Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 15 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i15.51327

PAPER

Comparative Evaluation of PD Detection Using Deep Learning on IMFCCs Extracted from VMD

Nouhaila Boualoulou[1](✉), **Benayad Nsiri**[2], **Taoufiq Belhoussine Drissi**[1]

[1]Laboratory Electrical and Industrial Engineering, Information Processing, Informatics, and Logistics (GEITIIL), Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco

[2]Research Center STIS, M2CS, National Higher School of Arts and Craft, Rabat (ENSAM), Mohammed V University in Rabat, Rabat, Morocco

boualoulounouha@gmail.com

ABSTRACT

This paper presents a new method for extracting vocal features for the diagnosis of Parkinson's disease (PD) via voice analysis applying variational mode decomposition (VMD). The classical method of extracting mel-frequency cepstral coefficients (MFCC) is compared to a new approach that generates coefficients named intrinsic mel-frequency cepstral coefficients (IMFCC). For this study, two audio databases were used: the SAKAR database containing 38 recordings and a PC-GITA database comprising 50 recordings. The signal preprocessing steps include frame segmentation, pre-emphasis, and filtering. The voice signal is then decomposed into intrinsic modes employing VMD. From these modes, the log-energy of specific components is calculated to extract the IMFCC. In this study, two types of classifiers were used: convolutional neural networks (CNN) and long short-term memory (LSTM). The results show that IMFCC provides a new perspective for representing vocal signals, capturing distinct features compared to classical MFCC. Notably, the IMFCC2 attained the highest accuracy of 100% adopting the CNN classifier. This approach could improve the performance of systems for identifying PD via voice analysis, offering a robust and complementary alternative to existing feature extraction methods.

KEYWORDS

Parkinson's disease (PD), intrinsic mel-frequency cepstral coefficients (IMFCC), mel-frequency cepstral coefficients (MFCC), long short-term memory (LSTM), convolutional neural networks (CNN), variational mode decomposition (VMD)

1 INTRODUCTION

Parkinson's disease (PD) represents a significant global health challenge due to its progressive nature and the substantial impact it has on patients' quality of life. The early stages of the disease often go undetected, which can delay crucial interventions and worsen outcomes. Therefore, early and accurate diagnosis is paramount for enabling timely treatment, improving patient prognosis, and enhancing overall disease management. In this context, the use of non-invasive, cost-effective biomarkers for PD detection has gained increasing attention in the research community.

Boualoulou, N., Nsiri, B., Drissi, T.B. (2024). Comparative Evaluation of PD Detection Using Deep Learning on IMFCCs Extracted from VMD. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(15), pp. 121–140. https://doi.org/10.3991/ijoe.v20i15.51327

Article submitted 2024-07-23. Revision uploaded 2024-10-02. Final acceptance 2024-10-02.

Recent advancements in signal processing and machine learning have provided promising opportunities to harness voice-based biomarkers for PD diagnosis. Voice, as a non-invasive and readily accessible indicator, has the potential to reflect subtle changes in motor control associated with the onset and progression of PD. By analyzing these voice patterns, it is possible to develop diagnostic tools that are both effective and easy to implement in clinical settings.

This study investigates the integration of advanced signal processing techniques with state-of-the-art machine learning algorithms to analyze voice recordings for PD detection. Specifically, we propose a comprehensive framework that utilizes variational mode decomposition (VMD) to decompose voice signals into their fundamental components. Following this, we compute intrinsic mel-frequency cepstral coefficients (IMFCC) to extract relevant features that capture the nuances of voice affected by PD. These features are then fed into deep learning models, specifically convolutional neural networks (CNN) and long short-term memory (LSTM) networks, for classification purposes. This integrated approach is designed to maximize the accuracy and efficiency of PD detection, thereby contributing to improved patient care and disease management.

The proposed methodology is applied to two distinct datasets: the Sakar dataset, which comprises 38 audio recordings from both PD patients and healthy controls, and the PC-GITA dataset, which includes 50 audio recordings. These datasets offer a diverse range of voice samples, making them suitable for evaluating the robustness and generalizability of the proposed models. To ensure rigorous model evaluation, the datasets are partitioned into training and testing sets using robust cross-validation techniques. Specifically, a holdout method is employed, where 20% of the data is reserved for testing, to prevent overfitting and to accurately assess the model's ability to generalize to unseen data.

The entire implementation process, including data preprocessing, feature extraction (MFCC and IMFCC), and model training (LSTM and CNN), is conducted using MATLAB. This environment provides the necessary tools to handle the complex computational tasks involved and ensures that each step is meticulously designed to uphold the reproducibility and rigor of the experimental setup. This attention to detail is crucial for validating the proposed methodology and for facilitating future research efforts in this domain.

The effectiveness of the proposed techniques is evaluated using a variety of performance metrics, including accuracy, specificity, and sensitivity. These metrics provide a comprehensive assessment of the models' ability to correctly classify PD and non-PD cases based on the extracted voice features. The results obtained from these evaluations are expected to highlight the potential of combining advanced signal processing and deep learning techniques in developing reliable and scalable diagnostic tools for Parkinson's disease.

Subsequent sections of the paper include: Section 2 discusses previous research; Section 3 offers a comprehensive overview of the database; Section 4 outlines the research methodology; Section 5 presents the results and includes an in-depth discussion and analysis. Lastly, Section 6 delivers the concluding remarks and recommendations for future research.

