



A NOVEL OPTIMIZATION BASED SQUEEZENET WITH DWARF MONGOOSE OPTIMIZATION ALGORITHM FOR FEATURE SELECTION AND DEEP LEARNING CLASSIFIER FOR PARKINSON'S DISEASE

V R Sadasivam^{*1}, R. Poornima², M.Nithya³, G. Nagarajan⁴

¹Department of Information Technology, K S Rangasamy College of Technology, Tamil Nadu, India.

²Department of Information Technology, RMK Engineering College, Tamil Nadu, India.

³Department of Biotechnology, K.S.Rangasamy College of Technology, Tamil Nadu, India.

⁴Department of AI & DS, Sri Shanmugha College of Engineering and Technology, Tamil Nadu, India.

¹<http://orcid.org/0000-0001-7443-8046>, ²<http://orcid.org/0000-0001-5879-6898>,

³<http://orcid.org/0000-0001-6634-4146>, ⁴<http://orcid.org/0000-0001-8359-6628>

Email: *sadasivam@ksrct.ac.in, poorniom@gmail.com, nithyam@ksrct.ac.in, nagu.gee@gmail.com

ARTICLE INFO

Article History

Received: March 11, 2026

Reviewed: April 11, 2026

Accepted: July 11, 2026

Published: August 31, 2026

Keywords:

Deep Neural Network (DNN),
SqueezeNet,
Dwarf Mongoose Optimization,
Parkinson's disease (PD),
Fuzzy Convolution Bi-Directional
Long Short-Term Memory (FCBi-
LSTMs).
Example five.

ABSTRACT

A neurological condition that affects nerve cells is called Parkinson's disease (PD). Numerous neurological disorders, such as tremors, limb stiffness, and changes in walking patterns, are brought on by PD. Preserving the health of individuals affected by PD requires early detection of the condition. However, early PD diagnosis is challenging due to the aging population. Deep learning (DL) models have been tested for accuracy and for this test, Parkinson's dataset and feature relevance analysis has been utilized. The DNN technique is utilized to evaluate and extract pressure, grip angle, and timestamp from spiral drawing inputs. The voice data in this case is sourced from Kaggle. In this study, a novel DL model for PD detection and prediction is presented. It combines an optimization technique for dwarf mongoose with a SqueezeNet. For classifications, SqueezeNet with DMOA, trained on FCBi-LSTMs generates the best features. This study's suggested model utilizes the learning repository at the University of California, Irvine (UCI). The values of Matthews correlation coefficients (MCCs), accuracy, and F-measure values are evaluated using Leave-One-Person-Out-Cross Validations (LOPO CVs).



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

A neuronal ailment that affects the central nervous system and impairs movement and coordination is called PD. Each year, more than 50,000 Americans are diagnosed with PD, according to the National Institutes of Health (NIH). A person with PD may experience lifelong impairment due to tremors, loss of posture, and balance issues. Parkinson's disease is the second most common disease, and World Health Organization (WHO) has established that one of the most difficult scenarios for the treating physician is the disease's extended duration. Parkinsonism affects 10 billion people worldwide, making it a threat. Early and appropriate intervention, however, can significantly lessen the impact of the illness and aid in the patient's recovery. Table I examines the different phases of PD [1]. People with PD may have distinguishable symptoms include tremors, limb rigidity, altered gait patterns, and balance problems. However, patients experience a slow development of these symptoms. In addition, PD symptoms fall into two different categories: Movement-related motor symptoms (i) and non-motor, motion-unrelated motor symptoms (ii). The symptoms that are not related to motor function are usually more severe and include sadness, erratic sleep patterns, olfactory impairment, and cognitive decline. PD affects more than 10 million people worldwide and more than 52 billion Americans annually. Key tactics in the fight against PD include early detection, grouping patients according to PD phases, starting therapy right once, and identifying medications and symptoms that can be minimized [2]. Thus, the advancement of possibly curative medications needs to be slowed down and PD must be stopped in its tracks. Even with a number of PD symptoms and diagnostic procedures, the precise diagnosis of PD is still unknown [3].

Together, these diagnostic markers are used to diagnose PD, and other biomarkers are examined to identify the disease at an early stage. The goal of PD therapy is to lessen symptoms rather than to stop or reverse the course of the illness in individuals. The experiment's dataset, which included a voice impaired by PD, was obtained from the UCI Machine Learning repository. The dataset contains 22 voice measurement-related features in addition to a "status" feature that indicates the patient's PD condition. The data's parallel coordinates display the largest range. An analysis of several ML algorithms revealed that SVM had the highest accuracy, at 88.9%; Random Forest came in second, at 90.26%; and the Naive Bayes method had the lowest accuracy, at 69.23%. Disparities in the number of clusters between patients who were well and those who were not were found using clustering techniques such as Hierarchical clustering and Self-Organizing Maps (SOM) [4]. As a result of the recent and ongoing increase in the variety of medical data, efficient categorization and prediction algorithms are needed. Numerous agents can affect a classification algorithm's accuracy, according to earlier research on ML [5]. Medical data sets and diagnostic issues are analyzed using ML techniques [6]. As a result, better medical judgments and treatments will be made, and financial expenses will go down [7]. Furthermore, feature selection is crucial to the understanding of medical data. The feature selection technique plays a crucial role in global combinatorial optimization within machine learning. It Compress the number of features from the original set, Eliminates redundant or irrelevant features without appreciably lowering the amount of information retained, and simplifies models to expedite training and improve comprehension [8].

