Breton et al., "Correlates of mental disorders in the Quebec general population of 6 to 14-year olds," *Journal of Abnormal Child Psychology*, vol. 28, no. 1, pp. 47–62, 2000.

[10] E. J. Costello, D. L. Foley, and A. Angold, "10-year research update review: the epidemiology of child and adolescent psychiatric disorders: II. developmental epidemiology," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 45, no. 1, pp. 8–25, 2006.

[11] E. Emerson, S. Einfeld, and R. J. Stancliffe, "The mental health of young children with intellectual disabilities or borderline intellectual functioning," *Social Psychiatry and Psychiatric Epidemiology*, vol. 45, no. 5, pp. 579–587, 2010.

[12] S. Patidar, R. B. Pachori, A. Upadhyay, and U. Rajendra Acharya, "An integrated alcoholic index using tunable- Q wavelet transform based features extracted from EEG signals for diagnosis of alcoholism," *Applied Soft Computing*, vol. 50, pp. 71–78, 2017.

[13] J. P. Amezquita-sanchez, N. Mammone, F. C. Morabito, and H. Adeli, "A New dispersion entropy and fuzzy logic system methodology for automated classification of dementia stages using electroencephalograms," *Clinical Neurology and Neurosurgery*, vol. 201, p. 106446, 2021.

[14] P. Durongbhan, Y. Zhao, L. Chen et al., "A dementia classification framework using frequency and time-frequency features based on EEG signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 826–835, 2019.

[15] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Identification of epileptic seizures from scalp EEG signals based on TQWT," *Advances in Intelligent Systems and Computing*, vol. 748, pp. 209–221, 2019.

[16] R. Satpathy, T. Choudhury, S. Satpathy, S. N. Mohanty, and X. Zhang, *Data Analytics in Bioinformatics: A Machine Learning Perspective*, John Wiley & Sons, 2021.

[17] A. Sharma, J. K. Rai, and R. P. Tewari, "Schizophrenia detection using biomarkers from electroencephalogram signals," *IETE Journal of Research*, vol. 68, no. 4, pp. 3056–3064, 2022.

[18] S. K. Khare, V. Bajaj, and U. R. Acharya, "Detection of Parkinson's disease using automated tunable Q wavelet transform technique with EEG signals," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 679–689, 2021.

[19] S. L. Oh, Y. Hagiwara, U. Raghavendra et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Computing and Applications*, vol. 32, no. 15, pp. 10927–10933, 2020.

[20] M. Saeedi, A. Saeedi, and A. Maghsoudi, "Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 3, pp. 1007–1018, 2020.

[21] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, "A major depressive disorder classification framework based on EEG signals using statistical , spectral , wavelet , functional connectivity , and nonlinear analysis," *Journal of Neuroscience Methods*, vol. 358, p. 109209, 2021.

[22] G. Alba, E. Pereda, S. Mañas, L. D. Méndez, A. González, and J. J. González, "Electroencephalography signatures of attention-deficit/hyperactivity disorder: clinical utility," *Neuropsychiatric Disease and Treatment*, vol. 11, pp. 2755–2769, 2015.

[23] S. Khoshnoud, M. A. Nazari, and M. Shamsi, "Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals," *Journal of Integrative Neuroscience*, vol. 17, no. 1, pp. 17–30, 2018.

[24] S. K. Loo and S. Makeig, "Clinical utility of EEG in attention-deficit/hyperactivity disorder: a research update," *Neurotherapeutics*, vol. 9, no. 3, 2012.

[25] R. Catherine Joy, S. Thomas George, A. Albert Rajan, and M. S. P. S. P. Subathra, "Detection of ADHD from EEG signals using different entropy measures and ANN," *Clinical EEG and Neuroscience*, vol. 53, no. 1, 2021.

[26] F. Ghassemi, M. Hassan, M. Tehrani-Doost, and V. Abootalebi, "Using non-linear features of EEG for ADHD/normal participants' classification," *Procedia Social and Behavioral Sciences*, vol. 32, pp. 148–152, 2012.

[27] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Rafieivand, M. Begol, and H. Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," *Biomedical Engineering Letters*, vol. 6, no. 2, pp. 66–73, 2016.

[28] R. Yaghoobi Karimu and S. Azadi, "Diagnosing the ADHD using a mixture of expert fuzzy models," *International Journal of Fuzzy Systems*, vol. 20, no. 4, pp. 1282–1296, 2018.

[29] Y. K. Boroujeni, A. A. Rastegari, and H. Khodadadi, "Diagnosis of attention deficit hyperactivity disorder using non-linear analysis of the EEG signal," *IET Systems Biology*, vol. 13, no. 5, pp. 260–266, 2019.

[30] A. Mueller, G. Candrian, V. A. Grane, J. D. Kropotov, V. A. Ponomarev, and G. M. Baschera, "Discriminating between ADHD adults and controls using independent ERP components

and a support vector machine: a validation study," *Nonlinear Biomedical Physics*, vol. 5, no. 1, pp. 1–18, 2011.

[31] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of BMD and ADHD patients using their EEG signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956–1963, 2011.

[32] O. Article, A. Allahverdy, A. K. Moghaddam, M. R. Mohammadi, and A. M. Nasrabadi, "Detecting ADHD children using the attention continuity as nonlinear feature of EEG," *Frontiers in Biomedical Technologies*, vol. 3, no. 1-5, pp. 28–33, 2016.

[33] E. Pereda, M. García-Torres, B. Melián-Batista, S. Mañas, L. Méndez, and J. J. González, "The blessing of dimensionality: feature selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronisation," *PLoS One*, vol. 13, no. 8, pp. 1–24, 2018.

[34] H. T. Tor, C. P. Ooi, N. S. Lim-Ashworth et al., "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Computer Methods and Programs in Biomedicine*, vol. 200, 2021.

[35] A. Ahmadi, M. Kashefi, H. Shahrokhi, and M. A. Nazari, "Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes," *Biomedical Signal Processing and Control*, vol. 63, 2021.

[36] A. Ekhlasi, A. M. Nasrabadi, M. Mohammadi, A. Motie Nasrabadi, and M. Mohammadi, "Classification of the children with ADHD and healthy children based on the directed phase transfer entropy of EEG signals," *Frontiers in Biomedical Technologies*, vol. 8, no. 2, pp. 115–122, 2021.

[37] M. Moghaddari, M. Z. Lighvan, and S. Danishvar, "Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG," *Computer Methods and Programs in Biomedicine*, vol. 197, 2020.

[38] M. Ahmadlou and H. Adeli, "Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD," *Clinical EEG and Neuroscience*, vol. 41, no. 1, pp. 1–10, 2010.

[39] A. Allahverdy, A. M. Nasrabadi, and M. R. Mohammadi, "Detecting ADHD children using symbolic dynamic of nonlinear features of EEG," 2022 https://ieeexplore.ieee.org/abstract/document/5955548.

[40] M. Ahmadlou and H. Adeli, "Functional community analysis of brain: a new approach for EEG-based investigation of the brain pathology," *NeuroImage*, vol. 58, no. 2, pp. 401–408, 2011.

[41] D. P. Dash and M. H. Kolekar, "Hidden Markov model based epileptic seizure detection using tunable Q wavelet transform," *Journal of Biomedical Research*, vol. 34, no. 3, pp. 170–179, 2020.

[42] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3560–3575, 2011.

[43] A. Bhattacharyya, R. B. Pachori, A. Upadhyay, and U. R. Acharya, "Tunable-Q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals," *Applied Sciences*, vol. 7, no. 4, 2017.

[44] S. T. George, M. S. P. Subathra, N. J. Sairamya, L. Susmitha, and M. Joel Premkumar, "Classification of epileptic EEG signals using PSO based artificial neural network and tunable-Q wavelet transform," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 709–728, 2020.

[45] W. He, Y. Zi, B. Chen, F. Wu, and Z. He, "Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform," *Mechanical Systems and Signal Processing*, vol. 54, pp. 457–480, 2015.

[46] M. Murugappan, W. Alshuaib, A. K. Bourisly, S. K. Khare, S. Sruthi, and V. Bajaj, "Tunable Q wavelet transform based emotion classification in Parkinson's disease using electroencephalography," *PLoS One*, vol. 15, no. 11, 2020.

[47] V. Bajaj, S. Taran, S. K. Khare, and A. Sengur, "Feature extraction method for classification of alertness and drowsiness states EEG signals," *Applied Acoustics*, vol. 163, 2020.

[48] A. R. Hassan, S. Siuly, and Y. Zhang, "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating," *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 247–259, 2016.

[49] S. K. Khare and V. Bajaj, "Constrained based tunable Q wavelet transform for efficient decomposition of EEG signals," *Applied Acoustics*, vol. 163, 2020.

[50] R. Acharya, O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, "Non-linear analysis of EEG signals at various sleep stages," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 1, pp. 37–45, 2005.

[51] K. Jindal, R. Upadhyay, and H. S. Singh, "Application of tunable-Q wavelet transform based nonlinear features in epileptic seizure detection," *Analog Integrated Circuits and Signal Processing*, vol. 100, no. 2, pp. 437–452, 2019.

[52] A. Rizal and R. D. Estananto, "Epileptic EEG signal classification using multiresolution Higuchi fractal dimension," *International Journal of Engineering Research & Technology*, vol. 12, no. 4, pp. 508–511, 2019, http://www.irphouse.com508.

[53] M. Bachmann, J. Lass, A. Suhhova, and H. Hinrikus, "Spectral asymmetry and Higuchi's fractal dimension measures of depression electroencephalogram," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 251638, 8 pages, 2013.

[54] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine*, vol. 18, no. 3, pp. 145–156, 1988.

[55] S. M. Fernandez-Fraga and J. Rangel, "Comparison of Higuchi, Katz and multiresolution box-counting fractal dimension algorithms on EEG waveforms signals based on visual evoked potentials," *Revista EIA/English Version*, vol. 14, no. 27, 2017.

[56] P. Ghaderyan, F. Moghaddam, S. Khoshnoud, and M. Shamsi, "New interdependence feature of EEG signals as a biomarker of timing deficits evaluated in attention-deficit/hyperactivity disorder detection," *Measurement*, vol. 199, 2022.

[57] M. Rezaeezadeh, S. Shamekhi, and M. Shamsi, "Attention deficit hyperactivity disorder diagnosis using non-linear univariate and multivariate EEG measurements: a preliminary study," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 577–592, 2020.

[58] L. Dubreuil-Vall, G. Ruffini, and J. A. Camprodon, "Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG," *Frontiers in Neuroscience*, vol. 14, p. 251, 2020.

[59] S. Kaur, S. Singh, P. Arun, D. Kaur, and M. Bajaj, "Phase space reconstruction of EEG signals for classification of ADHD and control adults," *Clinical EEG and Neuroscience*, vol. 51, no. 2, pp. 102–113, 2020.

[60] M. Altınkaynak, N. Dolu, A. Güven et al., "Diagnosis of attention deficit hyperactivity disorder with combined time and frequency features," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 927–937, 2020.

[61] M. Y. Chang, C. S. Ouyang, C. T. Chiang et al., "A new method of diagnosing attention-deficit hyperactivity disorder in male patients by quantitative EEG analysis," *Clinical EEG and Neuroscience*, vol. 50, no. 5, pp. 339–347, 2019.

[62] A. Bashiri, L. Shahmoradi, H. Beigy et al., "Quantitative EEG features selection in the classification of attention and response control in the children and adolescents with attention deficit hyperactivity disorder," *Future Science OA*, vol. 4, no. 5, pp. 11–13, 2018.

[63] J. C. Chow, C. S. Ouyang, C. L. Tsai et al., "Entropy-based quantitative electroencephalogram analysis for diagnosing attention-deficit hyperactivity disorder in girls," *Clinical EEG and Neuroscience*, vol. 50, no. 3, pp. 172–179, 2019.

[64] S. Khoshnoud, M. Shamsi, and M. A. Nazari, "Non-linear EEG analysis in children with attention-deficit/ hyperactivity disorder during the rest condition," in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, Iran, 2016.

[65] H. Helgadóttir, Ó. Ó. Gudmundsson, G. Baldursson et al., "Electroencephalography as a clinical tool for diagnosing and monitoring attention deficit hyperactivity disorder: a cross-sectional study," *BMJ Open*, vol. 5, no. 1, 2015.

[66] J. J. González, L. D. Méndez, S. Mañas, M. R. Duque, E. Pereda, and L. De Vera, "Performance analysis of univariate and multivariate EEG measurements in the diagnosis of ADHD," *Clinical Neurophysiology*, vol. 124, no. 6, pp. 1139–1150, 2013.

[67] A. D. Nazhvani, R. Boostani, S. Afrasiabi, and K. Sadatnezhad, "Classification of ADHD and BMD patients using visual evoked potential," *Clinical Neurology and Neurosurgery*, vol. 115, no. 11, pp. 2329–2335, 2013.

[68] B. Abibullaev and J. An, "Decision support algorithm for diagnosis of ADHD using electroencephalograms," *Journal of Medical Systems*, vol. 36, no. 4, pp. 2675–2688, 2012.

[69] T. Y. Wah, M. A. Mohammed, U. Iqbal, S. Kadry, A. Majumdar, and O. Thinnukool, "Novel DERMA fusion technique for ECG heartbeat classification," *Life*, vol. 12, no. 6, p. 842, 2022.

[70] A. A. Mutlag, M. K. Abd Ghani, M. A. Mohammed et al., "Multi-agent systems in fog–cloud computing for critical healthcare task management model (CHTM) used for ECG monitoring," *Sensors*, vol. 21, no. 20, p. 6923, 2021.

[71] A. A. Mutlag, M. K. A. Ghani, and M. A. Mohammed, "A healthcare resource management optimization framework for ECG biomedical sensors," in *In Efficient Data Handling for Massive Internet of Medical Things*, pp. 229–244, Springer, Cham, 2021.

[72] A. U. Rahman, M. Saeed, M. A. Mohammed, M. M. Jaber, and B. Garcia-Zapirain, "A novel fuzzy parameterized fuzzy hypersoft set and Riesz summability approach based decision support system for diagnosis of heart diseases," *Diagnostics*, vol. 12, no. 7, p. 1546, 2022.

[73] J. Prasanna, M. S. P. Subathra, M. A. Mohammed, R. Damaševičius, N. J. Sairamya, and S. T. George, "Automated epileptic seizure detection in pediatric subjects of CHB-MIT EEG database—a survey," *Journal of Personalized Medicine*, vol. 11, no. 10, p. 1028, 2021.

[74] M. S. P. Subathra, M. A. Mohammed, M. S. Maashi, B. Garcia-Zapirain, N. J. Sairamya, and S. T. George, "Detection of focal and non-focal electroencephalogram signals using fast Walsh-Hadamard transform and artificial neural network," *Sensors*, vol. 20, no. 17, p. 4952, 2020.

Hindawi
Journal of Sensors
Volume 2022, Article ID 1222692, 14 pages
https://doi.org/10.1155/2022/1222692

Research Article

Effective Preprocessing and Normalization Techniques for COVID-19 Twitter Streams with POS Tagging via Lightweight Hidden Markov Model

Senthil Kumar Narayanasamy,[1] **Yuh-Chung Hu**,[2] **Saeed Mian Qaisar**,[3] **and Kathiravan Srinivasan**[4]

[1]*School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India*
[2]*Department of Mechanical and Electromechanical Engineering, National Ilan University, Yilan 26047, Taiwan*
[3]*Electrical and Computer Engineering Department, Effat University, Jeddah, Saudi Arabia*
[4]*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, India*

Correspondence should be addressed to Kathiravan Srinivasan; kathiravan.srinivasan@vit.ac.in

Received 4 July 2022; Accepted 20 July 2022; Published 2 August 2022

Academic Editor: Mohit Mittal

The major focus of this research work is to refine the basic preprocessing steps for the unstructured text content and retrieve the potential conceptual features for further enhancement processes such as semantic enrichment and named entity recognition. Although some of the preprocessing techniques such as text tokenization, normalization, and Part-of-Speech (POS) tagging work exceedingly well on formal text, it has not performed well when it is applied into informal text such as tweets and short messages. Hence, we have given the enhanced text normalization techniques to reduce the complexity persist over the twitter streams and eliminate the overfitting issues such as text anomalies and irregular boundaries while fixing the grammar of the text. The hidden Markov model (HMM) has been pervasively used to extract the core lexical features from the Twitter dataset and suitably adapt the external documents to supplement the extraction techniques to complement the tweet context. Using this Markov process, the POS tags are identified as states of the Markov process, and words are the desired results of the model. As this process is very crucial for the next stage of entity extraction and classification, the effective handling of informal text is considered to be important and therefore proposed the most effective hybrid approach to deal with the issues appropriately.

1. Introduction

In recent years, the impact of social media sites is rampant and gaining huge popularity among social media users consistently. Particularly, Twitter has gained huge momentum among the users and providing an open platform for information exchange in a variety of events and situations. The events can be classified as political crisis, natural calamities, disasters, celebrations, etc. Recently, the tweets related to COVID-19 have been very pervasive and made a prominent impact on the government agencies to take immediate actions. Also, the information pertaining to coronavirus has been used for travelers and business people to take preliminary actions for their proposed plans. The information posted in Twitter are needs to be organized and classified according to its credibility score and further pave way for segregating them into primary and secondary information. Normally, the secondary information in a tweet is a retweet [1]. Nowadays, users prefers to use social media platforms such as Twitter for getting the latest news, and there are high chances of drawing wrong conclusions by reading false news. Hence, the demand for implementing the credible system that is capable of identifying the correct news and classifying them into the right emotions provides the right information for the decision-making processes.

Therefore, automatic detection of events such as people, organizations, locations, and other entities from unstructured content is challenging and has shown very poor performance due to its unorthodox content [2]. Similarly, many named

entity recognition (NER) research works have been carried out recently with respect to twitter streams such as [3, 4]. These research works have largely been aimed at augmenting the capabilities to extract the potential named entities from the tweets and focused on improving the state-of-the-art methods in detecting the Out-Of-Vocabulary (OOV) words. But due to the lack of contexts and noisy structure of the tweets, detecting the potential named entities from tweets poses a great challenge and gives huge difficulties to annotate the tweets with necessary POS tags. Figure 1 illustrates the open challenges in handling COVID-19 twitter streams. Table 1 provides the detailed information about the publically available named entity annotated tweets.

Besides, tweet tokenization has been a great challenge for many NER systems, and the existing methods such as Penn-Tree Bank (PTB), TweetMotif, TwokenizerTool [5], and TwitIE tokenizer [6] failed to address these issues effectively. Therefore, we provide the mechanism to solve the fundamental preprocessing techniques such as tokenization, normalization, and POS tagging of tweets. These normalization processes have reduced the complexity persist over the given datasets and addressed the overfitting issues on the informal text categorically. As this process is very crucial for the next stage of entity extraction and classification, the effective handling of informal text is considered to be important and therefore proposed the most effective hybrid approach to deal with the issues appropriately. In addition to that, informal text would have certain common open challenges that has been listed in Figure 1. This research work has profusely handled these open challenges in the effective way and able to outperform the results with good precision.

The major contributions of this paper are given below:

(i) A detailed discussion on text normalization techniques also delineated the difficulties of converting Out-Of-Vocabulary (OOV) words into In Vocabulary (IV) words

(ii) Proposed an enhanced tweet tokenization technique to the alternate of Stanford tokenizer and Penn Tree Bank Tokenizer

(iii) In order to extract the appropriate named entities from unstructured text, we have utilized the lightweight hidden Markov model (HMM) to filter the correct lexical features that has been generated through the POS tagging

1.1. Paper Structure. The rest of the paper is organized as follows: Section 2 discusses the collaborative work rendered by the researchers in the fields of tweet normalization and preprocessing. Also, the prominent tools and methods used for tweet tokenization and normalization have been discussed briefly. Section 3 gives a comprehensive idea about our proposed methods and techniques followed for tweet normalization. Besides, we have briefed about the Out-Of-Vocabulary conversion methods and some evaluation metrics to detect the OOV methods. Section 4 highlights the procedures for tweet tokenization and segmentation. This section has delineated the procedures to tokenize the tweets with some standard nomenclatures. In Section 5, the tweet preprocessing and normalization approach has been dealt with a novel algorithm and analyzed the existing dictionaries to detect the potential OOV words. In Section 6, POS tagging for tweets has been discussed and also introduced the lightweight hidden Markov model for extracting the lexical features generated from the standard POS tagging. Section 7 gives the empirical shreds of evidence of tweet normalization and effective preprocessing results. Also, we have given some error analyses of our proposed methods.

Figure 1: Open challenges in handling COVID-19 twitter streams.

2. Related Works

As most people use social media sites to post their messages daily, the amount of information stored on these websites get increased exponentially, and the messages are informal in nature due to its limited space constraints. Text normalization plays the seminal role in the process of detecting and removing the noisy text (i.e., tweet in this research) into standard words. Therefore, it has gained huge research attention in recent years and increasingly attracted many researchers to carry forward their research work in this domain. Besides, many academic conferences and workshops [7–9] have been conducted to gather the data related to informal texts. The Association for Computational Linguistics (ACL) [10] and North American Association for Computational Linguistics (NAACL) [11] have been encouraging the researchers and students to actively participate in their conferences and workshops to gain knowledge on both formal and informal text. Recently, the Text Retrieval Conference (TREC) has created a new web page related to informal languages used in social media sites and also conducted workshops [12] relevant to the field. In this

Table 1: Openly available social media annotated corpus.