2 RELATED WORK

Taha Khan et al. [1] presented cepstral separation distance characteristics, showcasing their effectiveness in detecting PD. They noted that these features performed well, with intra-class correlation coefficients exceeding 0.9. Orozco-Arroyave et al. [2] concentrated on PD diagnosis utilizing prolonged words and vowels, achieving an accuracy of up to 85% through the employment of cepstral and spectral features. Mehmet et al. [3] proposed an innovative method for PD detection from voice signals, utilizing pre-trained deep networks and LSTM models alongside Mel spectrograms obtained from filtered audio signals using VMD. Karan et al. [4] explored voice tremors in PD patients applying a combination of VMD and Hilbert spectrum analysis (HSA). For this purpose, they introduced a new set of characteristics termed Hilbert Cepstral Coefficients. In the work conducted by Sakar et al. [5], a range of voice signal processing algorithms were evaluated for PD evaluation. They presented an innovative tool named Q-factor Wavelet Transform (TQWT) and trained classifiers with diverse feature groups. Their findings revealed that TQWT and MFCC attained the greatest accuracy, underscoring their importance in PD classification. An average accuracy of 86% was attained using the support vector machines (SVM) classifier. Various additional studies have similarly employed DWT for the accurate recognition of PD [6], [7], [8]. In recent research, Asmae Ouhmida et al. [9] explored various machine learning techniques to diagnose PD by analyzing voice disorders. They assessed different classifiers such as SVM, K-nearest neighbors (KNN), and decision trees (DT) and used feature selection methods such as mRMR and ReliefF to boost performance. The findings indicated that the KNN classifier outperformed others, achieving an accuracy rate with an AUC of 98.26%. Rania Khaskhoussy et al. presented a novel approach for the automatic identification of PD using voice signal analysis. This study employs SVM and CNN as learning techniques for the classification of speech task-derived data. Two sets of input data are utilized: raw speech signal values and i-vector features with dimensions of 100, 200, and 300. Analysis of a test dataset consisting of 28 participants reveals 100% accuracy, affirming the efficacy of the suggested approach for detecting PD [10]. Ahmed Anter et al. introduced a powerful regression model designed to track PD patients through voice recordings. Their model incorporates a binary version of an ant lion optimizer (BALO) for selecting voice features and an extreme learning machine (ELM) using differential evaluation (DE) for continuous prediction of the Unified Parkinson’s Disease Rating Scale (UPDRS). In comparison to several meta-heuristic models and machine learning forecasting techniques, the BALO-DEELM approach shows notable efficiency in feature selection and achieves more precise predictions [11]. Tao Zhang and colleagues introduced a technique for extracting voice features using fractional attribute topology (FrAT). Initially, FrAT is applied to incorporate energy information into the voice spectrogram for time–frequency representation. Concurrently, feature extraction is carried out using formal concept analysis, establishing a formal context using statistical data in the fractional domain to develop FrAT. Subsequently, connected components that denote discrete FrAT degrees are extracted as CCF-FrAT features to enhance classification accuracy. The approach achieved top accuracies of 99.57%, 95.33%, and 94.13% across three datasets at p = 0.7 [12]. Renata et al. proposed using extreme learning machines (ELM) to expedite training time in the context of PD detection using voice signal spectrograms. This study examined five different pre-trained CNN models namely VGG-16, AlexNet, SqueezeNet, ResNet-50, and Inception V3. The findings indicated that the ELM-based classifier achieved a comparable level of accuracy to CNN models, while significantly reducing the training time [13]. Gaffari SeleK et al. proposed a new method, SkipCon-Net, which integrates DL and ML techniques for PD detection from speech signals. SkipCon-Net is tailored to extract critical features from voice signals. Moreover, the RF algorithm is used to forecast features derived from the SkipCon-Net architecture. The combined SkipCon-Net + RF approach showed excellent performance, achieving an accuracy of 98.30% on the PDO_Dataset and 99.11% on the PD_Dataset [14]. Tao Zhang et al. introduced co-occurrence direction attribute topology (CDAT), a method for describing voice features. Initially, CDAT establishes a formal context using statistically derived direction information within spectrogram sub-regions to depict relationships between energy points and their directional attributes. This method captures coupling among directional attributes. The number of connected domains in CDAT, indicating nodal coupling intensity, is extracted as structural features and validated across various classifiers. Experimental evaluations on Parkinsonian sustained vowel datasets (CPPDD and SPDD) achieved high accuracy of 95.84% and 93.90% with RF, respectively. The suggested approach emphasizes variations in energy direction derivatives in spectrograms via attribute topology, demonstrating robust classification accuracy across diverse native language datasets and in cross-corpora experiments, often outperforming or matching state-of-the-art PD classification methods [15]. In addition to previous research on PD, various studies have explored machine learning techniques for medical diagnostics across different conditions, further validating the application of such techniques to PD detection. Al-Nawashi et al. proposed a machine learning-based approach for breast cancer detection using CNNs for feature extraction and classifiers like RF, SVM, and logistic regression. Their model was tested on a dataset of over 3,000 mammograms and achieved high precision and accuracy, demonstrating the potential of machine learning in cancer diagnostics [16]. Khazaaleh et al. presented an unsupervised machine learning model for handling DNA malfunctions, highlighting the effectiveness of unsupervised learning in genomic data analysis [17]. In the field of diabetic retinopathy diagnosis, Al-Hazaimeh et al. combined artificial intelligence and image processing techniques to analyze retinal fundus images. Using CNNs for feature extraction, their model achieved an impressive 98.3% accuracy, showcasing the reliability of AI in the field of ophthalmology [18]. Gharaibeh et al. introduced a swin transformer-based segmentation model combined with a multi-scale feature pyramid fusion module for Alzheimer's disease detection. The swin transformer model is known for its robust performance in medical image segmentation tasks [19]. These diverse applications of machine learning in medical diagnostics illustrate the broad utility of these techniques, reinforcing their relevance for neurodegenerative diseases such as Parkinson's.

3 DATASET

Sakar dataset: The Sakar dataset includes 38 audio samples, with 20 recordings from individuals diagnosed with PD (10 males and 10 females) aged between 39 and 79 years and 18 recordings from healthy individuals (eight males and eight females) aged between 50 and 70 years. Participants were directed to pronounce the sustained vowel 'a' using a standard microphone operating at a sampling rate of 44.100 Hz. These recordings were captured in stereo mode using a desktop computer equipped with a 16-bit sound card and saved in WAV file format [20].

PC-GITA dataset: The PC-GITA dataset, collected using professional-grade equipment, comprises recordings from 50 healthy individuals and 50 individuals diagnosed with PD. Recordings were made at a resolution of 16 bits and a sampling rate of 44.1 kHz. The healthy group included women aged 43 to 76 years (mean age 61.4 years, SD 9.5 years) and men aged 31 to 86 years (mean age 60.5 years, SD 9.4 years). The PD group consisted of women aged 49 to 75 years and men aged 33 to 81 years [21].

4 METHODOLOGY

The recommended approach involves several critical stages, including initial handling of the captured data, signal breakdown using VMD into intrinsic modes, then extraction of MFCC and IMFCC features from each mode, and classification using CNN and LSTM. This systematic framework facilitates methodical and structured analysis of speech data, leading to increased accuracy in PD detection.

Figure 1 outlines the methodology suggested in this study, with a comprehensive explanation provided in the ensuing section.

