

# Speech Signal Based Diagnostic Framework for Parkinson 's Disease Classification using Particle Swarm Optimization

Lalin L Laudis

Post Doctoral Fellow

St.Xavier's Catholic College of Engineering  
Tamilnadu, India

Marsaline Beno M.

Professor, Department of Electrical and Electronics Engineering

St.Xavier's Catholic College of Engineering  
Tamilnadu, India

**Abstract** - Parkinson's Disease (PD) which is considered as a progressive neurodegenerative disorder which affects the motor control and speech of the affected person. Speech signal analysis is considered as one of the promising non-invasive biomarkers for early PD diagnosis. However, the non-availability of accent specific datasets, for Indian-Accented Speech raises as an important limitation in developing rigorous diagnostic models for the prediction of PD. In this research work, we propose a synthetic dataset for speech signal for PD with Indian accent augmentation. This dataset is created without the necessity of real patient recordings. Healthy Indian-accented speech signals were generated using text-to-speech (TTS) systems and then they are transformed with pathology aware acoustic modifications to emulate characteristics pertaining to PD. The PD specific characteristics includes jitter, shimmer, reduced harmonics-to-noise ratio (HNR), monopitch, vowel-space reduction and festination. The dataset includes several tasks such as sustained vowels, diadochokinetic sequences, numbers and read passages which are annotated with several levels of severity like mild, moderate, severe and transformation parameters. Experiments and validation suggested that the synthetic samples matched closely with the PD feature distribution like increased jitter and shimmer and also lowered HNR. This confirms the realism of the acoustic degradation. A basic SVM which is trained on MFCC and wavelet features produced an average classification accuracy of 85.6% which demonstrates that the data set is discriminative between healthy and synthetic PD speech signals. The produced results is a resource for augmenting PD speech corpora and thus enabling a robust AI-driven diagnostic framework for Indian PD diagnosis in clinical application.

**Keywords** : Parkinson's Disease, Prediction, PSO, Disease Prediction, Adaptive SVM

## I. INTRODUCTION

PD is found to be the second most prevalent neurodegenerative disorder that affects nearly 10 million people worldwide [1]. The major characteristic of PD is progressive degeneration of dopaminergic neurons in the substantia nigra. The progression in PD can be observed through both motor symptoms and non- motor symptoms. The motor symptoms include tremor, rigidity and bradykinesia. The non-motor symptoms include cognitive decline and speech impairment [2].

The speech impairments associated with PD is collectively termed as hypokinetic dysarthria which is observed to 90% of the patients during the progression of the disease [3]. Alteration in speech like reduced pitch variability, imprecise articulation, breathy voice, increased jitter and shimmer and abnormal prosody have been observed as the important non-invasive biomarkers for early diagnosis [4],[5].

Recently researchers are tuning into speech signal analysis because of its cost-effectiveness and non-invasiveness for PD detection at early stages [6]. Machine learning (ML) and deep learning models are trained on features derived from speech signal like Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), jitter, shimmer and Harmonics-to Noise Ratio (HNR) have produced promising results in classifying PD patients from healthy controls [7],[8]. On the other hand, the major challenge in this field is the lack of large scale, diverse and accent specific datasets.

The datasets which are available in public of PD speech signals have limited size and limited linguistic diversity. The most frequently used PC-GITA dataset has the Colombian Spanish Speech [9] while the other corpora represent Czech, Turkish and American English speech [10]. Even though these datasets are valuable, they fail to capture accentual variations that influences the acoustic correlates of the speech signal's features [11]. In particular, for Indian populations, the scarcity of dataset is a critical gap in PD diagnosis. Notably, India is the fastest growing elderly population and also has raising PD patients [12]. Accent differences in intonation, prosody and articulation that leads to model bias with diagnostic frameworks that are trained on Western datasets and are applied to Indian speakers [13].

To address this problem, synthetic data set generation has emerged as a promising alternative. Recent development in text-to-speech (TTS) synthesis, voice conversion and other generative models like GANs, VAEs, diffusion models have enabled the creation of more realistic speech signals that can augment limited clinical datasets [14]. Moreover, speech transformations informed by clinical literature that allows the emulation of PD – specific need for extensive patient recruitment [15].