Therefore, to increase processing speed and forecast accuracy, a strong feature selection technique is needed. There are three different kinds of feature selection methods: filter that collects features from the data set without any learning, hybrid which combines the feature selection stage and the classifier building, and wrapper which employs learning approaches to estimate valuable features [9]. DNNs are hierarchical layers that produce deep abstract representations of input information in applications. Numerous applications, including as speech recognition, picture classification, drug development, and genetic research, have made use of DLTs [10]. The key reason DNNs have been adopted by researchers for PD categorizations is their efficacy. Given their ability to replicate intricate and nonlinear data linkages, DNNs are particularly useful classifiers for dealing with PDs. Sensor activity [11] and EEG data [12] are examples of singular variables used as CNN inputs in prior research on the categorization of PD, but no attempt was made to employ discrete parallel layers for classification. In a study [13], voices were multiplied in order to increase the number of voice recordings used in techniques testing and training with CVs (cross-validations). This resulted to unbalanced performance measurements. Leave-One-Person-Out-Cross Validations were used to assess the performance of the suggested framework because the data comprised voice recordings of both healthy individuals and people with Parkinson's disease. LOPO-CVs excluded Samples from particular persons in test set Cycles and used other instances in training. The following is the contribution.

- To improve the prediction of PD, a thorough investigation into the effectiveness of various feature selection and DL techniques was conducted.
- As far as we can tell, we are the first to build a Voice and Spiral dataset including a variety of variables from both healthy and PD-affected volunteers. We are also incorporating the new idea of the speech and spiral dataset into the identification and classification of PD.
- This study proposes a feature-choice SqueezeNet architecture based on agreements. Ensembles of feature selection techniques, such as Dwarf Mongoose Optimization, seek to replace single feature selection methods with multiple feature selection strategies, whereas SqueezeNet uses optimal features to train fuzzy convolution bi-directional long short-term memory classifiers.
- This suggested strategy's performance confirmed with LOPO-CVs after it was trained on datasets from the DL repository. This approaches introduces a model depends on DL repositories, measures the approaches' accuracy, F-measure values, and MCCs using LOPO-CVs.

II. LITERATURE SURVEY

In the very 1st architecture, multiple feature sets were merged before that are being fed as inputs into nine-layered CNNs. In their 2nd architecture, feature sets were fed directly to convolution layers in parallel. Various feature sets were generated which are then blended [14] with the help of vocal data to detect PDs using the dual frameworks of CNNs. Therefore, before each parallel branch merged into a layer, its deep features were acquired. As they used parallel convolutions to learn deep features, their second model demonstrated extremely encouraging test results. Also, to clearly identify patients with PD from healthy individuals, the retrieved deep features were effective in boosting the classifiers' capacity for discrimination. By [15] combined ensemble learning with CART to classify PDs. CART was utilized in the study to choose the best training vocal samples with broad range of variation. To ascertain whether training data was ideal, the evaluation employed Assemblies, such methods as random forests, support vector machines, and ELMs are included. We used ensemble- Education systems that had been trained to classify test data. In addition to improving PD predictions with voice data categorizations, this research indicated that CART and RF Composition were maintained stability when compare to alternative techniques. In turn [16] predicted that speech problems, the disorder's initial symptom, would be used to identify Parkinson's disease. For classification, they employed the SoftMax method and DNNs using stacked autoencoders.

The superiority of DNN classifiers over alternative classification algorithms was shown by the simulation results from their two datasets. According to [17] suggested using enhanced fuzzy FKNNs in conjunction with CBFOs using Gauss mutations on audio data as a rapid way to find PDs. Their CBFO FKNN successfully modified the FKNN's parameters through the use of an advanced instance-based learning technique. The study employed PD datasets to extensively evaluate their suggested method in term of Segmentation specificity, accuracy, sensitivity, and Receiver operating characteristic (ROC) area , or AUCs. Results delivered more accurate clinical diagnosis conclusions to medical professionals. By [18] used ANNs with MLPs to classify PDs on datasets from the UCI ML library. They collected both control groups and voice recordings of patients with PD. The study trained between 10,000 and 6,000 neurons across many networks; the number of neurons increased tenfold in the buried layers. ANNs' analysis of speech-related features could be utilized to evaluate how PDs are affecting patients. With the availability of biological data, MLTs can also diagnose various neurological illnesses.

Well-known speech characteristics like shimmers, jitters, fundamental frequency parameters, and harmonicity factors were used by [19] in their PD research to describe the disorders. PPEs, DFAs, and RPDEs were used to evaluate the PDs. PDs have been categorized via XGBoost algorithm with the help of the identified baseline features. Feature significance plots were used to guide feature selection in order to boost the model's proficiency. loc Shimmer attributes that resulted have been removed . XGBoost's evaluations the feature relevance enhanced the features' effectiveness to raise classification accuracy. Unlike the initial method, which combines distinct feature selections [20] and feeds them into a CNN with nine layers, the Features from concurrent input layers directly input into convolution layers according to the second model design. The deep features of parallel branches obtain information simultaneously before merging data through each parallel connection in the merge layer. The suggested models undergo Leave-One-Person-Out Cross Validation (LOPO CV) to validate their findings after being trained using UCI ML data. We examine our data using the The F-score and Matthews correlation coefficient (MCC), and accuracy due to the unequal class distribution. Experimental results demonstrate that this second model shows great potential because it uses concurrent convolution layers to learn features from each batch. The model learns representation features from each feature batch through concurrent convolution layers

Problem Statement:

One major problem for PD patients is early diagnosis of the disease. If individuals receive an early diagnosis, they can improve their quality of life even if their health worsens. Another problem is that diagnosing PD involves multiple processes, such as asking the patient about their neurological history and testing their motor skills in different settings. Most recent research works on the homo dataset, which includes text, speech, video, and image data. The proposed study has identified issues with multi-data handling processes and dataset modification. After a certain dataset is examined, the efficacy of disease prediction is controlled. The application of ML-based methods for multivariate data processing allows for more real-time solutions. Multi-variate vocal data analysis (MVDA) aims to identify PD based on features from many datasets by applying machine learning algorithms. This work aims to explore potential enhancements to multi-variate and multimodal data processing to improve the rate of disease identification. For the purpose of analyzing Parkinson's data based on voice features, research is now being conducted on several machine learning techniques at once, such as support vector machines, K-NN, naïve Bayes, and artificial neural networks. Based on these studies, the MVDA uses large datasets and ML techniques to enhance disease identification. It is recommended that the multivariate acoustic features of several patients be incorporated into the proposed MVDA. With the use of suggested ML approaches, the subjective disease has been identified under the MVDA system.