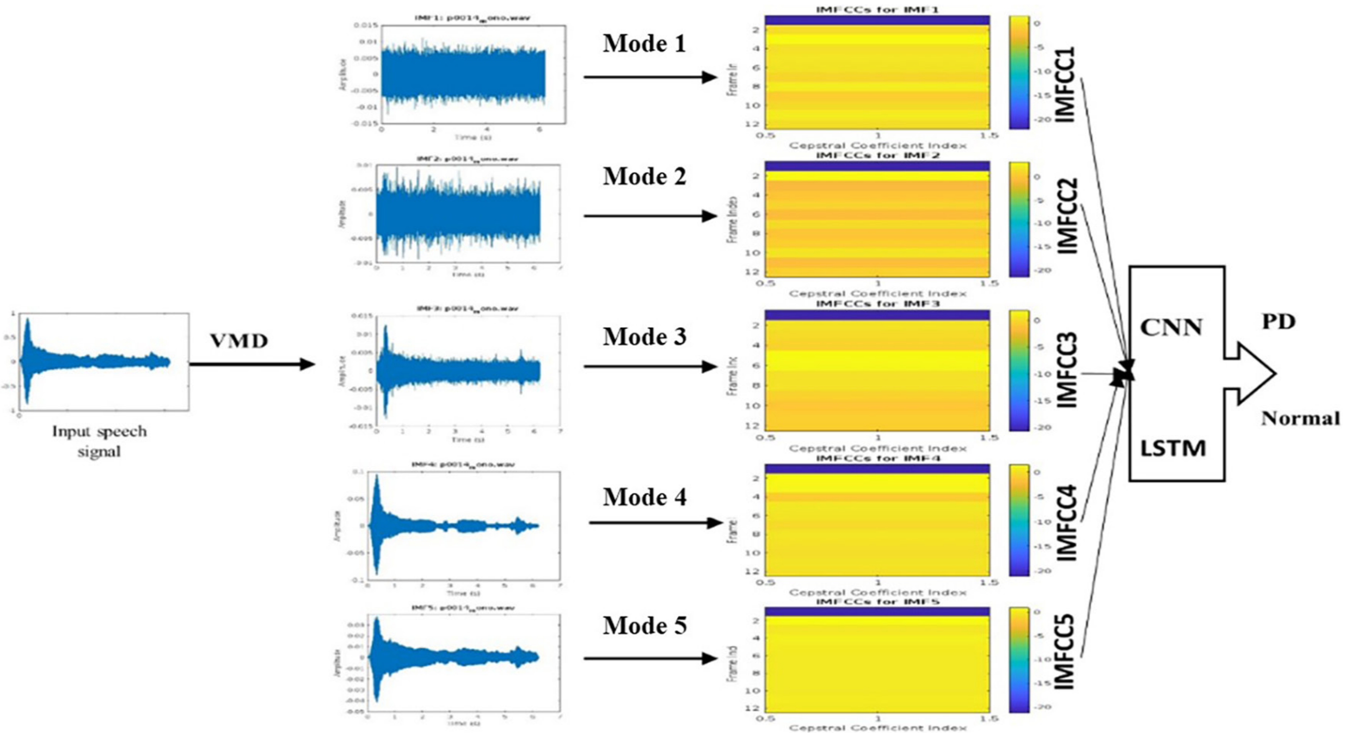

Fig. 1. The proposed approach

Algorithm 1: Algorithm of the Suggested Method

```
1. Input: A set of speech signals (x1, x2, …, xn).
2. Output: Five matrices for five modes, where each mode contains a set of 12 IMFCC coefficients.
3. Load Audio Data:
   – Load the dataset containing the speech signals (audio files).
4. For each audio file in the dataset:
   1. Load the speech samples.
   2. Apply VMD (Variational Mode Decomposition) to decompose the speech signal into modes.
   3. Select the first 5 modes from the decomposed signal.
   4. For each of the 5 modes:
      1. Extract the mode's signal.
      2. Compute the Energy: Square each value of the mode's signal to get its energy.
      3. Apply Logarithm: Take the logarithm (base 10) of the energy to get the log-energy.
      4. Compute IMFCCs:
         – Apply the Discrete Cosine Transform (DCT) to the log-energy to obtain 12 IMFCC coefficients
           for the mode.
      5. Store IMFCCs:
         – Save the 12 IMFCC coefficients for each mode.
   5. Compile Results:
      – Store the 12 IMFCC coefficients for each of the 5 modes separately in a matrix.
5. End For.
6. Output: Five matrices for five modes, where each mode contains a set of 12 IMFCC coefficients.
```

4.1 Feature extraction

This section introduces an innovative method for identifying PD using voice analysis. The method combines VMD and the extraction of IMFCC. The approach aims to extract robust and precise audio features essential for identifying vocal markers associated with Parkinson's disease.

Variational mode decomposition is an innovative algorithm for signal processing, presented by Dragomiretskiy [22] to address the lack of a theoretical basis for the empirical mode decomposition (EMD) algorithm. VMD is an adaptive signal decomposition technique that decomposes a signal into a finite number of sub-signals, referred to as modes, each possessing unique and narrowband frequency characteristics. Unlike traditional approaches such as EMD, VMD is framed as a variational problem, enhancing its robustness and effectiveness in decomposing complex signals.

Variational mode decomposition aims to disaggregate an incoming signal x(t) into a set of modes $\{u_k(t)\}$ Equation 2, where each mode $u_k(t)$ is concentrated around a central frequency w_k.

Each mode $u_k(t)$ is assumed to be amplitude modulated and frequency modulated, formulated as (Equation 1):

$$u_k(t) = A_k(t)\cos(\varphi_k(t)) \tag{1}$$

Where $A_k(t)$ is the amplitude envelope and $\varphi_k(t)$ is the phase

$$x(t) = \sum_{k=1}^{k} u_k(t) + \mathrm{res}(t) \tag{2}$$

Where $u_k(t)$ is decomposed mode and res(t) is residual component.

Intrinsic mode function cepstral coefficient: An advanced method for feature extraction that derives cepstral coefficients from modes obtained via VMD. This novel technique, introduced in this study, involves applying the DCT to the logarithm of the energy of each mode (Equation 3). Equation 4 details the computation formula for IMFCCs [23], [24].

$$E_{mode_i} = \frac{1}{N}\sum_{i=1}^{N}(mode_i)^2 \tag{3}$$

$$IMFCC_{mode_i} = DCT\left(\log\left(E_{mode_i}\right)\right) \tag{4}$$

With i denoting the mode number.

Mel frequency cepstral coefficient (MFCCs): MFCCs are calculated from the modes obtained using VMD. The process begins by applying VMD to the original audio signal, decomposing it into several modes. For each selected mode, MFCCs are computed to capture cepstral characteristics derived from the spectral information of the mode. The MFCC extraction involves pre-emphasizing the mode, framing the signal, and applying the short-time Fourier transform (STFT) to compute the magnitude spectrum. A mel-frequency filter bank is then applied to this spectrum, followed by the computation of the logarithm of the filter bank energies. Finally, the DCT is applied to these log energies to produce the MFCCs. These coefficients represent the cepstral information, which emphasizes the spectral envelope of the selected mode, enhancing the analysis of audio features, particularly for tasks such as voice signal analysis and disease detection. The transformation formula from linear frequency to mel frequency, detailed in Equation 5.

$$\mathrm{Mel}(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right) \tag{5}$$

4.2 Method explanation

The initial stage of the method entails preparing the sound signal for analysis. The vocal signal is segmented into frames of a fixed duration of 25 milliseconds (ms) with an overlap of 10 ms between frames to ensure continuity and detailed capture of temporal variations. A pre-emphasis filter is applied to amplify the high frequencies, compensating for the typical drop in energy in these frequencies and improving the clarity of the extracted features.

Once the signal is preprocessed, VMD is applied to break down the vocal signal into several intrinsic modes. VMD is an advanced signal decomposition technique that solves the variational minimization problem to extract a set of modes, each centered on a specific frequency. This decomposition allows precise separation of the different frequency components of the vocal signal, essential for capturing fine characteristics associated with vocal anomalies in Parkinson's disease.

After decomposition, the first five modes are selected, as they generally contain the high-frequency components of the signal (see Figure 2). These modes are crucial for analysis because they capture the fine variations and anomalies in the voice that may be linked to Parkinson's disease.

For each selected mode, the energy is calculated by squaring the mode values. This energy measure quantifies the amplitude of oscillations in each mode, providing a clear indication of each frequency band's contribution to the vocal signal. To compress the dynamic range of the energy values and mitigate the impact of extreme variations, a base-10 logarithmic transformation is implemented to the energy values. Then, a DCT is applied to the log-energy values to obtain the cepstral coefficients. These coefficients, called IMFCC, capture the main spectral characteristics of the vocal signal.

The number of cepstral coefficients extracted is set to 12, corresponding to the first few coefficients that contain most of the relevant spectral information. These coefficients are then averaged for each segment of the vocal signal, producing a representative and stable feature vector. The IMFCC feature vectors for all frames of the vocal signal are aggregated to form a complete set of audio descriptors. These descriptors are then analyzed to identify vocal markers associated with PD. By comparing the IMFCC extracted from healthy individuals to those from Parkinson's patients, specific anomalies in vocal characteristics can be identified.

4.3 Classification

Long short-term memory: An architecture of recurrent neural network (RNN) built to solve the issue of diminishing gradients in conventional RNNs. LSTM networks excel in capturing prolonged dependencies and are widely applied in tasks involving sequential data, for instance, natural language processing and speech recognition. Their ability to retain and utilize information over extended sequences makes them indispensable in modeling complex temporal relationships [25].

LSTM network definition: The network consists of a sequence of key layers:

- Input layer (sequence input layer): Each sequence of 12 coefficients is treated as a temporal sequence with a dimension of one at each time step.

- LSTM layer: Incorporating 200 hidden units, this layer is optimized to handle long-term relationships in sequences, extracting meaningful features from IMFCC data over time. The output at each time step t is determined as (Equation 6):

$$h_t = \mathrm{LSTM}(x_t, h_{t-1}, c_{t-1}) \tag{6}$$

Where c_{t-1} is the previous cell state, h_{t-1} is the previous hidden state, and x_t is the input at time step t.

