SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V10I11P102

Volume 10 Issue 11, 11-26, November 2023

Original Article

Optimized Hybrid Model for Enhanced Parkinson's Disease Classification Using Feature Fused Voice Signal

S. Sharanyaa[1], M. Sambath[2]

[1]*Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Tamilnadu, India.*
[2]*Department of Information Technology, Hindustan Institute of Technology and Science, Tamilnadu, India.*

[1]*Corresponding Author : Rnsharanyaa@gmail.com*

Received: 05 September 2023 Revised: 17 October 2023 Accepted: 08 November 2023 Published: 30 November 2023

Abstract - *Parkinson's Disease (PD) is a common neuro disorder that leads to reduced nerve function in the brain as a result of decreased dopamine generation. The disease is progressive, and patients may have difficulty speaking, resulting in speech variations. Hence, it is essential to detect the disease at an early stage, and through proper diagnosis, the effect of Parkinson's Disease can be controlled. This work aims to detect and classify PD based on a vocal feature set using a hybrid CNN-ALSTM model. The model is trained with Spectral, Acoustic, and Mel-Spectrogram features obtained from denoised voice signals. This proposed work involves four phases. In the first phase, voice signals are extracted from the voice input data, and de-noising is done using Improved Optimized Variational Mode Decomposition (IO-VMD). In the second phase, the Mel-Spectrograms are generated from the pre-processed data, where deep features are extracted and trained using Custom CNN, EfficientNetB0, and Inceptionv3 models. In the third phase, a metaheuristic Squirrel Search Water cycle Algorithm (SSWA) is applied to the feature vectors, where SSWA is used for feature selection and hyper parameter tuning. Finally, the spectral and acoustic features extracted from voice signals are concatenated with the mel spectrogram feature vectors, trained, and classified using the Attention based Long Short Term Memory (ALSTM) model. A comparative analysis of models like CNN-ALSTM, Inceptionv3-ALSTM, and EfficientNetB0-ALSTM is carried out to classify PD. From the result analysis, the SSWA algorithm with a proposed hybrid EfficientNetB0-ALSTM model achieves an accuracy of 96.8% and performs better than the other models.*

Keywords - *Neural network, Optimization algorithm, Spectrogram, Transfer learning, Voice signal.*

1. Introduction

PD is a weakening nerve de-generative sickness. Even though the reason for the sickness is yet to be identified, it is recognized by the slow decline of dopamine-carrying neurons in the affected person's brain. The early diagnosis of PD with minimal cost will help in the treatment of those already suffering and those at risk of being affected by it. The statistics also reveal that PD is diagnosed more frequently in men compared to women [1].

The current state of PD treatment consists of symptoms that can be reduced or made inactive based on medication. Motor and non-motor symptoms are the two different kinds of PD symptoms. Generally, motor symptoms involve difficulty in muscle movement, resulting in severe movement problems, namely bradykinesia, firmness, trembling, and cognitive impairment. In contrast, non-motor symptoms result in sleeping disorders, weakness, dysphagia, depression, etc [2].

Though there are various symptoms categorized, the primary sign of the disease is speech difficulty, which occurs in 90% of people [3] with variations in the jitter feature, shimmer feature, and other baseline acoustic characteristics in the voice. These variations lead to hoarseness, roughness, and strain in the voice. Speech impairment occurs due to a lack of muscle coordination in the vocal cord. It is widely accepted that changes in acoustic elements in a person's voice greatly reflect PD progression, which makes many researchers work on voice processing and analysis to develop an automated data-driven framework for PD detection [4].

Speech impairment is evaluated at present primarily through clinician examinations and conducting questionnaire sessions with patients and caretakers as well concerning the patient's nature of work, day-to-day routine, hobbies, sleeping time and other difficulties and to rate them based on the Unified PD Rating Scale (UPDRS) ranging between 0-4. Though clinicians use this method for identifying the disease in most cases, it has some limitations, including the differences in disease progression that could be seen between two patients with the same score [5].

Detecting and managing neurodegenerative diseases in their early stages is crucial for improving quality of life and

slowing disease progression. Though it is true that many neurodegenerative disorders currently have no cure, an early diagnosis and intervention can help control symptoms, delay progression, and provide patients with better care and support. Diagnosis allows healthcare providers to initiate appropriate treatment and interventions at an earlier stage of the disease. This can help control the motor symptoms of PD more effectively and delay the onset of complications [6].

In the current era, AI models perform better at diagnosing Parkinson's Disease, which helps clinicians with further treatment. Convolutional Neural Network (CNN) models help diagnose PD by training the image data, mostly gait images, brain MRI and CT scan images, and extracting the features at higher level layers. CNN improves the feature learning ability of the model and detects the disease accurately in most cases [7]. PD detection using Recurrent Neural Network (RNN) is familiar using time-frequency analysis with a deep learning shrinkage network and EEG signals. The model yields a better classification if more information is present in the data [8].

Recent studies show that feature selection techniques combined with the deep learning model enhance the accuracy of PD detection, including minimum redundancy, maximum relevance, genetic algorithm [9, 13], multiple fine-tuning approaches with CNN and voice features, mel-frequency spectral coefficients, Principal Component Analysis [10, 11], etc. are used with deep neural networks for better performance. This work proposes hybrid deep learning models with the IO-VMD and SSWA optimization algorithms. The main contributions of the current study are:

- Pre-processing the input voice signals using the IO-VMD algorithm
- Generation of Mel-Spectrogram images from pre-processed data.
- Deep feature extraction using the Custom CNN, Inceptionv3, and EfficientNetB0 models.
- Use the Squirrel Search Water cycle optimization Algorithm (SSWA) to select the optimal features and fine-tune the hyper parameters.
- Integration of optimized features with extracted baseline acoustic and spectral features to make a single hybrid feature vector and fed to the ALSTM model for PD classification.
- A comparative analysis of hybrid models CNN-ALSTM, Inceptionv3-ALSTM, and EfficientNetB0-ALSTM, along with SSWA, is carried out to increase the classification performance.

This research work is prepared as follows: In section 2, associated works are discussed, and in section 3, the dataset used and workflow methodology are given. In section 4, implementation results with discussions are provided. The conclusion part and the future scope are given in section 5.

2. Related Works

Voice data has indeed gained attention in diagnosing Parkinson's Disease (PD) because vocal impairment is a common and often early symptom in some patients. Early detection of PD can be challenging, as the symptoms can be subtle and easily missed during routine clinical evaluations. Specifically, the latest research indicates that speech abnormalities can appear up to ten years before the symptoms of a cardinal motor deficit [12]. The statistical parameters are calculated via a Support Vector Machine (SVM) with a genetic algorithm and a discrete wavelet transform. Statistical methods can be used to determine the pattern difference between Parkinson's sufferers' speech patterns and those of normal individuals [13].

In a recent study [14], a model-driven and data-driven approach was developed as a decision-making process for pathological voice and addressed gender difference issues. Using the Chisquare L1 normalization SVM algorithm, a multiple-level feature selection strategy is employed to identify PD from features, including voice recordings. KNN, SVM, and DT were used for PD classification, with KNN yielding the potential outcome [15].

An end-to-end strategy employing a CNN-multi layer perceptron neural network achieves 68.6% accuracy, whereas a feature-derived model using SVM achieved 67.9% accuracy using the PC-Gita speech dataset [16]. The authors in [17] used the Intrinsic Mode Function Cepstral Coefficient feature (IMFCC) for detecting PD from speech signals but with minimal dataset samples of 25 PD and 20 healthy controls. In [13], discrete wavelet transform is applied by an evolutionary genetic algorithm with a support vector machine to detect PD. However, the classification performance is not satisfactory.

The authors in [18] have developed an automated diagnosis of PD by using Bidirectional LSTM and Wavelet Scattering Transform to identify speech disorders in patients having central nervous issues. Comparable research was done using spectrogram-derived deep features [11, 18].

The study [19] used a variety of machine learning classifiers to analyze mPower data with shimmer and glottal quotients features for hand-held voice analysis for Parkinson's Disease (PD) detection, achieving 71% accuracy. The study [20] uses MDVP audio data in a random forest, KNN, and logistic regression models. A random forest model with relatively little (30 PD data) achieved 91.8% of the classification accuracy.

Voice analysis and machine learning models can potentially provide a non-invasive and objective method for identifying early signs of PD. Apart from speech signals, many research studies were carried out using gait symptoms, handwriting, circle drawing images, and memory games for

PD detection [21]. Early diagnosis is a crucial step in effectively managing Parkinson's Disease. It allows for timely intervention, personalized treatment plans, and proactive measures to balance symptom control with preventing complications like levodopa-induced dyskinesias [22]. However, most published works use very limited datasets and collections of features (often less than 50 instances), and no statistical comparison has been made over a more extensive range of illness phases, medication information, or impairment severity. Table 1 gives the recent literature studies on the detection of PD with the methodology used, dataset details and accuracy obtained.