III MATERIALS AND METHODS

The spiral drawings and voice recordings of both healthy and ill individuals are used in the suggested methodology for PD prediction. The two datasets came from Kaggle. Speech datasets are filtered using the MFCC feature extraction approach to remove noise and clean the data. The data was used to generate a SqueezeNet model. This model uses voice data to predict PD. Whether a voice is that of a person with PD or not can be determined using this model. The spiral drawings made by both healthy individuals and PD sufferers are examined, and the images are run through several CNN layers in order to identify and identify patterns that can be used for categorization. To develop the model, the transfer learning method methodology is used. Because it is more accurate, the trained model VGG16 is employed. The spiral drawing dataset is used to train this neural network, and the model's performance is evaluated against the testing dataset. The model can determine if a person has PD or not when an image is uploaded. An accurate model of the PD prediction system is produced by connecting the two models mentioned above. The system as a whole can determine a person's PD status by utilizing their speech and spiral drawing.

Voice Dataset:

To develop a DLmodel for voice recognition-based PD prediction, a substantial and diverse dataset of speech recordings from both healthy and PD-affected individuals would be needed. Along with speaker demographic data, the dataset should contain audio samples spanning a variety of speaking tenors and emotional states. Making sure the dataset is well-balanced and curated that is, that there are a similar number of samples from people with and without PD is crucial. Our Voice Dataset, which we gathered via Kaggle, served as the main input. After that, The data is split for training and evaluation purposes . 2) MFCC Feature Extraction: In audio analysis and voice processing, MFCC (Mel-Frequency Cepstral Coefficient) is a frequently utilized feature extraction technique. A collection of coefficients known as the MFCCs serves as a compact representation of a speech signal's spectral envelope. MFCC feature extraction from a voice dataset is displayed in Figure 1.

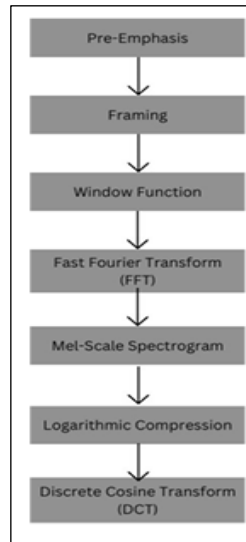


Figure 1: MFCC feature extraction from a voice dataset.
Source: Authors, (2026).

- **Pre-emphasis:** To highlight high-frequency components in the input audio stream, a high-pass filter is used.
- **Framing:** To produce a series of brief windows, the input audio signal is split into overlapping or non-overlapping frames.
- **Window Function:** To minimize spectral leakage, each frame is assigned a window function (similar to the Hamming window).
- **FFT:** The time-domain data is converted into the frequency domain for every frame using the Fourier Transform.
- **Mel-Scale Spectrogram:** This represents the non-linear frequency resolution of the human auditory system by converting the magnitude of the FFT coefficients into the Mel scale.
- **Logarithmic compression:** To lower the dynamic range, the Mel-scale spectrogram is converted into the logarithmic domain.
- **Discrete Cosine Transform (DCT):** Using the logarithmic Mel-scale spectrogram as a starting point, the DCT generates the MFCC coefficients, which are a condensed representation of the spectral envelope of the audio source. A machine-learning model for voice recognition, speaker identification, or other speech-related tasks can employ the derived MFCC features as input. The MFCCs capture the significant spectrum fluctuations that are pertinent to speech analysis and provide a compact representation of the spectral information in the spoken signal.

Spiral drawing Dataset:

Spiral drawing datasets are collections of spiral drawings that have been captured on camera or recorded. In the fields of computer vision and ML, this kind of dataset is widely used for developing and testing motion analysis and pattern recognition algorithms. A spiral drawing dataset could contain a variety of spiral variations, including spirals drawn with different instruments (e.g., pencil, pen, stylus), spirals with varied curvatures and diameters, and spirals drawn in a clockwise or counter clockwise direction. The dataset might additionally contain participant demographic data, like handedness, gender, and age. ML models can be trained on the photos or recordings in the dataset to identify and categorize spirals according to a variety of characteristics, including the spiral's size, direction, and drawing speed. Subsequently, the models can be applied to examine fresh spiral illustrations and forecast spiral characteristics. Making sure the spiral drawing dataset is representative of the target population, balanced, and diversified is crucial. This will make it more likely that the ML models created with the dataset will be precise and able to be applied to fresh spiral designs. Figure 2: spiral sketching dataset displayed

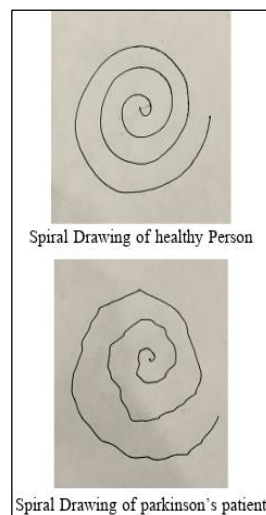


Figure 2: Spiral drawing dataset.
Source: Authors, (2026).