Annotated corpus	Number of tokens	Entity schema
Finin et al. [58]	7 K	Person, location, and organization
Ritter et al. [59]	46 K	Freebase
Liu et al. [3]	12 K	Person, location, product, and organization
Rowe et al. [60]	29 K	Person, location, misc, and organization
Derczynski et al. [61]	165 K	Person, location, and organization

research work, we have analyzed the works demonstrated by various researchers and their research findings and shortcomings in detail.

Earlier, the researchers [13, 14] have taken their research work only on normalizing the spelling errors produced from the web sources. They have used the *n*-gram model to assess the probability of each word within its context and estimate the relative frequency of the word in the given sequence. Their *n*-gram model has mapped the words using Many-to-One (*N*-to-1) cardinality, and the real word substitutions such as word usage with its context and grammatical structure of the sentence were detected and converted successfully, but their research work faced some serious lapses in dealing with unstructured content and failed to retain the accuracy rate attributed by the many prominent researchers [15–17].

Meanwhile, [18, 19] has demonstrated their research work on microtexts such as Twitter and SMS for detecting phonetic misspellings, standard acronyms, and contractions. Generally, misspelled words were detected by Natural Language Processing (NLP) systems using the mult-channel models which effectively find the lexical variance on some factors such as contextual wounding of the word, phonetic similarity, orthographic factors, and expansion of acronym using the standard dictionary. As suggested by previous researchers [20–23], they have utilized the Aspell spell corrector to detect the misspelling on Twitter as well as on SMS datasets.

Later, [24] has developed a spelling corrector which uses Google Style-based spell corrector, and it just find the proximity of a word and recognizes the correct spelling for the given word. The algorithms work on the conditional probability of the word based on edit distance measure and choose the word which has less edit distance (i.e., less number of deletion, replacement, insertion, or transposition used to convert the word into correct form). They have set the threshold limit of edit distance is less than or equal to 2. But due to textual sparseness in informal text, most of the misspelled words in the informal text require more edits and demand more comparison concerning the context of the words. Again, the Norvig system [25] works exceptionally well on standard orthodox text but failed to get the precision on informal text.

Finding the Out-Of-Vocabulary word is very challenging in social media sites; particularly, it has been prevalently used in twitter streams. The OOV word is defined as unorthodox words, and it has not been presented in the standard dictionary for reference. To tackle this issue, many research works had been carried out [26–30] and attained some considerable accuracy rates with respect to the BLEU score. The researchers [31] have used the classifier to detect the OOV words as ill-formed words based on the similarity measures such as phoneme and grapheme score and converted the ill-formed words into standard English words. In their approach, they had used the dictionary, context support for the word, and similarity measure to predict the correct form of the OOV words and finally attained the *F*-score of 68.30%. Even though the result is considerably good at some aspects, it had not performed well on noisy tweets and yields poor results if there was no context support for the ill-formed words.

Later, the author [32–34] has used a hybrid approach to deal with OOV words present in social media sites and prepared heuristic approach such as string similarity measure, edit-distance function, and subsequence overlap function to detect the OOV words and converted them into its appropriate In-Vocabulary words. The correct candidate word was selected based on the *n*-gram model and used the confusion matrix to find the proximity score which is likely to be the correct English form. This approach has reduced the burden of previous researchers and yields a good accuracy rate of up to 72.15%.

3. Proposed Method

In this research work, we have downloaded the Twitter datasets related to COVID-19 from 6th Workshop on Noisy User-generated Text (W-NUT) [35] for our analysis. It has manually annotated almost 10,000 tweets related to COVID-19 and built a corpus called *COVIDKB* that is a well-structured knowledge base to support the SPARQL queries. To extract the structured knowledge from the tweets, our primary task in this preprocessing step is to remove the usernames, special symbols, retweets, hashtags, and emoticons from the tweets and take only the original tweets for the next level of processing.

3.1. Problem Definition. "Given the tweet corpus T, eliminate the tweets which do not convey much information regarding the event E and remove or replace noisy tokens in the tweets with normalized tweets."

The basic regular expression followed to remove the special symbols, retweets, and other emoticons is given below:

```
def process_text():
""" " Remove emoticons, usernames, retweets etc. and returns list of cleaned tweets. """
data=pull_tweets()
regex_remove="(@[A-Za-z0-9]+)|([^0-9A-Za-z  t])|(\w+:\/\/\S+)|^RT|http.+?"
```

```
    stripped_text=[re.sub(regex_remove, “, tweets).strip()
for tweets in data]
    return ". ".join(stripped_text)
```

Once the tweets are cleaned using the above regular expression snippet, we need to perform the tokenization of tweets to fix the proper tagging of words and identify the proper nouns and pronouns for effective entity extraction and disambiguation. Each of the following methods helps to solve the ambiguity that persists over the tweets and identifies the entities with proper references in the external document sources. The three basic components of this research, i.e., tokenization, normalization, and POS tagging are considered to be noncore components but they are very crucial in this research because the informal nature of tweets has condensed the words and give space for ambiguity. Hence, we have proposed a novel method to deal with the issues and remove the ambiguity with the support of external document.

Before we go into the next phase of the preprocessing pipeline as given in Figure 2, we have taken two types of dictionaries to correct the misspelled words in the tweets and fix the correct word form to it. In this case, at first, we have benchmarked some of the standard online spell correction dictionaries for our analysis such as Norvig's Spell Corrector (https://norvig.com/spell-correct.html), BK-Tree (https://issues.apache.org/jira/browse/LUCENE-2230) (Burkhard–Keller Tree), SymSpell (https://symspellpy.readthedocs.io/en/latest/) (Symmetric Delete Spell correction algorithm), LinSpell (https://github.com/wolfgarbe/LinSpell) (Linear Search Spell Correction), and PyEnchant Dictionary(https://pyenchant.github.io/pyenchant/tutorial.html) as given in Table 2. From the analysis, we have observed that the PyEnchant Dictionary is suitable for informal text processing and compatible with all the programming environments. Besides, PyEnchant Dictionary is faster than the above four algorithms, and the indexing method is found very effective for searching the words. The PyEnchant dictionary has been getting updated frequently on every year and enriched its gazetteer words. Hence, we have used the PyEnchant Dictionary for the misspelled words and identified the OOV words in the tweets if any.

Second, we have created our own slang dictionary for converting the slang word into its correct English word form and fixing the correct meaning for the slang word. In this context, we have searched the online slang dictionary application sites such as NoSlang Dictionary (https://www.noslang.com/), Urban Dictionary (https://www.urbandictionary.com/), Translit (https://www.translit.ie/), and a few more web sources. Further we manually extracted the slang words such as contractions, abbreviations, slangs, unorthodox word forms, and canonical words from the above listed online sources and gathered their equivalent English meanings appropriately. Then, we have listed all the slang words with their corresponding English meanings into separate files such as abbreviation, slang words, contractions, and emoticons. Later, we formatted each of the above files and removed the duplicate entry if present in each file (Table 3). Eventually, we checked the absolute meaning of each token in the file and ordered them alphabetically for easy processing of search operation. The first column in each file contains the slang words or abbreviations or contractions, and the second column gives the corresponding meaning or abbreviation of that token.

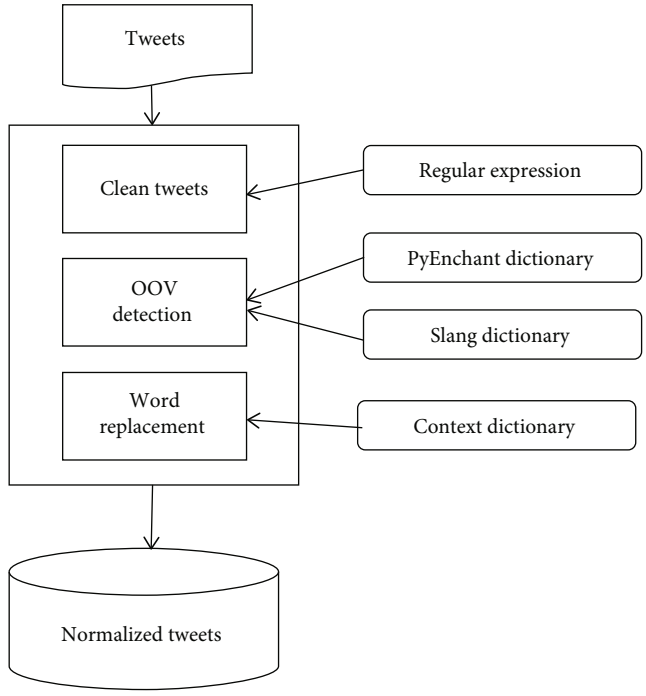


Figure 2: Tweet normalization and word replacement techniques.

3.2. Tweet Normalization. The first step in the tweet normalization approach is tokenization. This is the basic preprocessing step followed in all the natural language processing as well as in the information extraction projects. The proposed approach for tweet normalization has been given in the following Figure 2.

Mainly, the process of tweet normalization takes three critical analysis: (i) detect the candidate tokens on the mutual comparison of standard vocabulary sets; (ii) identify the symbols, emoticons, and OOV words from the tweets with respect to word contortions such as spelling mistakes and displacement of the grammatical structure of words; and (iii) discard the OOV words from the tweets using the standard corpus. All the above-described steps are completely language independent and work exceedingly well with language specifies resources (see Figure 3). For the OOV words, it has been largely dependent on standard abbreviations and acronyms to filter out the In Vocabulary (IV) words and produce the candidate list of IV words for POS tagging.

The main challenge in processing informal text such as tweets is that it gives difficulty in dividing the tweets into multiple tokens and categorically identifies the potential named entities from the divided tokens. The major task of classifying the tokens into IV words and OOV words would be a serious implication in the process of tweet normalization. The standard dictionary (i.e., in this case, we have used PyEnchant Dictionary for word comparison and dictionary lookup) is more than enough to identify the IV words from the tweets and appropriately categorize those into one of the predefined

Table 2: Conventional text normalization methods and techniques.

Technique	Abbreviations	Repeated characters	Misspelled words
Regular expression	X	√	X
Replace() function using WordNet	X	√	√
Expanding abbreviations by CSV file replacement	√	X	X
Probability model using edit distance	X	√	√
Spell correction using TextBlob	X	X	√
NLTK library	X	√	√
Phonetic edit distance	X	√	√
PyEnchant library	X	√	√

Table 3: Proposed slang dictionary for unorthodox content.

Type of dictionary	Total entry
Abbreviations	1346
Contractions	131
Slang words	1296
Emoticons	164
Total	2937

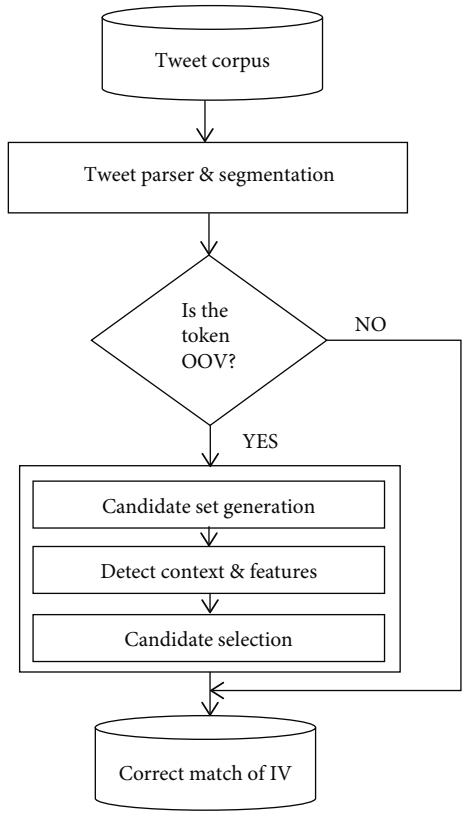


Figure 3: General framework for detecting IV and OOV words.

categories sets based on the POS tags assigned to it, but the remaining nonstandard tokens (i.e., OOV words) need to be compared and find the appropriate candidate list to fit it into the correct word match.

3.2.1. Statistical Rules. So far, we have discussed handling the OOV words and the problem of choosing the appropriate candidate words for the given nonstandard token in the tweets/sentences. We have identified some of the implicit traits of the nonstandard tokens (i.e., OOV words) after running through a huge tweet corpus downloaded from COVIDKB [35] and defined the following basic procedure to tackle the OOV words if any present on the user-generated content.

Here are the following examples:

Type 1: in the OOV words, the social users may have missed the spaces either knowingly or unknowingly and stretched into two or more standard words. Example: stywith (stay with) and cometogether (come together)

Type 2: the OOV words are framed upon the sound of the words rather than the lexical structure of the words. Example: c u agn (see you again)

Type 3: the OOV words are constructed based on the first letters of the standard words or the phonetic positioning of the words. Example: u r (you are) and thx (thanks)

For these types of errors in the OOV words, we have constructed the Slang dictionary (Table 3) as mentioned in the previous Section 3 and detect the possible candidate set of words to disambiguate the OOV words. We also find the exact fit of matching words to the given token based on the context given by the language model. We have reduced the 1-to-*N* mapping of the candidate list for the OOV words into 1-to-1 mapping and increased the efficiency of the tweet normalization.

3.2.2. Multiple Character Reduction. The OOV words have occurred at many places in the tweets and disturb the process of transforming them into standard English words. The major problem faced in handling the OOV words was that it contains many nonword tokens and repetition of multiple characters to express the inherent emotions to the reader. This was also explained in [36]. The characters which occurred more than once will be deduced to single characters, and then PyEnchant Dictionary has been utilized subsequently to prevent the further mistakes caused by the earlier reduction of multiple characters. For example, the words such as speed, speech, and breed have contained the same character that appears more than once in the word and still gives the correct English meanings. If we reduce those multiple occurrences of characters to a single character, then it would cause spelling mistake and leads to bad normalization. To get the correct form of the word, we have proposed the appropriate method where it has taken the

utmost care to tokenize the OOV words and split the OOV words based on some designed patterns. Besides, multiple punctuation symbols posed similar difficulties, and we reduced the multiple punctuation marks into single punctuation using the defined regular expression.

3.3. Evaluation Metrics. To assess the quality of results for the OOV words, the following metrics have been used and have been evaluated the performance of the system.

3.3.1. Miss Rate (MR). It measures that the number of OOV words was missed with respect to OOV words detected.

$$\mathrm{MR} = \frac{\#\text{OOVs references} - \#\text{OOVs detected}}{\#\text{OOVs references}}. \quad (1)$$

3.3.2. False Alarm Rate (FAR). It measures the number of IV words that had been falsely reported as OOV words.

$$\mathrm{FAR} = \frac{\#\text{OOVs reported} - \#\text{OOVs detected}}{\#\text{IVs references}}. \quad (2)$$

3.3.3. Word Error Rate (WER). It measures the number of errors that occurred during the substitution, deletion, or insertion of characters by the proposed system.

$$\mathrm{WER} = \frac{\#\text{Substitution errors} + \#\text{deletion errors} + \#\text{insertion errors}}{\#\text{Words in references}}. \quad (3)$$

3.3.4. Precision. It measures that the number of OOV words has been detected correctly by the proposed system.

$$\mathrm{Precision} = \frac{\#\text{OOVs detected}}{\#\text{OOVs reported}}. \quad (4)$$

3.3.5. Recall. It measures the number of OOV words detected with respect to the OOV references.

$$\mathrm{Recall} = \frac{\#\text{OOVs detected}}{\#\text{OOVs references}}. \quad (5)$$

3.3.6. F1. It measures both the precision and recall of the OOV words detected by the proposed system.

$$\mathrm{F1} = \frac{2.\text{Precision X recall}}{\text{Precision} + \text{recall}}. \quad (6)$$

3.4. Experimental Analysis. We have manually analyzed the tweet normalization for the COVIDKB since there is no gold standard dataset followed to assess this language model and hence assessed the performance of COVIDKB which contains 10,000 tweets. For each tweet, we have considered all the modifications that happened in the process of normalization by the proposed system and validated them. The four major tweet normalization operations such as insertion, deletion, substitution, and tokenization have been monitored manually, and the correctness of the results has been measured through the F1-score produced by the proposed system. Table 4 shows the number of OOV words detected, pronunciation accuracy, and identical pronunciation score of the system.

Table 4: Accuracy rate of COVIDKB Twitter dataset and OOV detected.

COVIDKB Twitter dataset	Accuracy rate
# of detected OOVs	3728
Pronunciation accuracy	67%
Identical pronunciation	72%

Based on Table 4, the precision, recall, and F1 score for the OOV words have been measured, and it has been given in Table 5. Figure 4 illustrates the ROC curve and statistical analysis of investigating the OOV words in tweets.

This ROC curve depicts the accuracy rate, and sensitivity of OOV words presents in the tweets and was able to identify the missed percentage of Part-of-Speech tags for the tokenized tweets.

In addition to finding the OOV words present in the tweets, there are also some other factors to be considered for effective normalization such as stemming, lemmatization, stop word removal, and emoticons detections [37–39]. Extra supervision is required to handle these preprocessing methods and further, these methods help to provide contextual support for sentiment analysis, word cluster, information extraction, entity detection, and many more (see Table 6). As we discussed finding the potential named entities in the tweets, these features help to solve the ambiguity that persists over the text and largely support the contextual score for the proposed system.

The performance metrics and combinational score of the preprocessing methods have been given in Table 7 for further comprehension and enhancement for accuracy. Figure 5 portrays the ROC curve for text normalization and statistical summary of preprocessing steps. It has become evident that the blend of emotions, lemmatization, and stopwords removal has given the performance marginally high and outperfomed with the other integrated approaches.

4. Tweet Tokenization and Segmentation

As tokenization is a first step in the pipeline, the major aim of tweet tokenization is to split the tweet into some meaningful chunks (i.e., semantic tokens) that can be words, word phrases, or any cardinals. Due to the informal nature of the tweets, tweet tokenization process gives difficulties in handling the informal text and comparatively challenging than formal text processing operations [40]. Hence, it requires some sophisticated techniques to solve the issues and effectively perform the tokenization processes. In this connection, we have analyzed some of the techniques followed by earlier researches for tokenization of normal text content. Since the formal text has been supported with well-structured context and language grammar [41], it had performed well on all the grounds, and the major tokenization approach that was followed by researchers was Stanford tokenization. Normally, the Stanford tokenizer utilized the JFlex lexical analyser [42] for tokenization of sentences and produced the results for the given formal text. In some cases,

Table 5: Evaluation of OOV words and F1 score.

OOV operations	Precision	Recall	F1 score
Insertion	0.872	0.801	0.834993
Deletion	0.857	0.782	0.817784
Substitution	0.839	0.768	0.801932
Tokenization	0.894	0.852	0.872495

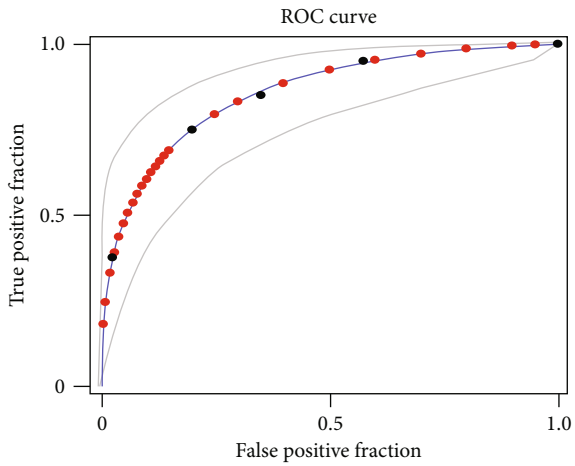


Figure 4: ROC curve and statistical analysis of investigating the OOV words in Tweets.

Table 6: Preprocessing techniques and method used.

Preprocessing techniques	Methods deployed
Stop–word removal	Rainbow list
Stemming	Snowball stemmer
Emoticon	Regular expression
Tokenization	Unigram, bigram, and N-gram
Weighting scheme	TF-IDF

Table 7: Preprocessing impact of text normalization.

Preprocessing methods	Precision	Recall	F1 score
Stop word	0.547	0.521	0.533684
Stemming	0.649	0.592	0.619191
Emoticons	0.723	0.684	0.702959
Stem+stop	0.812	0.769	0.789915
Emo + Lem + stop	0.875	0.837	0.855578
Emo + stem+stop	0.836	0.795	0.814985

the researchers have even used Penn Treebank Tokenizer [43] to tokenize the content which is using the specific regular expression written in SED script. The above tokenizers have been commonly used for most formal text processing and yield a good accuracy rate for all the instances. But when the tweet is informal in nature and mostly unorthodox, then the above tokenizers would have been a bad choice in this regard and produce inappropriate results.