- Fully connected layer: Provides a final interpretation of features extracted by the LSTM layer, reducing dimensions while preserving information richness.
- Softmax layer: Normalizes outputs into probabilities, providing a probability distribution over potential class.
- Classification layer: Defines the final output by categorizing input sequences into one of the two classes, suited for the specific task of binary classification.

Network training options: Training parameters are adjusted to optimize convergence and model performance:

- Adam optimizer: Chosen for its effective handling of gradients. Parameter updates are computed as (Equation 7):

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \tag{7}$$

Where θ_t are the network parameters, ε is a small constant for numerical stability, $\hat{m}_t$ is the estimated gradient $\hat{v}_t$ is the estimated second moment of the gradient and η is the learning rate.

- 200 training epochs: Determines how many times the entire training dataset is traversed during model learning.
- Mini-batch of 70: Subset size employed for calculating gradients and adjusting network weights, facilitating accelerated and more efficient learning.
- Gradient threshold of 1: Limits the size of gradients updated at each backpropagation step, preventing gradient explosions that could compromise learning stability.
- Training progress visualization: Monitors model performance over time, facilitating diagnosis of potential learning issues or overfitting.

Convolutional neural network: An advanced DL framework designed to automatically extract meaningful features from input data using convolutional layers. These neural networks are renowned for their ability to discern intricate patterns in complex datasets, making them indispensable in diverse fields, including speech recognition and speaker identification. Their hierarchical feature learning capabilities and robust performance have positioned CNNs at the forefront of modern signal processing and machine learning methodologies [25].

Defining CNN Architecture:

- Input layer: The input layer expects data of size [1, 12] representing each IMFCC sequence.

- Convolutional layers:
  - First convolutional layer: Apply a convolution operation with a filter size of 3 and 8 filters, Equation 8.

$$Y[i]=\sum_{j=0}^{2} X[i+j]\cdot W[j]+b \tag{8}$$

Where Y[i] is the output, X is the input sequence, W is the filter, and b is the bias term

  - Batch normalization: Normalize the activations to stabilize and accelerate training, Equation 9.

$$\hat{X}=\frac{X-\mu}{\sqrt{\sigma^2+\varepsilon}} \tag{9}$$

Where μ and σ are the mean and the variance of X, and ε is a small constant for numerical stability

  - ReLU activation: Apply the ReLU activation function to incorporate non-linearity, Equation 10.

$$\mathrm{ReLU}(x)=\max\,(0,x) \tag{10}$$

  - Max Pooling: Down-sample the feature maps to reduce dimensionality, Equation 11.

$$Y_{POOL}=\max_{j=0}^{k-1} X[i+j] \tag{11}$$

Where K is the size of the pooling window.

  - Second and third convolutional layers: Apply additional convolutional layers with increasing numbers of filters (16 and 32, respectively), Subsequently, batch normalization, ReLU activation, and max pooling are applied after every layer.
- Flatten layer: Convert the 3D output from the convolutional layers into a 1D vector to prepare for the fully connected layer, Equation 12.

$$\text{Flattened output}=\text{reshape}(Y) \tag{12}$$

- Fully connected layer: Use a fully connected layer with two output neurons corresponding to the number of classes, Equation 13.

$$Z=W\,.\,Y+b \tag{13}$$

Where Z is the output of the fully connected layer, W is the weight matrix, and b is the bias vector.

Softmax and classification layers:

- Softmax layer: Converts CNN output into class probabilities, ensuring they sum to 1 for classification.
- Classification layer: Uses softmax probabilities to make final predictions and computes categorical cross-entropy loss during training.

Training process

- Optimizer (SGDM): Utilizes Stochastic Gradient Descent with Momentum for weight updates, enhancing convergence speed.
- Batch size: Mini-batch size set to 40 samples, optimizing memory usage and training efficiency.
- Max epochs: Trains over 1000 epochs, iterating through the entire training dataset multiple times to improve model accuracy.
- Regularization techniques: Includes batch normalization to stabilize training and prevent overfitting, ensuring robust generalization.
- Visualization and monitoring: Tracks training progress with visual plots, aiding in performance evaluation and hyperparameter tuning for optimal MFCC sequence classification.

4.4 Classification explanation

- **For CNN:** The 12 extracted IMFCCs from the audio recordings are used as inputs to CNN for the classification of PD. This classification process involves several key steps, enabling the network to learn and leverage the features of the IMFCCs to differentiate between the voices of individuals with PD and those who are healthy. The CNN is structured to optimize learning of spectral characteristics from the IMFCC vectors. The sequential input layer accepts the IMFCC vectors, each sized 1 × 12, corresponding to the 12 extracted coefficients. These vectors represent the spectral features of voice segments. The convolutional layers of the CNN capture local patterns in the IMFCC data. The first convolutional layer uses filters of size 3 to scan the IMFCC vectors, producing eight distinct feature maps. Batch normalization is applied after each convolution to stabilize and accelerate learning, followed by a ReLU activation to introduce non-linearity, allowing the network to model complex relationships in the data. Max pooling layers decrease the dimensionality of the feature maps while preserving the key information. This is followed by a second convolutional layer with sixteen additional filters for a deeper analysis of local features. Another round of batch normalization and ReLU activation improves the stability and non-linearity of the model. A second max pooling layer further reduces dimensionality. The third convolutional layer applies thirty-two filters, enabling finer feature extraction. A final batch normalization and ReLU activation stabilize the network further. The flatten layer converts the resulting 2D feature maps into a 1D vector, processing the data for the fully connected layers. These fully connected layers process the flattened vector and learn complex non-linear combinations of the characteristics derived from convolutional layers. Finally, the softmax layer generates a probability distribution over the classes (PD or healthy), and the classification layer uses these probabilities to predict the most likely class. During training, the CNN adjusts its weights based on the error between predictions and actual labels, using a stochastic gradient descent algorithm with momentum. The convolutional filters learn to detect relevant patterns in the IMFCCs that indicate PD. The fully connected layers aggregate these patterns to make the final predictions.
- **For LSTM:** The 12 extracted IMFCCs from the audio recordings are used as inputs to the LSTM network for the classification of PD. This process involves diverse key steps, enabling the network to learn and leverage the features of the IMFCCs to discriminate between PD patients and healthy ones. The IMFCC vectors, each containing 12 coefficients, are formatted into cell arrays suitable for input into the LSTM network. The labels corresponding to the training and test sets are also prepared, with the training labels being converted to categorical format for the classification

task. The LSTM network is designed with an input size of 1 and consists of several layers: a sequence input layer, an LSTM layer with 200 hidden units configured to output only the last element in the sequence, a fully connected layer with two output classes (indicating the presence or absence of PD), a softmax layer to generate probability distributions over the classes, and a classification layer to make the final prediction. The network is trained using the 'adam' optimization algorithm over 200 epochs, with a mini-batch size of 70. The training process adjusts the network's weights based on the error between predictions and actual labels, optimizing the network to improve its classification accuracy. Once trained, the network is used to classify the test set IMFCCs. The predicted labels are compared to the actual test labels to evaluate the network's performance. A confusion chart is created to view the classification results, showing the distribution of TN, TP, FN, and FP predictions. The overall accuracy of the network is computed as the ratio of correctly predicted instances to the total number of instances in the test set. By leveraging the temporal dynamics captured by the LSTM network, this classification approach aims to enhance the accuracy and reliability of PD recognition via voice analysis, offering a significant tool for automated diagnosis and clinical monitoring.