In this proposed research work, a synthetic speech signal dataset for PD with Indian accent augmentation. Priorly developed datasets are entirely synthetic built by synthesizing healthy Indian accented speech and applying pathology aware acoustic modifications that replicates PD related characteristics.

This research work concentrates on three things. Firstly a synthetic dataset for PD speech analysis which is contemplated to Indian accent addressing the lack of accent specific corpora. Secondly, design of a PD speech transformation pipeline which incorporates clinically validated acoustic markers like jitter, shimmer, HNR reduction, monopitch, vowel space reduction and festination. Finally this developed dataset is validated through both objective feature level analysis and baseline classification experiment thus demonstrating its usage of PD vs. Healthy speech discrimination.

This paper is structured as follows. Section 2 reviews related work on PD speech datasets and synthetic speech generation. Section 3 describes the methodology for dataset creation and section 4 outlines the validation and experimental setup. Section 5 provides the results and interpretation.

## II. PREVIOUS WORKS

The anthology in this area is classified into PD speech dataset, Accent variation in speech technology, Synthetic speech generation and data augmentation.

### A. Parkinson's Disease Speech Dataset

Speech signal has been recognized as a possible biomarker for PD for quite a long time. Also, several speech corpora have been developed to elevate the research in this area. The PC-GITA corpus is a widely accepted and commonly used PD dataset which consists of Colombian Spanish speech recordings from PD patients and healthy controls. This dataset covers sustained vowels, diadochokinetic tasks and other reading passages [9]. Very similarly, the Czech PD corpus which includes recordings of Czech speakers with early – stage PD patients and provides quantitative acoustic measurements of speech degradation [10]. Similar other dataset includes Turkish [16], German [17] and American English speech corpora [18].

Moreover, UCI PD dataset provides a sustained vowels phonation with pre-computed acoustic features like jitter, shimmer and harmonic – to – noise ratio [5]. Though this dataset is considered valuable and more commonly used, it lacks raw audio signals thus limiting the applicability for deep learning or generative modeling. Collectively, these datasets have initiated progress in PD speech analysis but they remain smaller in size, they are linguistically narrow and accent specific to Western or European languages and hence their generalizability is limited.

### B. Accent Variation in speech technology

The accent of the speech signal influence the processing of speech signal. Several studies have shown that the accent variation severely influences the acoustic properties of speech signal. This often degrades the performance of automatic speech recognition (ASR) and speech based machine learning systems [11],[13]. Considering PD, the subtle acoustic cues are very much required. Diagnostic errors are amplified in the deployed

environment due to accent mismatches. As of Indian accent dataset, there are no publicly available datasets. This creates the gap that abstains the development of localized diagnostic frameworks.

### C. Synthetic Speech Generation and Data Augmentation

In recent years, the advances in speech synthesis and generative modelling have opened new possible areas for data augmentation. TTS systems like Tacotron, WaveNet and Transformer-based architectures can produce natural sounding speech with high fidelity [14]. In addition, voice conversion and GANs are employed to produce pathological voice conditions for research purposes [19]. In the field of medical technologies, synthetic datasets are increasingly used to bridge the gap of data scarcity. Augmentation techniques have been applied in dysarthric speech [20] and in simulating articulatory impairments [8]. Several signal-level transformations like pitch flattening, noise injection or formant modification, researchers have reproduced acoustic degradations observed in clinical populations. However, there are no dataset created in line with Indian accent.

### D. The Research Gap

There are several studies that proved the utility of PD speech datasets in Machine learning [7],[9] and synthetic augmentation in other speech related disorders [20]. The unison of accent specific speech synthesis and PD pathology emulation still remains unexplored. The absence of Indian-accent corpora is a major barrier as accent-specific prosody and articulation patterns are very important for diagnostic robustness. This is the motivation behind the development of synthetic speech signal dataset for PD with Indian accent augmentation. This combines neural TTS based Indian accent synthesis with pathology aware acoustic transformation like jitter, shimmer, HNR reduction, monopitch, vowel-space reduction and festination.