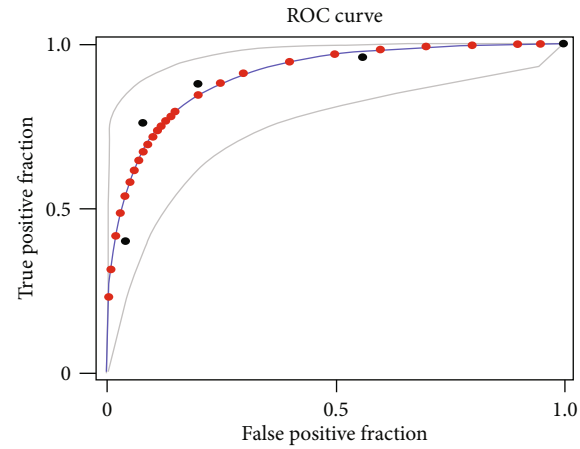


Figure 5: ROC curve for text normalization and statistical summary of preprocessing steps.

4.1. Proposed Approach for Tokenization. Like the procedure followed for formal text tokenization, some researchers followed the same on informal text tokenization [44]. In the formal text tokenization methods, they have divided the token into meaningful chunks if they encountered any spaces or specific delimiters present in the sentence. This method has resulted in poor performance in POS tagging and made the entity detection processes complicated for informal text such as tweets. Therefore, we have considered the key phrases of the token up to the length of 4 (i.e., as almost all the named entities can be restricted within that range) and split the tokens based on the following patterns:

(1) (Noun)+: for the given tweet, the tokenizers find more than one continuous nouns that can be clubbed into one key phrase and considered as one single token. Example: Samsung Galaxy Phone

(2) (Adjective) + (noun)+: if the noun started with one or more adjectives, again, it is considered to be a single token, and division has been made accordingly by the tokenizer. Example: Fantastic Donald Trump and Digital Camera

(3) (Noun) + [CD]: one or more nouns followed with some cardinals. Example: James Bond 007 and iPhone 8i

To support the above patterns and filter the tokens from the tweets, we have emulated the tokenizer called ARKTweetNLP [45] which is an open-source module for download and infused our pattern into the above package to effectively filter out the key phrases for the next phases of POS tagging. The main reason to choose this module over other tokenizers is the fact that this tokenizer has been designed by [46] in considering the Twitter-specific regular expressions, covered a wide range of emoticons, and achieved the good performance on tweet tokenization. We have given some examples in Table 8 that the proposed tokenizer has been able to split the tweets into some meaningful chunks successfully.

Table 8: Proposed method results on tokenized tweets.

Tweet	Tokenized tweet
Prince Charles met Albert of Monaco just days before he tested positive for coronavirus	Prince Charles/met/Albert of Monaco/just days/before/he/tested/positive/for coronavirus
Just heard Boris Johnson has tested positive for coronavirus. The man is not fit to lead out country.	Just heard/Boris Johnson/has tested/positive/for coronavirus./The man/is not fit/to lead out/country.
Tom Hanks and his wife Rita Wilson have both tested positive for coronavirus, the US actor said Wednesday	Tom Hanks/and/his wife/Rita Wilson/have/both/tested/positive/for coronavirus/, the US actor/said/Wednesday/.
I finally personally know someone that has tested positive for COVID-19. Interesting timing. As they are now getting more tests made, more tests coming back positive.	I/finally/personally/know/someone/that/has tested/positive/for COVID-19/. Interesting timing/. As/they/are now getting/more tests made/, more tests/coming back/positive.
Former movie producer Harvey Weinstein, who is serving a prison sentence for sexual assault and rape, has tested positive for coronavirus	Former/movie producer/Harvey Weinstein/, who/is serving/a prison sentence/for sexual assault/and/rape/, has tested/positive/for coronavirus.
Redmond-based Nintendo of America has confirmed that an employee has tested positive for coronavirus.	Redmond-based Nintendo/of America/has confirmed/that/an employee/has tested/positive/for coronavirus.

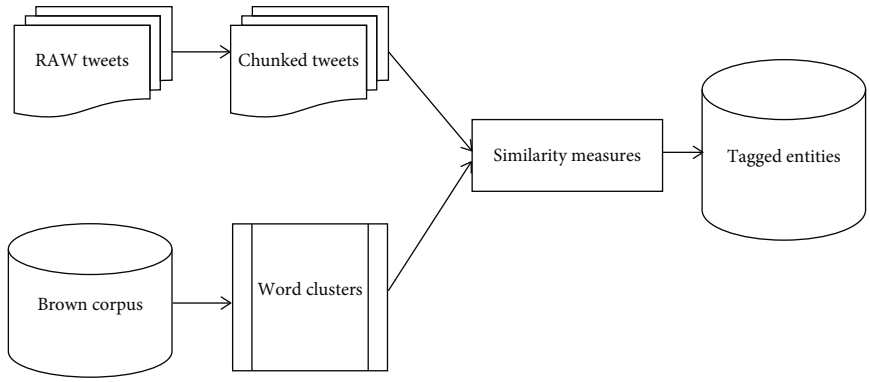


Figure 6: Proposed method for text normalization using word cluster.

5. Tweet Normalization and Processing

The next step in our pipeline is the normalization process that is used to identify the tokens as either Out-Of-Vocabulary or not and convert them into their Standard English word or word phrase. Normalization is very important for text processing because it will help to solve the ambiguity that persists on any token in the text [47, 48]. As explained earlier, many researchers have used statistical machine translations, phrase-based statistical model, character-level edit distance, dependency parsers, and even built-in parallel corpora to train the model to generate the possible candidates for the Out-Of-Vocabulary (OOV) words. Some have even tried to use language models and phonemic edit distance measures to deal with the problems differently, but normalization problem is quite challenging in itself and poses great difficulty to the informal text. For instance, abbreviations and slang words are very difficult to map in the existing spell correctors.

5.1. Proposed Approach for Normalization. We have used the supervised normalization technique for the proposed system and also utilized online resources such as Brown Corpus, PyEnchant Dictionary, and Microsoft Web *N*-gram model. We have used the Brown Corpus for text normalization because of the fact that similar words occur in a similar context. That is, similar words would have the same set of distributions and arrangements of words on either side, i.e., left and right. With that assumption, we have clustered the Brown Corpus and trained it for our supervised classifier to normalize the words if it encounters any OOV words. As the Brown Corpus has already clustered 47 Million tweets and almost produced more than 1200 word clusters (i.e., on each cluster, it has arranged the tweets with similar context), we have effectively utilized these clusters in this normalization process and transform the OOV words successfully as depicted in Figure 6.

Besides, we have used the PyEnchant Dictionary for mapping the OOV words in its dictionary and performed two string similarity measures such as Levenshtein and metaphone edit distance. We used both the string similarity measures for the candidate word selection because some of the words have given the same phonetic sound, for example, know and no and pork and fork. To disambiguate this conundrum, we have taken both the measures and choose the best one using the Microsoft Web *N*-gram model. Out of all the suggestions listed from both the Brown Cluster and PyEnchant Dictionary, the Microsoft Web *N*-gram model will choose the correct word which has the high score based on the conditional probabilities of candidate words given in the context (i.e., word combinations before and after the words as suggested by Brown Cluster and String Similarity Measure). The word

```
Input: Tweet Dataset
Output: Normalized Tweet
BEGIN
        Normalized Tweet ⟵ {}
        FOREACH token IN Tweet DO
                IF token NOT EQUAL TO Noun THEN
                        IF token IN OOV word THEN
                                Token⟵OOV word
                ELSE
                Token ⟵not OOV word
        IF token IN BROWN Cluster THEN
                Cand-Token ⟵ Fetch the candidate words from the BROWN Cluster
                                and perform the Levenshtein and Metaphone Edit distance
                Normalized Tweet ⟵ Append the Cand-token with highest frequency score
        ELSE
                Token⟵ Not in BROWN Cluster
        Sug-token ⟵ Retrieve the suggestion for the token using PyEnchant dictionary
        FOREACH Suggestion from Sug-token DO
                Score=(Prob(Prev-token + Suggestion) + Prob(Suggestion + Next-token))/2
                                                                [Using Microsoft Web N-Gram]
Normalized Tweet ⟵ Append the Suggestion with highest score
RETURN Normalized Tweet
END
```

Algorithm 1: Normalization Algorithm.

with the highest score can be returned as output by the system. The algorithm has been given below for the whole normalization process and how the system has returned the normalized output successfully.

6. Tweet Part-of-Speech Tagging

After normalizing the tweets using the hybrid approach, we need to perform the Part-Of-Speech (POS) tagging, and this process is very crucial for entity extraction and classification. The entity extraction has been performed on the tweets based on the POS tagging and extracts the entities which have been attributed as nouns, proper nouns, pronouns, or any objects as nouns [49]. So, POS tagging of tweets would determine the grammatical structure and category for each token that is segmented on the normalized tweets. Many words in the tweets would have no syntactic features such as hashtags, URLs, emoticons, and @mentions. The dependency parser has taken more time in processing than these nonsyntactic features and consumes time unnecessarily. But without appropriate utilization of standard annotators, with the use of a simple rule-based filter, it can extract and annotate the #hashtags, @mentions, punctuations, and retweet tokens effectively. Next, for the multiword expressions, we have considered two types of approaches. First, the proper nouns have been compounded together for information extraction and assigned the single tag for the compound words. Second, the lexical idioms, such as "stuck in the crowd" and "hay in the stack," have been manipulated with shallow parsing and clubbed into a single token for tagging. The same approach has been followed even for the idiomatic relationships and performs the internal analysis of multiword tokensdependency parsers.

6.1. Proposed Approach on Part-of-Speech (POS) Tagging. To effectively attribute the tags to every token divided on the given tweets, we have implemented the supervised learning approach to train the model and tagged the tokens on the linguistic features followed by the natural language processing toolkits. To assign the tag to the segmented tokens, there are many features considered such as capitalization, surrounding words, tags on the surrounding words, and presence of any cardinal on the word. Upon scrutinizing the model features, appropriate tags have been assigned to the token correctly. Since the tweets lack grammatical structures and a dearth of context around the words, assigning appropriate tags to the segmented tokens has become critical and sometimes failed to attribute the correct tagging. Also, as the tweets contain more slang words, OOV words, spelling errors, and abbreviations, it has become challenging to assign the appropriate POS tagging on the tokens and need extra supervision for picking the features for the proposed model. Particularly, capitalization has not been considered to be a good feature in informal text processing because many social users have not followed the proper capitalization of words. In addition to that, they have used more adjectives to extend their greetings and emphasized more on the thoughts which become a cumbersome task for the POS tagging (see Table 9). Hence, we have used the lightweight hidden Markov model to predict the context and assign the correct tags to the tokens.

6.2. Lightweight Hidden Markov Model–POS Tagging. The lightweight hidden Markov model used the supervised word clusters trained from Brown Corpus and extracted other lexical features generated from Stanford POS tagging. These word clusters have been used to train on the unlabeled tweets and filter out the word clusters again generated from the new labeled

Table 9: General POS tag and its descriptions.

Tag	Description
DT	Determiner
PRP	Person pronoun
VB	Verb, base form
VBP	Verb
IN	Preposition or conjunction
NN	Noun, singular, or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural

Table 10: COVIDKB Twitter dataset and its annotated ground truth.

Tweets	Tokens	Repeated characters	Abbreviations	Misspelled words	Total
1000	13487	672	308	1128	2108

datasets. These tagging features extract only the conventional tagging features such as words, surrounding words, surrounding tags, and use of cardinals, and also, it extracts the Twitter-specific features such as hashtags, usernames, and emoticons. This distributional-based word similarity feature finds useful for twitter streams and outperforms the Stanford tagger on the datasets given. The word-tag probability is considered to be a stochastic model in which the tagger is deemed as a Markov process with unobservable states and yields the observable outputs. In simpler terms, the POS tags are the states of the Markov process, and words are the desired results of the model. The conditional probability of the lightweight hidden Markov model has been given below:

The POS tagger comprises of the following:

(i) $P_s(T_i)$: probability of the sequence beginning in tag T_i

(ii) $P_t(T_j \mid T_i)$: Probability of the sequence changing from tag T_i to tag T_j

(iii) $P_E(W_j \mid T_i)$: Probability of the sequence terminating word W_j on tag T_i

The tagger makes two straightforward suppositions:

(i) The likelihood of a word depends just on its tag, i.e., given its tag, it is autonomous of different words and different labels

(ii) The likelihood of a tag depends just on its past labels, i.e., given the past labels, it is autonomous of next labels and labels before the past labels

In noun phrases, a noun acts like a subject or an object to a verb or an adjective. To create a noun phrase Chunker, we define a noun phrase Chunker that indicates how the tweet to be chunked.

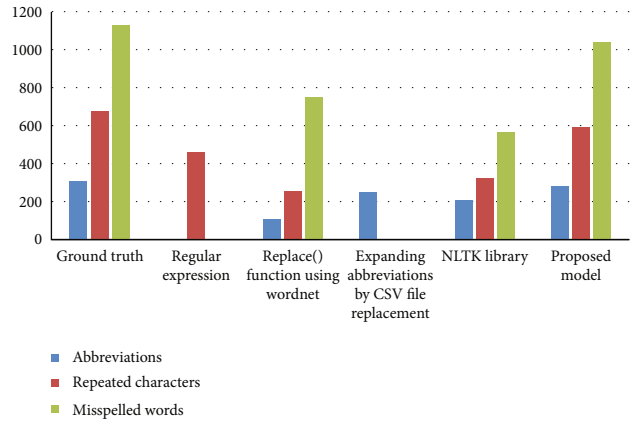


Figure 7: Statistical analysis of text normalization processes.

$$\text{Grammar} = \text{“NP}: \{<\text{DET}>?<\text{ADJ}>*<\text{Noun}>\}\text{”}. \quad (7)$$

Here are the following examples: tweet: “I am at Bicycle Ranch in Scottsdale, AZ”. POS tagging of tweet: (I/PRP, am/VBP, Bicycle/NNP, Ranch/NNP, in/IN, Scottsdale/NNP, AZ/NNP).

7. Experimental Evaluation and Analysis

For our experiment analysis, we have downloaded the COVIDKB Twitter datasets provided by [35]. The COVIDKB Twitter datasets consist of 10,000 annotated tweets with ground truth values. But for our experimental analysis, we have taken 1000 tweets as the test sets due to the condition of the system and to save the computational time (see Table 10). We have considered this COVIDKB Twitter datasets because it has been manually annotated with research students and contains a wide range of entities such as a person, location, travel, and contacts. Therefore, we benchmarked this dataset for our experimental analysis and observed the state of the performance of the proposed system.

7.1. Result Analysis. We have compared our proposed model with existing preprocessing techniques and found that our proposed model has outperformed all other conventional models and yielded a better performance and accuracy rate. For the analysis of the results, we have considered the basic normalizing features such as repeated characters, abbreviations, and misspelled words. As these three components are very crucial in the tweet normalization, it has been compared with regular expression and replace function using WordNet, NLTK Library, and replacement CSV file. Our proposed model has converted the informal words into correct English words with a much higher accuracy rate and almost 20-25% increase in the precision score. As the noisy token has been hindered the performance of POS tagging and further block the entity extraction, the proposed approach has paved the better way for the POS tagging at this level and helps to solve the entity detection and recognition in the next level of processing as it was witnessed in Figure 7. A detailed view of the comparison has been given in the following Table 11.

Table 11: Proposed method on normalizing repeated characters, abbreviations, and misspelled words.

Techniques	Abbreviations	Repeated characters	Misspelled words
Ground truth	308	672	1128
Regular expression	—	462	—
Replace() function using WordNet	108	253	749
Expanding abbreviations by CSV file Replacement	247	—	—
NLTK library	210	319	561
Proposed model	281	590	1036

Table 12: Proposed method accuracy rate on POS tagging of tweets.

Techniques	BLEU (%)	WER (%)
Regular expression	81.26	8.91
Replace() function using WordNet	80.19	10.73
Expanding abbreviations by CSV file Replacement	78.22	12.18
NLTK library	79.14	10.98
Proposed model	83.25	8.23

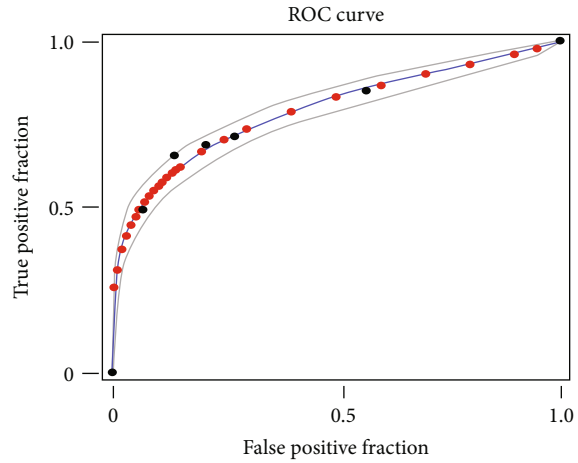


Figure 8: ROC curve for evaluated POS tagging of tweets.

Moreover, Table 12 presents the accuracy rate of the proposed method on POS tagging of tweets.

The experiments are evaluated based on the BLEU score and WER score. While analyzing the output with other comparison models, it has become evident that the proposed method has effectively removed the repeated characters and misspelled words and normalized the appropriate text forms for the given tweets. The calculated p values from the BLEU and WER have been comparatively less than 0.005 at the 95% CI level, 0.0029 and 0.0028, respectively. Therefore, the proposed method has invariably shown the consistent precision with its BLEU score less than 85% or the WER score greater than 8%. Figure 8 depicts the ROC Curve for evaluated POS tagging of tweets.

This ROC curve has been obtained after the comparative analysis over the techniques mentioned in Table 12. It has become very apparent that the proposed model has been yielded with good accuracy rate as it was measured with BLEU and WER techniques.

The overall analysis for the tweet normalization, preprocessing, and assignment of POS tags to the tokenized tweets has been given in Table 13. The accuracy rate and the sensitivity rate of the ROC analysis have been gradually improved after employing the proposed method. Eventually, the fitted ROC accuracy has been increased considerably and proved that the proposed method has outperformed other preprocessing steps carried through for the unstructured datasets.

Table 13: Overall plotting parameters for ROC utilized for tweet preprocessing.

Plotting parameters	Investigating OOV words in tweets	Tweet normalization and preprocessing	Evaluated POS tagging for tweets
Number of cases	2108	450	950
Number of correct	1907	412	702
Accuracy	77.5%	84%	73.9%
Sensitivity	75%	88%	68.6%
Specificity	80%	80%	79.2%
Positive cases missed	107	77	149
Negative cases missed	94	53	99
Fitted ROC curve	0.85	0.90	0.79

8. Conclusions

Natural language processing (NLP) allows digital gadgets and devices to comprehend the semantics in languages. Usually, the NLP can be broadly characterized into two categories such as data preprocessing and model development [50–57]. There are several text normalization strategies proposed by eminent researchers to solve the impending issues and reduce the error rate considerably [50–53]. However, they have certain confinements and still do not accomplish great outcomes when it has come to informal text processing. Rather than normalizing one kind of ill-formed word, we have considered a wide range of ill-formed words found on the tweets datasets and cleaned them under three primary classifications: incorrectly spelled words, contractions, and repeated letters. The primary motivation behind why we have sorted these unorthodox words is because we might want to guarantee that all subcategories of these three fundamental issues are standardized into their right form by the most appropriate procedures. Hence, the target of this exploration is to locate the best standardization approach with the end goal to productively and precisely clean tweets containing incorrect spellings, shortened forms, and repeated characters. The future scope of this research work would add tremendous advantage if it has been resourcefully utilized for natural disasters and the imminent threat of future disease outbreaks. The major limitation of this work has been restricted only with twitter streams in general and deals particularly on preprocessing techniques of the tweet tokenization and POS tagging. Further, it needs fine-grained datasets for identifying the OOV words if the domain-specific research work has been carried out in future.

Data Availability

The article's original contributions generated for this study are included; further inquiries can be directed to the corresponding author.

Conflicts of Interest

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

References

[1] D. Pla Karidi, Y. Stavrakas, and Y. Vassiliou, "Tweet and followee personalized recommendations based on knowledge graphs," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 6, pp. 2035–2049, 2018.

[2] L. Derczynski, A. Ritter, S. Clark, and K. Bontcheva, "Twitter part-of-speech tagging for all: overcoming sparse and noisy data," in *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pp. 198–206, Hissar, Bulgaria, 2013.

[3] F. Liu, F. Weng, B. Wang, and Y. Liu, "Insertion, deletion, or substitution? Normalizing text messages without pre-categorization nor supervision," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 71–76, Portland, Oregon, USA, 2011.

[4] L. Derczynski, D. Maynard, G. Rizzo et al., "Analysis of named entity recognition and linking for tweets," *Information Processing & Management*, vol. 51, no. 2, pp. 32–49, 2015.

[5] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: linking text sentiment to public opinion time series," in *Fourth international AAAI conference on weblogs and social media*, Washington, DC, USA, 2010.

[6] K. Bontcheva and L. Derczynski, "Extracting information from social media with gate," in *Working with Text*, pp. 133–158, Chandos Publishing, 2016.

[7] J. Tabassum, S. Lee, W. Xu, and A. Ritter, "WNUT-2020 task 1 overview: extracting entities and relations from wet lab protocols," 2020, https://arxiv.org/abs/2010.14576.

[8] Y. Hu, S. Gao, D. Lunga, W. Li, S. Newsam, and B. Bhaduri, "GeoAI at ACM SIGSPATIAL," *Sigspatial Special*, vol. 11, no. 2, pp. 5–15, 2019.