4.5 Evaluation

To assess the effectiveness of the classifier, the confusion matrix is utilized, which is a tabular representation of predicted versus actual classifications. This matrix aids in calculating various performance metrics that gauge the classifier's accuracy in identifying individuals with and without PD based on its predictions.

The confusion matrix is structured as follow (refer to Table 1):

Table 1. The structure of the confusion matrix

	Predicted Negative (No Parkinson's)	Predicted Positive (Parkinson's)
Actual Negative (No Parkinson's)	**TN**	**FP**
Actual positive (Parkinson's)	**FN**	**TP**

From the confusion matrix, the following performance metrics accuracy, sensitivity, and specificity, respectively in Equation 14, Equation 15, and Equation 16 can be derived

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \tag{14}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{15}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{16}$$

True positives (TP) pertain to accurately identified individuals without PD. True negatives (TN) correspond to correctly identified individuals afflicted with PD. False positives (FP) signify individuals without PD who were erroneously classified as having the disease. False negatives (FN) denote individuals with PD who were incorrectly classified as not having the disease.

5 RESULTS AND DISCUSSION

The proposed method for detecting PD is based on extracting IMFCC from voice signals. First, VMD is applied to break down the voice signals into distinct modes, each representing different frequency components of the signal, enabling more detailed analysis. Following this, MFCC features are extracted from these modes, with parameters such as a frame duration of 25 ms, frame shift of 10 ms, and 20 filter bank channels. These features capture key spectral characteristics of the voice, which are essential for distinguishing between healthy individuals and those with PD. The IMFCCs are obtained by applying the DCT to the logarithm of the energies of the intrinsic modes.

These extracted features provide a compact and informative representation of the voice signals, which are then used for classification. For this, CNNs and LSTM networks are employed. CNNs are used to capture important local features from the IMFCCs through convolution and pooling layers, which reduce data dimensionality while preserving critical information. LSTMs are utilized to capture long-term temporal dependencies in the voice signals, modeling complex time variations. Model performance is evaluated using holdout cross-validation, with the dataset split into 80% for training and 20% for testing. Models are optimized on the training data, and their generalization ability is assessed on the unseen test data.

5.1 Comparative study of feature patterns

This section focuses on a comparative analysis of the first five modes between individuals with PD and healthy individuals. The energy distribution across these modes reveals compelling differences: energy levels are notably lower in PD patients compared to healthy subjects, as depicted in Figure 2. This highlights the diagnostic potential of the initial modes in distinguishing between the two groups. The distinct energy patterns observed in these modes underscore their relevance in capturing essential vocal characteristics associated with Parkinson's disease. Importantly, energy distribution in healthy individuals show higher levels compared to those affected by the disease. This suggests that crucial vocal signal information is concentrated within the first five modes.

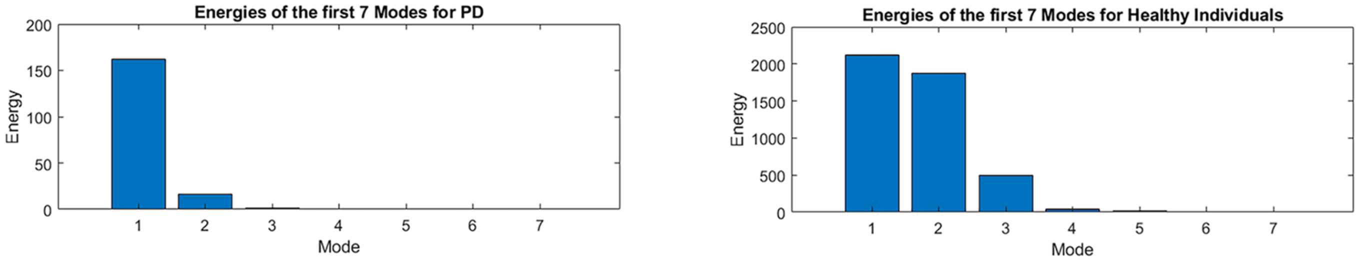

Fig. 2. Comparative analysis of the energies of the first seven modes derived from VMD for individuals with PD and healthy individuals

5.2 Comparative analysis of IMFCC and MFCC energies

The comparison between the energies of IMFCC and MFCC, presented in Figure 3, highlights key differences across the first five modes. For mode one, the IMFCC energy reaches 2976, while the MFCC energy is 1439. Similarly, for mode two, IMFCC records 2923 compared to MFCC's 1072. The trend continues with mode three showing IMFCC at 3170 and MFCC at 937, mode four with IMFCC at 3627 and MFCC at 859, and mode five with IMFCC at 3776, while MFCC peaks at 634.

These observations underline a consistent pattern where IMFCC energies are significantly higher than MFCC energies across all modes, suggesting that IMFCC provides a richer and more detailed representation of vocal information. This enhanced energy profile, particularly in the IMFCC spectrum, suggests a superior capability for capturing the subtle vocal characteristics crucial for PD recognition. The substantial difference in energy levels implies that IMFCC encodes more comprehensive vocal features, which could lead to improved diagnostic accuracy and sensitivity in clinical applications.

5.3 Statistical analysis of IMFCC and MFCC coefficients

The statistical analysis compares the correlation coefficients and p-values between IMFCC and MFCC across two databases, Tables 2 and 3. In Database SAKAR, significant correlations are observed primarily in the first few coefficients of both IMFCC and MFCC. Specifically, IMFCC one exhibits a correlation coefficient of 0.5585 with a highly significant p-value of 0.0003, indicating a strong association. Similarly, MFCC one shows a correlation coefficient of 0.7820 with an equally significant p-value of 0.0000. As we progress through the coefficients, the correlations fluctuate with varying degrees of significance. Notably, IMFCC 3 and IMFCC nine show moderate correlations with coefficients of 0.3937 ($p = 0.0050$) and 0.2585 ($p = 0.1171$), respectively. Meanwhile, MFCC three demonstrates a weaker correlation coefficient of –0.1366 ($p = 0.3431$), indicating less pronounced association in this context. In Database PC-GITA, a similar trend is observed where IMFCC one exhibits a strong correlation coefficient of 0.8122 ($p = 0.0000$), aligning closely with MFCC one, which shows a correlation coefficient of 0.9471 ($p = 0.0000$). This robust correlation continues across several coefficients, highlighting the consistency in vocal feature representation between IMFCC and MFCC. Overall, the results underscore the effectiveness of both IMFCC and MFCC in capturing vocal characteristics, with IMFCC often demonstrating comparable or stronger correlations compared to MFCC across various coefficients. These findings support the potential of IMFCC as a valuable feature set in the examination and categorization of vocal signals, particularly in applications such as PD detection.