Pre-processing:

During the pre-processing stage, the complete filter technique procedure takes place outside of the model that are ignored by the filter. The main considerations of filter methods are the internal relationships and the distribution correlation of the data. Consequently, the benefit of filter approaches is their ease of computation. Because filter approaches are simple to apply and compute quickly, they are often used to identify this condition. The following is a list of various popular filtering methods. The minimum redundancy and maximum relevancy (mRMR) technique selects features that are not identical to each other yet show a strong "correlation" with the classification variable. The wrapper method decides whether to accept or reject a feature based on the functioning update of a classifier. By taking specific classifiers into consideration, the wrapper technique generates a carefully designed subset. As a result, for wrapper methods, there is less chance of finding the local maximum. Due to its notable efficiency improvement, the wrapper approach is used in ML diagnostics. However, it suffers from a number of drawbacks, including as high computational costs and overfitting. Wrapper-based feature selection algorithms use a classifier to generate several predictor variable machine learning models. They then pick the variable subset that yields the best model. The correlation between the predictor and independent variables is computed using statistical methods in filter-based approaches, which do not require the assistance of a learning algorithm. Assigning scores takes into consideration how relevant the predictor factors are to the target variable. After then, the variables with better scores are used to build the ML model. Consequently, the aim of this research is to Determine the most relevant features utilizing a filter-based feature selection method for improved PD diagnosis.

Feature Extraction:

Mel Frequency Cepstral Coefficients (MFCCs), wavelet transforms (WTs), tunable Q-factor wavelet transforms (TQWTs), vocal fold features, baseline and temporal frequency features, were all included in the dataset.

- i. Baseline features: speaking impairments caused by PD have been effectively utilized to assess PDs and monitor the progression of the illness after medication treatments. Studies have made extensive use of the baseline features [13] for speech-based Parkinson's disease research. The basic frequency parameters (#5), DFAs (detrended fluctuation analyses) (#1), harmonicity parameters (#2), PPEs (#1), and RTDEs (recurrence time density entropies) (#1) are a few specific examples.
- ii. Time-frequency features: features include bandwidth (#4), formant frequencies (#4), and intensity parameters (#3).
- iii. MFCCs: MFCC-based extractions integrate spectral domain partitions and cepstral analyses through the use of triangular overlapped filter banks. When PD affects the tongue and lips, articulators that move quickly deteriorate. MFCCs are able to identify this. The vocal folds, the log energies and 1st/2nd order derivatives of the signal, the mean deviation & standard deviation of the first 13 MFCCs, and other information were used to generate the 84 MFCC-related features (#84) that were included in the dataset [13].
- iv. WTs: In general, WTs are employed to decide on signals, particularly those with little variations at the regional level. PD has been diagnosed in a number of studies using WT characteristics derived from the raw fundamental frequencies (F0) of a speech sample. This work yielded 182 WT attributes, including energies, Teager-Kaiser energies, Shannon's and log energy entropies, using both detailed coefficients and approximations.
- v. Using feature extraction with TQWTs: To improve the signal's qualities, TQWTs alter three parameters: numerous levels (J), redundancy (r), and Q-factors (Q), depending upon the attribute of the signal. The decomposed layer counts are represented by the symbol J while Q-factors directly correspond to time domain oscillations. The J + 1 sub-bands arise from high-pass filtering together with one low-pass filtered output while Q-factors measure the time domain signal oscillations. A J high-pass filter results in J + 1 sub-bands and one single low-pass filtered output during decomposition. During decomposition the signal decomposes into one low-pass filtered output and J + 1 sub-bands derived from high-pass filtering. Wavelet's temporal localizations are made possible by ringing, which is controlled by. The dataset contained 432 TQWT-related features that were found by the tests employed in this work [13].
- vi. Vocal fold features: The researchers generated features from vocal fold vibrations to analyze how noises affect them. The study produces additional information about noise reactions in vocal folds with enhanced detail through sonic analysis. Glottis quotients (#3), glottal-to noise excitations (GNEs) (#6), vocal fold excitation ratios (VFERs) (#7), and empirical mode EMD decompositions (#13) comprised part of the research data collection.
- vii. Concat features: the features like vocal fold, baseline, & time-frequency are combined to create concat features.

Feature Selection using SqueezeNet with Dwarf Mongoose Optimization:

PD is the result of a malfunction in the Parkinson's actual processing mechanism. In general, different DL approaches are used to identify Parkinson's disorders. SqueezeNet, which is trained using the DMOA approach, is used in this study to predict PD. For the purpose of predicting cardiac disease, the SqueezeNet model chooses $V_k \times z$ as the input. The SqueezeNet approach has the advantage of better detection outcomes at minimal building costs. Subsequently, the structure of SqueezeNet is explained in the next section.

Structure of SqueezeNet:

Many fire modules, each with an expansion layer and a squeeze convolution layer, make up SqueezeNet. Next, the transmission of squeeze convolution layer's output is done to the last convolutional layer. This transmission is initiated with a solitary convolution layer which is being tracked by 8 number of fire modules. "Bm" represents SqueezeNet model's output. Figure 2 illustrates that the max-pooling procedure is executed in 2 steps by SqueezeNet.

SqueezeNet Training Using DMOA:

This section provides more explanation of the DMOA, which is used to train the SqueezeNet employed in this investigation. The ability of dwarf mongooses to forage is the basis for the main concept of DMOA. DMOA is a metaheuristic model that is used to solve optimization problems. For the aforementioned optimization problems, DMOA can produce and enhance a potential solution. By employing this tactic to investigate different areas of the problem search space, the dwarf monkeys are hopping from one food source to another. DMO further uses a single tuning parameter. The DMOA algorithm consists of the following phases.

- (1) Initialization: The algorithmic solutions and constraints are first set up. This is the action that results in the best outcome.
- (2) Fitness Measure: The Mean Squared Error (MSE) is used to determine the best answer.

$$b_{min} = \frac{1}{w} \sum_{m=1}^w (B_m^* - B_m)^2 \quad (1)$$

Where B_m^* stands for the predicted outcomes, B_m for the SqueezeNet categorized outcome, and w for the total number of samples.