[9] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, "Semeval-2019 task 6: identifying and categorizing offensive language in social media (offenseval)," 2019, https://arxiv.org/abs/1903.08983.

[10] C. Zong, F. Xia, W. Li, and R. Navigli, *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, Association for Computational Linguistics, Online, edition, 2021.

[11] K. Toutanova, A. Rumshisky, L. Zettlemoyer et al., *Proceedings of the 2021 conference of the north American chapter of the Association for Computational Linguistics: human language technologies*, Association for Computational Linguistics, 2021.

[12] E. M. Voorhees and D. Harman, "Overview of the sixth text retrieval conference (TREC-6)," *Information Processing & Management*, vol. 36, no. 1, pp. 3–35, 2000.

[13] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[14] D. G. Lee, H. C. Rim, and D. Yook, "Automatic word spacing using probabilistic models based on character n-grams," *IEEE Intelligent Systems*, vol. 22, no. 1, pp. 28–35, 2007.

[15] Z. Xue, D. Yin, and B. D. Davison, "Normalizing microtext," in *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*, San Francisco, USA, August 2011.

[16] M. Kaufmann and J. Kalita, "Syntactic normalization of twitter messages," in *International conference on natural language processing*, Kharagpur, India, January 2010.

[17] A. Garain, S. K. Mahata, and S. Dutta, "Normalization of numeronyms using nlp techniques," in *2020 IEEE Calcutta Conference (CALCON)*, Kolkata, India, 2020.

[18] E. Mapa, L. Wattaladeniya, C. Chathuranga et al., "Text normalization in social media by using spell correction and dictionary based approach," *Systems learning*, vol. 1, pp. 1–6, 2012.

[19] L. Besacier, E. Barnard, A. Karpov, and T. Schultz, "Automatic speech recognition for under-resourced languages: a survey," *Speech Communication*, vol. 56, pp. 85–100, 2014.

[20] B. Han, P. Cook, and T. Baldwin, "Lexical normalization for social media text," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 1, pp. 1–27, 2013.

[21] P. Sosamphan, V. Liesaputra, S. Yongchareon, and M. Mohaghegh, "Evaluation of statistical text normalisation techniques for twitter," in *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, vol. 1, pp. 413–418, Porto, Portugal, 2016.

[22] A. Sarker, "A customizable pipeline for social media text normalization," *Social Network Analysis and Mining*, vol. 7, no. 1, pp. 1–13, 2017.

[23] J. Kim, E. Lee, T. Hong, and P. Kim, "Correcting Misspelled Words in Twitter Text," in *International Conference on Big Data Technologies and Applications*, pp. 83–90, Springer, Cham, 2017.

[24] C. Napoles and C. Callison-Burch, "Systematically adapting machine translation for grammatical error correction," in *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 345–356, Copenhagen, Denmark, 2017.

[25] R. Beckley, "Bekli: a simple approach to twitter text normalization," in *Proceedings of the Workshop on Noisy User-generated Text*, pp. 82–86, Beijing, China, July 2015.

[26] J. J. Jung, "Online named entity recognition method for microtexts in social networking services: A case study of twitter," *Expert Systems with Applications*, vol. 39, no. 9, pp. 8066–8070, 2012.

[27] Y. Tsvetkov and C. Dyer, "Lexicon stratification for translating out-of-vocabulary words," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 125–131, Beijing, China, July 2015.

[28] N. F. Liu, J. May, M. Pust, and K. Knight, "Augmenting statistical machine translation with subword translation of out-of-vocabulary words," 2018, https://arxiv.org/abs/1808.05700.

[29] M. Savargiv, B. Masoumi, and M. R. Keyvanpour, "A new ensemble learning method based on learning automata," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 3467–3482, 2020.

[30] V. K. Sharma, N. Mittal, and A. Vidyarthi, "Context-based translation for the out of vocabulary words applied to hindi-english cross-lingual information retrieval," *IETE Technical Review*, vol. 39, no. 2, pp. 276–285, 2020.

[31] E. Egorova and L. Burget, "Out-of-Vocabulary word recovery using fst-based subword unit clustering in a hybrid asr system," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5919–5923, Calgary, AB, Canada, April 2018.

[32] N. K. Singh, D. S. Tomar, and A. K. Sangaiah, "Sentiment analysis: a review and comparative analysis over social media," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 97–117, 2020.

[33] E. Elakiya and N. Rajkumar, "RETRACTED ARTICLE: In text mining: detection of topic and sub-topic using multiple spider hunting model," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3571–3580, 2021.

[34] K. Liu, *Incorporate out-of-Vocabulary Words for Psycholinguistic Analysis Using Social Media Texts-An OOV-Aware Data Curation Process and a Hybrid Approach*, The Claremont Graduate University, 2021, (Doctoral dissertation.

[35] S. Zong, A. Baheti, W. Xu, and A. Ritter, *Extracting a Knowledge Base of COVID-19 Events from Social Media*, arXiv, 2021.

[36] R. Ferreira, R. D. Lins, S. J. Simske, F. Freitas, and M. Riss, "Assessing sentence similarity through lexical, syntactic and semantic analysis," *Computer Speech & Language*, vol. 39, pp. 1–28, 2016.

[37] D. Maynard, I. Roberts, M. A. Greenwood, D. Rout, and K. Bontcheva, "A framework for real-time semantic social media analysis," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 44, pp. 75–88, 2017.

[38] H. Mulki, C. B. Ali, H. Haddad, and I. Babaoğlu, "Tw-star at semeval-2018 task 1: preprocessing impact on multi-label emotion classification," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 167–171, New Orleans, Louisiana, USA, June 2018.

[39] S. K. Narayanasamy, K. Srinivasan, S. Mian Qaisar, and C. Y. Chang, "Ontology-enabled emotional sentiment analysis on COVID-19 pandemic related twitter streams," *Frontiers in public health*, vol. 9, 2021.

[40] A. Pinto, H. Gonçalo Oliveira, and A. Oliveira Alves, "Comparing the performance of different NLP toolkits in formal and social media text," in *5th Symposium on Languages, Applications and Technologies (SLATE'16)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[41] M. A. Hassonah, R. Al-Sayyed, A. Rodan, A. Z. Ala'M, I. Aljarah, and H. Faris, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on twitter," *Knowledge-Based Systems*, vol. 192, article 105353, 2020.

[42] D. Bollegala, R. Kiryo, K. Tsujino, and H. Yukawa, "Language-independent tokenisation rivals language-specific tokenisation for word similarity prediction," 2020, https://arxiv.org/abs/2002.11004.

[43] R. Dridan and S. Oepen, "Tokenization: returning to a long solved problem—a survey, contrastive experiment, recommendations, and toolkit—," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 378–382, Jeju Island, Korea, July 2012.

[44] D. Yogish, T. N. Manjunath, and R. S. Hegadi, "Review on Natural Language Processing Trends and Techniques Using

Nltk," in *International Conference on Recent Trends in Image Processing and Pattern Recognition*, pp. 589–606, Springer, Singapore, 2018.

[45] S. Chatterji, R. K. Rahul, and A. Arora, "A hybrid approach for identifying sentiments around aspects," in *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, pp. 33–37, Kolkata, India, July 2015.

[46] S. Medeiros, F. Mascarenhas, and R. Ierusalimschy, "From regexes to parsing expression grammars," *Science of Computer Programming*, vol. 93, pp. 3–18, 2014.

[47] N. Hanafiah, A. Kevin, C. Sutanto, Y. Arifin, and J. Hartanto, "Text normalization algorithm on twitter in complaint category," *Procedia computer science*, vol. 116, pp. 20–26, 2017.

[48] V. C. Tran, N. T. Nguyen, H. Fujita, D. T. Hoang, and D. Hwang, "A combination of active learning and self-learning for named entity recognition on twitter using conditional random fields," *Knowledge-Based Systems*, vol. 132, pp. 179–187, 2017.

[49] Y. Jamoussi and A. Y. Nouira, "An extracting model for constructing actions with improved part-of-speech tagging from social networking texts," in *2017 11th international conference on intelligent systems and control (ISCO)*, pp. 77–81, Coimbatore, India, January 2017.

[50] S. K. Narayanasamy and M. Dinakaran, "An algorithmic approach to rank the disambiguous entities in Twitter streams for effective semantic search operations," *Sādhanā*, vol. 45, no. 1, p. 29, 2020.

[51] C. Iwendi, C. G. Huescas, C. Chakraborty, and S. Mohan, "COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients," *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–21, 2022.

[52] A. Kumar, K. Srinivasan, W.-H. Cheng, and A. Y. Zomaya, "Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data," *Information Processing & Management*, vol. 57, no. 1, article 102141, 2020.

[53] S. Chauhan, R. Banerjee, C. Chakraborty, M. Mittal, A. Shiva, and V. Ravi, "A self-congruence and impulse buying effect on user's shopping behaviour over social networking sites: an empirical study," *International Journal of Pervasive Computing and Communications*, vol. 17, no. 4, pp. 404–425, 2021.

[54] A. Shabbir, M. Shabbir, A. R. Javed, M. Rizwan, C. Iwendi, and C. Chakraborty, "Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals," *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–28, 2022.

[55] P. A. Pattanaik, M. Mittal, M. Z. Khan, and S. N. Panda, "Malaria detection using deep residual networks with mobile microscopy," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 1700–1705, 2022.

[56] A. R. Javed, M. U. Sarwar, S. Khan, C. Iwendi, M. Mittal, and N. Kumar, "Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition," *Sensors*, vol. 20, no. 8, p. 2216, 2020.

[57] S. Ramaneswaran, S. Vijay, and K. Srinivasan, "TamilATIS: dataset for task-oriented dialog in Tamil," in *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pp. 25–32, Association for Computational Linguistics, Dublin Ireland, 2022.

[58] T. Finin, W. Murnane, A. Karandikar, N. Keller, J. Martineau, and M. Dredze, "Annotating named entities in Twitter data with crowdsourcing," in *Proceedings of the NAACL Workshop on Creating Speech and Text Language Data With Amazon's Mechanical Turk*, pp. 80–88, Los Angeles, 2010.

[59] A. Ritter, C. Cherry, and B. Dolan, "Data-driven response generation in social media," in *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 583–593, Edinburgh, Scotland, UK, 2011.

[60] M. Rowe and H. Alani, "Mining and comparing engagement dynamics across multiple social media platforms," in *Proceedings of the 2014 ACM conference on Web science*, pp. 229–238, New York, NY, USA, June 2014.

[61] L. Derczynski, K. Bontcheva, and I. Roberts, "Broad twitter corpus: A diverse named entity recognition resource," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 1169–1179, Osaka, Japan, December 2016.

Hindawi
Journal of Sensors
Volume 2022, Article ID 8980198, 10 pages
https://doi.org/10.1155/2022/8980198

Research Article

Research on Working Memory States Based on Weighted K-Order Propagation Number Algorithm: An EEG Perspective

Yao Chen,[1] Yuhong Zhang,[2] Weiwei Ding,[1] Fachang Cui,[1] and Liya Huang[1,3]

[1]*College of Electronic and Optical Engineering & College of Flexible Electronics, Nanjing University of Posts and Telecommunications, Nanjing, China*
[2]*College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, China*
[3]*National and Local Joint Engineering Laboratory of RF Integration and Micro-Assembly Technology, Nanjing, China*

Correspondence should be addressed to Liya Huang; huangly@njupt.edu.cn

Received 27 April 2022; Accepted 26 June 2022; Published 11 July 2022

Academic Editor: Rocio Perez de Prado

Working memory (WM) is considered the mental workplace that retains and manipulates information. This study investigates the internal mechanism in WM states from an electroencephalography (EEG) network perspective. Firstly, we devised a novel letter-sequence version of the n-back experiment to collect EEG data, analyzed the neural oscillations in the theta and gamma bands, and then constructed Phase Lock Value (PLV) grounded brain networks to examine the synchronizations among dissimilar brain regions. The complex topology properties (e.g., global efficiency, local efficiency, and small-worldness) were scrutinized as well. Additionally, we presented an original algorithm, the Weighted K-Order Propagation Number (WKPN) algorithm, to extract the important brain regions associated with WM processes. The simulation revealed that the frontal and posterior regions were activated in two WM states, i.e., update and readout states. Throughout the readout, brain networks performed better in efficiency and resistance to interference. Furthermore, the right prefrontal and parietooccipital regions became more prominent in the completion of extra difficult WM tasks. In summary, these EEG-based results can be taken as promising evidence to understand and improve WM.

1. Introduction

Working memory (WM) is defined as a system that temporarily processes and stores information with restricted capacity in the human brain [1]. It plays an important role in human intelligence and complex cognitive activities such as learning, comprehension, and reasoning. Standard WM states include update, maintenance, and readout [2]. As the objects being remembered are constantly varying, WM adapts to these alterations by updating the stored information with new ones. Another important function of WM is to accurately read information from related brain regions for further cognitive tasks. Studies have shown that during these two states of WM (e.g., update and readout), the brain exhibits abnormal patterns in patients with neurological disorders, such as Attention Deficit Hyperactivity Disorder (ADHD) [3] and Posttraumatic Stress Disorder (PTSD) [4]. Therefore, the study of these two states is of great importance.

Due to the noninvasive and high temporal resolution characteristics, electroencephalography (EEG) has become a popular technique used in the study of human cognitive function [5, 6]. A body of evidence indicates that EEG oscillations in the theta and gamma bands are of particular relevance to WM states. For example, Itthipuripat et al. [7] concluded that the theta power of the frontal cortex increases during the updating state. Polanía et al. [8] decoded contents of visual WM information within high-gamma oscillations in the human prefrontal cortex (PFC) during encoding (that is, updating) and maintenance periods. The research of Semprini et al. [9] likewise validated that theta and gamma oscillations of frontal and posterior areas are associated with updating of memory information, the maintenance of WM, and readout. However, there is

no consensus on the specific oscillation mechanisms corresponding to WM states.

There is increasing recognition that human functional states depend not only on the oscillations of the independent cortices but also on the interactions between a large number of neurons or different regions [10]; consequently, scholars have applied network methods to study information exchange and communication of brain regions [11, 12]. By considering the EEG electrodes as nodes and the statistical relationships between them as edges, a complex brain network can be implemented. In particular, statistical relationships can be quantified using functional connectivity metrics such as Mutual Information [13], Pearson's correlation [14], and Phase Locking Value (PLV) [15]. Since PLV can reflect phase synchronization of signals and interactive coupling and control relationships of different brain regions in a superior degree, it is more suitable for studying the EEG data [16]. As a result, in this study, we construct PLV-based brain networks to investigate the underlying mechanism of WM.

After constructing brain networks, researchers have applied graph theory to analyze network features and probe the human cognitive mechanism and pathogenesis of neurological diseases. As representing the global transmission of information and functional separation in networks, the global efficiency and local efficiency of directed functional networks have been analyzed in Alzheimer's disease in alpha and beta bands [17]. Furthermore, it has been demonstrated that the neural network has a small-world topology [18] that supports simultaneous global and local information processing. The small-worldness index has also been calculated to extract topology descriptors of brain networks to understand WM phases [19].

These indicators are effective in assessing overall network topography but are not at local important brain regions where activities are generally associated with cognitive tasks. For example, Haque et al. [20] suggested that the prefrontal region is vital for WM improvement from the perspective of electrical stimulation. However, evidence of concrete locations of important brain regions is still needed. Currently, algorithms for evaluating important regions of networks include, but are not limited to, weighted degree centrality (WDC) algorithm [21] and weighted betweenness centrality (WBC) algorithm [22]. The WDC algorithm only considered the impact of adjacent nodes on the importance of a node, which cannot reflect the global characteristics. The WBC algorithm considered the global properties but ignored the influence of the node number on the transmission efficiency of the network. To further bridge the local and global characteristics, our team proposed the Weighted K-Order Propagation Number (WKPN) algorithm [23], which introduced a scale factor K that manifests the variation of network properties and enables multiscale analysis of structural changes in networks.

In this paper, we investigate the update and readout states under different loads and study the mechanisms with the WKPN algorithm. Different from the old-fashioned experimental paradigm of memorizing Arabic numerals, we first proposed an original WM paradigm combining alphabet letters with a classical n-back experiment [24]. Then, the neural oscillations and activation brain regions in theta and gamma bands were analyzed. To study the interactions among brain regions, we calculated the PLV of electrode pairs to construct brain networks. Their general topology features were examined by computing global efficiency and local efficiency to assess the small-worldness property. Lastly, the WKPN algorithm was applied to extract the node importance features of brain networks to locate specific sites of key brain regions. The findings are taken as a contribution from an EEG perspective to WM investigations in the future.

2. Materials and Methods

The analysis of WM states can be divided into the following six steps: (1) Designing an experiment to obtain EEG data. (2) Preprocessing raw signals to get analyzable data. (3) Determining the investigated frequency band by time-frequency analysis. (4) Constructing brain networks based on PLV in the determined frequency band. (5) Calculating the properties (e.g., global efficiency and local efficiency) of brain networks. (6) To find the key nodes affecting network properties, the WKPN algorithm is used to locate important brain regions.

2.1. Paradigm. For the purpose of inducing WM with diverse levels of difficulty and load, we devised two n-back experiments with a letter-sequence version (in Figure 1). The lengths of the letter sequence for them were 2 (low WM load) and 4 (high WM load), respectively, and each included 1-back, 2-back, and 3-back tasks. The letters presented were randomly selected from the English alphabet. The presentation of the paradigm was implemented by E-Prime v2.0 (Psychology Software Tools) in this study.

The paradigm started with a cue sign (a black cross) presented in 2 seconds. Subsequently, three groups of letter sequences were displayed in the center of the screen. Subjects are asked to memorize them in order. The duration of every two letters was 3 seconds, and for better memorization, every four letters were presented for 5 seconds. The update state was defined as shown in Figure 1(a). This was followed by a blank picture that lasted 0.5 seconds, which corresponds to the maintenance state. A task cue (a number) of n-back remained on the screen to remind the subjects to recall and record the letter sequence, and this period was considered the readout state. All subjects were required to complete the experiment for the 2 letters condition, then take a 10-minute break to finish the 4 letters condition (in Figure 1(b)). Furthermore, subjects were asked to respond as accurately and quickly as possible, and their reaction time and accuracy were recorded to analyze behavioral performance. Before the formal experiment, a practicing session was prepared for subjects to ensure they were familiar with the entire procedure.

2.2. EEG Recording. The EEG data were recorded with a 500 Hz sampling rate using the NEUROSCAN electrode

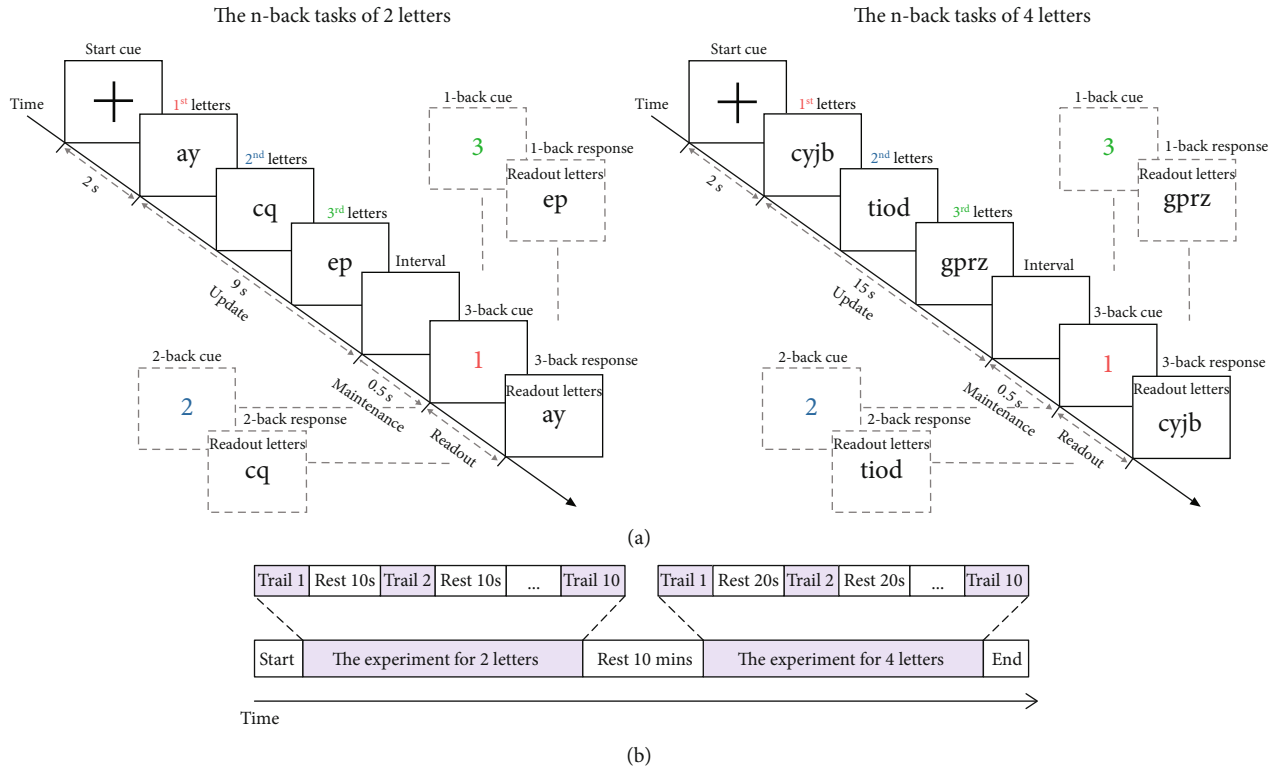


Figure 1: Experimental paradigms and procedure. (a) Paradigms for the n-back ($n = 1, 2, 3$) tasks of 2 letters and 4 letters conditions. (b) The whole experimental procedure.

cap, the SynAmps amplifier, and Curry 7.0 software. The 64 Ag/AgCl electrodes were placed in accordance with the international 10-20 system [25], and the ground electrode was placed on the FZ electrode. The reference electrodes, M1 and M2, were located in the Bilateral mastoid for measurement of myoelectricity, and the additional electrodes HEO and VEO at the outer left and right canthus and up and down the left eye, respectively, for the recording of the movement of the eye. The impedances were kept below 10 kiloohms.