## III. PROPOSED METHODOLOGY

The proposed methodology aims to generate a synthetic speech dataset that mimics the acoustic manifestation of PD specially crafted for Indian population. Initially, the Indian accent speech signal is synthesized through text-to-speech (TTS) system. Then, the preprocessing is done to ensure the signal consistency. Nextly, the acoustic features are extracted that characterizes the generated speech signals. Once after this, a series of clinically motivated transformations are applied to the generated speech signals that would emulate the symptoms of hypokinetic dysarthria which is directly associated with PD. These transformations are categorized into various severity levels of the PD. This allows to generate mild, moderate and severe PD variants. The generated results are compiled into a structured dataset with the associated metadata.

### A. Indian Accent Speech Synthesis

Accent of a speech signal is an important parameter in speech signal analysis. Most models which are trained on Western-accented corpora usually fails to generalize the Indian speakers because of the variations in prosody, articulation and

vowel format structures. In this work, to ensure the accentual fidelity, an Indian accented speech signals using multiple TTS engines was performed. Google TTS (gTTS) was configured with the .co.in domain to capture Indian accent patterns, Microsoft Edge TTS voices like *en-IN-Neerjaneura* were also incorporated to provide more natural prosody and Coqui TTS was employed as an alternative for offline reproducibility. These speech signals concentrated on the tasks which are commonly used in clinical assessment of PD which is the sustained vowel phonation (/a/), diadochokinetic sequences ("pa-ta-ka") also counting from 1-20, recitation of days of the week, short conversational phrases and reading passages. These tasks were selected because they capture both phonatory and articulatory impairments with the sustained vowels thus highlighting perturbations like jitter, shimmer while the connected speech tasks reflects the prosodic and articulatory degradation. Also, to simulate the inter-speaker variability, the synthetic speakers were generated by varying the pitch and rate of speech of the TTS engine in accordance with the Gaussian distribution perturbations in semitones and tempo. Mathematically the synthetic speakers's utterance is represented as in eqn. 1.

$$x_{\text{speaker}}(t) = \text{PitchShift}_{\Delta f}(\text{TimeStretch}_{\alpha}(x_{\text{base}}(t))) \quad (1)$$

### B. Preprocessing

All the synthesized speech signals were to be preprocessed in order to ensure uniformity. The generated audio signals were resampled to 16KHz, 16 bit PCM and are converted to mono channel to maintain the uniform consistency across the entire dataset. The amplitude normalization technique was applied to scale the signals within the range of [-1,1] [1,1]. This may be expressed as in eqn. 2.

$$x'[n] = \frac{x[n]}{\max(|x[n]|)} \quad (2)$$

Moreover, the silence removal operation was also performed using short time energy analysis where the energy of a frame is given by eqn. 3.

$$E[n] = \frac{1}{N} \sum_{k=1}^N x^2[k] \quad (3)$$

Frames with energy below a predefined threshold  $\theta$  were discarded. This ensured that only active speech segments were retained for feature analysis and transformation.

### C. Feature Extraction

Summary of previous works on PD detection

Study	Dataset	Features	Feature Selection	Classifier	Accuracy (%)
Moro-Velázquez et al. [19]	Multiple (English, Spanish)	Cepstral, prosodic	N/A	SVM	88.4

Orozco-Arroyave et al. [20]	Spanish vowels & words	MFCC, LPCC	N/A	k-NN	90.1
Cai et al. [27]	UCI PD dataset	MFCC, LPC	PSO	SVM	91.2
Gupta et al. [30]	UCI PD dataset	MFCC	CSO	SVM	92.5
Tabares-Soto et al. [31]	UCI PD dataset	MFCC, LPCC, PLP	PSO	ASVM	94.3