Table 1. Recent literature works for Parkinson's Disease classification based on voice features

Objective	Methodology Used	Dataset Used	Model Performance
PD detection with an ensemble approach [1]	Ensemble model with Optimal Features and Sample dependant base Classifiers (EOFSC) with DNN	Speech, voice and vowel phonation data - 20 disease and 20 Healthy from the UCI Repository	Accuracy 87.5%
Ensemble classification using voice samples [4]	Sequential backward selection with wrapper method for feature selection and ensemble classifiers	UCI archives with 32 samples (23 PD and 8 healthy)	Accuracy 90.0%
Augmentation of PD voice data [23]	Deep CNN model - Xception is used for training and to evaluate the performance, Xception model is used	PC-GITA dataset is used	Accuracy 82.12% for word set and 92.3 for vowel /a/
PD detection using optimization algorithm [24]	Chronological Smart Sun Flower Optimization Algorithm (CSSFOA) for feature selection with ZF net model	mPower voice data	Accuracy 94.5%
ML vs DL comparison for PD [25]	CNN, KNN, SVM and Naïve Bayes	266 healthy and 160 PD from Italian native speakers	82.2% accuracy
PD based on smartphone data [26]	RBF kernel SVM with MFCC	mPower voice data (1000 samples)	90% accuracy
PD detection using time series data [27]	Time series features with 1D and 2D – Convolutional Neural Network	i. Gyenno Science Research Center ii. PC-PITA dataset	81.6% and 92.0% accuracy
PD using multi-class data [28]	Genetic Algorithm with Naïve Bayes and MLP	111 healthy and 51 PD	95% accuracy
Dysphonic voice in PD diagnosis [29]	Boruta wrapper technique for feature selection and classification using various machine learning algorithms, including KNN, DT, RF, LDA, XGBoost, etc.	176 Healthy and 178 PD samples from Italian, Spanish, Czech language	91% accuracy on average
PD detection using semi-supervised competitive learning [30]	Correlation pattern analysis using Pearson's correlation coefficients and PCA with KNN and SVM classifier	UCI machine learning repository with 80 subjects PD 40, Healthy 40	83.8% accuracy by SSCL method.

2.1. Challenges

The existing literature shows the performance of various models in detecting Parkinson's Disease. Even though models of neural networks can extract features at higher layers, very few studies incorporate optimization strategies to pick relevant components for deep neural network training, and only a restricted number of attributes are studied. Hand-crafted features were most prominently used for PD diagnosis using traditional ML algorithms. But did not provide a satisfactory outcome. Moreover, processing voice samples directly into a neural network is not effective. Because one second of audio, when processed at 44.1KHZ sampling rate, would give 44100 sample points (features), which is enormous for the neural network to handle. So, the proposed study shows a better and more prominent feature selection method using the meta-heuristic SSWA optimization algorithm for enhanced classification accuracy of PD in the hybrid CNN-ALSTM Model.

3. Methodology

The proposed work depicts PD detection and classification using voice input data. The input signals are pre-processed using Improved Optimized-VMD and filtered to eliminate the noise in the data. From the denoised voice signals, Mel-Spectrogram images are extracted for further processing. The images are trained using a custom CNN model [31], Inceptionv3 [32], and EfficientNetB0 [33] models, followed by hyperparameter tuning using a novel SSWA optimization algorithm. From the initial population, the fitness function is computed based on squirrel migration and the feature vectors are generated until stop criteria are reached. The highly correlated features are extracted from the spectrogram images. The temporal features are extracted from the optimization result, and baseline acoustic and spectral features where the embedded vectors are generated. Highly correlated components are selected for reconstructing the original signal, and the proposed architecture is given in Figure 1.

3.1. Dataset Description

In the proposed work, the mPower dataset [34] obtained from the synapse is used along with a demographic survey. The dataset contains audio recording samples of healthy subjects and disease-affected individuals, in which 12,300 samples of healthy controls and 12,100 PD voice recordings were considered for this study for training from 65022 samples. These samples are fine-tuned based on the demographic survey, as shown in Figure 2. The mPower dataset requires participants to record a sustained phonation by pronouncing /aa/ into the iPhone microphone for 10 seconds. The recordings of voice are collected based on the medicine taking time, like right after or before taking Parkinson's medicine, I don't use Parkinson's medication, and it has no value.

3.2. Data Pre-Processing

Each audio sample is 10 seconds long. The signals are converted into frames and stored as frames for feature extraction, and the frame length is chosen as 30 ms. 40% over-binning is fixed for smooth transitions.

$$P = \{p_1, p_2, \ldots., p_i, \ldots. p_j\} \quad (1)$$

The sequence of voice signals is represented as P, as given in Equation 1, where it denotes the i^{th} voice signal and j denotes the total count of input voice signals. Pre-processing and de-noising using IO-VMD are performed to lessen the noise in the input speech signals.

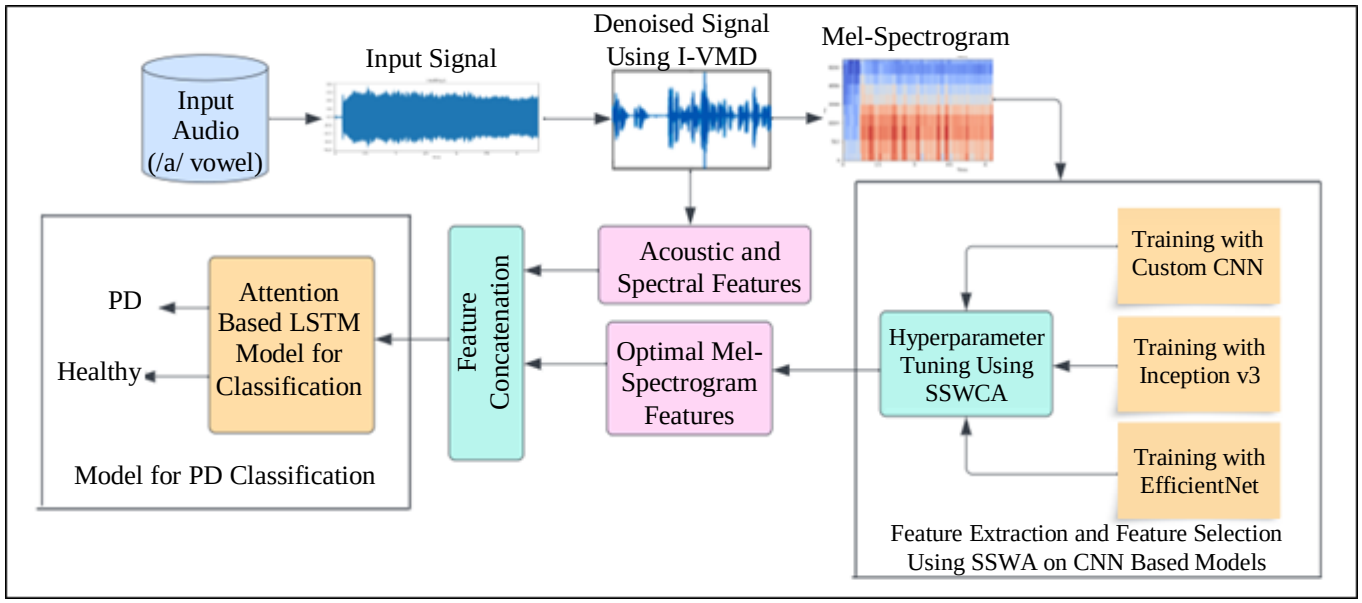

Fig. 1 Architecture diagram of the proposed hybrid CNN-ALSTM model for PD classification

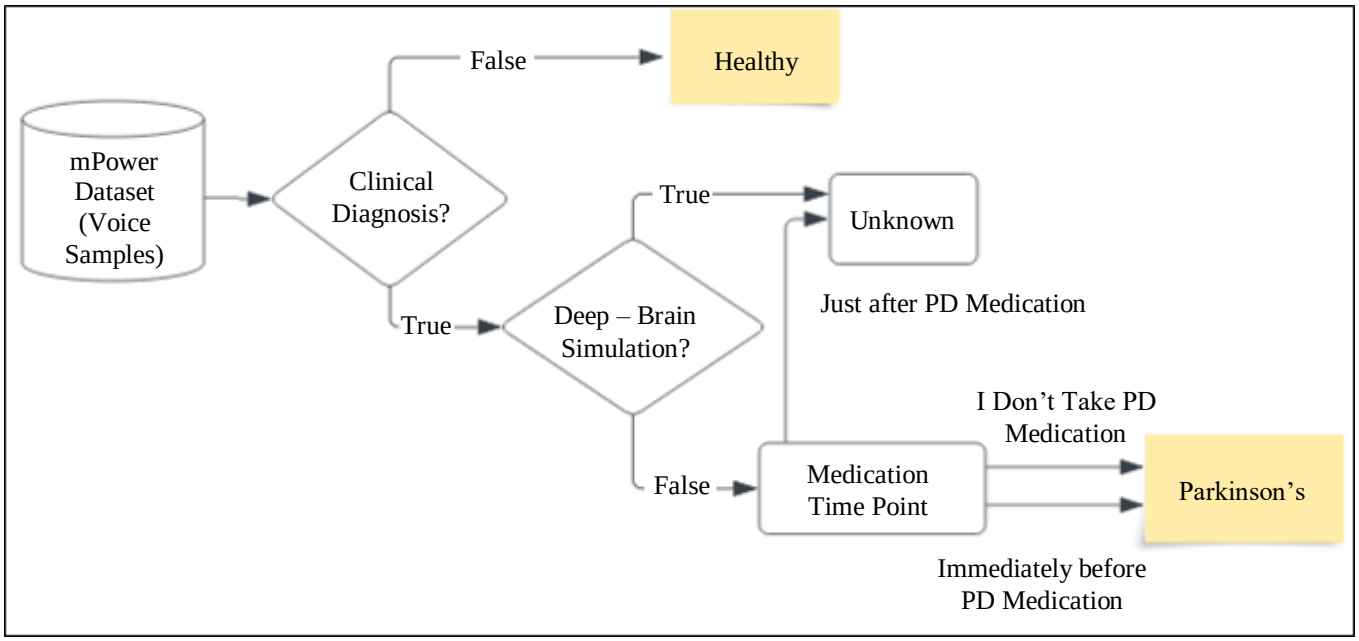

Fig. 2 Fine-tuned mPower voice dataset

3.3. Improved Optimized Variational Mode Decomposition (IO-VMD) for De-Noising

Variational Mode Decomposition (VMD) is a pre-processing method used in signal processing applications to decompose the non-stationary signals [35], F(t), into discrete modes or sub-signals NK(t). Each method of the signal NK(t) is localized on a central frequency WK(t), and the result of applying the Lagrangian multiplier for converting to non-constraint optimization is,