Sixteen healthy students (2 female, age: $M = 23.6$ years old, $SD = 1.2$) were recruited for the experiments, and their participation was rewarded for the better experimental performance. These subjects were right-handed and had no personal or family history of neurological disease. Their memory abilities were all at the same normal level. Before the formal experiment, subjects were prohibited from consuming alcohol and were required to remain sober. Written and oral forms of informed consent were provided, and all subjects chose the latter for convenience. Only subjects who gave their consent were allowed to perform the experiment.

2.3. Data Preprocessing Analysis. Data preprocessing analysis was performed using Brainstorm [26]. The continuous raw data were first filtered with a 0.5-60 Hz bandpass filter to exclude the extra low- and high-frequency component, and with a 50 Hz notch filter to eliminate the power-line interference. The signal artifacts were removed in two steps: first, we manually deleted a small number of signal fragments with significant multiple-channel noise. Then, independent component analysis (ICA) [27] was applied to decompose the EEG data into twenty components. Components of the artifact related to eye movements, muscle activities, and heartbeats were removed. Finally, the artifact-free data were segmented into the update epoch and the readout epoch according to Figure 1. Prior to analysis, the epoch data were removed linear trend and completed baseline correction, where the baseline was the signal from -2 to -0.002 sec for update epoch analysis. For readout epoch analysis, the baseline depended on the length of readout epoch. Readout epochs with incorrect responses were excluded.

2.4. Time-Frequency Analysis. To analyze the time-frequency representation of epoch EEG data in theta and gamma bands, the complex Morlet wavelet transform was performed. The continuous-time signal $x(t)$ was convolved with the complex Morlet wavelet function, $w(t, f)$, to acquire the power spectrum:

$$TF(t, f) = |w(t, f) * x(t)|^2, \quad (1)$$

where t is time and f is frequency. $w(t, f)$ has the shape of a sinusoid, weighted by a Gaussian kernel [28], and it can be expressed as

$$w(t, f) = Ae^{\left(-t^2/2\sigma_t^2\right)} e^{i2\pi ft}, \quad (2)$$

where σ_t is the standard variation of the Morlet wavelet in the time domain, $A = (\sigma_t\sqrt{\pi})^{-1/2}$, $i = \sqrt{-1}$. The resolution

of complex Morlet wavelet transform is given in units of Full-Width Half Maximum of the Gaussian kernel, both in time and frequency. The frequencies were reorganized into two frequency bands (theta (4-8 Hz) and gamma (30-60 Hz)) closely related to the WM process.

We divided epoch data into several trials and averaged the time-frequency spectra of the subject across all trials to conclude which brain regions were activated at specific frequency bands. Statistics on power changes were performed by one-way ANOVA (significant level is 5%). The significant frequency bands in WM states were analyzed and chosen for the follow-up brain network construction.

2.5. Complex Brain Network Analysis

2.5.1. Constructing Brain Network. The human brain can be seen as a complex network formed with specific connectivity patterns [29]. We created PLV brain networks in theta and gamma bands. At first, information in the two bands of epochs data (baseline excluded) was extracted, and then the Hilbert Transform [30] was performed on a pair of electrode signals $x(t)$ and $y(t)$ to obtain their analytic signals $z_x(t)$ and $z_y(t)$, according to the equations (3) and (4),

$$z_x(t) = x(t) + iHT(x(t)), \tag{3}$$

$$z_y(t) = y(t) + iHT(y(t)). \tag{4}$$

The relative phase $\Delta\phi(t)$ between $z_x(t)$ and $z_y(t)$ was also computed, and the corresponding PLV [31] was calculated as,

$$PLV_{xy} = \left| \frac{1}{n} \sum_{k=1}^{n} e^{i\Delta\phi(t_k)} \right|, \tag{5}$$

where n is the total number of data points and t_k is a data point. Note that according to its definition, PLV always takes values between 0 and 1. The larger the PLV, the stronger the phase synchronization of the two signals, and 1 signifies that one signal completely follows the other. Finally, the electrodes were considered nodes, and the PLV values were viewed as the weights of edges to construct the weighted brain networks with 62 nodes (the M1 and M2 electrodes were excluded).

2.5.2. Global Efficiency and Local Efficiency. Global efficiency reflects the efficiency of information exchange of a network, and local efficiency corresponds to network robustness. According to the literature [32], small-world networks have both high global efficiency and local efficiency. Therefore, after constructing the PLV networks, we selected the two indices to evaluate the small-worldness to describe the general topological properties of the networks. Global efficiency and local efficiency were calculated using the Brain Connectivity Toolbox developed for the MATLAB environment.

Since the efficiency of the communication between nodes v_i and v_j can be expressed as the inverse of the shortest path, $1/d_{ij}$, between them, the global efficiency of the brain network is defined as the average of the inverse of the shortest paths among all node pairs [33],

$$E_{\text{global}} = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \tag{6}$$

where $n = 62$. Note that if v_i is not connected to v_j, then d_{ij} is equal to ∞. The value of E_{global} ranges from 0 to 1, and the larger the value, the better the ability to exchange information within the network.

Rows and columns corresponding to node v_i ($i = 1, 2, \cdots, n$) in adjacency matrix A were deleted, and then n subnetworks S_i were constructed. Local efficiency of the network is introduced as the average global efficiency of all subnetworks [34],

$$E_{\text{local}} = \frac{1}{n} \sum_{i=1}^{N} E_{\text{global}}(S_i). \tag{7}$$

The local efficiency assesses the ability of information transmission of the network in the absence of a certain node.

Any changes in global and local efficiency related to the different conditions (update, readout, 2 letters, 4 letters, n-back tasks) were evaluated by means of statistical comparison, a one-way ANOVA method. This test was taken for all subjects and the significance level was set at 5%.

2.5.3. Weighted K-Order Propagation Number Algorithm. Considering that the WKPN algorithm was proposed based on the comprehensive examination of the local and global properties of networks, we used it to assess the specific locations of important brain regions.

The WKPN algorithm abstracts the disease transmission based on the network topology. Each node was set as the source of infection separately, and after a certain propagation time, the node importance was obtained based on the number of infected nodes in the network. When the value of propagation time is small, the result reflects the local characteristic of the network, and as the time value increases, it corresponds to the global property. Simulations showed that this method can thoroughly portray the impact of long-distance connections on information transmission in small-world networks and can also increase the importance of bridge nodes [35].

The steps of the WKPN algorithm are explained in detail as follows. Based on the adjacency matrix A (62×62) of the network, the shortest paths of all node pairs were calculated. The propagation time K is defined as the values of the shortest paths, namely, $K \in [0, d]$, and d is the diameter of the weighted network. The number of nodes that a node v_i can reach within K is considered the K-order propagation number $N_{v_i}^K$, and it was combined with the information entropy to calculate the K-order structural entropy of the brain network,

$$H^K = -\sum_{i=1}^{n} \frac{N_{v_i}^K}{\sum_{j=1}^{n} N_{v_j}^K} \log \left(\frac{N_{v_i}^K}{\sum_{j=1}^{n} N_{v_j}^K} \right), \tag{8}$$

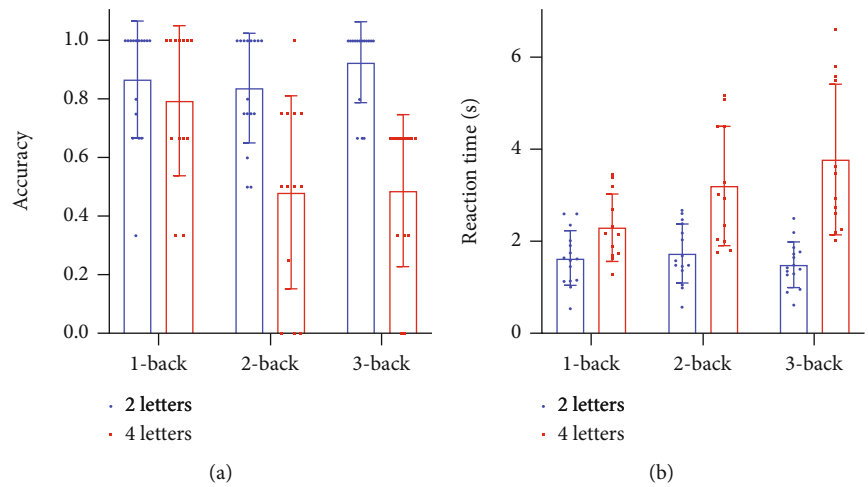


FIGURE 2: Behavioral performance of all subjects. (a) Response accuracy. (b) Reaction time. Statistics were performed using the paired sample t test with a 95% confidence interval. A single data point represents a subject. Error bars represent standard deviation (SD).

where n is the number of nodes, equal to 62 based on previous work. Subsequently, the node importance of v_i is defined by considering the comprehensive evaluation from $K = 0$ to $K = d$, according to the formula,

$$Q_{v_i} = \sum_{K=0}^{d} \left(1 - \frac{H^K - \min(H)}{\max(H) - \min(H)}\right) \left(\frac{N_{v_i}^K - \min\left(N^K\right)}{\max\left(N^K\right) - \min\left(N^K\right)}\right), \tag{9}$$

where $H = \{H^0, H^1, \cdots, H^d\}$ and $N^K = \{N_{v_1}^K, N_{v_2}^K, \cdots, N_{v_n}^K\}$. Eventually, the nodes of the PLV networks were ranked from highest to lowest importance to obtain vital brain regions. The algorithm has been validated to be effective for the assessment of important nodes in complex networks [35].

3. Results and Discussion

3.1. Behavioral Results. We recorded and analyzed the behavioral performance of all subjects during the experiments. As shown in Figure 2, response accuracy usually decreased and reaction time increased as the difficulty of the n-back task and WM load increased. However, interestingly, in the 3-back task, the accuracy was higher ($p = 0.085$) and the reaction time was shorter ($p < 0.05$) as compared to the 2-back task in the 2 letters condition. This is contrary to the perception that the harder the task, the lower the response accuracy. We speculated that the memory target of the 3-back is the initial appearance which has an impressive effect on the brain to elicit better WM behavioral results under the low load condition.

3.2. Results of EEG Power. The topographical maps of time-averaged power in theta and gamma bands are reported in Figure 3. For the low WM load condition (in Figure 3(a)), the midfrontal and posterior regions were activated at the theta band in both update and readout states. Theta power was generally higher in the update state than in readout ($p < 0.05$, $p < 0.05$), which is consistent with the fact that theta oscillations in these two regions have a vital impact on WM update[36]. During update and readout, the gamma power was increased in the occipital lobe, possibly indicating that the posterior region plays an important role in integrating and processing visual stimuli in our experiment.

Under the high WM load condition (in Figure 3(b)), the frontal and posterior regions were activated more strongly in theta band compared to under the low load condition during the update ($p < 0.05$, $p < 0.05$). It means that the oscillations were strengthened with WM load increasing. However, we did not find a clear pattern in the power changes of the activated regions during readout. Considering that the response accuracy in the 2-back and 3-back tasks had been as low as 50%, we believed that 12 letters may have exceeded the WM capacities of subjects, and thus for them, there was no difference among the n-back tasks.

The differences of theta power in readout state under low WM load were calculated at a representative electrode (FZ) of the midfrontal region where the power peaked in the n -back tasks (in Figure 4). The theta power of the FZ electrode in the 3-back task was lower than that in the 2-back task ($p < 0.05$), which is contrary to the conclusion related to the difficulty factor of the tasks. This trend echoes the behavioral results, and our conjecture that the memory content with first presence makes a strong impression has been verified.

3.3. Topological Properties of PLV Brain Networks. The global efficiency and the local efficiency of all brain networks are demonstrated in Figure 5. As shown in the box plots, the global efficiency and local efficiency were generally higher in readout as compared to those in the update state. It indicates that brain networks have higher small-world properties; namely, they are more capable of transmitting information globally and more resistant to interference. Furthermore, we found that global efficiency and local efficiency in readout were higher under low load condition. Given the lower response accuracy in behavioral results, we suggest that the

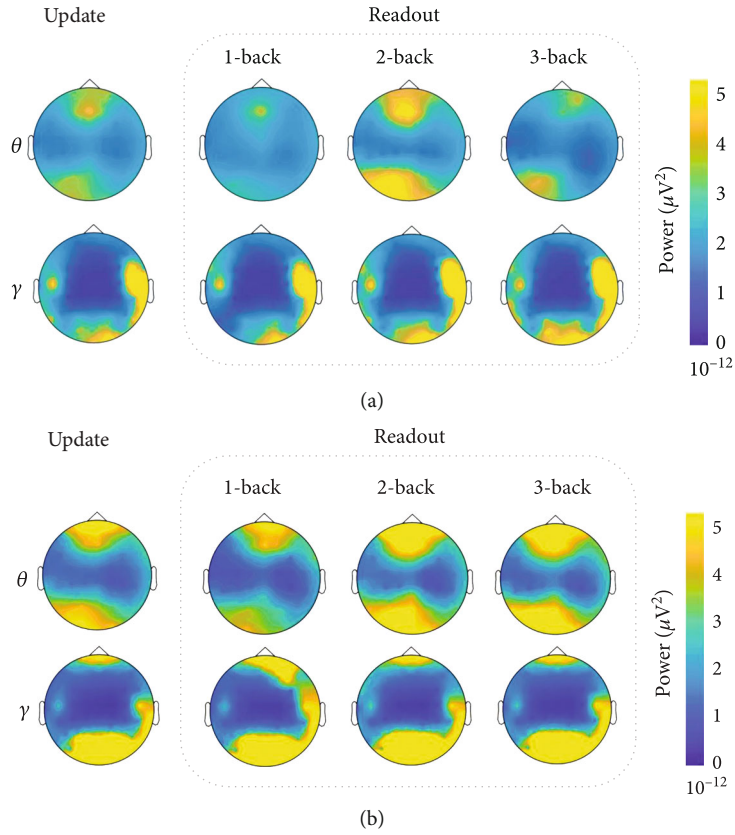


FIGURE 3: Topographic maps of the theta and gamma power in update and readout states for an example subject. (a) 2 letters condition. (b) 4 letters condition.

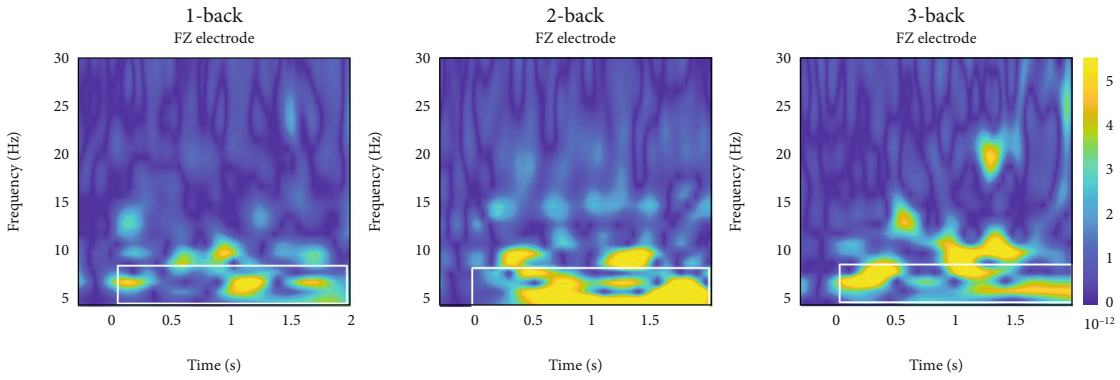


FIGURE 4: Time-frequency power plots (after baseline correction) from a representative midfrontal electrode (FZ) for 1-back, 2-back, and 3-back tasks for an example subject. The area marked in white is the theta band.

small-worldness of the brain is reduced under high WM load condition, especially when it exceeds the individual's capacity.

We ranked the importance of all nodes in order from the highest to the lowest based on the WKPN algorithm. The top five important nodes of the PLV networks and their first five edges with the highest weights are shown in Figure 6. Figure 6(a) illustrates that important brain regions in both update and readout states are concentrated in the frontal and parietooccipital lobes for low WM load condition. It is consistent with the results of time-frequency analysis and reflects the processing of visual information in the occipital lobe [37] and the executive control of the frontal region [38] during the WM task in terms of synchrony. In the theta band, the frontal regions were more tightly connected in readout. With the increase of difficulties of n-back tasks, the node importance of the right prefrontal electrodes became greater, especially the AF4 electrode, and the number of important nodes in the parietooccipital region was growing. It indicates that the interactive coupling and control relationships in the two regions intensified with the increase of task complexity. In particular, it has been

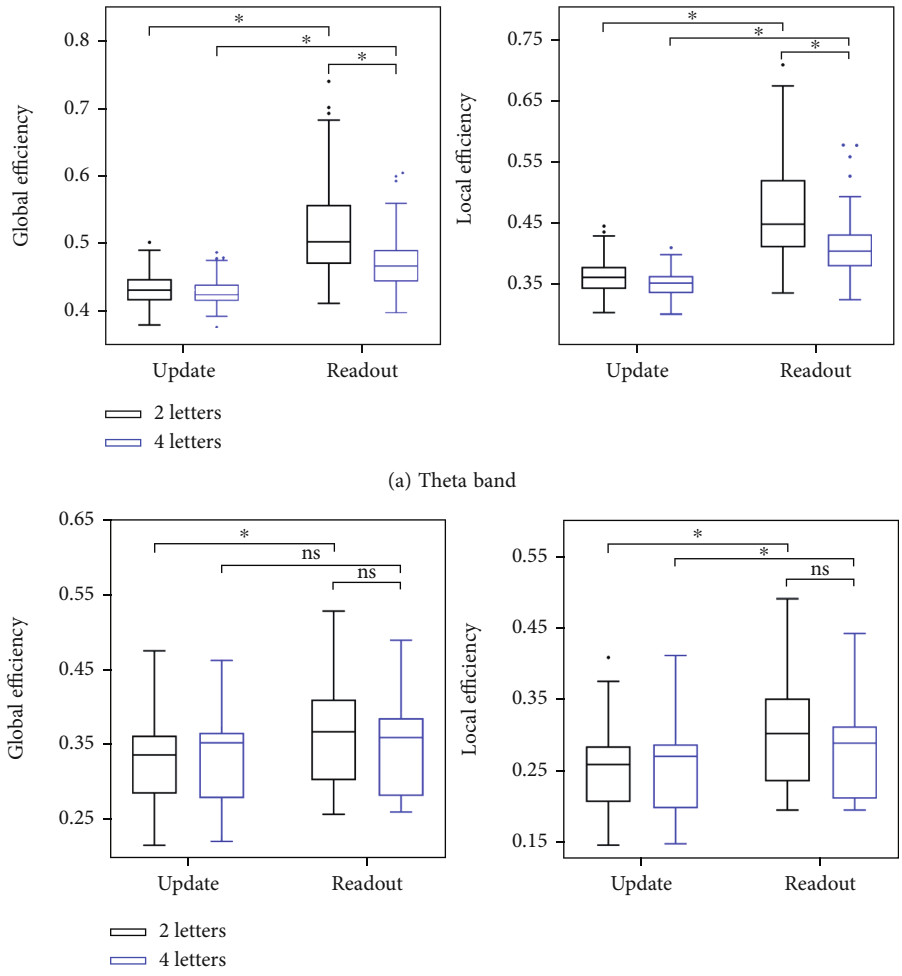


Figure 5: The box plots for 2 and 4 letters conditions. Global efficiency and local efficiency of all brain networks relative to the update and readout states were computed at (a) the theta band and (b) the gamma band. A single data point represents a brain network. The asterisks indicate $p < 0.05$ and the ns means $p > 0.05$.