- (3) Alpha Group. The initial Group is Alpha. The efficacy of the entire strategy is assessed once the population has begun in which selection of Alpha female will be chosen based on the values of likelihood that are determined by the following as:

$$k = \frac{(b_{min})_p}{\sum_{p=1}^t (b_{min})_p} \quad (2)$$

In this case, b_{min} indicates the fitness function, and t indicates the mongoose count in k . The solution's upgrading approach is provided as:

$$s_{p+1} = s_p + \alpha_p * Q \quad (3)$$

In which, Q represents the vocalization of the lead female, and s_p is the solution for this iteration's solution. The distributed random integer is represented as α_p . The sleeping mound, which is determined after each loop, is given by

$$L_r = \frac{(b_{min})_{p+1} - (b_{min})_p}{\max[(b_{min})_{p+1} - (b_{min})_p]} \quad (4)$$

Additionally, the sleeping mound's average count is calculated as

$$\omega = \frac{\sum_{p=1}^t L_r}{t} \quad (5)$$

In which, the sleeping mound is indicated by L_r , and the total number of sleeping mounds is indicated by t . The DMOA algorithm moves on to the scouting phase after meeting the babysitting exchange requirement.

- (4) Scout Group: This is the phase where the mongoose moves around in the best sleeping mound, and the family travels a good distance. As a result, Scout Mongoose is defined as:

$$s_{p+1} = \begin{cases} s_p - D * \alpha_p * z |s_p - \vec{S}|, & \text{if } \omega_{p+1} > \omega_p \\ s_p + D * \alpha_p * z |s_p - \vec{S}|, & \text{otherwise} \end{cases} \quad (6)$$

In this case, the value of D and the value of S can be calculated as,

$$D = \left(1 - \frac{U}{H_U}\right)^{2 * \frac{U}{H_U}} \quad (7)$$

$$\vec{S} = \sum_{p=1}^t \frac{s_p * L_r}{s_p} \quad (8)$$

In which, z denotes the random value between (0,1). As inferior group members, babysitters are typically young adults who are intent on rousing the feminine alpha to carry out the daily hunting. The DMOA pseudocode is displayed in Algorithm 1.

- (5) A Second Look at Feasibility. The solution's viability is assessed in relation to the fitness value calculation. To replace the subpar solution with the superior one iteratively, the optimal solution in this case is the one with the lowest MSE value.
- (6) Finalization. The aforementioned procedures are all carried out repeatedly until the ideal outcome is reached. Algorithm 1 shows the DMOA algorithm's pseudocode.

Algorithm 1: DMOA's Pseudocode.

```

Initiate the algorithmic constraints
While  $U < H_U$  do
  For (p=1 to k) do
    Estimate the fitness of mongoose
    Set the time counter
    Estimate the value of alpha by equation (5)
    Compute the best solution by equation (6)
    Evaluate the sleeping mound using equation (7)
    Evaluate the mean value of the sleeping mound using equation (8)
    Compute the movement vector using equation (9)
    Execute the scout mongoose for a successive solution using equation (10)
  End for
   $c = c + 1$ 
End while
Get the best solution

```

Source: Authors, (2026).

Classification of PDs Using FCBi-LSTMs:

FCBi-LSTMs were utilized in this paper to classify PD's. The computes fuzzy methods weights by modifying membership values in order to extract the most relevant features related to PDs. The chosen attributes from PD datasets are analyzed by CNNs and FCBi-LSTMs. Convolution layers, which make up CNNs, are fed outputs from pooling levels to convolution layers that come after them. Given their knowledge of the local perception of biological vision cells, CNNs use partial filters for convolutions, which provides considerable advantages in feature extraction. The convolution layer's represents the chosen Properties from the PD dataset is enhanced by filtering it into multiple output matrices, each with a range of $(Nm+1)$. The CNN pooling layer serves as a method regarding lowering a matrix's dimension without sacrificing the essential connections between its elements. Average pooling layers, or pooling layers, are fed inputs from convolution layers. As an intermediate variable, the final convolution layer's output is used in the Bi-LSTM data analysis technique. Therefore, there are more applications for LSTM than only changing loop cells and adding nonlinear elements to input. Bi-directional LSTMs perform better than uni-directional LSTMs in terms of capturing structural information, and fuzzy weights are generated using Gaussian membership functions. CNN's convolution layers process the Bi-LSTMs' final outputs to diagnose Parkinson's illness. Bi LSTM and CNN processed features are integrated by multimodal factorized bilinear pooling (MFB).

IV. RESULTS AND DISCUSSION

In comparison to other methods like SVM, CNN, and FCLSTM-CNN (fuzzy convolution long short-term memory-based convolution neural networks), this section explains the experimental results that the suggested FCBi-LSTM classifier produced. Since the test sets had fewer samples, the individual occurrences that remained in the training set were used to assess the performance of LOPO CVs. The class The MIT-BIH arrhythmia database served to investigate arrhythmia recognition and classification systems through the use of MATrix LABoratory R2016a (MATLAB R2016a). The following system specifications have been used in the implementation process: A Windows 8.1 Pro 64-bit operating system, a 1TB hard drive, an Intel (R) Core™i3-4160T CPU@3.10GHz 3.09GHz processor, and 4.00 GB RAM are included.

Dataset

The Voice and Spiral dataset's results are displayed in Figures 3 and 4.

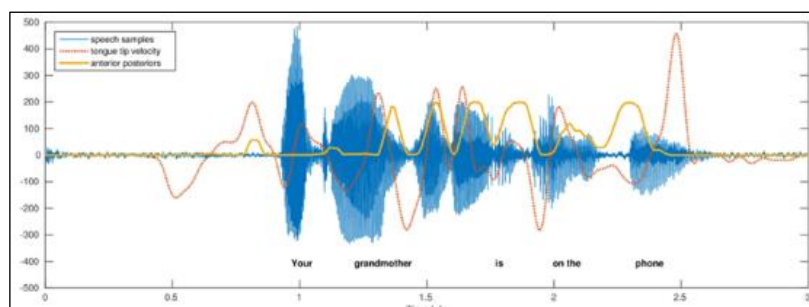


Figure 3: Voice dataset Result.

Source: Authors, (2026).