$$L\{NK, WK, \lambda\} = \alpha \sum_{K=1}^{k} ||\partial t \left[\left(\partial t + \frac{j}{\pi t}\right) * NK(t)\right] e^{-jWK}|| + F(t) - \sum_{n=1}^{K} NK(t)|| + \langle \lambda(t), F(t) - \sum NK(t)\rangle \quad (2)$$

In Equation 2, λ is the Lagrangian multiplier, ∂ is the time derivative, *K ->{1, 2,..., k}* represents the number of IMFs and * denotes the convolution operation. The noisy signals are converted to Intrinsic Mode Functions (IMFs) to decrease the low-frequency and high-frequency noise signals.

The issues revolve around tuning the hyperparameters, precisely the parameters K and penalty α, and the time-consuming process of identifying their optimal values. In an Improved Optimized Variational Mode Decomposition (IO-VMD), the Shannon entropy measure of the signal is used to determine these parameters automatically.

The low Signal-to-Noise Ratio (SNR) tends to exhibit higher entropy, which implies that a minimal set of modes may suffice due to the masking effect of noise on the discriminative features of the signal. Consequently, the number of modes in IO-VMD is inversely related to the signal's entropy. α parameter can be identified using the entropy measure, considering the distortion effects on the frequency characteristics of the signal.

Integrating the Shannon entropy measure into the pre-processing step to automatically determine parameters could streamline the optimization process and mitigate the computational burden associated with parameter tuning.

3.3.1. Algorithm for Improved Optimized Variational Mode Decomposition

Input: f -> signal to be analyzed,
n -> number of modes, e -> level of tolerance
Output: K, $\propto$

1. Initialize: $\propto, K_{min}, K_{max}$

   Let $K = K_{min}$. Set $K_{min} = 1$ and $K_{max} = 15$
2. Compute the energy loss coefficient denoted by e.

   $$e = \|f - \sum u_k\| / \|f\|$$

   Where, u_k is the Kth mode function
3. If $e < \mu_1$, calculate Shannon entropy of every mode μ_k

   $$K = \lfloor K_{max} - (S(f_i(t)/S_{max}) / (K_{max}) + K_{min} \rfloor \quad (3)$$

   $$\propto = (|((S(f_i(t) - S_{min})/ (s_{max} - s_{min})|) * \propto_s) + \propto_b \quad (4)$$

   $S(f_i(t))$ is the Shannon entropy and S_{min} and S_{max} are the smallest and highest Shannon entropy values in the dataset. $\propto_s$ represents the step value and $\propto_b$ denotes the base value, which is set as 1000.
4. Else, iterate K= K+1.
5. To terminate the execution, check the condition given in Equation 5,

$$\frac{\sum_k \|\mu_k^{n+1} - \mu_k^n\|}{\|\mu_k^n\|} < e \quad (5)$$

Till the above condition is satisfied, iterate the process. The optimal output is obtained using IO-VMD.

6. Return μ_k of each mode, K and penalty parameter$\propto$. IO-VMD estimated the optimal K value as 12 and $\propto =$ 10000. The maximum value of $\propto$ is 20000.

3.4. Feature Extraction

Baseline acoustic and spectral features are extracted using the librosa audio library and implemented in Python. The tool's versatility is beneficial for both developing the proposed model and managing the essential characteristics of audio signals for the classification of PD. Different types of vocal feature sets are extracted for the proposed work from mPower dataset samples.

3.4.1. Generation of Mel Spectrogram Features

Mel-Spectrograms are well suited for audio analysis. It is a 2D representation of an audio signal over frequency and time [7]. Spectrogram images can be deeply analyzed using the CNN model's convolutional layers, pooling layers, and other architectural elements. It is obtained by computing a Fast Fourier Transform on each window to move from the time to the frequency domain, generating the signal's power spectrum. Then, a mel filter bank is applied to simulate human auditory perception.

For many voice recognition applications, the Mel frequency is used, which scales up to 1 kHz and increments logarithmically for greater values. It is the rate of the tone chosen by the human ear. A colour map is used to depict each frame's power and energy spectrum, with the intensity of the colours denoting the signal's higher energies. Table 2 shows the parameter settings for generating spectrograms from the speech input signals. Figure 3 gives the process of developing the Mel-Spectrogram.

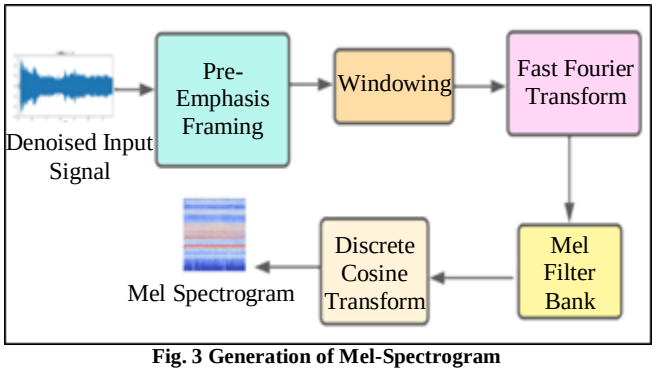

Fig. 3 Generation of Mel-Spectrogram

Table 2. Parameter settings for generating Mel-Spectrogram

Parameter Used	Value
Sampling rate	44.1kHz
Frame length	30ms
Window length	2048
Frame shift	11.61ms

3.4.2. Spectral and Acoustic Features

Table 3 describes extracted spectral and acoustic features. Five jitter variants, five shimmer variants, two mean harmonic parameters, six spectral characteristics, and 39 MFCC coefficients and 39 GTCC coefficients (0 order, 1st order, and second order delta co-efficient values with energy coefficient factor) are computed from voice signals using the librosa python package and voice box toolkit.

Table 3. Extracted baseline acoustic & spectral features

Parameter	Features	Description
Frequency	LocalJitter, local absolute jitter, RapJitter, ppq5Jitter, ddpJitter	Variations in cycle-to-cycle duration with respect to frequency
Amplitude	Local shimmer, local dB shimmer, apq3Shimmer, apq5Shimmer, ddaShimmer	Variations in cycle to cycle duration with respect to amplitude
Harmonicity	Noise to harmonic, harmonic to noise	Expressed in dB. It indicates the harmonics and noise present in the signal
Spectral features	Spectral flux, spectral kurtosis, spectral centroid, spectral spread, tonal power ratio, fluctuation index	Frequency domain representation indicates the frequency distribution of voice signal performed using Fourier analysis
Mel Frequency Cepstral Coefficients and Gammatone Cepstral Coefficients	Zero order, first order and second order derivatives	MFCC coefficients and GTCC coefficients gives the spectral characteristics of voice signal

3.5. SSWA Optimizer for Feature Selection Process

SSWA is a fusion of SSA [36] and WCA [37] meta-heuristic optimization algorithms. WCA addresses various optimization issues and provides optimal solutions. WCA is capable of solving combinatorial optimization problems with less computational effort.

The SSWA algorithm is devised by employing the characteristics of the food nature of squirrels and the flow of water through the river and streams towards the sea. It is more robust and scalable for solving real-time optimization problems and provides global optimal solutions.

$$Q = \{Q_1, Q_2, \ldots\ldots, Q_i, \ldots\ldots, Q_j, 1 \le i \le j\} \quad (6)$$

In Equation 6, j represents the total solution and Q_i shows the ith solution.

3.5.1. Computing Fitness Function

It is computed using the Canberra distance metric for finding out the optimal parameters and is defined as,

$$C_{nb} = \sum_{i=1}^{n} \frac{\left|F_i^j - F_i^{j+1}\right|}{\left|F_i^j\right| - \left|F_i^{j+1}\right|} \quad (7)$$

In Equation 7, F_i^j and F_i^{j+1} denote the next consecutive features.

3.5.2. Finding New Update Solutions

The new positions of the stream and river populations are updated using WCA, and the update equation is given in Equation 8.

$$Q_{stream}^{j+1} = Q_{stream}^{j} + rand\ X\ l\ X\ (Q_{sea}^{j} - Q_{stream}^{j}) \quad (8)$$

Where, l denotes the cost estimate function, r and indicates the random distribution between [0,1], Q_{sea}^{a} shows the current position of the sea population and Q_{stream}^{a} denotes the exact position of the stream population.