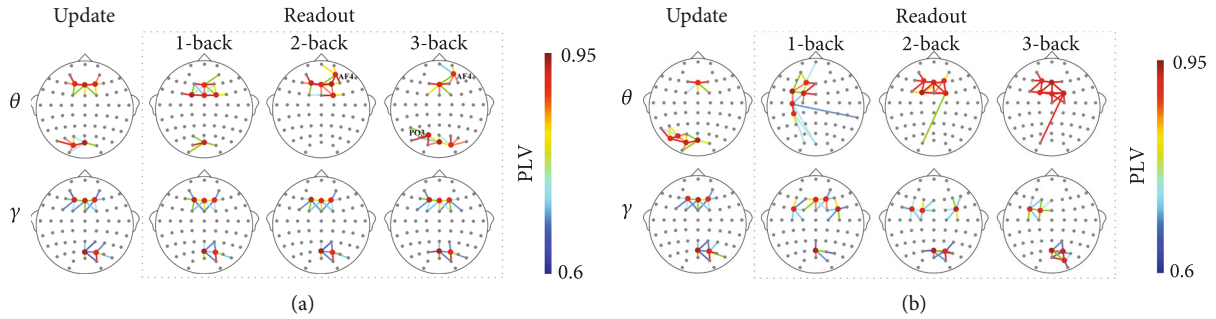


Figure 6: The top five important nodes and their first five edges in theta and gamma bands were chosen to represent the significant brain regions and the connections in update and readout states for (a) 2 letters condition and (b) 4 letters condition. In the readout of the 4 letters condition, only the brain network map of an example subject is shown because of the large inter-subject variation in reading effects. The others are maps of the brain network averaged across subjects.

proposed that the left hemisphere is relative to verbal information and logical reasoning, while the right is responsible for spatial thinking [39]. Therefore, it can be assumed that as the difficulties of WM tasks increase, the brain may convert verbal information into spatial information to better extract memory content. In the gamma band, there was no

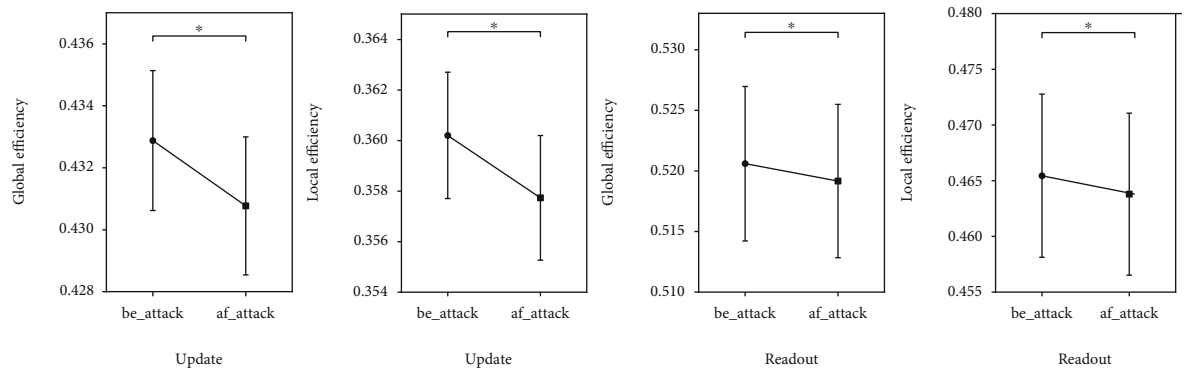


Figure 7: Global efficiency and local efficiency in theta band for the 2 letters condition before and after attacks on the networks. The be_attack represents before the attack and the af_attack after. Error bars represent the standard error of mean (SEM) of properties in brain networks for all subjects. The asterisks indicate $p < 0.05$.

significant difference between the important brain regions which were mainly located in the frontal and right parietooccipital regions in the two states.

For the high load condition, the parietooccipital region became more vital in update in the theta band (Figure 6(b)), which is perhaps related to the increased visual information in our paradigm. However, in readout state, there was no clear trend in the distribution and connection of important brain regions. This further validates the tasks in 4 letters condition are so challenging that there is no difference between the *n*-back tasks for subjects.

Based on the above findings, we attacked the important node (FZ) of brain networks in update and AF4 in readout, as well as calculated global efficiency and local efficiency again. As shown in Figure 7, these two indexes dropped significantly after the networks were attacked. This implies that these key nodes play a significant role in strengthening the efficiency of information transfer of the brain network. Thus, the stronger small-worldness of brain networks in readout might be related to the enhanced importance of the right-hand nodes.

4. Conclusions

In this paper, EEG data collected with a novel letter-sequence version of *n*-back pattern were used to research WM update and readout states. Time-frequency analysis was applied to investigate activated brain regions and the conversion of their power. The PLVs of electrode pairs were calculated to construct complex brain networks, and then the topological characteristics of the PLV networks were considered by calculating global and local efficiency to assess the small-worldness. Ultimately, the WKPN algorithm was leveraged to divulge the specific location of crucial brain regions from the perspective of the synchronization and control relationships.

Based on the aforementioned approaches, the principal conclusions can be summarized as follows: (1) In terms of both power and synchrony, the frontal and parietooccipital regions each have an important impact on the WM update and readout states. (2) The brain networks have higher global and local efficiency during WM readout. (3) The importance of the right prefrontal region (such as AF4) and the parietooccipital region (such as PO3) increases in the readout state under more difficult WM tasks. (4) The first memory target achieves higher accuracy in a lower load condition.

The WKPN algorithm is verified to be effective by locating and attacking network critical nodes. This means that we provide a new way from a brain network perspective to track important regions at the sensor level. It is helpful for some intensive research, such as transcranial electrical stimulation (tES) [40]. Selection of stimulus points is based on the locations obtained from the algorithm rather than a priori knowledge, which can improve the effectiveness of stimuli. Moreover, considering that the EEG technique acquires scalp signals, source localization [41] is expected to be applied to survey the deeper level of brain mechanisms in future work.

Data Availability

Raw data are available via a request to the lead author with the need of a formal data-sharing agreement.

Conflicts of Interest

The authors declare that the research was conducted in the absence of commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (grant number 61977039) and the 2019 Research Project of University Education Informatization (grant number 2019JSETKT009).

References

[1] A. Baddeley, "Working memory: looking back and looking forward," *Nature Reviews Neuroscience*, vol. 4, no. 10, pp. 829–839, 2003.

[2] B. Rypma and M. D'Esposito, "Isolating the neural mechanisms of age-related changes in human working memory," *Nature Neuroscience*, vol. 3, no. 5, pp. 509–515, 2000.

[3] H. A. D. Keage, C. R. Clark, D. F. Hermens et al., "ERP indices of working memory updating in AD/HD: differential aspects of development, subtype, and medication," *Journal of Clinical Neurophysiology*, vol. 25, no. 1, pp. 32–41, 2008.

[4] C. Galletly, C. R. Clark, A. C. McFarlane, and D. L. Weber, "Working memory in posttraumatic stress disorder—an event-related potential study," *Journal of Traumatic Stress*, vol. 14, no. 2, pp. 295–309, 2001.

[5] S. Ahrens, J. D. Twanow, J. Vidaurre, S. Gedela, M. Moore-Clingenpeel, and A. P. Ostendorf, "Electroencephalography technologist inter-rater agreement and interpretation of pediatric critical care electroencephalography," *Pediatric Neurology*, vol. 115, pp. 66–71, 2021.

[6] S. Singla, G. E. Garcia, G. E. Rovenolt et al., "Detecting seizures and epileptiform abnormalities in acute brain injury," *Current Neurology and Neuroscience Reports*, vol. 20, no. 9, p. 42, 2020.

[7] S. Itthipuripat, J. R. Wessel, and A. R. Aron, "Frontal theta is a signature of successful working memory manipulation," *Experimental Brain Research*, vol. 224, no. 2, pp. 255–262, 2013.

[8] R. Polanía, W. Paulus, and M. A. Nitsche, "Noninvasively decoding the contents of visual working memory in the human prefrontal cortex within high-gamma oscillatory patterns," *Journal of Cognitive Neuroscience*, vol. 24, no. 2, pp. 304–314, 2012.

[9] M. Semprini, G. Bonassi, F. Barban et al., "Modulation of neural oscillations during working memory update, maintenance, and readout: an hdEEG study," *Human Brain Mapping*, vol. 42, no. 4, pp. 1153–1166, 2021.

[10] C. W. Lynn and D. S. Bassett, "The physics of brain network structure, function and control," *Nature Reviews Physics*, vol. 1, no. 5, pp. 318–332, 2019.

[11] A. P. Alivisatos, A. M. Andrews, E. S. Boyden et al., "Nanotools for neuroscience and brain activity mapping," *ACS Nano*, vol. 7, no. 3, pp. 1850–1866, 2013.

[12] J. M. Bernabei, T. C. Arnold, P. Shah et al., "Electrocorticography and stereo EEG provide distinct measures of brain connectivity: implications for network models," *Brain Communications*, vol. 3, no. 3, 2021.

[13] R. A. Thuraisingham, "Estimating electroencephalograph network parameters using mutual information," *Brain Connectivity*, vol. 8, no. 5, pp. 311–317, 2018.

[14] J. Gao, W. Wang, and J. Zhang, "Explore interregional EEG correlations changed by sport training using feature selection," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 6184823, 2016.

[15] A. M. Gong, J. P. Liu, L. Lua, G. R. Wu, C. H. Jiang, and Y. F. Fu, "Characteristic differences between the brain networks of high-level shooting athletes and non-athletes calculated using the phase-locking value algorithm," *Biomedical Signal Processing and Control*, vol. 51, pp. 128–137, 2019.

[16] Z. M. Wang, Y. Tong, and X. Heng, "Phase-locking value based graph convolutional neural networks for emotion recognition," *IEEE Access*, vol. 7, pp. 93711–93722, 2019.

[17] S. Afshari and M. Jalili, "Directed functional networks in Alzheimer's disease: disruption of global and local connectivity measures," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 949–955, 2017.

[18] N. Franzmeier, J. Hartmann, A. N. W. Taylor et al., "The left frontal cortex supports reserve in aging by enhancing functional network efficiency," *Alzheimer's Research & Therapy*, vol. 10, no. 1, p. 28, 2018.

[19] J. Toppi, L. Astolfi, M. Risetti et al., "Different topological properties of EEG-derived networks describe working memory phases as revealed by graph theoretical analysis," *Frontiers in Human Neuroscience*, vol. 11, p. 637, 2017.

[20] Z. Z. Haque, R. Samandra, and F. A. Mansouri, "Neural substrate and underlying mechanisms of working memory: insights from brain stimulation studies," *Journal of Neurophysiology*, vol. 125, no. 6, pp. 2038–2053, 2021.

[21] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245–251, 2010.

[22] H. J. Wang, J. M. Hernandez, and P. Van Mieghem, "Betweenness centrality in a weighted network," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 4, article 046105, 2008.

[23] P. C. Tang, C. C. Song, W. W. Ding, J. K. Ma, J. Dong, and L. Y. Huang, "Research on the node importance of a weighted network based on the K-order propagation number algorithm," *Entropy*, vol. 22, no. 3, p. 364, 2020.

[24] W. K. Kirchner, "Age differences in short-term retention of rapidly changing information," *Journal of Experimental Psychology*, vol. 55, no. 4, pp. 352–358, 1958.

[25] H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.

[26] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: a user-friendly application for MEG/EEG analysis," *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 879716, 2011.

[27] C. Jutten and J. Herault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[28] J. Yordanova, V. Kolev, and A. Rothenberger, "Event-related oscillations reflect functional asymmetry in children with attention deficit/hyperactivity disorder," *Supplements to Clinical Neurophysiology*, vol. 62, pp. 289–301, 2013.

[29] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in Systems Neuroscience*, vol. 9, p. 175, 2015.

[30] M. Le Van Quyen, J. Foucher, J.-P. Lachaux et al., "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *Journal of Neuroscience Methods*, vol. 111, no. 2, pp. 83–98, 2001.

[31] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[32] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, article 198701, 2001.

[33] T. P. Zhang, B. Fang, and X. Y. Liang, "A novel measure to identify influential nodes in complex networks based on network global efficiency," *Modern Physics Letters B*, vol. 29, no. 28, pp. 1550168–1550442, 2015.

[34] M. L. Stanley, S. L. Simpson, D. Dagenbach, R. G. Lyday, J. H. Burdette, and P. J. Laurienti, "Changes in brain network efficiency and working memory performance in aging," *PLoS One*, vol. 10, no. 4, article e0123950, 2015.

[35] Y. H. Zhang, Y. Liao, Y. D. Zhang, and L. Y. Huang, "Emergency braking intention detect system based on K-order propagation number algorithm: a network perspective," *Brain Sciences*, vol. 11, no. 11, p. 1424, 2021.

[36] R. J. Zhu, Y. M. Luo, Z. Y. Wang, and X. Q. You, "Modality effects in verbal working memory updating: transcranial direct current stimulation over human inferior frontal gyrus and posterior parietal cortex," *Brain and Cognition*, vol. 145, p. 8, 2020.

[37] A. Sugiura, B. H. Silverstein, J.-W. Jeong et al., "Four-dimensional map of direct effective connectivity from posterior visual areas," *NeuroImage*, vol. 210, article 116548, 2020.

[38] B. Griesmayr, B. Berger, R. Stelzig-Schoeler, W. Aichhorn, J. Bergmann, and P. Sauseng, "EEG theta phase coupling during executive control of visual working memory investigated in individuals with schizophrenia and in healthy controls," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 14, no. 4, pp. 1340–1355, 2014.

[39] Z. T. Ren, Y. Zhang, H. He, Q. Y. Feng, T. Y. Bi, and J. Qiu, "The different brain mechanisms of object and spatial working memory: voxel-based morphometry and resting-state functional connectivity," *Frontiers in Human Neuroscience*, vol. 13, p. 248, 2019.

[40] K. Ergo, E. De Loof, G. Debra, B. Pastotter, and T. Verguts, "Failure to modulate reward prediction errors in declarative learning with theta (6 Hz) frequency transcranial alternating current stimulation," *PLoS One*, vol. 15, no. 12, article e0237829, 2020.

[41] J. M. Schoffelen and J. Gross, "Source connectivity analysis with MEG and EEG," *Human Brain Mapping*, vol. 30, no. 6, pp. 1857–1865, 2009.

Hindawi
Journal of Sensors
Volume 2022, Article ID 5841630, 11 pages
https://doi.org/10.1155/2022/5841630

Research Article

Secure MRI Brain Image Transmission Using IOT Devices Based on Hybrid Autoencoder and Restricted Boltzmann Approach

S. Aruna Deepthi,[1] **E. Sreenivasa Rao**,[2] **and M. N. Giriprasad**[1]

[1]*JNTUA Anantapuramu, Anantapuramu, India*
[2]*Vasavi College of Engineering, Hyderabad, India*

Correspondence should be addressed to S. Aruna Deepthi; sadeepthi@hotmail.com

Received 18 April 2022; Revised 9 May 2022; Accepted 17 May 2022; Published 29 May 2022

Academic Editor: Jaroslav Frnda

In recent times, the medical image processing solves several clinical issues by inspecting the visual images, which are generated in the clinical health care units. The main objective of the research is to gain valuable information from the images for better clinical diagnosis. In the biomedical engineering domain, the medical image analysis is an emerging research topic. In the recent decades, the use of medical images is highly growing, which are acquired from different image modalities; so, there is a necessity for data compression for transmission, storage, and management of digital medical image datasets. Hence, the machine learning methods are effective for medical image analysis, where the deep learning models are used in the machine learning tools for automatically learning the feature vectors from the huge medical datasets. The automated deep learning models are effective compared to the conventional handcrafted features. In addition to this, a wireless sensor network (WSN) is used to create a primary health care scheme, which brings patient data together and expands the medical conveniences, whereas the WSN design must comprise of sensor nodes, because it consumes less power and resources at a relatively low cost; so, it is essential for implementing the Raspberry Pi-based WSN nodes. The sensor nodes are important for limited battery capacity and to transmit the vast amount of medical data. The proposed work is broadly classified into two categories such as (i) the medical image compression algorithm is developed using the deep learning model based on autoencoders and restricted Boltzmann machines (RBM) and (ii) implementation of the WSN sensors nodes with Raspberry Pi and Messaging Queue Telemetry Transport (MQTT) Internet of Things (IoT) protocol for secure transmission of the medical images. The experimental results are evaluated using the standard performance metrics like peak signal to noise ratio (PSNR) and presented a Real-Time Linux (RTL) implementation of the design. The proposed model showed 10 dB to 15 dB improvement in the PSNR value, while transmitting the medical images, which is better compared to the existing model.

1. Introduction

Magnetic resonance imaging (MRI) or computerized tomography (CT) scan medical imaging creates digital images of the human body. While these imaging techniques generate enormous quantities of data, compression is needed for storage and transmission [1]. Most current compression schemes achieve a high compression rate at the expense of significant quality loss. Maintaining image quality only in the area of interest, i.e., in medically relevant regions, may be required in some areas of medicine. A standard 12-bit medical X-ray has a dimension of 2048 pixels by 2560 pixels. This equates to 10,485,760 bytes in file size. A standard 16-bit mammogram image could be 4500 pixels by 4500 pixels, with a file size of 40,500,000 bytes (40 MB) [2]. This has implications for disc storage as well as image transmission time. Despite the fact that disc capacity has steadily increased, the amount of digital imagery provided by hospitals and their new filmless radiology departments has increased even more rapidly. Even if there was limitless storage, the issue of transmitting the images would still exist.

Most hospitals have remote clinics or satellite centers in small towns and remote areas to make it easier for patients who have difficulty travelling long distances to the clinic,

particularly for diagnostic procedures. These facilities use applications, which enable clinic staff to work without the presence of a radiologist. A clinic technician or basic radiologist will take the X-ray and send it to the hospital via a network connection, where a diagnostic radiologist will read it and send back a diagnosis. Although this may seem to be reasonable, keep in mind that the patient is often asked to remain in the imaging machine until the radiologist certifies that the data is sufficient. Not only does compression affect storage costs but it also affects transmission times, MRI apparatus use, and patient safety and security [3].

Compression techniques can help improve overall treatment by reducing file size and transmission time. The redundancy that occurs in images is exploited by image compression techniques. Different forms of redundancy exist. Each compression technique can take advantage of one of these redundancies. Spatial, temporal, and spectral redundancies are the three forms of redundancies [4]. Deep learning is a branch of machine learning that is based on a neural network that processes data and mimics the thought process by layering algorithms. Deep learning uses a deep architecture consisting of several layers of transformations to simulate the functioning of the human brain [5]. This is close to how the human brain processes knowledge. Traditional machine learning techniques performed poorly when faced with high-dimensional data, necessitating a preliminary feature extraction process to transform the most informative representation of the raw data into a normalization matrix. Deep learning got rid of the difficult task of sophisticated automated feature extraction without sacrificing the data's sense. Autoencoder is a form of unsupervised artificial neural network that learns the most efficient data. The goal of an autoencoder is to learn a representation called encoding, which is then used to reduce dimensionality by training the network. We have considered both the autoencoder [6] and RBM [7] as the compression techniques for medical images on the belief that they can be used to select the most discriminative features in an image in a most efficient manner leading to smaller footprint of the image without losing quality.

WSN [8] is a communication network that does not rely on wires or other electronic links between them. WSNs are used in industries like manufacturing, forest fire detection, transportation, construction, and office space surveillance and monitoring. WSNs are usually deployed closer together in inaccessible locations where battery replacement is nearly impossible. The WSN's key challenges include limited power, limited processing ability, and an open climate. We choose to implement the system design for wireless sensor networks with Raspberry pi (RPi) [9], a small footprint hardware device with wireless, and Bluetooth capabilities for communication with other similar Raspberry Pi units that can act as sensor nodes to which the compressed medical images can be transmitted. The platform comprises of four WMSN nodes built with the Raspberry Pi (RPi) [9]; we introduce a technique where every source node retains its data transmission rate uniformly and periodically apprises its information sending rate based on the other node's congestion level. Each node is identified by a unique IP address so that it can create a routing path to the sink using multihop communication; we also propose the MQTT [10] protocol for the transmission mechanisms, where the packets lost can be stored and retransmitted when the network bottleneck has cleared. The suggested platform can be adaptable, accessible, and appropriate for wireless monitoring of constructions, open spaces, remote locations, etc., which is depicted in Figure 1.

Largely, the information handling utilizes far less power than transmitting it in a wireless channel. Hence, by compressing the images before transmitting them will considerably reduce the net energy consumed across the sensor nodes. It is also feasible for image compression to be sustained at a high compression ratio without evident degradation of quality in the reconstructed images [11] The image compression schemes proposed and developed in this work could have some parameters such as minimalism in coding, minimal memory demand, low computational need, and maximum compression rate. The image compression process in WSN is stated in Figure 2.

The major contributions of this research are given below:

(1) Developed a medical image compression algorithm based on the autoencoders and restricted Boltzmann machines. An effective image algorithm consumes less space on the hard-drive and effectively retains the same physical size

(2) WSN sensor's nodes with Raspberry Pi and MQTT IoT protocol is implemented for secure transmission of the medical images, and then the proposed model effectiveness is validated by using the performance measures like PSNR

This research manuscript is structured as follows: the Section 2 surveys the related research literature articles and previous published papers on the topic "medical image compression and WSN.". The Section 3 covers the general performance metrics used to analyze the compression standard and then the relevant research. In addition, the compression methodologies used in this work are discussed effectively in Section 3. The Section 4 provides information about the WSN design with Raspberry Pi, and the Section 5 provides the results achieved with the compression algorithms with the original and the reconstructed images. The RTL schematic diagram is provided in Section 5. The Section 6 discusses the scope of the work and its implications for future research in the field.