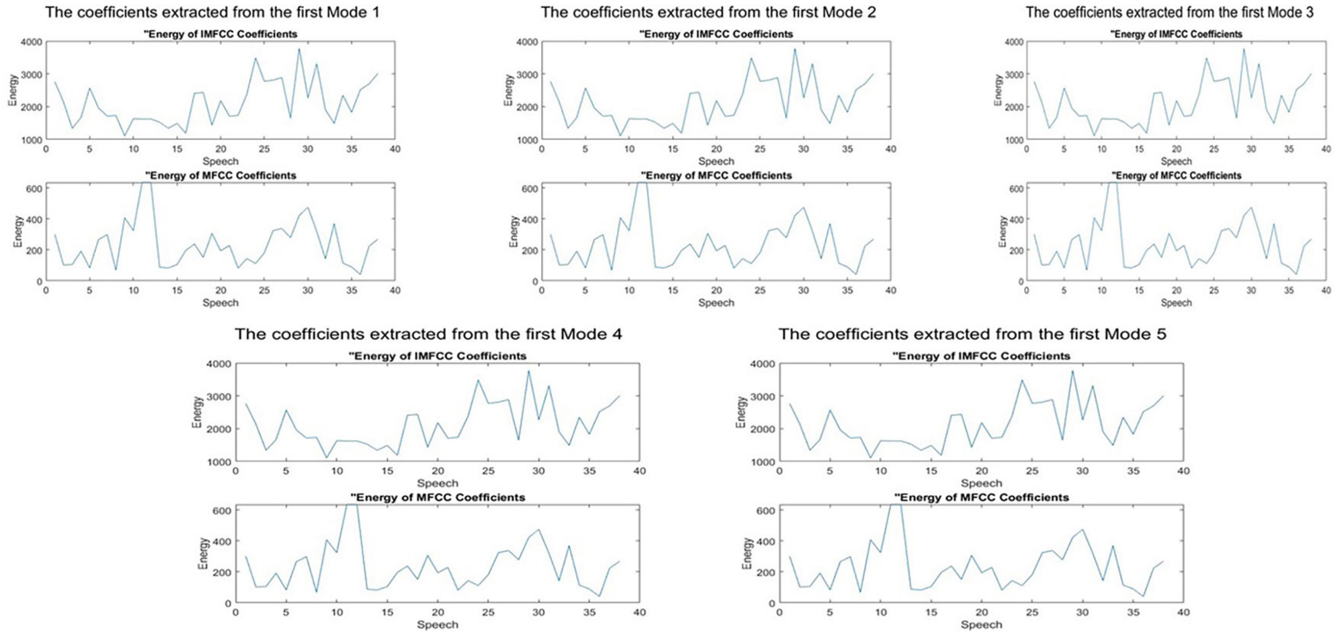

Fig. 3. Energy comparison for MFCC and IMFCC coefficients extracted from modes through VMD

Table 2. Analysis of Spearman correlation in proposed feature Sakar database

Sakar Database					
IMFCC Coefficients	Correlation Coefficients	P-Value	MFCC Coefficients	Correlation Coefficients	P-Value
IMFCC 1coefficients	0.5585	0.0003	**MFCC 1coefficients**	0.7820	0.0000
IMFCC 2coefficients	–0.0151	0.9283	**MFCC 2coefficients**	0.2489	0.1319
IMFCC 3coefficients	–0.0149	0.9293	**MFCC 3coefficients**	0.3824	0.0178
IMFCC 4coefficients	–0.2257	0.1731	**MFCC 4coefficients**	0.3169	0.0525
IMFCC 5coefficients	0.0278	0.8684	**MFCC 5coefficients**	0.3316	0.0420
IMFCC 6coefficients	0.3176	0.0520	**MFCC 6coefficients**	0.2707	0.1002
IMFCC 7coefficients	–0.2235	0.1775	**MFCC 7coefficients**	0.3169	0.0525
IMFCC 8coefficients	–0.0468	0.7801	**MFCC 8coefficients**	0.0867	0.6049
IMFCC 9coefficients	0.2585	0.1171	**MFCC 9coefficients**	0.3955	0.0140
IMFCC 10coefficients	0.2154	0.1941	**MFCC 10coefficients**	0.3344	0.0402
IMFCC 11coefficients	–0.3095	0.0587	**MFCC 11coefficients**	–0.1615	0.3326
IMFCC 12coefficients	–0.0131	0.9376	**MFCC 12coefficients**	0.1370	0.4121

Table 3. Analysis of Spearman correlation in proposed feature PC-GITA database

PC-GITA Database					
IMFCC Coefficients	Correlation Coefficients	P-Value	MFCC Coefficients	Correlation Coefficients	P-Value
IMFCC 1coefficients	0.8122	0	**MFCC 1coefficients**	0.9471	0
IMFCC 2coefficients	0.6441	0.0000	**MFCC 2coefficients**	0.2118	0.1395
IMFCC 3coefficients	0.3937	0.0050	**MFCC 3coefficients**	–0.1366	0.3431
IMFCC 4coefficients	0.4466	0.0013	**MFCC 4coefficients**	0.1382	0.3376
IMFCC 5coefficients	0.4504	0.0012	**MFCC 5coefficients**	0.2799	0.0493
IMFCC 6coefficients	0.1309	0.3636	**MFCC 6coefficients**	0.1382	0.3376
IMFCC 7coefficients	0.0906	0.5303	**MFCC 7coefficients**	0.1311	0.3629
IMFCC 8coefficients	0.0894	0.5360	**MFCC 8coefficients**	0.2279	0.1113
IMFCC 9coefficients	0.4821	0.0005	**MFCC 9coefficients**	–0.0228	0.8749
IMFCC 10coefficients	0.2356	0.0995	**MFCC 10coefficients**	0.1370	0.3417
IMFCC 11coefficients	0.5406	0.0001	**MFCC 11coefficients**	0.1850	0.1977
IMFCC 12coefficients	0.6452	0.0000	**MFCC 12coefficients**	0.3624	0.0101

5.4 Results and comparison with previous study

Table 4. The outcome of the MFCC using Sakar database

Sakar Dataset						
MFCCs	**CNN**			**LSTM**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
MFCC1	71.42	100	50	63.63	60	66.66
MFCC2	71.42	100	50	72.72	60	83.33
MFCC3	57.14	66.66	50	54.54	60	50
MFCC4	71.42	100	50	54.54	40	66.66
MFCC5	57.14	33	75	45.45	20	66.66

Table 5. The outcome of the IMFCC using Sakar database

Sakar Dataset						
IMFCCs	**CNN**			**LSTM**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
IMFCC1	57.14	66.66	50	72.72	100	50
IMFCC2	**90**	100	80	54.54	20	83.33
IMFCC3	57.14	66.66	50	63.63	20	100
IMFCC4	42.85	33	50	36.36	40	33.33
IMFCC5	42.85	33	50	36.36	40	33.33

Table 6. The outcome of the MFCC using PC-GITA database

PC-GITA Dataset						
MFCCs	**CNN**			**LSTM**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
MFCC1	70	40	100	80	100	60
MFCC2	70	80	60	70	80	60
MFCC3	50	20	80	80	100	60
MFCC4	60	20	100	50	60	40
MFCC5	50	60	40	50	0	100

Table 7. The outcome of the IMFCC using PC-GITA database

PC-GITA Dataset						
IMFCCs	**CNN**			**LSTM**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
IMFCC1	90	100	80	90	80	100
IMFCC2	**100**	100	100	80	60	100
IMFCC3	70	60	80	80	60	100
IMFCC4	70	60	80	70	60	80
IMFCC5	70	80	60	80	80	80

The results show that performance varies according to the coefficients used (MFCC vs. IMFCC) and the classification models (CNN vs. LSTM). For the Sakar dataset, the IMFCC coefficients with the CNN model (specifically IMFCC2) achieve a high accuracy of 90%, as demonstrated in Tables 4 and 5. For the PC-GITA dataset, using IMFCC coefficients with a CNN model (specifically IMFCC2) yields the best performance, with an accuracy of 100%, as shown in Tables 6 and 7.

It is important to note that the PC-GITA dataset contains 50 audio samples, while the Sakar dataset contains 38. The number of samples can also influence model performance, with a higher number of samples potentially offering better generalization and increased accuracy.

These results suggest that the IMFCC approach combined with CNNs can offer superior performance for detecting PD from voice signals. However, the effectiveness of the models also depends on the specific dataset, as demonstrated by the performance differences between the Sakar and Spanish datasets. Therefore, the most suitable approach may vary, and it is crucial to select the method based on the context and characteristics of the available data.