The new update position is computed in Equation 9 as,

$$Q_{river}^{j+1} = Q_{river}^{j} + rand\ X\ l\ X\ \left(Q_{sea}^{j} - Q_{river}^{j}\right) \quad (9)$$

The new updated position of SSWA is given as,

$$Q_{stream}^{new} = \frac{a_c B_d}{a_c B_d - 1}\left[\sqrt{\mu} rand\ f(1, E_{var}) - \frac{Q^j(1 - a_c B_d)}{a_c B_d}\right] (10)$$

In Equation 10, μ denotes the search position coefficient, E_{var} denotes design variables, rand f shows the normal distribution function that generates random numbers. a_c shows random gliding distance and B_d random number with a default value of 0.9 and Q_{stream}^{new} denotes the position of the new stream. By integrating the optimal solution of the SSA and WCA algorithm, the new update equation of SSWA is obtained. For each solution update, the fitness function is calculated, and after reaching maximum iterations, the best solution has arrived, and the optimal features of PD are extracted.

3.6. Deep Feature Extraction Using CNN Based Architectures

This study uses the architectures of custom CNN, adaptive transfer learning Inceptionv3 and EfficientNetB0 CNN models for deep feature extraction.

3.6.1. Custom Convolutional Neural Networks (CNN)

The custom CNN model uses five convolutional layers, three Fully Connected (FC) layers, and four max pooling layers. The Convolutional layer extracts highly intense and prominent regions from the input. The kernel size is set as three, and the padding value is one, followed by a pooling layer for downsampling. The max pooling layer has a stride value of 2 and a learning rate 0.0001. In contrast, the dropout layer 0.5 enhances model accuracy by eliminating local optimum solutions with the number of epochs to 50. The convolution process is given in Equation 11. C represents the output of the convolution operation. P is an image, and K shows kernel size. a, b denotes row and column matrix.

$$C[a,b] = (P * K)[a,b] = \sum_i \sum_j K[i,j](P)[a-i, b-j] \quad (11)$$

3.6.2. Inceptionv3 Model

Inception-V3 is a commonly used model for image classification tasks. It consists of symmetric and asymmetric components encompassing layers like convolution, average pooling, max pooling, dropout layers, concatenation layers and Fully Connected layers. A batch normalization process is used in the model, which improves the accuracy. The model includes the functionality of label smoothing convolutions, which are factorized and classifiers capable of propagating the label's information down to all the layers of the network structure.

3.6.3. EfficientNet Model

EfficientNet is a Convolutional Neural Network architecture that uniformly scales all dimensions using depth, resolution, and other metrics. EfficientNetB0 is the baseline architecture which uses the fixed scaling coefficients. The resolution, width, depth, and other parameters are scaled down using compound coefficients.

The receptive field and the number of channels are increased depending on the size of the input image. Here, the input image is resized to 224*224*3. More fine-grained patterns from the data are extracted well, and the network is tuned for good accuracy. The architecture blocks of the Efficient Net model for the variant B0 are given in Figure 4.

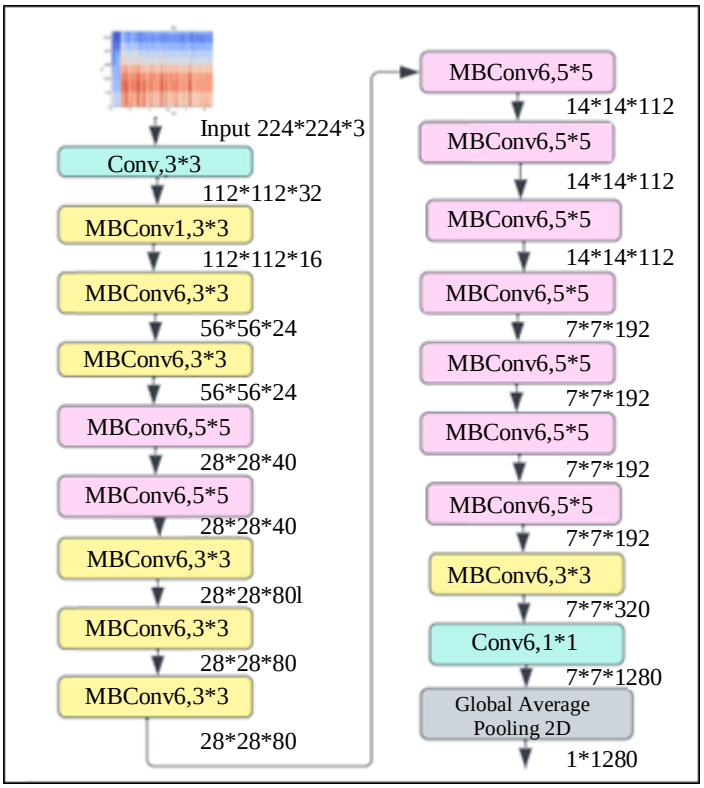

Fig. 4 Architecture blocks of EfficientNet-B0 model for feature extraction

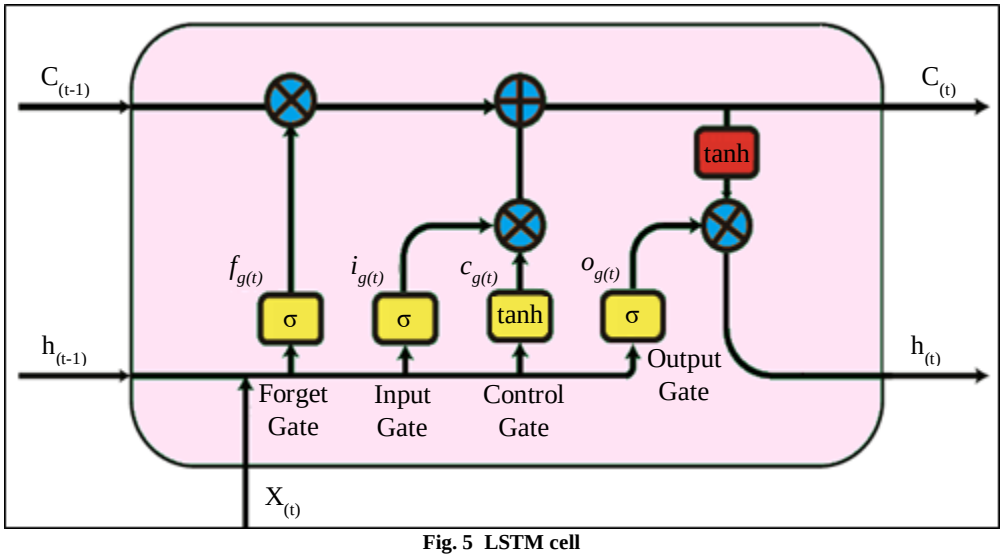

Fig. 5 LSTM cell

3.7. Attention-Based LSTM Model (ALSTM) for PD Classification

LSTMs are well-suited for processing data sequences and are commonly used in tasks involving time series or sequential data [38]. The main reason for using the LSTM model is that it can learn long-range dependencies. But could not hold or retain for a more extended period of time.

Hence, along with LSTM layers, an attention mechanism is added with attention weights to capture the most prominent and best informative feature from the input sequence, thereby improving accuracy. In the proposed work, CNN features are fed to the attention-based LSTM model, and it is made of an input layer with several units representing the total features in the input data, two LSTM layers with variable LSTM units that carry input from the first layer and produce the target value and an attention layer. Input Gate ig(t), Output Gate Og(t), Forgot Gate fg(t), and Memory gate constitute each LSTM unit and are given in Figure 5. LSTM model structure for classifying PD is shown in Figure 6.