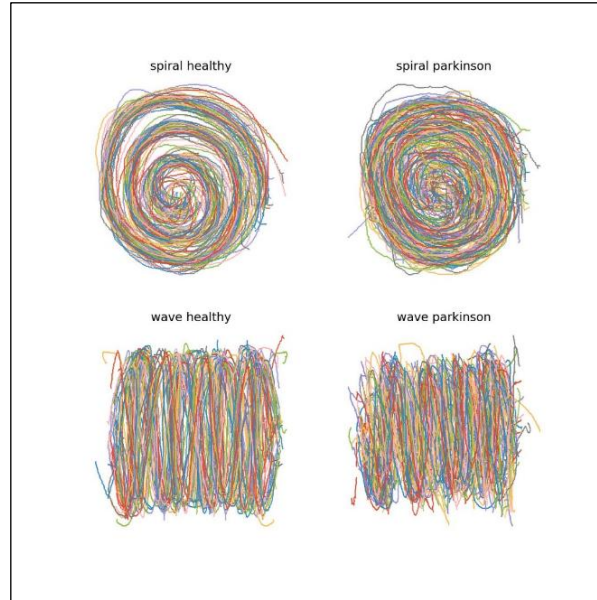


Figure 4: Spiral dataset Result.
Source: Authors, (2026).

Metrics Utilized for Evaluation

Evaluation Metrics: Evaluation measures exist to determine the predictability of classifiers. While accuracy is a commonly used criterion, conclusions drawn from data with an unbalanced class distribution may be deceptive. The effectiveness of a classifier in differentiating between classes, even when there is a class imbalance, can be assessed using assessment metrics such as MCCs and F-measure. Include the number of instances in each class that are correctly and incorrectly classified in order to enable the confusion matrix shown in in Table 1 to depict binary classification. These letters in the confusion matrix, TP, FP, TN, and FN stands for True Positive, False Positive, True Negative, and False Negative respectively. The following were calculated: recall, F-measure ,precision, accuracy and error using the formulas depending on these counts.

$$Precision = \frac{tp}{tp+fp} \quad (9)$$

$$recall = \frac{tp}{tp+fn} \quad (10)$$

$$F - measure = \frac{2*Precision*Recall}{precision+recall} \quad (11)$$

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad (12)$$

$$error = 100 - Accuracy \quad (13)$$

Another statistic that can be used to assess the reliability of binary classifications is the MCC, which accounts for the TP, FP, FN, and TN Counts. Generally speaking, it is seen as an equilibrium measure that may be used even when there is an uneven distribution of classes. The correlation coefficients between expected and actual occurrences are simply called MCCs, and they have a range of -1 to +1. The Score +1 represents perfect predictions between actual labeling and forecast while score -1 shows differences between these two variables.

Table 1: Confusion matrix for two-class classification.

Actual/predicted as	Negative	Positive
Positive	FN	TP
Negative	TN	FP

Source: Authors, (2026).

Comparison of Results:

Three different feature types were used in connection to accuracy, error, F-measure, and MCC in the experimental evaluations of classifiers. The accuracy rate of MFCCs, Wavelets, and Concated features in the CNN combination was 94.1752%, whereas the accuracy rate of MFCCs, Wavelets, and Concated features in the SVM combination was 88.1294%. The accuracy value of TQWT+MFCC+Wavelet is 93.047%, whereas the accuracy value of TQWT+Wavelet+Concat is 93.0854%. Similarly, the accuracy value of TQWT+MFCC+Concat, whereas the accuracy value of MFCC+Wavelet+Concat is found to be 95.1557% for FCLSTM-CNN. By using MFCCs+Wavelets+Concated combinations, the proposed FCBi-WLSTM classifier Reached the best accuracy scores of 98.7720% (F-measure rate of 98.5010% and 71.400% for MCC). The results of the TQWT+MFCC+Wavelet combination were 98.3100%, 96.6381%, 74.300%, and 3.3619%, in that order.

The highest results of f-measures, accuracies, MCCs, and errors were likewise produced by these feature set combinations. These results are displayed in Figures 5-8. At the first feature level combination, the F-measure values of CNN is 85.86%, SVM is 82.91%, and FCLSTM-CNN is 94.22%. With the first level feature combination the proposed FCBi-LSTM achieved an F-measure score of 98.31%. Figure 5 compares the F-measure outcomes of four distinct feature-level combinations using different classifiers.

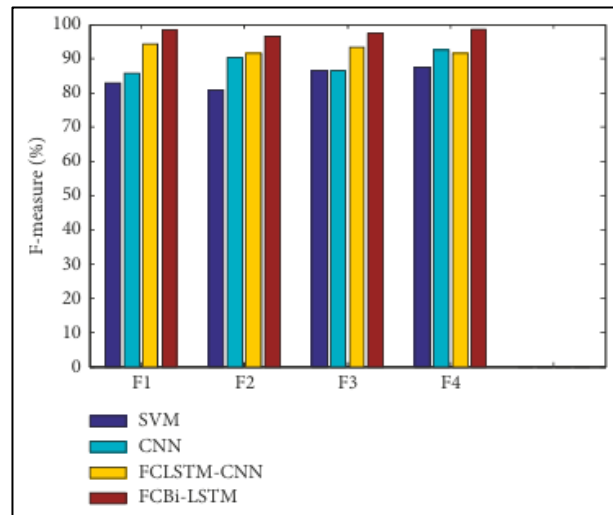


Figure 5: F-measure results comparing classifiers to feature level combinations.
Source: Authors, (2026).

From Figure 6, it is clearly seen that x-axis accuracy is measured with feature-level combinations on classifiers. Compared to SVM, FCLSTM-CNN, and CNN which obtained final feature level combinations with accuracy of 88.1294%, 94.175%, and 95.1557%, respectively, FCBi-LSTM obtained 98.772% accuracy.