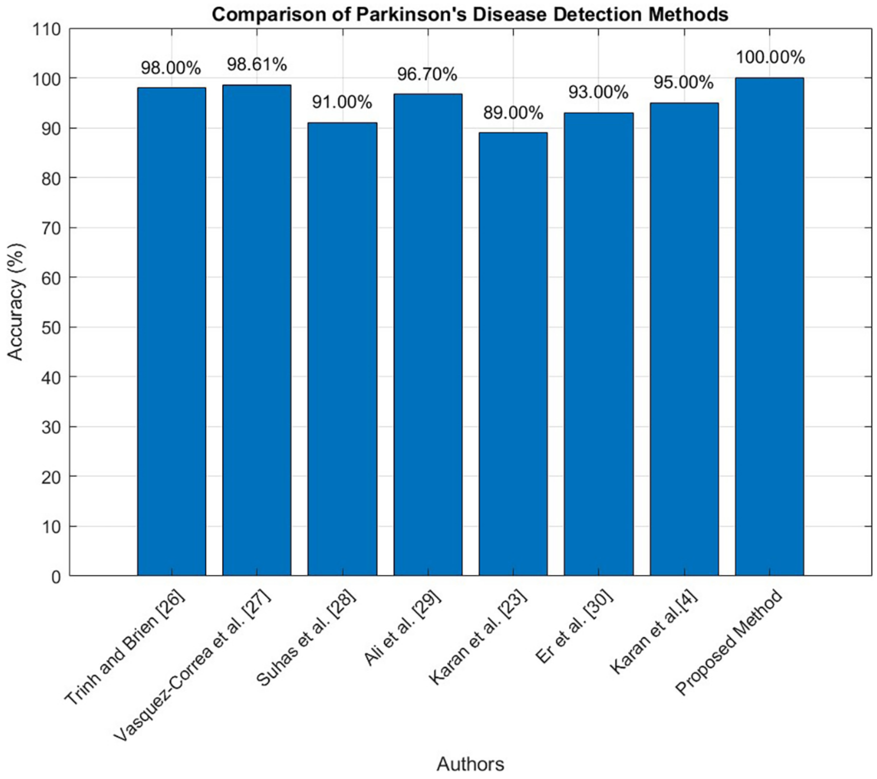

Fig. 4. Comparison between the proposed method and other approaches

In the domain of PD detection, various methods have been explored, as shown in Figure 4, each utilizing specific datasets and techniques. Trinh and Brien focused on the SPDD dataset, employing the CNN approach to achieve an accuracy rate of 96.7% [26]. Vasquez-Correa et al., utilizing a Spanish dataset, applied CNN and reached an accuracy of 89% [27]. Suhas et al. used the NIMHANS dataset, applying CNN and achieving 93% accuracy [28]. Ali et al., worked with the Sakar dataset, employing a neural network method to achieve 95% accuracy [29]. Karan et al., applying EMD-IMFCC approach to represent Parkinson's speech characteristics effectively. These features were evaluated using two datasets: dataset-1 and dataset-2, each comprising 20 individuals with normal speech and 25 individuals affected by PD. An improvement of 98% in accuracy was observed for IMFCC-SVM [23]. Er et al. utilized the PC-Gita database and applied the ResNet 101+LSTM method, achieving an impressive accuracy of 98.61% [3]. Karan et al. also used the PC-Gita database, applying the MLP method with an accuracy of 91% [4].

The results indicate that the proposed approach in this study (VMD-IMFCC-CNN) outperforms other methods in terms of accuracy, achieving a perfect score of 100%. This superior accuracy may be attributed to the specific combination of methods used in this study, indicating its potential efficacy in PD detection, particularly when compared to the results of the other studies are illustrated in Figure 4.

In the context of the results, it is evident that the IMFCC extracted from the VMD exhibits superior performance, and it also outperforms the conventional MFCC. When comparing the results obtained in this paper to those of previous studies, it becomes apparent that VMD-IMFCC yields the best results; this indicates that this approach has the potential to greatly enhance the accurate diagnosis of Parkinson's disease.

This manuscript presents an innovative approach for PD detection using IMFCCs derived from voice signals. The incorporation of VMD for mode extraction followed by the use of CNNs and LSTM networks for classification represents a significant advancement over traditional methods. The use of VMD provides a more precise decomposition of voice signals into intrinsic modes compared to EMD, improving the quality of features extracted and enhancing classification accuracy. By focusing on the first five modes, the study benefits from capturing the most significant components of the voice signal, as indicated by their higher energy levels compared to other modes. This targeted feature extraction enhances the model's ability to differentiate between healthy individuals and those with PD, leading to more reliable diagnostic results. The performance of the proposed method is rigorously evaluated using holdout cross-validation, ensuring that the results are robust and generalizable. However, the study's effectiveness is contingent on the quality and size of the dataset used. While the dataset includes a sufficient number of samples, expanding it to include a more diverse range of subjects could further validate and generalize the findings. Additionally, the combination of VMD, IMFCC extraction, CNNs, and LSTMs introduces a level of complexity that may not be easily implementable in all clinical settings. Simplifying the model or developing a more user-friendly interface could address this limitation. Finally, while focusing on the first five modes is justified by their higher energy, further exploration into the impact of other modes and their potential contributions to detection accuracy would provide a more comprehensive understanding of feature relevance.

6 CONCLUSION

The analysis of vocal signals for detecting PD has been advanced through the application of log-energy IMFCC extracted using VMD. By leveraging data from the PC-GITA and SAKAR databases, this method facilitated the extraction of distinctive and pertinent features crucial for disease detection. Experimental results demonstrated that utilizing IMFCC, specifically IMFCC2, from the PC-GITA database with a CNN classifier achieved a remarkable accuracy of 100% in Parkinson's disease detection. This highlights the effectiveness and potential of this novel approach in extracting vocal characteristics, presenting a promising and complementary alternative to conventional techniques like mel-frequency cepstral coefficients.

Future work could focus on further validating these findings across larger and more diverse datasets, as well as exploring the integration of advanced machine learning techniques to enhance real-time diagnostic capabilities based on voice analysis.

7 REFERENCES

[1] T. Khan, J. Westin, and M. Dougherty, "Cepstral separation difference: A novel approach for speech impairment quantification in Parkinson's disease," *Biocybern. Biomed. Eng.*, vol. 34, no. 1, pp. 25–34, 2014. https://doi.org/10.1016/j.bbe.2013.06.001

[2] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words," *Expert Systems*, vol. 32, no. 6, pp. 688–697, 2015. https://doi.org/10.1111/exsy.12106

[3] M. B. Er, E. Isik, and I. Isik, "Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with variational mode decomposition," *Biomed. Signal Process. Control.*, vol. 70, p. 103006, 2021. https://doi.org/10.1016/j.bspc.2021.103006

[4] B. Karan and S. Sekhar Sahu, "An improved framework for Parkinson's disease prediction using variational mode decomposition-Hilbert spectrum of speech signal," *Biocybern. Biomed. Eng.*, vol. 41, no. 2, pp. 717–732, 2021. https://doi.org/10.1016/j.bbe.2021.04.014

[5] C. O. Sakar *et al.*, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019. https://doi.org/10.1016/j.asoc.2018.10.022

[6] T. B. Drissi, S. Zayrit, B. Nsiri, and A. Ammoummou, "Diagnosis of Parkinson's disease based on wavelet transform and Mel frequency cepstral coefficients," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 3, pp. 125–132, 2019. https://doi.org/10.14569/IJACSA.2019.0100315