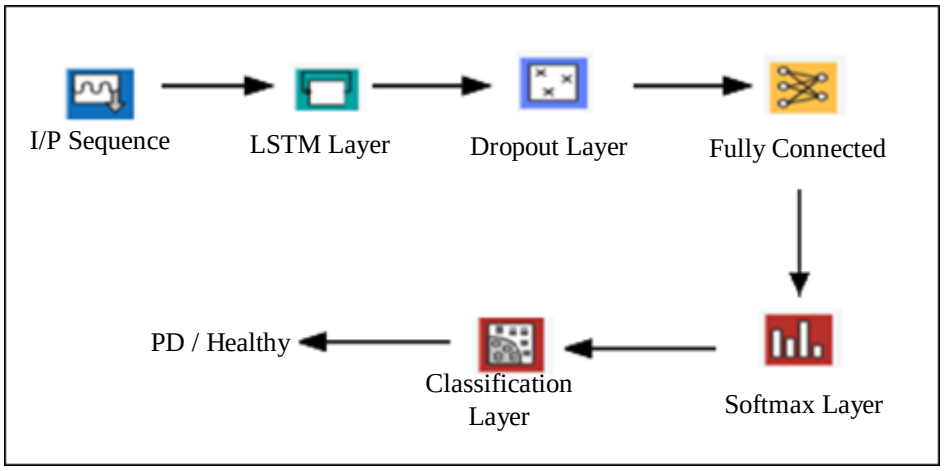

Fig. 6 LSTM model structure

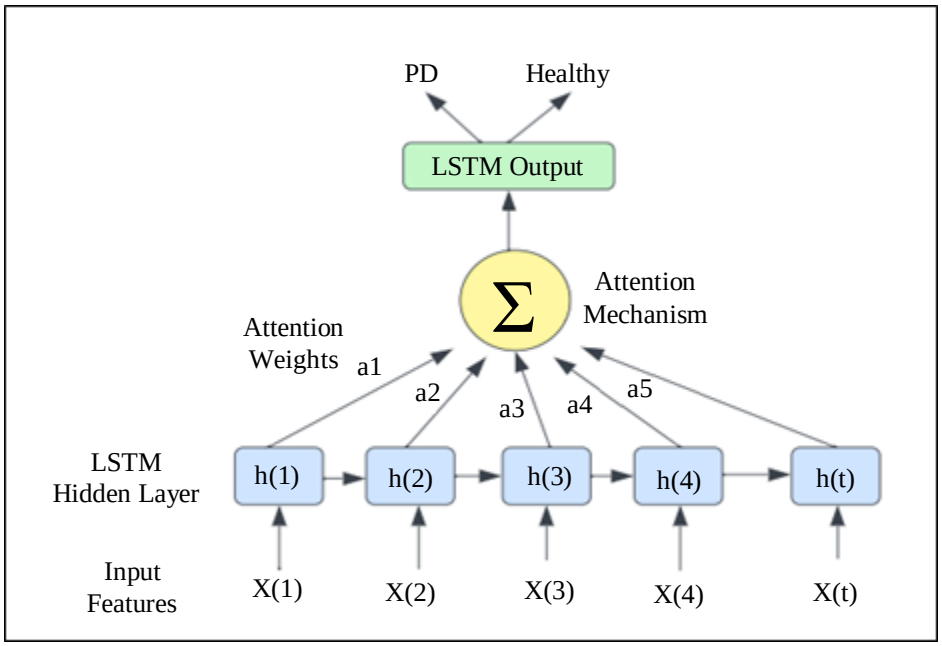

Fig. 7 Architecture of attention-based LSTM for PD classification

Forget gate gives the cell's needs information and is presented in Equation 12. $X_{(t)}$ is the current input to the ALSTM cell, $h_{(t-1)}$ is the previous output, σ denotes the sigmoid function, W represents weight inputs, and b is the bias value. The sigmoid layer is known as the input gate layer, and it determines whether or not the value has to be updated (0 or 1), as given in Equation 13. The tanh layer generates a vector and transforms them between -1 and 1, as given by $C_{(t)}$ in Equation 14.

Equation 15 gives the output gate, where the sigmoid function decides which part of information passes to output, and this output is multiplied with $C_{(t)}$ from tanh function. Network parameters such as hidden units, batch size, and epochs were fine-tuned for better results. The dropout layer was also utilized to keep the model from being overfit.

$$f_{g(t)} = \sigma\left(W_f[h_{(t-1)}, X_{(t)}] + b_f\right) \quad (12)$$

$$i_{g(t)} = \sigma\left(W_i[h_{(t-1)}, X_{(t)}] + b_i\right) \quad (13)$$

$$c_{(t)} = \tanh(W_c[h_{(t-1)}, X_{(t)}] + b_c) \quad (14)$$

$$o_{g(t)} = \sigma(W_o[h_{(t-1)}, X_{(t)}] + b_0) \quad (15)$$

$$h_{(t)} = o_t . \tanh\left(C_{(t)}\right) \quad (16)$$

Attention-based LSTM adds an attention layer with weights to the traditional LSTM model. The architecture of

Attention-based LSTM for PD classification is given in Figure 7. At every time step t, the hidden layer is denoted as shown in Equation 17,

$$h_{(t)} = LSTM\left(h_{(t-1)}, X_{(t)}\right) \tag{17}$$

The weighted sum of encoded hidden vectors for the attention layer is computed by,

$$Wa = \tanh(W_{hl} y_{(t)} + b_v) \tag{18}$$

In Equation 18, W_{hl} represents the weight vectors among hidden layers.

Where, $y_{(t)} = \sum_{t=1}^{n} a_{(t)}\ h_{(t)}$

Here, a(t) represents the attention weights, $h_{(t)}$ denotes the hidden layer, and n represents the size of the input vectors. The attention weight $a_{(t)}$ is given in Equation 19 as,

$$a_{(t)} = \frac{\exp(q_{(t)})}{\sum_{t=1}^{n} \exp(q_{(t)})} \tag{19}$$

Where, $q_{(t)} = \tanh(W_A h_{(t)} + b_q)$

Here, W_A is the parameter matrix, b_q is the bias value. Softmax function is used in the output layer and given in Equation 20, where e^{Xi} is the exponential function of the input feature vector and e^{Xj} is the exponential function of the output feature vector. FC layers are connected to the softmax layer attached to the classification layer with two output classes, PD or healthy.

$$\text{SoftMax(Xi)} = \frac{e^{Xi}}{\sum_{j=1}^{K} e^{Xj}} \tag{20}$$

4. Results and Discussion

4.1. Implementation Results

Figures 8(a) and 8(b) show the waveforms of healthy and PD samples. The Mel-Spectrogram images are generated by applying the mel scale from the denoised signals after pre-processing and are given in Figure 9. The generated Mel-Spectrogram image of healthy and PD samples is given in Figures 10(a) and 10(b). The custom CNN-ALSTM model obtained a classification accuracy of 91.1% and a validation accuracy of 86.32%. As the epoch increases, the loss value decreases during training and is given in Figure 11. The Inceptionv3- ALSTM model obtained a classification accuracy of 94.3% and a validation accuracy of 88.68% with a learning rate of 0.1. The number of iterations is 48; 6 iterations per epoch is used, and the number of epoch is set as 8.