2. Related Research Works

In today's healthcare systems, medical imaging is a necessary tool. With applications in tumor segmentation, cancer identification, classification, image driven therapy, medical image description, and restoration, machine learning plays a critical role in CADx. Since any redundant information is lost during compression, lossy techniques will not be able to restore the original image from the reconstructed image. The lossless procedure will precisely recreate the actual image from the restored image.

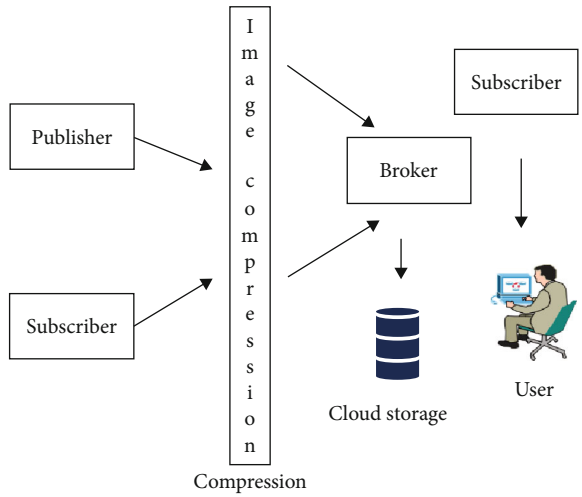


Figure 1: Raspberry Pi WSN with four nodes for compressed image transmission using MQTT.

Information can be compressed using transform-based coding strategies like DCT [12], DWT [13], SVD [14], PCA [15], and wavelet-based compression techniques like EZW [16], SPIHT [17], WDR [18], and JPEG2000 [19]. On satellite images, Walker et al. [20] used PCA and neural network algorithms to compress the data. Finding the covariance matrix, eigenvalue, and eigenvectors of an input image using the PCA method is extremely difficult. PCA is used to compress images, but the compressed result is not considered to be appropriate. The ANN algorithm, which produces better results, will boost the outcome.

Gaidhane et al. [21] found that image compression using the wavelet transform produces the best results as compared to DCT for ultrasound and angio images. Blocking objects plagued the DCT process. Puniene et al. [22] used DCT and SPIHT-based compression techniques to compress medical images. DCT is used to decompose the medical image, and then SPIHT is used to compress the coefficients. Singular value decomposition (SVD) and wavelet difference reduction techniques are used by Antonini et al. [14] to propose a lossy image compression method (WDR). The image quality is better with SVD compression, but the compression ratio is poor. As a result, the SVD output was compressed once more using WDR. At high compression ratios, WDR achieves excellent image quality.

Angadi and Somkuwar [23] use a combination of SVD and the embedded zero tree wavelet approach to compress ECG signals. The SVD method, followed by the EZW method, has been tested and proven to improve the efficiency of the reconstructed signal. Kumar et al. [24] used DCT and DWT techniques to compress images in wireless sensor networks. As compared to the discrete cosine transform, the discrete wavelet transform has a higher PSNR value and a faster compression process. Each node in a sensor network has extremely limited resources, such as memory, electricity, and processing power. To address these limitations, image compression algorithms rely on DCT and DWT that are used to reduce memory space and storage usage. They evaluate the results of DCT and DWT procedures using several performance metrics, and the results showed that the discrete wavelet transform outperforms the discrete cosine transform in terms of PSNR.

Artificial intelligence approaches, specifically in the context of computer vision, imaging, voice recognition, and natural language comprehension [25, 26], are just one of those directions that can help resolve the limitations of traditional image compression standards. Deep learning got ahold of the difficult task of sophisticated automated feature extraction without sacrificing the data's context. As a result, deep learning for feature extraction and classification of medical images from a variety of diseases has gained a reputation for exceeding expectations and producing more than satisfactory results.

After training two stacked denoising autoencoders to obtain a reduced version of the data dimension, Xing et al. [27] used a theory components neural network to classify data. They identified the areas of the brain that distinguish ASD from conventional controls (TC) with approximately 70% accuracy across the entire dataset. Heinsfeld et al. [28] reduced multivariate data using a variational autoencoder (VAE) model and discovered the most discriminative features. Choi [29] added spatiotemporal information in the fMRI to the 3D conventional neural network to detect spatially useful features (CNN). The researchers then devised a voting system based on these characteristics to decide whether or not each subject has ASD.

In [30], the authors proposed a DNN-based novel feature election method from fMRI images and then used it to obtain whole-brain function communication patterns using multiple trained sparse autoencoders. They developed a DNN-FS classification model that had an accuracy of 86.36 percent on a sample of 55 ASD and 55 TC.

WSN technology can be used to develop realistic Health Care WSNs that meet the main system design requirements of secure connectivity, node mobility, multicast technology, energy efficiency, and timely data delivery. Long data transmission routes, vast volumes of data, and limited battery power all reduce WSN lifetime. The optimization of energy consumption effectively extends the lifespan of the network, which is important. To get more energy optimization, Guo et al. [31] proposed an energy efficient clustering hierarchy protocol for WSN. Mann and Singh [32] presented a routing protocol for WSNs that was established to preserve a reasonable level of scalability, energy efficiency, and reliability. A fuzzy logic and genetic dependent clustering approach to optimize network energy was defined in the paper presented by Sim and Lee [33]. Saeedian et al. [34] employs a Digital Signal Processing-based Wireless Sensor Network Platform to achieve high compression efficiency of physiological data for telemedicine applications.

With the advancement in rapid hardware prototyping technology and the advent of microcontroller-based board such as Arduino, ESP32, and Raspberry Pi, WSNs with smaller size footprint, better energy efficiency, and low cost are a reality enabling researchers experiment with them. Hsu et al. [35] created a new model for saving, exchanging, and archiving patient health records using a Raspberry Pi board and a hard drive. The disc that is accessible in the local

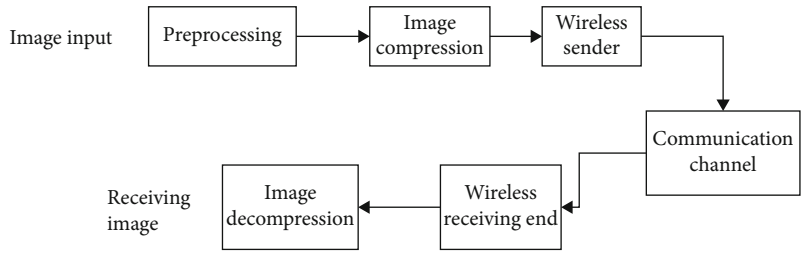


Figure 2: Image compression in a WSN.

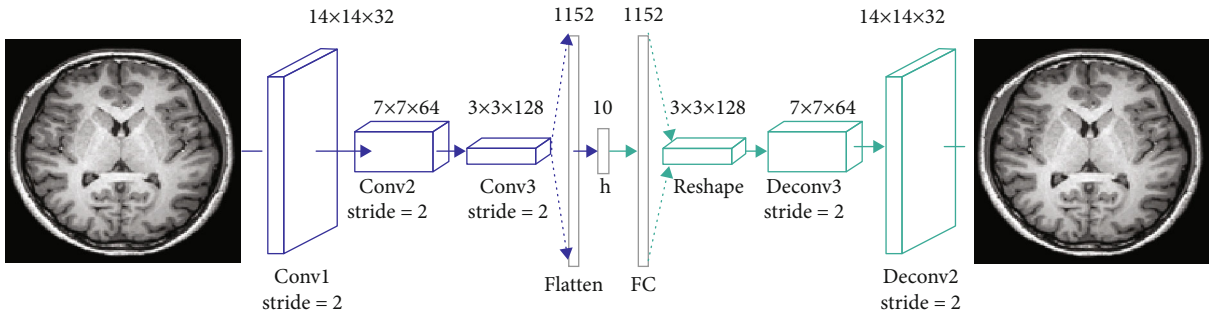


Figure 3: The structure of proposed convolutional autoencoders (CAE) for MNIST.

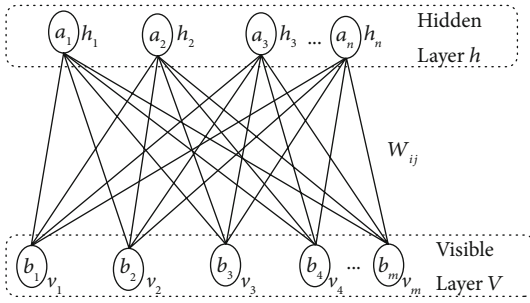


Figure 4: Basic structure of a restricted Boltzmann machine.

cloud and can be shared with other public clouds that are farther out, such as Google Drive, Azure Cloud, and other similar services, are available. They also safeguarded the medical details by introducing a new security protocol.

Elhoseny et al. [36] used fast bilateral filter for noise removal in the medical images, because it has better edge preservation ability. Further, the image segmentation is carried out by using the canny edge detector. In addition, the developed fast bilateral filter algorithm is implemented in Raspberry Pi by utilizing open CV software. Kumar and Gupta [37] have implemented a fast and secure encryption technique for the medical images on the basis of one dimensional logistic map, which is associated with pseudo random numbers. In the resulting phase, the proposed technique is validated on the standard medical datasets under the conditions of noise and differential attacks. Ahmed and Salah [38] applied a fast sub pixel registration technique to achieve high-resolution image registration on the basis discrete wavelet transform and the convolutional neural network. In this literature, the classification result of convolutional neural network and genetic algorithm is used for MRI image registration.

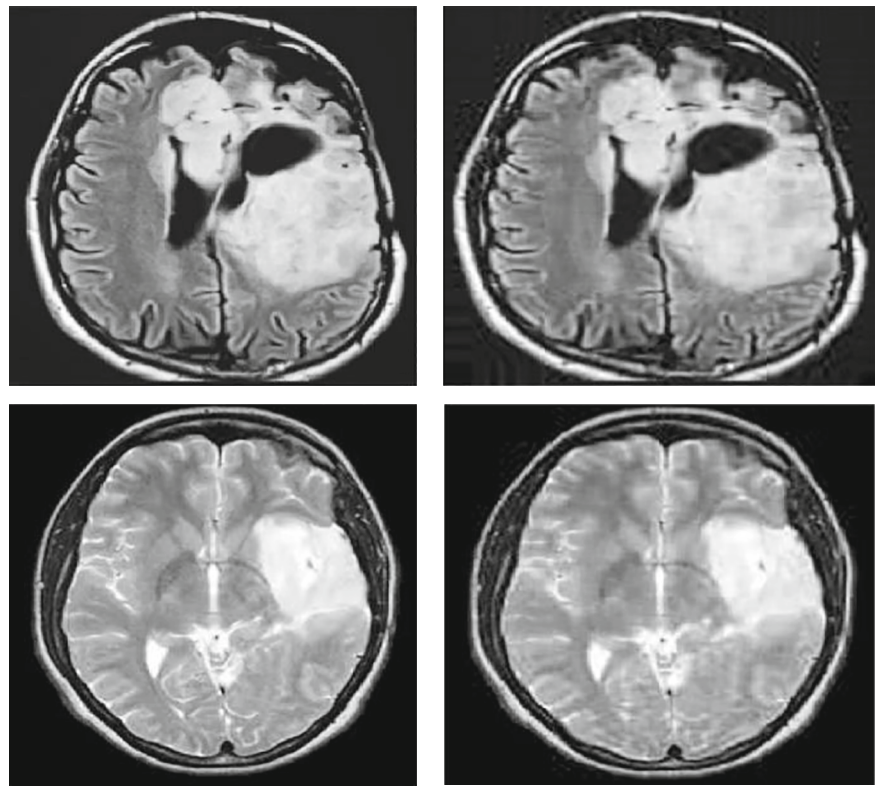
3. Image Compression and Performance Metrics

3.1. Image Compression. Image comprises of pixels that are extremely related to each other. Consequently, it holds a sizable quantity of redundancy that utilizes substantial memory for image storage space, which in turn can reduce the transmission bandwidth. These redundancies can be grouped into two classes: (1) spatial and (2) temporal redundancy. In the spatial class, nearby pixels are correlated, whereas in temporal redundancy, there exists similarities between the two subsequent frames. Hence, to eliminate the redundancies, image compression has to be done for a reduced amount of storage area and smaller bandwidth [14, 39, 40].

Image compression can further be categorized into two types: lossy and lossless. Lossy image compression is commonly applied in WMSN owing to its gains over lossless compression standards, namely, higher compression rates, that reduces the bytes needed to be transmitted over the WMSN and less power consumption for transmitting the images.

Additionally, lossy image compression has the advantage of taking less time for encoding/decoding the image transmitted in relation to lossless compression. This paper focuses its approach on the lossy compression algorithm because of our need to increase the bandwidth availability in the sensor nodes for transmission without any congestion, which can subsequently reduce the transmission delay across the network and provide a more streamlined transfer suitable for live image or video feeds [41].

Figure 5: MQTT architectural block diagram.



(a) (b)

Figure 6: Autoencoder compression for sample test MR images with tumor. (a) Actual images. (b) Compressed images.

3.2. Performance Metrics. We have studied the compression ratio and peak signal to noise ratio (PSNR) readings to analyze the performance of the compression algorithms considered here. We have decided not to consider the processing time as it would not be fair to compare it between the simulation-based experimental setup versus the Raspberry Pi-based hardware implementation. The mathematics behind the performance metrics used to evaluate the compression schemes is presented below for better understanding of the concepts.

3.2.1. Mean Square Error (MSE). MSE is defined in Equation (1).

$$\text{MSE}\,\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - y_n)^2, \tag{1}$$

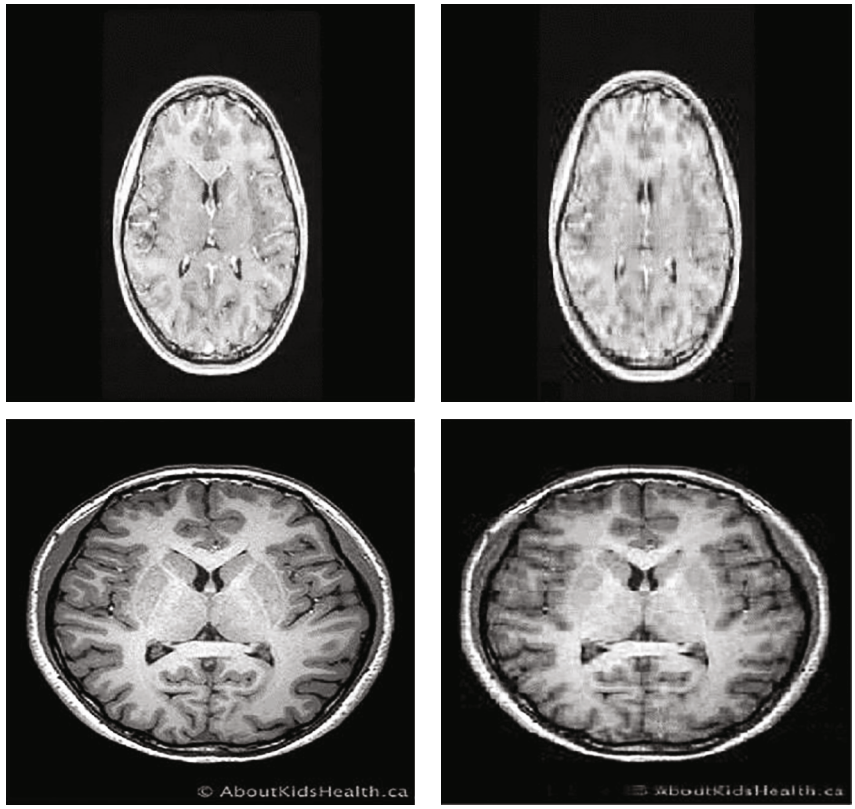
where N is the data sequence length, and x_n represents the input data sequence, while y_n corresponds to the reconstructed data sequence.

3.2.2. Peak Signal to Noise Ratio (PSNR). The quality of signal representation is influenced by the relationship between the maximum feasible signal value and the distortion capacity, which can be used as a comparison of compressed and original image quality. The better the quality of the compressed or reconstructed image, the higher the PSNR. The decomposed image quality is better when the PSNR value is higher [42]. PSNR is defined as the size of the error in relation to the signal's peak value x_{peak} (for 8-bit pixel x^2_{peak} equals 255) and is calculated using following equation (2).

$$\text{PSNR} = 10\log_{10}\frac{x^2_{\text{peak}}}{\sigma_d^2}, \tag{2}$$

where σ_d^2 is the MSE of (σ^2).

3.2.3. Compression Ratio (CR). Compression ratio (CR) is a phrase used to refer to the ratio of binary sequence length of the compressed output image (B1) to binary sequence length of the original uncompressed input image (B0) and



(a) (b)

Figure 7: Autoencoder compression for sample test MR images without tumor. (a) Actual images. (b) Compressed images.

measured as bits per pixel (bpp) as given in Equation (3).

$$\mathrm{CR} = \frac{B0}{B1}. \tag{3}$$

It is generally used to find how good the compression efficiency is as a higher value means better compression [42].

3.2.4. Power Consumption. For all practical implementation cases of the WSN, power consumption must be considered as the most important metric, which is highly influenced by the above discussed metrics. The transmission power and the energy dissipation across the nodes can be greatly reduced by adapting to less complex processing units and by minimizing the data size.

To summarize, less transmission error is a direct product of low MSE values and has an inverse relation with PSNR, which proportionately increases, which in turn points out the noise in the compressed image is lower and this can help with the better reconstruction of the image.

4. Image Compression Methodologies

4.1. Autoencoders. An auto encoder is a fine example of the unsupervised neural network learning algorithm, and it is graphically depicted in Figure 3. Commonly, it is used during the back propagation of the neural network model where the target values are selected to be equal to the inputs such that $y(i) = x(i)$. It makes use of the convergent and divergent layers where convolution and deconvolution take place, and the features are compressed in the middle layers to generate the desired output.

An autoencoder can be mathematically characterized as follows to describe the encoder and the decoder sections of the model [6], as mentioned in the equations (4)–(6).

$$\phi : X \longrightarrow F, \tag{4}$$

$$\psi : F \longrightarrow X, \tag{5}$$

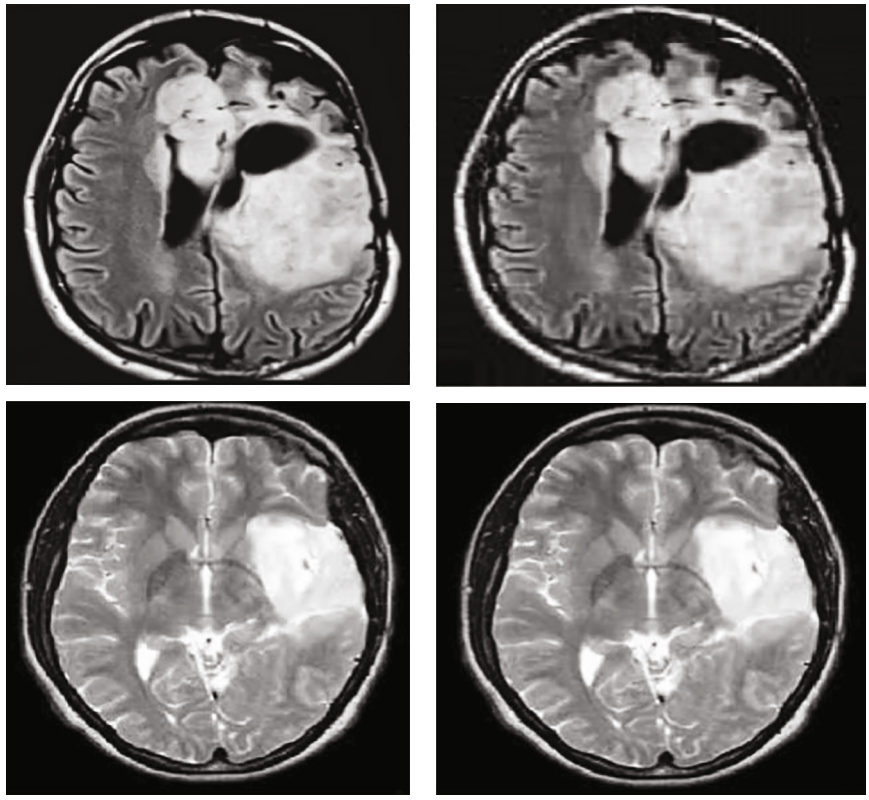
$$\phi, \psi = \underset{\phi,\psi}{\operatorname{argmin}} \|X - (\psi\, o\, \phi)X\|^2. \tag{6}$$

As can be seen from the equations (4)–6, presented above, the encoder and decoder parameters are enhanced in a manner that minimizes the reconstruction error, which is the error between the original input image and the reconstructed compressed image.

The parameter settings of the convolutional autoencoder are given below [12]:

Number of batches: the original image is divided into training set and further, it is divided into many batches to perform the stochastic gradient descent optimization of the model. In this experimentation, the batch size is considered as 100

Learning rate: at any time, the model weight is changed at every iteration, a hyperactive parameter that restricts how to adjust the model in response to the predicted error.



(a) (b)

Figure 8: RBM Compression for sample test MR images with tumor. (a) Actual images. (b) Compressed images.