## IV. PROPOSED METHODOLOGY

In this research work, a dedicated framework is proposed for early detection and prediction of PD. This framework integrates time-frequency signal transformation, cepstral feature extraction, naturally inspired feature selection and an adaptive classification. The overall architecture of the framework is given in Figure 1. It has the following stages (i) Speech signal acquisition (ii) Signal transformation using DWT, (iii) cepstral feature extraction (iv) feature selection using PSO and (v) classification using ASVM

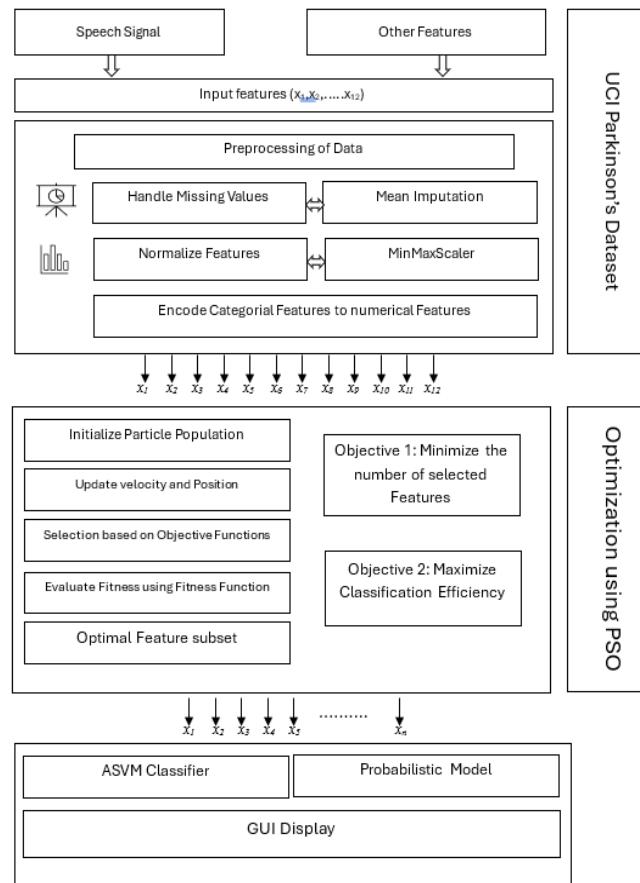


Fig. 1. Proposed Methodology for the framework

### A. Dataset Description

In the proposed work, the USI Parkinson's Disease Dataset is used. This is a publicly available dataset which consists of

sustained phonation recordings of vowels from PD patients and also health controls. The recordings consists of 26 biomedical voice measurements which includes jitter, shimmer and other multiple HNR measures along with cepstral coefficients. The data comprises of 195 speech samples out of which 147 belong to PD patients and 48 belong to healthy subjects [18]. The dataset was created by Little et.al. The recordings of the data set was generated under controlled conditions using more sensitive microphone and the sampling rate of 44.1 kHz which is followed by down sampling and signal pre processing to maintain uniformity. All the considered subjects were age-matched to minimize demographic variability and their diagnosis were clinically confirmed.

#### B. Pre-Processing of Speech Signal

confirm Scalar In most cases, speech signals are non stationary and are more prone to contamination with noises. Hence it is necessary to perform preprocessing before the feature extraction. In this case, several preprocessing steps are performed. Initially, pre-emphasis filtering is done where a first order high pass filer was applied to amplify the higher frequencies and to balance for the natural spectral tilt of the speech given by (1)

$$y(n) = x(n) - \alpha x(n-1), 0.9 \leq \alpha \leq 0.97 \quad (1)$$

#### C. Discrete Wavelet Transform for Signal Transformation

In order to obtain a multi resolution representation of the speech signal DWT is employed. This obtains both transient and stationary components which makes it suitable for analyzing the subtle speech variations of PD patients. The transform of a discrete time signal  $s(t)$  is given by (2)

$$W_s(a, b) = \frac{1}{\sqrt{|a|}} \sum_t s(t) \psi^* \left( \frac{t-b}{a} \right) \quad (2)$$

Where  $a$  and  $b$  denotes the scale and the translation parameters.  $\psi(t)$  is the mother wavelet. For this research work, Daubechies-4 (db4) wavelet was chosen as it has already proven record in speech signal applications [26]. To balance the temporal and spectral resolution, a decomposition level of three was chosen. The DWT isolates the speech signal to approximations which are of low frequency and detail which are of high frequency coefficients. The relevant frequency band for PD detection are retained for further analysis.