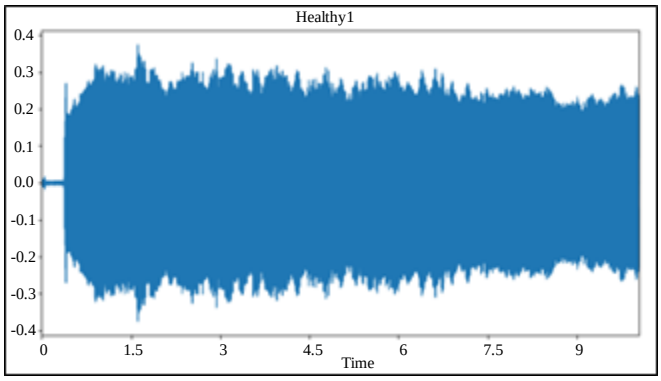

Fig. 8(a) Waveforms - healthy sample

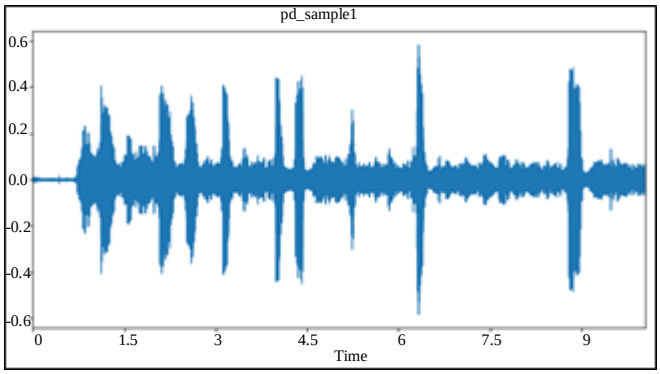

Fig. 8(b) Waveforms – PD sample

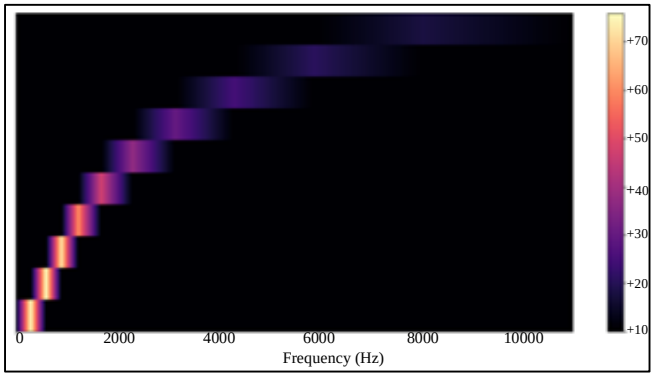

Fig. 9 MEL scale

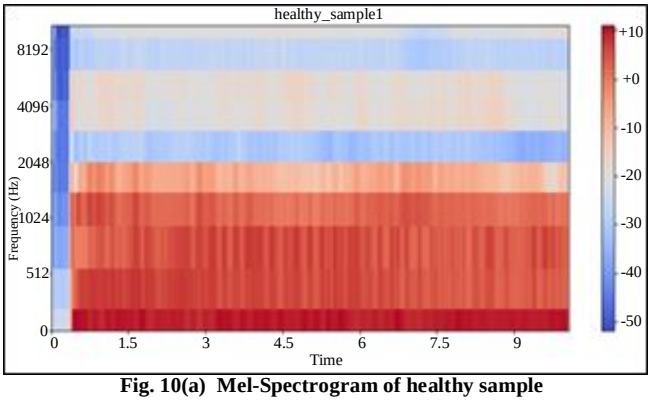

Fig. 10(a) Mel-Spectrogram of healthy sample

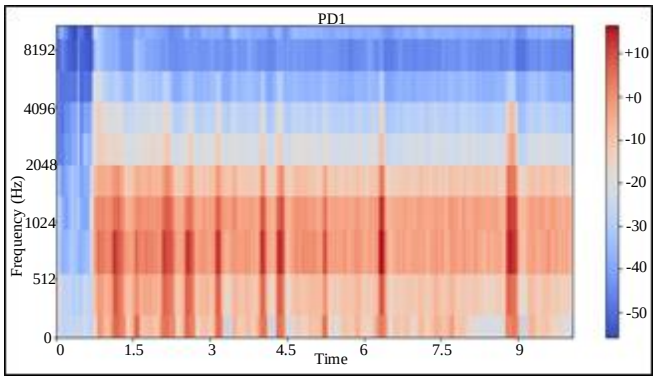

Fig. 10(b) Mel-Spectrogram of PD sample

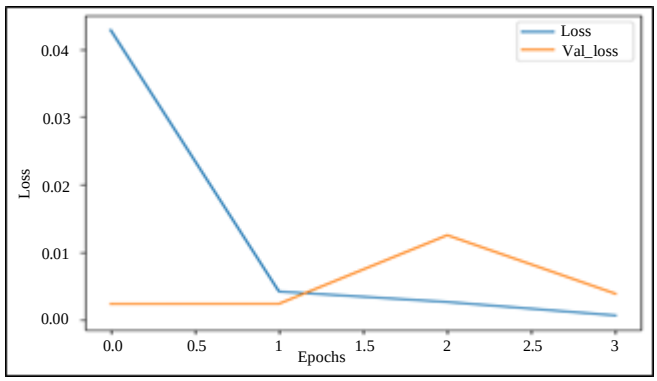

Fig. 11 Loss value of custom CNN model

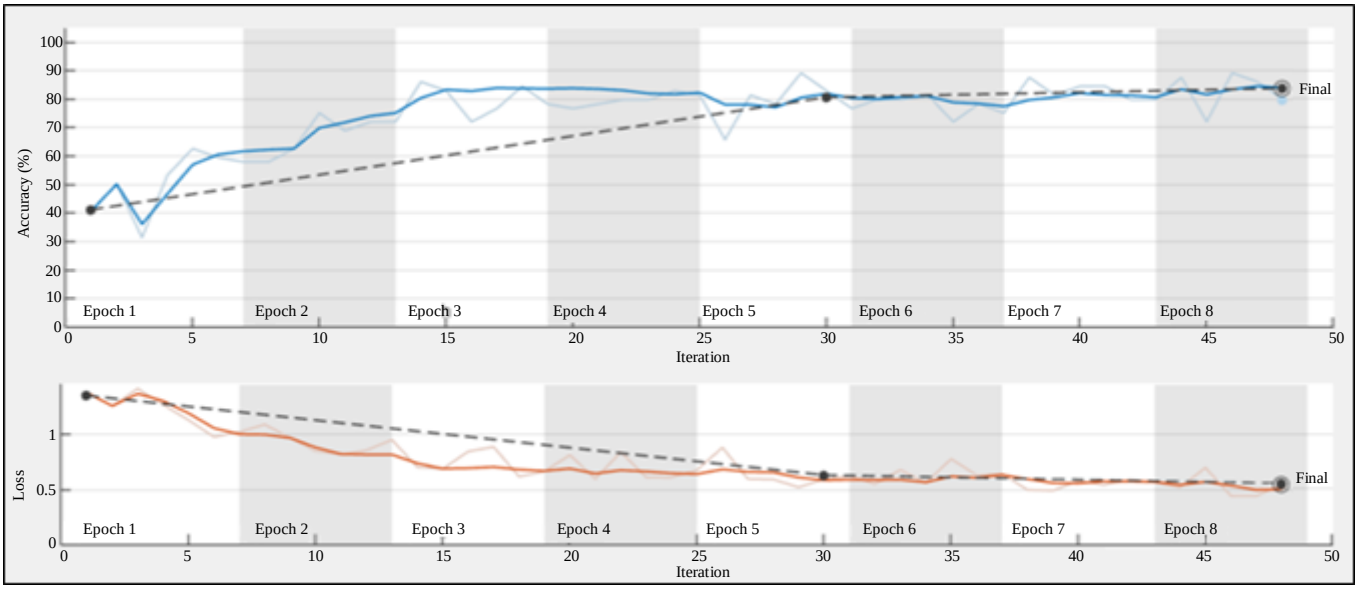

Fig. 12 Accuracy, loss graph of Inceptionv3-ALSTM hybrid model

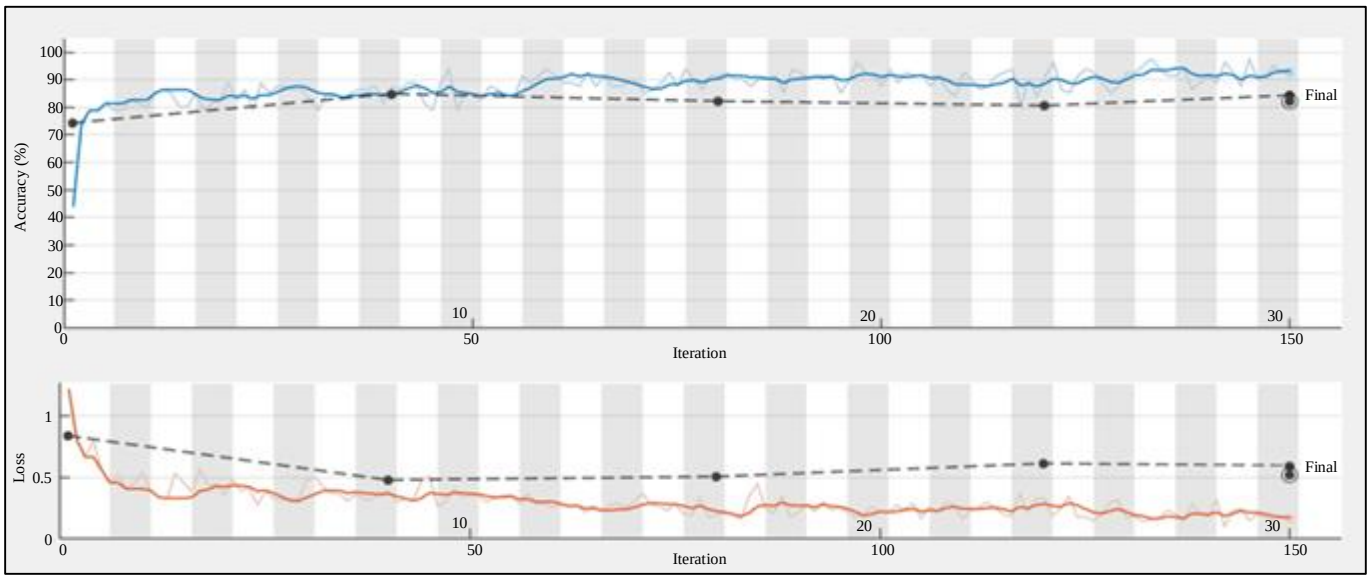

Fig. 13 Accuracy, loss graph of EfficientNetB0-ALSTM hybrid model

The loss value is 0.5. The sensitivity value is 95.2%, and the specificity obtained is 95.6%. The EfficientNetB0 - ALSTM model has a classification accuracy of 96.8% and a validation accuracy of 94.18% with a learning rate of 0.001.

The number of iterations is 150, and the number of epochs is 30 and 5 iterations per epoch are used. The loss value is 0.3. The sensitivity and specificity values obtained are 97.12% and 97.69%, respectively. Figures 12 and 13 gives the accuracy and loss graph of the EfficientNetB0-ALSTM hybrid model and the Inceptionv3-ALSTM hybrid model.

4.2. Discussion Based on Model Performance

The performance metrics estimate the model's accuracy, sensitivity and specificity. After IO-VMD and SSWA are applied, the maximum accuracy rate is precise at 96.8%, sensitivity at 97.12%, and specificity at 97.69% for the EfficientNetB0 - ALSTM Model with batch size 32 with a learning rate of 0.001.

For batch size 32 and a learning rate 0.1, the maximum accuracy rate was acquired at 94.3%, sensitivity 95.2%, and specificity 95.6% for the Inceptionv3 model. For custom CNN, 91.1% accuracy is obtained, which underperforms the other adaptive transfer learning models.A comparative analysis uses CNN-based architectures, solely considering the Mel spectrogram feature. The performance of CNN-based models using Mel-Spectrogram input is higher in PD detection after de-noising using IO-VMD. In addition, it is discovered that the performance is lower than the feature-fused model.

Table 4 gives the performance analysis metrics of CNN-based models with Mel-Spectrogram images. In the end, a PD classification model that combines CNN-based architecture with ALSTM using concatenated features as input is used for comparison. It is found that the EfficientnetB0 model, when combined with the ALSTM model, yields a greater accuracy of 96.8%.

On implementing the SSWA algorithm, both Mel spectrogram and concatenated features provide good classification accuracy. Table 5 gives the standard metrics for performance evaluation on hybrid models with concatenated features.