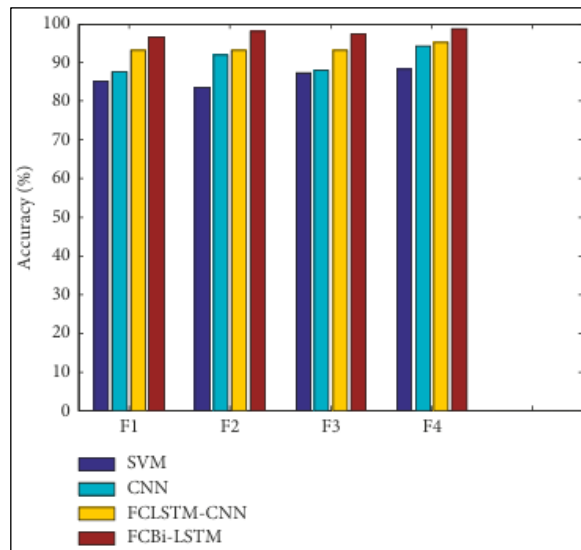


Figure 6: Feature level combination accuracy results versus classifiers.
Source: Authors, (2026).

Comparisons of classifiers with four different feature-level combinations are shown in Figure 7. Figure 4 illustrates that the final feature level combinations for SVM, CNN, and FCLSTM-CNN displayed increased error values of 11.8706 %, 5.8248 %, and 4.8443 %, respectively, but the FCBi-LSTM findings generated lower error values of 1.2280 %. The results presented in Figure 7 contrast classifier performances using four different feature level combinations. Figure 4 shows FCBi-LSTM yielded lower error values of 1.2280 % in its results. However, the final feature level combinations for CNN exhibited the error value of 5.8248 %, whereas SVM exhibited the error value of 11.8706 %, and FCLSTM-CNN exhibited the error value of 4.8443 %.

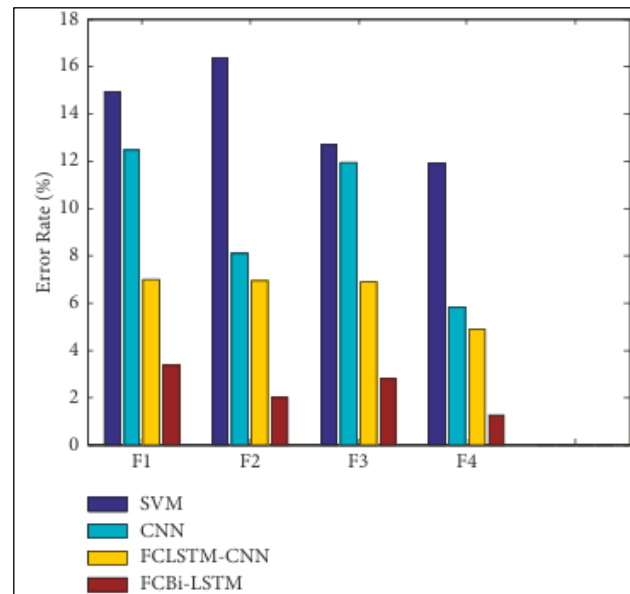


Figure 7: Feature level combination error results compared to classifiers.
Source: Authors, (2026).

A noticeably higher result of 74.30 % is obtained for the developed technique and 57.2007% for CNN, 56.20 % for SVM, and 67.6669 % for FCLSTM CNN classifiers, respectively for the first feature based combination when Figure 8 compared the MCC results, the TQWT+MFCC+ Concat Attribute set. The suggested approach produces better MCC results for all classifiers since it makes advantage of feature selection.

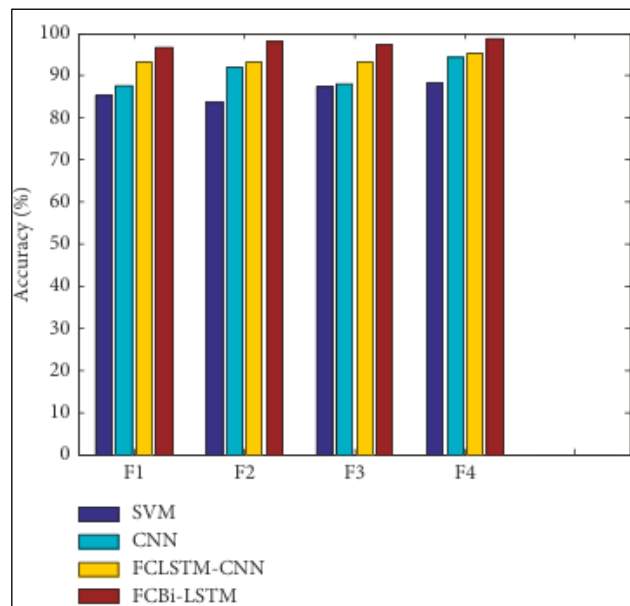


Figure 8: Feature level combination vs. classifiers MCC outcomes.
Source: Authors, (2026).

V. CONCLUSIONS

The second most common neurological condition, PD significantly impairs quality of life, is incurable, and has no known cure. Using neuro-protective and early treatment strategies for PD requires an early diagnosis. In this work, a multiclass classification problem for PD analysis is presented using a feature selection method. FCBi-LSTM and SqueezeNet with DMOA are offered for PD analysis. After that, PD is diagnosed using the FCBi-LSTM classifier. Medical practitioners may be able to better aid PD patients in their recovery and recuperation thanks to this dependable and precise model that accurately detects the illness in its early stages. Precision, recall, accuracy, F-measure, MCC, and F-measure were used to gauge each algorithm's efficacy in the classification process when DL libraries were being evaluated. The accuracy of the suggested model is evident when comparing the results with those of other approaches that are currently in use. Engineering and medicine have a bright future thanks to deep learning. The objective of upcoming work is to creatively expand on previous findings. Multiple data types can enter the network at once thanks to the suggested CNN's Simultaneous convolution layers. Using multimodal data, It allows us to categorize Parkinson's illness. In addition, the writers want to apply further deep learning algorithms to the classification process.