It should be minimal, but not too small, as this could lead to a lengthy training phase, where the learning rate is considered as 0.01, and the number of iteration is assumed as 2000

Max epochs: this cannot be a definite value based on logic but just a backstop against endless iterations. It can be easily changed to a suitable value later if the model seems to overfit or underfit the learning experience. The maximum epoch of 2000 is considered for training the model

4.2. Restricted Boltzmann Machines (RBM). Boltzmann machines, as with any other neural network, consist of an input layer and several hidden layers. The neurons make stochastic decisions [43] like when to turn on depending on the data fed during the training process and based on the minimization of the cost function. The Boltzmann machine upon training attains the learning to deduce some interesting features from the dataset on which it is trained that can help the model to learn the complex fundamental relationships and patterns inherent in the data [22].

From Figure 4, it can be seen that the weight, $w_{ij} \ni W$, is connected to the visible unit, V, to the hidden unit, h, where $W \ni R_{mxn}$ is the super set of all the weights, considering both the hidden and visible units. The visible unit, V, biases can be signified as $b_i \ni b$, while the hidden unit, h, biases are represented as $c_j \ni c$.

We can assume that the joint distribution of a visible layer vector, V, and a hidden layer vector, h, is proportional to the exponential of the negative energy of the configuration as shown in Equation (7) based on [24]: the Boltzmann distribution from statistical physics:

$$P(v, h) \propto e^{-E(v,h)}. \tag{7}$$

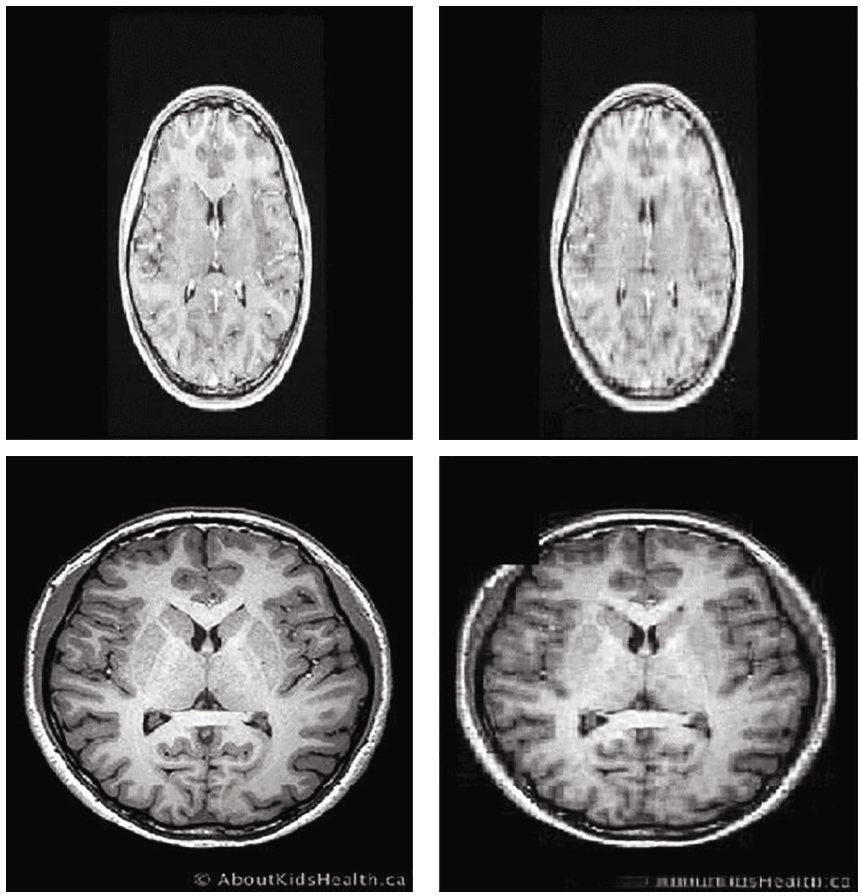
Before designing the RBM model using deep learning approach, we can reset the parameters to the original ones or leave them as defaults reasonable for most image processing applications. The parameters setting of restricted Boltzmann machine [43] is discussed below:

Random initialization iterations: initially, a set of random number of trail weights is considered for each layer of the network in order to use it as a good starting point for testing the model. Here, a stochastic gradient descent training is applied, and the number of iterations is selected as 2000

Number of batches: the training set is divided into batches to reduce the processing time and memory of hardware, where the batch size is considered as 100

Learning rate: in the case of autoencoder, the learning rate should be small for reducing the training process length that helps in the faster generation of the learning model, and the learning rate is considered as 0.1

Max epochs: the maximum number of epochs is selected by trial and error, and the number is adjusted based on the learning rate, whether it is overfitting or underfitting. The maximum epochs is considered as 100, but it is dynamically changed based on the above scenarios



(a) (b)

Figure 9: RBM Compression for sample test MR images without tumor. (a) Actual images. (b) Compressed images.

Table 1: Autoencoder–PSNR values for test sample of two images with and without tumor.

Images with tumor (dB)	Images without tumor (dB)
11.5639	11.8199
13.5750	6.8321

Table 2: RBM–PSNR values for test sample of two images with and without Tumor.

Images with tumor (dB)	Images without tumor (dB)
11.874	10.785
9.572	9.264

Table 3: Autoencoder Xilinx ISE device utilization factor.

Logic utilization	Used	Available	Device utilization
No. of slices	284	4656	6%
No. of slices in flip flops	60	9312	0%
No. of 4 input LUTs	430	9312	4%
No. of bonded IOBs	65	92	70%
No. of GCLKs	1	24	4%

5. Raspberry Pi WSN Implementation

The hardware implementation was done using Raspberry Pi, and MQTT (Message Queuing Telemetry Transport) protocol was selected as the de facto method to transfer the compressed images across the Raspberry Pi WMSN [9]. MQTT was an opt choice for wireless networks where high latency is an issue due to its bandwidth constraints and unpredictable network downtimes. In case of the connection gets broken when a subscribing client tries to access the transmitted image from the broker, the broker has the ability to buffer the lost messages back to the subscriber again when the network becomes online. Similarly, when the publisher node loses the connection with the broker, the broker can initiate the process to close the connection; before it can send all the subscribed nodes in the network, the cached message received earlier from the subscriber.

This process is explained clearly in Figure 5 with a block diagram. Here, we have used four Raspberry Pi nodes which can behave as publisher, broker, and subscriber. A publisher, by design, can behave as both a publisher and a subscriber; so, every other Raspberry Pi node can transmit and receive images except the broker which can only facilitate the communication between the subscriber and the publisher. Any Raspberry Pi node can initiate the transmission across the

Table 4: RBM Xilinx ISE device utilization factor.

Logic utilization	Used	Available	Device utilization
No. of slices	447	4656	9%
No. of slices in flip flops	56	9312	0%
No. of 4 input LUTs	861	9312	9%
No. of bonded IOBs	65	92	70%
No. of GCLKs	1	24	4%

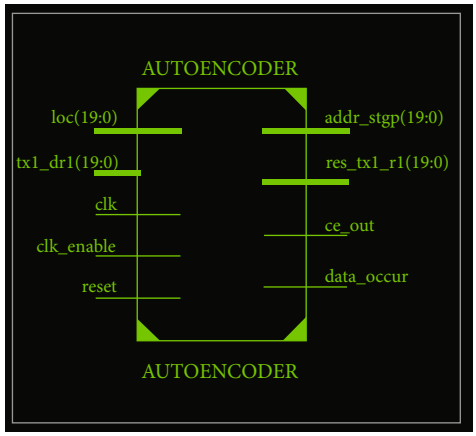


Figure 10: RTL schematics of autoencoder design.

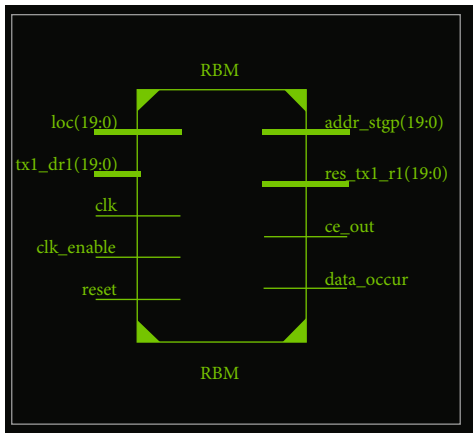


Figure 11: RTL schematic of restricted Boltzmann machines (RBM).

WMSN by acting as a publisher by transmitting images to the broker, while the other nodes can subscribe to receive the images from the broker, which can then serve the images to the subscriber.

6. Results and Discussion

We have used different approaches in the evaluation of results by analyzing the performance metrics of the image compression and WSN using both simulation and hardware implementations. The simulation analysis is done on the system configuration with 128 GB random access memory, 4 TB hard disk, windows 10 operating system, and Intel core i9 processor.

The hardware implementation was done using Raspberry Pi, and MQTT (Message Queuing Telemetry Transport) protocol was selected as the de facto method to transfer the compressed images across the Raspberry Pi WMSN [9]. MQTT was the apt choice for wireless networks where high latency is an issue due to its bandwidth constraints and unpredictable network downtimes. In case of the connection gets broken when a subscribing client tries to access the transmitted image from the broker, the broker has the ability to buffer the lost messages back to the subscriber again when the network becomes online. Similarly, when the publisher node loses the connection with the broker, the broker can initiate the process to close the connection; before it can send all the subscribed nodes in the network, the cached message received earlier from the subscriber.

Our experimental process is concerned with analyzing the performance of the neural network-based image compression schemes by comparing their PSNR values when the images are transmitted through the WMSN. We have used the MRI images from the brain tumor dataset gathered from [44] for our experiment for its simplicity and availability of large number of images.

Test images included T1-weighted MR images with a TR of 1740 and an echo time (TE) of 20, T2-weighted MR images with a TR of 5850 and an echo time (TE) of 130, and FLAIR-weighted MR images with a TR of 8500 and an echo time (TE) of 130.

A 3 Tesla Siemens Magnetron Spectra MR computer was used to create these test images. The total number of slices for all channels was 15, resulting in 135 images at 9 slices or images per patient with a field of view of 200 mm, a 1 mm interslice distance, and voxel sizes of 0.78 mm × 0.78 mm × 0.5 mm. The proposed technique is tested on a real dataset that includes 512 × 512 pixel brain MR images, which was converted to gray-scale before processing using autoencoder/RBM. The autoencoder/RBM compression for sample test MR images with and without tumor is given in Figures 6–9. The images consisted of both the images with tumors detected and tissues without tumor (healthy tissues). We adopted this approach to analyze the performance of the image compression algorithms on both the MR images with and without tumor and identify if the tumor issues are visible after compression.

The RTL schematic for autoencoder and RBM was implemented with Xilinx ISE development platform version 14.1 of family Virtex 6-XC6VLX757, and the device utilization summary is given in Tables 1–4. RTL schematic is shown below in Figures 10 and 11. Comparing the RTL design of RBM with autoencoder, we can easily infer that RBM utilizes far lesser flip flops in its design for the same image compression, which attests for its superior performance when compared with autoencoder. It also performs better other factorization-based image compression methods by a significant factor.

7. Discussion

The proposed model's performance is validated by comparing it with the existing model developed by Elhoseny et al.

[36]. In the existing work, the author developed fast bilateral filter for noise removal in the medical images. The fast bilateral filter has better edge preservation ability. Then, the canny edge detector is developed for segmenting the brain tissues. Lastly, the fast bilateral filter algorithm is implemented in Raspberry Pi by utilizing open CV software. Compared to the existing model, the proposed model achieved showed 10 dB to 15 dB improvement in the PSNR value, while transmitting the medical images. In addition, the effectiveness of the security protocol is as follows: MQTT is validated in terms of accuracy, precision, and recall. In our test case, the MQTT protocol achieved 95.60% of accuracy, 96.90% of recall, and 95.92% of precision, which is better compared to the existing technique [38], where the existing technique achieved only 90.90% of precision and recall and 94% of accuracy.

8. Conclusion

This work was chiefly carried out to analyze the performance of the deep learning algorithms in compressing medical MRI images and the efficiency of the Raspberry Pi WSN in transmitting the compressed images across the WSN nodes. Since there are many requirements in the medical field where the need to stream the medical images across the WSN with greater efficiency for immediate presentation and diagnosis of doctors, it calls for optimized compression of images for preserving the data transmission bandwidth and lesser transmission time at the same time not losing any considerable loss of image quality as it is impair the prognosis. The deep learning neural network implementations of autoencoder and RBM coded on Raspberry Pi with MQTT as the transmission protocol for additional security in the WSN performed as per the expectations with minimal power loss and latency. The RTL schematic implementation was done for all the image compression schemes used in the paper as a means to find out the device utilization in the WSN. This can be used as the foundation for the further development of the work on custom FPGA boards that can offer more control over the power performance of the WSN.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] M. L. Rajaram, E. Kougianos, S. P. Mohanty, and U. Choppali, "Wireless sensor network simulation frameworks: a tutorial review: MATLAB/Simulink bests the rest," *IEEE Consumer Electronics Magazine*, vol. 5, no. 2, pp. 63–69, 2016.

[2] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M. Marcellin, and A. Bilgin, "The current role of image compression standards in medical imaging," *Information*, vol. 8, no. 4, p. 131, 2017.

[3] S. Sridevi, V. R. Vijayakuymar, and R. Anuja, "A survey on various compression methods for medical images," *International Journal of Intelligent Systems and Applications*, vol. 4, no. 3, pp. 13–19, 2012.

[4] S. E. Hussein and S. M. Badr, "Healthcare cloud integration using distributed cloud storage and hybrid image compression," *International Journal of Computer Applications*, vol. 80, no. 3, pp. 9–15, 2013.

[5] R. Agarwal, C. S. Salimath, and K. Alam, "Multiple image compression in medical imaging techniques using wavelets for speedy transmission and optimal storage," *Biomedical and Pharmacology Journal*, vol. 12, no. 1, pp. 183–198, 2019.

[6] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.

[7] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional auto encoder-based lossy image compression," in *2018 Picture Coding Symposium (PCS)*, pp. 253–257, San Francisco, CA, 2018.

[8] S. Saravanan and J. Sujitha, "Deep medical image reconstruction with autoencoders using deep Boltzmann machine training," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, no. 24, article 166360, 2020.

[9] R. Krishnamoorthy, T. Jayasankar, S. Shanthi, M. Kavitha, and C. Bharatiraja, "Design and implementation of power efficient image compressor for WSN systems," *Materials Today: Proceedings*, vol. 45, pp. 1934–1938, 2021.

[10] B. F. N. Mohsin Alabassby, J. F. Mahdi, and M. A. Kadhim, "Design and implementation WSN based on Raspberry Pi for medical application," *IOP Conference Series: Materials Science and Engineering*, vol. 45, no. 2, pp. 1934–1938, 2019.

[11] V. Kumar, G. Sakya, and C. Shankar, "WSN and IoT based smart city model using the MQTT protocol," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, no. 8, pp. 1423–1434, 2019.

[12] H. Zain Eldin, M. A. Elhosseini, and H. A. Ali, "Image compression algorithms in wireless multimedia sensor networks: a survey," *Ain Shams Engineering Journal*, vol. 6, no. 2, pp. 481–490, 2015.

[13] P. Telagarapu, V. J. Naveen, A. L. Prasanthi, and G. V. Santhi, "Image compression using DCT and wavelet transformations," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 4, no. 3, pp. 61–74, 2011.

[14] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[15] A. M. Rufai, G. Anbarjafari, and H. Demirel, "Lossy image compression using singular value decomposition and wavelet difference reduction," *Digital Signal Processing*, vol. 24, pp. 117–123, 2014.

[16] J. S. Taur and C. W. Tao, "Medical image compression using principal component analysis," in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 1, pp. 903–906, Lausanne, Switzerland, 1996.

[17] R. Sudhakar, R. Karthiga, and S. Jayaraman, "Image compression using coding of wavelet coefficients–a survey," *ICGST Journal of Graphics, Vision and Image Processing*, vol. 5, no. 6, pp. 25–38, 2005.

[18] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE*

Transactions on Circuits and Systems for Video Technology, vol. 6, no. 3, pp. 243–250, 1996.

[19] S. P. Raja and A. Suruliandi, "Image compression using WDR & ASWDR techniques with different wavelet codecs," in *Proceedings of International Conference on Advances in Computer Engineering*, pp. 102–105, 2011.

[20] J. S. Walker, Y. J. Chen, and T. M. Elgindi, *Comparison of the JPEG2000 Lossy Image Compression Algorithm with WDR-Based Algorithms*, University of Wisconsin–Eau Claire, 2005.

[21] V. Gaidhane, V. Singh, and M. Kumar, "Image compression using PCA and improved technique with MLP neural network," in *International Conference on Advances in Recent Technologies in Communication and Computing*, pp. 106–110, Kottayam, India, 2010.

[22] J. Puniene, V. Punys, and J. Punys, "Ultrasound and angio image compression by cosine and wavelet transforms," *International Journal of Medical Informatics*, vol. 64, no. 2–3, pp. 473–481, 2001.

[23] S. Angadi and A. Somkuwar, "Medical image compression using DCT and SPIHT algorithm," *International Journal of Electronics, Communication & Soft Computing Science and Engineering*, vol. 2, no. 5, pp. 10–13, 2014.

[24] R. Kumar, A. Kumar, and G. K. Singh, "Hybrid method based on singular value decomposition and embedded zero tree wavelet technique for ECG signal compression," *Computer Methods and Programs in Biomedicine*, vol. 129, pp. 135–148, 2016.

[25] T. Sheltami, M. Musaddiq, and E. Shakshuki, "Data compression techniques in wireless sensor networks," *Future Generation Computer Systems*, vol. 64, pp. 151–162, 2016.

[26] V. Subbaraju, M. B. Suresh, S. Sundaram, and S. Narasimhan, "Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging : a spatial filtering approach," *Medical Image Analysis*, vol. 35, pp. 375–389, 2017.

[27] X. Xing, J. Ji, and Y. Yao, "Convolutional neural network with element-wise filters to extract hierarchical topological features for brain networks," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 780–783, Madrid, Spain, 2018.

[28] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.

[29] H. Choi, "Functional connectivity patterns of autism spectrum disorder identified by deep feature learning," 2017, https://arxiv.org/pdf/1707.07932.

[30] X. Li, N. C. Dvornek, X. Papademetris et al., "2-channel convolutional 3D deep neural network (2CC3D) for fMRI analysis: ASD classification and feature learning," in *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1252–1255, Washington, DC, 2018.

[31] X. Guo, K. C. Dominick, A. A. Minai, H. Li, C. A. Erickson, and L. J. Lu, "Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method," *Frontiers in Neuroscience*, vol. 11, article 460, 2017.

[32] P. S. Mann and S. Singh, "Improved artificial bee colony metaheuristic for energy-efficient clustering in wireless sensor networks," *Artificial Intelligence Review*, vol. 51, no. 3, pp. 329–354, 2019.

[33] I. Sim and J. Lee, "Routing protocol with scalability, energy efficiency and reliability In WSN," *Intelligent Automation & Soft Computing*, vol. 16, no. 4, pp. 567–577, 2010.

[34] E. Saeedian, M. Niazi Torshiz, M. Jalali, G. Tadayon, and M. M. Tajari, "CFGA: clustering wireless sensor network using fuzzy logic and genetic algorithm," in *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, Wuhan, China, 2011.

[35] S.-J. Hsu, C.-H. Chen, S.-H. Chen, W.-T. Huang, Y.-J. Chang, and Y.-Y. Chen, "Conserving bandwidth In a wireless sensor network for telemedicine application," *Intelligent Automation & Soft Computing*, vol. 16, no. 4, pp. 537–551, 2010.

[36] L. C. Manikandan, R. K. Selvakumar, S. A. H. Nair, and K. P. Sanal Kumar, "Hardware implementation of fast bilateral filter and canny edge detector using Raspberry Pi for telemedicine applications," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4689–4695, 2021.

[37] M. Kumar and P. Gupta, "A new medical image encryption algorithm based on the 1D logistic map associated with pseudo-random numbers," *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 18941–18967, 2021.

[38] A. S. Ahmed and H. A. Salah, "The IoT and registration of MRI brain diagnosis based on genetic algorithm and convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 1, p. 273, 2022.

[39] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *World congress on medical physics and biomedical engineering*, pp. 183–189, Springer, Singapore, 2018.

[40] D.-N. Tran, T. N. Nguyen, P. C. P. Khanh, and D.-T. Trana, "An IoT-based design using accelerometers in animal behavior recognition systems," *IEEE Sensors Journal*, pp. 1–1, 2021.

[41] M. Elhoseny, G. Ramirez-Gonzalez, O. M. Abu-Elnasr, S. A. Shawkat, N. Arunkumar, and A. Farouk, "Secure medical data transmission model for IoT-based healthcare systems," *IEEE Access*, vol. 6, pp. 20596–20608, 2018.

[42] N. Abbas and F. Yu, "Design and implementation of a video surveillance system for linear wireless multimedia sensor networks," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 524–527, Chongqing, China, 2018.

[43] S. A. Deepthi, E. S. Rao, and M. N. G. Prasad, "RTL Implementation of image compression techniques in WSN," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 3, p. 1750, 2019.

[44] H. P. Medeiros, M. C. Maciel, R. Demo Souza, and M. E. Pellenz, "Lightweight data compression in wireless sensor networks using Huffman coding," *International Journal of Distributed Sensor Networks*, vol. 10, no. 1, Article ID 672921, 2014.