Table 4. Performance evaluation metrics of CNN models

Mel Spectrogram Feature (without IO-VMD and SSWA)	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Custom CNN	70.3	71.38	72.82
Inceptionv3	79.8	76.09	77.11
EfficientNet-B0	80.8	80.3	81.21
Mel-Spectrogram Feature (with IO-VMD and SSWA)	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Custom CNN	89.78	90.12	92.97
Inceptionv3	90.45	91.52	91.16
EfficientNet-B0	93.52	92.13	94.81

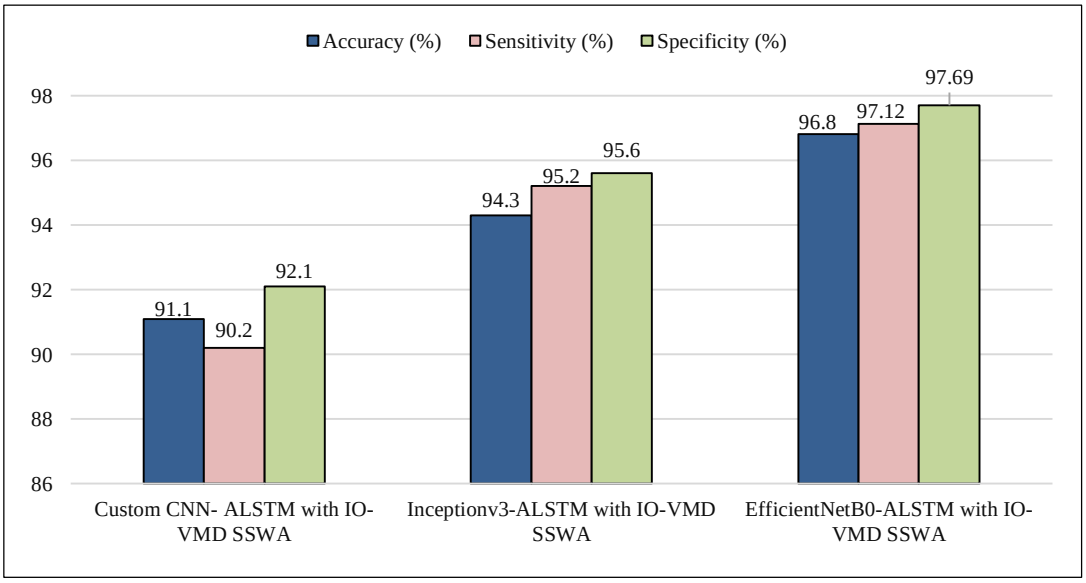

Fig. 14 Overall comparative analysis of proposed model

Table 5. Performance evaluation metrics of hybrid models using concatenated features

Proposed Model	Batch-Size	Learning Rate	Accuracy (%)	Sensitivity (%)	Specificity (%)
Custom CNN-ALSTM with SSWA	32	**0.1**	**91.1**	**90.2**	**92.1**
		0.01	92.7	93.1	91.4
		0.001	90.2	89.7	88.11
	64	0.1	89.20	88.40	84.32
		0.01	80.89	83.91	89.11
		0.001	80.21	81.19	81.43
	128	0.1	91.25	89.83	93.72
		0.01	88.76	90.80	91.43
		0.001	86.43	91.12	90.12
Inceptionv3-ALSTM with SSWA	32	**0.1**	**94.3**	**95.2**	**95.6**
		0.01	93.6	94.1	94.9
		0.001	91.8	92.3	92.11
	64	0.1	92.01	93.04	94.23
		0.01	87.8	89.09	90.12
		0.001	83.12	86.89	87.34
	128	0.1	93.52	94.13	94.81
		0.01	92.45	93.52	94.16
		0.001	91.78	93.12	93.97
EfficientNetB0-ALSTM with SSWA	32	0.1	95.30	96.43	96.41
		0.01	95.12	96.64	97.43
		0.001	**96.8**	**97.12**	**97.69**
	64	0.1	93.85	95.02	95.90
		0.01	95.87	96.93	97.02
		0.001	95.23	96.43	96.25
	128	0.1	95.64	96.42	97.21
		0.01	94.32	95.67	95.72
		0.001	93.18	95.41	96.21

4.3. Discussion Based on Similar Studies in the Literature (mPower Dataset)

Recent Literature has examined using the mPower dataset for PD diagnosis from voice samples. In [39], the authors used a time-frequency analysis algorithm on spectrogram images in a CNN model using the mPower dataset and achieved the best accuracy of 90.4%. However, the number of samples used by the authors is limited to 500 healthy and 500 PD. The authors of [27] obtained 99% accuracy on a limited 1000 random speech samples from the mPower dataset, with diagnosis relying solely on a single dependent MFCC feature. In addition, the work demonstrates a lack of knowledge regarding noise removal in the pre-processing step. 85% PD detection accuracy is achieved in the cited paper [40] using the AVEC and GEMAPS feature sets on the mPower dataset using the gradient boost algorithm, and the same algorithm is used in [20] and obtained the lowest accuracy of 71% for PD detection with a high computational period of 13.5 minutes.

Table 6. Comparison of proposed study with existing state of art methods

Performance Metrics	IMFCC+NN	SVM+GA	BiLSTM	SSWA+ALSTM	Proposed EfficientnetB0 + Attention LSTM
Accuracy	0.714	0.754	0.874	0.925	96.8
Sensitivity	0.772	0.795	0.895	0.934	97.12
Specificity	0.705	0.745	0.865	0.914	97.69

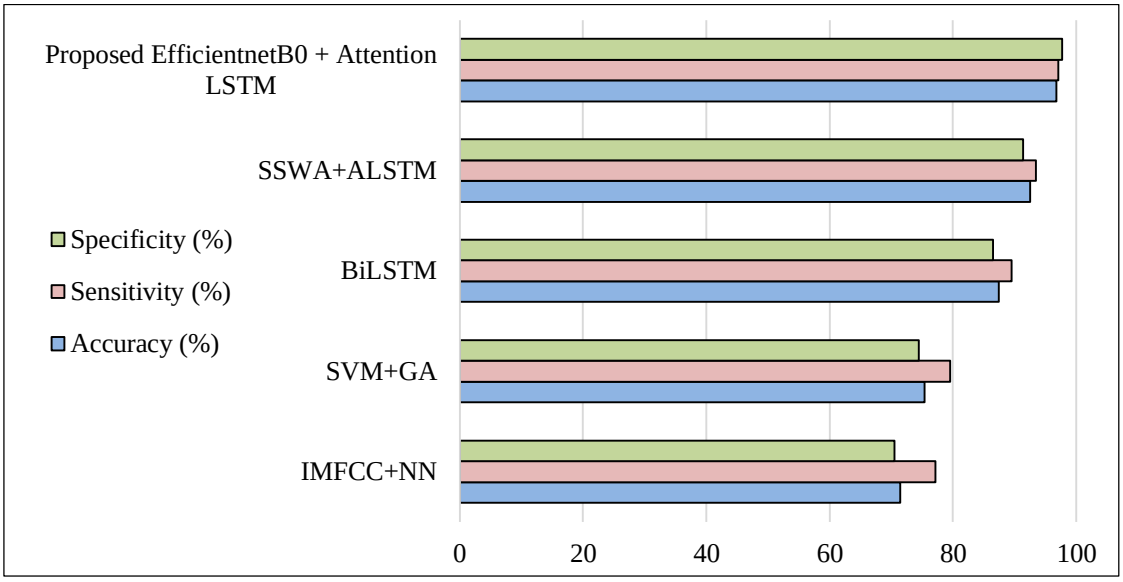

Fig. 15 Analysis based on the performance of proposed model with existing models

In [41], transfer learning CNN models were used to reduce computation time for medical diagnosis applications, including Squeezenet, Densenet, and Resnet for PD detection, and it was reported that densenet161 obtained the highest accuracy of 89.7% and sensitivity of 91.5% on the same mPower voice dataset. On comparing the earlier cited papers, this proposed work gives a promising result of 96.8% accuracy in classifying Parkinson's Disease. To the best of the information, the mPower dataset is the largest dataset for PD audio data.

The performance of the proposed work is also compared with the state-of-the-art techniques, including Intrinsic mode function cepstral coefficients with neural network [17], Support Vector Machine with Genetic Algorithm [13], Bidirectional LSTM [18], SSWA with attention-based LSTM and the Proposed IO-VMD and SSWA based EfficentnetB0 – ALSTM model for Parkisnon's disease classification and given in the Table 6. An Analysis of the proposed work compared with existing research literature work is shown in Figure 15.

5. Conclusion

The study used the mPower dataset, which contains many voice samples for training and classifying individuals into two categories (likely PD or not). After pre-processing, mel-Spectrogram images are generated from voice samples and used as input data. The FC layer of Inceptionv3 and EfficientNetB0 models was preferred to extract relevant features from these images. These pre-trained CNNs are known for extracting high-level features from images. To capture sequential information from the components extracted by the CNNs, an ALSTM network was designed and used.

Features extracted from CNN-based models, along with spectral and acoustic features, are concatenated as a single feature vector and given to the ALSTM model for PD classification. According to the study's results, the EfficientNetB0 + ALSTM model with IO-VMD achieved the highest classification performance, with an accuracy of 96.8%.