VI. ACKNOWLEDGMENTS

The author would like to appreciate the effort of the editors and reviewers.

VII. REFERENCES

- [1] Z. Wang, B. Hu, W. Zhou, M. Xu, and D. Wang, "Hopf bifurcation mechanism analysis in an improved cortex-basal ganglia network with distributed delays: An application to PD," *Chaos Solitons Fractals*, vol. 166, p. 113022, 2023.
- [2] A. Dovonou, C. Bolduc, V. S. Linan, C. Gora, M. R. Peralta, and M. Levesque, "Animal models of PD: Bridging the gap between disease hallmarks and research questions," *Transl. Neurodegener.*, vol. 12, p. 36, 2023.
- [3] P. Wilmschurst, S. Bewley, and P. Murray, "Hyperbaric oxygen therapy for the treatment of long COVID," *Clin. Med.*, vol. 23, pp. 99–100, 2023.
- [4] P. Mensah-Kane and N. Sumien, "The potential of hyperbaric oxygen as a therapy for neurodegenerative diseases," *GeroScience*, vol. 45, pp. 747–756, 2023.
- [5] P. Paplomatas, M. G. Krokidis, P. Vlamos, and A. G. Vrahatis, "An ensemble feature selection approach for analysis and modeling of transcriptome data in Alzheimer's disease," *Appl. Sci.*, vol. 13, p. 2353, 2023.
- [6] World Health Organization, "Parkinson disease," 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease>. Accessed: Jan. 31, 2023.
- [7] D. Larson and T. Simuni, "New dopaminergic therapies for PD motor complications," *Neuropharmacology*, vol. 204, p. 108869, 2022.
- [8] B. F. Mogharbel, M. A. Cardoso, A. C. Irioda, P. E. F. Stricker, R. C. Slompo, J. M. Appel, N. B. de Oliveira, M. C. Perussolo, C. S. Sacaki, N. N. da Rosa, et al., "Biodegradable nanoparticles loaded with levodopa and curcumin for treatment of PD," *Molecules*, vol. 27, p. 2811, 2022.
- [9] L. Smith and A. H. V. Schapira, "GBA variants and Parkinson disease: Mechanisms and treatments," *Cells*, vol. 11, p. 1261, 2022.
- [10] D. K. Khatri, K. Preeti, S. Tonape, S. Bhattacharjee, M. Patel, S. Shah, P. K. Singh, S. Srivastav, D. Gugulothu, L. Vora, et al., "Biotechnological advances for nose to brain delivery of therapeutics to improve the Parkinson therapy," *Curr. Neuropharmacol.*, vol. 21, pp. 493–516, 2022.
- [11] Y. Su, J. Liu, S. Tan, W. Liu, R. Wang, and C. Chen, "PLGA sustained-release microspheres loaded with an insoluble small-molecule drug: Microfluidic-based preparation, optimization, characterization, and evaluation in vitro and in vivo," *Drug Deliv.*, vol. 29, pp. 1437–1446, 2022.
- [12] A. C. Correia, A. R. Monteiro, R. Silva, J. N. Moreira, J. M. Sousa Lobo, and A. C. Silva, "Lipid nanoparticles strategies to modify pharmacokinetics of central nervous system targeting drugs: Crossing or circumventing the blood–brain barrier to manage neurological disorders," *Adv. Drug Deliv. Rev.*, vol. 189, p. 114485, 2022.
- [13] Z. L. Shi, Z. Y. Fan, H. Zhang, S. T. Li, H. Yuan, and J. H. Tong, "Localized delivery of brain-derived neurotrophic factor from PLGA microspheres promotes peripheral nerve regeneration in rats," *J. Orthop. Surg. Res.*, vol. 17, p. 172, 2022.
- [14] K. Saalwächter, "Grand challenges in polymers," *Front. Soft Matter*, vol. 2, p. 18, 2022.
- [15] R. Deb, S. An, G. Bhat, H. Shill, and U. Y. Ogras, "A systematic survey of research trends in technology usage for PD," *Sensors*, vol. 22, p. 5491, 2022.
- [16] N. E. Allen, C. G. Canning, L. R. S. Almeida, B. R. Bloem, S. H. Keus, N. Löfgren, A. Nieuwboer, G. S. Verheyden, T. P. Yamato, and C. Sherrington, "Interventions for preventing falls in PD," *Cochrane Database Syst. Rev.*, vol. 6, pp. 12–38, 2022.
- [17] W. Liu, T. Tung, C. Zhang, and L. Shi, "Systematic review for the prevention and management of falls and fear of falling in patients with PD," *Brain Behav.*, vol. 12, p. e2690, 2022.
- [18] R. Vitorio, M. Mancini, P. Carlson-Kuhta, F. B. Horak, and V. V. Shah, "Should we use both clinical and mobility measures to identify fallers in PD?" *Park. Relat. Disord.*, vol. 106, p. 105235, 2023.
- [19] D. P. Lima, S. B. De-Almeida, J. D. C. Bonfadini, A. H. S. Carneiro, J. R. G. de Luna, M. S. de Alencar, A. B. Viana-Júnior, P. G. B. Rodrigues, I. D. S. Pereira, J. D. S. Roriz-Filho, et al., "Falls in PD: The impact of disease progression, treatment, and motor complications," *Dement. Neuropsychol.*, vol. 16, pp. 153–161, 2022.
- [20] E. Maranesi, E. Casoni, R. Baldoni, I. Barboni, N. Rinaldi, B. Tramontana, G. Amabili, M. Benadduci, F. Barbarossa, R. Luzi, et al., "The effect of non-immersive virtual reality exergames versus traditional physiotherapy in PD older patients: Preliminary results from a randomized-controlled trial," *Int. J. Environ. Res. Public Health*, vol. 19, p. 14818, 2022.