[7] B. Nouhaila, B. D. Taoufiq, and N. Benayad, "An intelligent approach based on the combination of the discrete wavelet transform, delta delta MFCC for Parkinson's disease diagnosis," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 4, pp. 562–571, 2022. https://doi.org/10.14569/IJACSA.2022.0130466

[8] Z. Soumaya, B. Drissi Taoufiq, N. Benayad, K. Yunus, and A. Abdelkrim, "The detection of Parkinson disease using the genetic algorithm and SVM classifier," *Applied Acoustics*, vol. 171, p. 107528, 2021. https://doi.org/10.1016/j.apacoust.2020.107528

[9] A. Ouhmida, A. Raihani, B. Cherradi, and O. Terrada, "A novel approach for Parkinson's disease detection based on voice classification and features selection techniques," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 10, pp. 111–130, 2021. https://doi.org/10.3991/ijoe.v17i10.24499

[10] R. Khaskhoussy and Y. Ben Ayed, "Improving Parkinson's disease recognition through voice analysis using deep learning," *Pattern Recognition Letters*, vol. 168, pp. 64–70, 2023. https://doi.org/10.1016/j.patrec.2023.03.011

[11] A. M. Anter, A. W. Mohamed, M. Zhang, and Z. Zhang, "A robust intelligence regression model for monitoring Parkinson's disease based on speech signals," *Future Generation Computer Systems*, vol. 147, pp. 316–327, 2023. https://doi.org/10.1016/j.future.2023.05.012

[12] T. Zhang, L. Lin, and Z. Xue, "A voice feature extraction method based on fractional attribute topology for Parkinson's disease detection," *Expert Systems Application*, vol. 219, p. 119650, 2023. https://doi.org/10.1016/j.eswa.2023.119650

[13] R. Guatelli, V. Aubin, M. Mora, J. Naranjo-Torres, and A. Mora-Olivari, "Detection of Parkinson's disease based on spectrograms of voice recordings and extreme learning machine random weight neural networks," *Engineering Applications of Artificial Intelligence*, vol. 125, p. 106700, 2023. https://doi.org/10.1016/j.engappai.2023.106700

[14] G. Celik and E. Başaran, "Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals," *Applied Acoustics*, vol. 211, p. 109476, 2023. https://doi.org/10.1016/j.apacoust.2023.109476

[15] T. Zhang, L. Lin, J. Tian, Z. Xue, and X. Guo, "Voice feature description of Parkinson's disease based on co-occurrence direction attribute topology," *Eng. Appl. Artif. Intell.*, vol. 122, p. 106097, 2023. https://doi.org/10.1016/j.engappai.2023.106097

[16] M. M. Al-Nawashi, O. M. Al-Hazaimeh, and M. Kh. Khazaaleh, "A new approach for breast cancer detection-based machine learning technique," *Applied Computer Science*, vol. 20, no. 1, pp. 1–16, 2024. https://doi.org/10.35784/acs-2024-01

[17] M. Kh. Khazaaleh *et al.*, "Handling DNA malfunctions by unsupervised machine learning model," *J Pathol Inform*, vol. 14, p. 100340, 2023. https://doi.org/10.1016/j.jpi.2023.100340

[18] O. M. Al-hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-Nawashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 13, pp. 131–151, 2022. https://doi.org/10.3991/ijoe.v18i13.33985

[19] N. Gharaibeh, A. A. Abu-Ein, O. M. Al-hazaimeh, K. M. O. Nahar, W. A. Abu-Ain, and M. M. Al-Nawashi, "Swin transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 4, pp. 22–50, 2023. https://doi.org/10.3991/ijoe.v19i04.37677

[20] B. E. Sakar *et al.*, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013. https://doi.org/10.1109/JBHI.2013.2245674

[21] J. R. Orozco-Arroyave, J. D. Arias-Londõ No, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *LREC.*, 2014, pp. 342–347.

[22] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014. https://doi.org/10.1109/TSP.2013.2288675

[23] B. Karan, S. S. Sahu, and K. Mahto, "Parkinson disease prediction using intrinsic mode function-based features from speech signal," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 249–264, 2020. https://doi.org/10.1016/j.bbe.2019.05.005

[24] N. Boualoulou, M. Miyara, B. Nsiri, and T. B. Drissi, "Voice-based detection of Parkinson's disease using empirical mode decomposition, IMFCC, MFCC, and deep learning," in *Artificial Intelligence, Data Science and Applications, ICAISE 2023. Lecture Notes in Networks and Systems*, Y. Farhaoui, A. Hussain, T. Saba, H. Taherdoost, and A. Verma, Eds., 2024, vol. 838, pp. 144–150, Springer, Cham. https://doi.org/10.1007/978-3-031-48573-2_21

[25] N. Boualoulou, T. Belhoussine Drissi, and B. Nsiri, "CNN and LSTM for the classification of Parkinson's disease based on the GTCC and MFCC," *Applied Computer Science*, vol. 19, no. 2, pp. 1–24, 2023. https://doi.org/10.35784/acs-2023-11

[26] N. H. Trinh and D. O'brien, "Pathological speech classification using a convolutional neural network," in *Irish Machine Vision and Image Processing 2019*, 28–30 Aug 2019, Dublin, Ireland, 2019.

[27] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, “Convolutional neural network to model articulation impairments in patients with Parkinson’s disease,” in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, International Speech Communication Association, Sweden, 2017, pp. 314–318. https://doi.org/10.21437/Interspeech.2017-1078

[28] B. Suhas *et al.*, “Speech task based automatic classification of ALS and Parkinson’s disease and their severity using log Mel spectrograms,” in *2020 International Conference on Signal Processing and Communications (SPCOM)*, 2020, pp. 1–5. https://doi.org/10.1109/SPCOM50965.2020.9179503

[29] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, “Automated detection of Parkinson’s disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019. https://doi.org/10.1109/JTEHM.2019.2940900

8 AUTHORS

Nouhaila Boualoulou graduated in Electronics, Electrotechnics, Automatic, and Industrial Computing from Faculty of Science Ain Chock. University Hassan II – Casablanca, Morocco. She was a research student in Research Laboratory in Industrial and Electrical Engineering, Information Processing, Informatics and Logistics (GEITIIL). Currently she is a Faculty of Science Ain Chok, University Hassan II – Casablanca, Morocco. Her interests are in speech processing for detecting people with neurological disorders (E-mail: boualoulounouha@gmail.com).

Benayad Nsiri holds an MBI degree in computer sciences from Telecom Bretagne (2005), and did Ph.D. degree in signal processing from Telecom Bretagne in 2004. He received D.E.A (French equivalent of a M.Sc. degree) in electronics from Occidental Bretagne University, in 2000. Currently, he is a Full Professor in the National School of Arts and Crafts of Rabat (ENSAM),), Mohammed V University; a member of Research Center STIS, M2CS, Mohammed V University; and a member associate in Researcher, Industrial Engineering, Data Processing and Logistic Laboratory, Hassan II University. He was a Professor in the Faculty of Sciences Ain Chock, Hassan II University. Benayad NSIRI has advised and co-advised more than 15 Ph.D. thesis and contributed to more than 80 articles in regional and international conferences and journals. His research interests include but are not restricted to computer science, telecommunication, signal, and image processing, adaptive techniques, blind deconvolution, MCMC methods, seismic data, and higher-order statistics (E-mail: nsiri2000@yahoo.fr).

Taoufiq Belhoussine Drissi graduated with a Ph.D. degree in acoustics in 2009 at the university of le Havne (France), since 2011, he has been an assistant professor at the sciences faculty of Ain chock University Hassan II, Casablanca. His scientific interest lies in the research of non-destructive testing and signal treatment (E-mail: belhoussine2014@gmail.com).