This shows that the suggested technique effectively and efficiently detects PD from sound recordings. The research compares the proposed model with cutting-edge models, demonstrating that it achieves good efficiency in PD detection. This suggests that combining IO-VMD for pre-processing, CNN-based feature extraction, ALSTM for classification, and SSWA optimization for hyperparameter tuning is a promising approach for this specific task.

The findings of this research are significant because they contribute to the ongoing efforts to develop accurate and non-invasive methods for the early diagnosis of Parkinson's Disease. Early diagnosis can lead to better management and treatment outcomes for individuals affected by the disease. However, it's important to note that while this method shows promising results, further validation and testing on diverse datasets are typically required to assess its robustness and generalizability. Additionally, the clinical application of such models would require careful consideration of ethical and privacy concerns associated with patient data. Nonetheless, advancements in deep learning techniques hold great potential for improving the accuracy and accessibility of medical diagnostics. Future research can be conducted using different datasets, such as those related to gait images, memory patterns, handwriting images, etc., and it will be possible to identify different illness phases by considering the UPDRS scale.

References

[1] Liaqat Ali et al., "A Novel Sample and Feature Dependent Ensemble Approach for Parkinson's Disease Detection," *Neural Computing and Applications*, vol. 35, pp. 15997-16010, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Carlo Ricciardi et al., "Using Gait Analysis' Parameters to Classify Parkinsonism: A Data Mining Approach," *Computer Methods and Programs in Biomedicine*, vol. 180, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Lucijano Berus et al., "Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks," *Sensor*, vol. 19, no. 1, pp. 1-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Marjolein A.E. van Stiphout et al., "Oral Health of Parkinson's Disease Patients: A Case-Control Study," *Parkinson's Disease*, vol. 2018, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] M. Abdar, and M. Zomorodi-Moghadam, "Impact of Patients' Gender on Parkinson's Disease Using Classification Algorithms," *Journal of AI and Data Mining*, vol. 6, no. 2, pp. 277-285, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Rania Khaskhoussy, and Yassine Ben Ayed, "Improving Parkinson's Disease Recognition through Voice Analysis Using Deep Learning," *Pattern Recognition Letters*, vol. 168, pp. 64-70, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Zhi-Jing Xu et al., "Parkinson's Disease Detection Based on Spectrogram-Deep Convolutional Generative Adversarial Network Sample Augmentation," *IEEE Access*, vol. 8, pp. 206888-206900, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Luiz C.F. Ribeiro, Luis C.S. Afonso, and João P. Papa, "Bag of Samplings for Computer-Assisted Parkinson's Disease Diagnosis Based on Recurrent Neural Networks," *Computers in Biology and Medicine*, vol. 115, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Afzal Hussain Shahid, and Maheshwari Prasad Singh, "A Deep Learning Approach for Prediction of Parkinson's Disease Progression," *Biomedical Engineering Letters*, vol. 10, pp. 227-239, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Atiqur Rahman et al., "Parkinson's Disease Diagnosis in Cepstral Domain Using MFCC and Dimensionality Reduction with SVM Classifier," *Mobile Information Systems*, vol. 2021, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Wondimu Lambamo, Ramasamy Srinivasagan, and Worku Jifara, "Analyzing Noise Robustness of Cochleogram and Mel Spectrogram Features in Deep Learning Based Speaker Recognition," *Applied Sciences*, vol. 13, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jan Hlavnička et al., "Automated Analysis of Connected Speech Reveals Early Biomarkers of Parkinson's Disease in Patients with Rapid Eye Movement Sleep Behaviour Disorder," *Scientific Reports*, vol. 7, no. 1, pp. 1-13, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Zayrit Soumaya et al., "The Detection of Parkinson Disease Using the Genetic Algorithm and SVM Classifier," *Applied Acoustics*, vol. 171, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Pedro Gómez-Vilda et al., "The Role of Data Analytics in the Assessment of Pathological Speech-A Critical Appraisal," *Applied Sciences*, vol. 12, no. 21, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Fatih Demir et al., "A Simple and Effective Approach Based on a Multi-Level Feature Selection for Automated Parkinson's Disease Detection," *Journal of Personalized Medicine*, vol. 12, no. 1, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] N.P. Narendra, Björn Schuller, and Paavo Alku, "The Detection of Parkinson's Disease from Speech Using Voice Source Information," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1925-1936, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Biswajit Karan, Sitanshu Sekhar Sahu, and Kartik Mahto, "Parkinson Disease Prediction Using Intrinsic Mode Function Based Features from Speech Signal," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 249-264, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Andrius Lauraitis et al., "Detection of Speech Impairments Using Cepstrum, Auditory Spectrogram and Wavelet Time Scattering Domain Features," *IEEE Access*, vol. 8, pp. 96162-96172, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Javier Carrón et al., "A Mobile-Assisted Voice Condition Analysis System for Parkinson's Disease: Assessment of Usability Conditions," *Biomedical Engineering Online*, vol. 20, pp. 1-24, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Aditi Govindu, and Sushila Palwe, "Early Detection of Parkinson's Disease Using Machine Learning," *Procedia Computer Science*, vol. 218, pp. 249-261, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Nada R. Yousif et al., "A Generic Optimization and Learning Framework for Parkinson Disease via Speech and Handwritten Record," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 8, pp. 10673-10693, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] K. Kamalakannan, G. Anandharaj, and M.A. Gunavathie, "Performance Analysis of Attributes Selection and Discretization of Parkinson's Disease Dataset Using Machine Learning Techniques: A Comprehensive Approach," *International Journal of System Assurance Engineering and Management*, vol. 14, pp. 1523-1529, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Máté Hireš et al., "Voice-Specific Augmentations for Parkinson's Disease Detection Using Deep Convolutional Neural Network," *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Poprad, Slovakia, pp. 213-218, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Dharani M.K., and R. Thamilselvan, "Hybrid Optimization Enabled Deep Learning Model for Parkinson's Disease Classification," *The Imaging Science Journal*, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Giovanni Costantini et al., "Artificial Intelligence-Based Voice Assessment of Patients with Parkinson's Disease Off and On Treatment: Machine vs. Deep-Learning Comparison," *Sensors*, vol. 23, no. 4, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Sanjana Singh, and Wenyao Xu, "Robust Detection of Parkinson's Disease Using Harvested Smartphone Voice Data: A Telemedicine Approach," *Telemedicine and E-Health*, vol. 26, no. 3, pp. 327-334, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Changqin Quan et al., "End-to-End Deep Learning Approach for Parkinson's Disease Detection from Speech Signals," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 556-574, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Valerio Cesarini et al., "Voice Disorder Multi-Class Classification for the Distinction of Parkinson's Disease and Adductor Spasmodic Dysphonia," *Applied Sciences*, vol. 13, no. 15, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Sabrina Scimeca et al., "Robust and Language-Independent Acoustic Features in Parkinson's Disease," *Frontiers in Neurology*, vol. 14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Guidong Bao et al., "Classification of Dysphonic Voices in Parkinson's Disease with Semi-Supervised Competitive Learning Algorithm," *Biosensors*, vol. 12, no. 7, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Brett Koonce, *EfficientNet Convolutional Neural Networks with Swift for Tensorflow*, Image Recognition and Dataset Categorization, 1st ed., Apress Berkeley, CA, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] P. Ravi Prakash et al., "A Novel Convolutional Neural Network with Gated Recurrent Unit for Automated Speech Emotion Recognition and Classification," *Journal of Control and Decision*, vol. 10, no. 1, pp. 54-63, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[33] N. Dong et al., "Inception V3 Based Cervical Cell Classification Combined with Artificially Extracted Features," *Applied Soft Computing*, vol. 93, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[34] mPower Public Researcher Portal, mPower: Mobile Parkinson Disease Study, 2015. [Online]. Available: https://www.synapse.org/#!Synapse:syn4993293/wiki/247859

[35] Hua Li et al., "An Optimized VMD Method and Its Applications in Bearing Fault Diagnosis," *Measurement*, vol. 166, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[36] Mohit Jain, Vijander Singh, and Asha Rani, "A Novel Nature-Inspired Algorithm for Optimization: Squirrel Search Algorithm," *Swarm and Evolutionary Computation*, vol. 44, pp. 148-175, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[37] Hadi Eskandar et al., "Water Cycle Algorithm–A Novel Metaheuristic Optimization Method for Solving Constrained Engineering Optimization Problems," *Computers & Structures*, vol. 110-111, pp. 151-166, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[38] K. Sangeetha, and D. Prabha, "Sentiment Analysis of Student Feedback Using Multi-Head Attention Fusion Model of Word and Context Embedding for LSTM," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 4117-4126, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[39] Hanbin Zhang et al., "DeepVoice: A Voiceprint-Based Mobile Health Framework for Parkinson's Disease Identification," *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, USA, pp. 214-217, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[40] Timothy J. Wroge et al., "Parkinson's Disease Diagnosis Using Machine Learning and Voice," *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Philadelphia, PA, USA, pp. 1-7, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[41] Onur Karaman et al., "Robust Automated Parkinson Disease Detection Based on Voice Signals with Transfer Learning," *Expert Systems with Applications*, vol. 178, 2021. [CrossRef] [Google Scholar] [Publisher Link]