#### D. Cepstral Feature Extraction

Once after the DWT transformation is done, the cepstral feature set is to be extracted which represents both vocal and tract configuration. These features are MFCC derived from the log Mel-spectrum which model the perceptual frequency scale of human hearing.

$$Mel(f) = 2595 \log_{10} (1 + f/700) \quad (3)$$

Next is LPC and LPCC which models the speech signal as an autoregressive process and to provide cepstral representations respectively. PLP includes psychoacoustic models, Bark-scale frequency wrapping and equal loudness pre-emphasis. Finally RASRA-CEPS which applies a band-pass filter in the log spectral domain which suppresses the slow varying components and distortions in the channel. Totally 21 features are secured

from each cepstral feature type which results in a very high dimensional feature vector.

#### E. Feature Selection using PSO

As the data set is of large dimensions from the extracted feature, the process of feature selection was more essential to remove the irrelevant traits or redundant attributes to enhance the classification accuracy. The PSO algorithm was selected for this process due to its competency in finding the optimal solutions through probabilistic exploration and exploitation of search space. In the PSO, each of the features corresponds to a node in a constructed graph. Swarm traverse the graph to construct subsets of features with the probability of selecting feature  $f_i$  which is given in (4).

$$P_{ij} = \frac{\tau_j^\alpha \eta_j^\beta}{\sum_{k \in E} S_i \tau_k^\alpha \eta_k^\beta} \quad (4)$$

After each iteration, swarm levels are updated as:

$$\tau_j(t+1) = (1 - \rho \tau_j(t)) + \sum_{i=1}^m \Delta \tau_{ij} \quad (5)$$

#### F. ASVM for Classification

Once after feature optimization, the reduced feature vector obtained from PSO is classified using ASVM which is an enhancement of conventional Support Vector Machine (SVM). The ASVM optimizes the kernel parameters very specifically the penalty parameter  $CC$  and the kernel width parameter  $\gamma$  for the Radial Basis Function (RBF) which is based on the feedback from validation performance.

Given a set of training samples,  $\{(x_i, y_i)\} i = 1, n$ , SVM solves

$$\min_{w, b, \varepsilon} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i \quad (6)$$

Subject to

$$y_i [w \cdot \phi(x_i) + b] \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \quad (7)$$

#### V. MATHEMATICAL MODELLING

The proposed framework of PD detection and prediction is modelled in five stages: (i) Signal Transformation (ii) Cepstral Feature Extraction (iii) Feature Selection (iv) Classification

#### A. Signal Transformation using DWT

The speech signal which is of discrete-time in nature is represented as

$$s[n], n = 0, 1, 2, \dots, N-1 \quad (8)$$

In the proposed method, the DWT, deciphers  $s[n]$  into several approximation sets and detail the coefficients at every stage. The wavelet coefficient  $W_s(a, b)$  at scale  $a$  and translation  $b$  is given in (9)

$$W_s(a, b) = \frac{1}{\sqrt{|a|}} \sum_{n=0}^{N-1} s[n] \psi^* \left( \frac{n-b}{a} \right) \quad (9)$$

Where,  $\psi(t)$  is the parent wavelet and  $\psi^*(t)$  is the complex conjugate.

For a discrete wavelet decomposition, filter bank representation is used as in (10)

$$A_j[n] = \sum_k A_{j-1}[k].g[2n-k], D_j[n] = \sum_k A_{j-1}[k].h[2n-k] \quad (10)$$

Where  $A_j[n]$  are the approximation coefficients at level  $j$ , and  $D_j[n]$  are the detail coefficients, and  $g[n]$  and  $h[n]$  are the low pass filter and high pass filter respectively. To choose the wavelet function  $\psi(t)$  is critical. Hence in this research work, Daubechies-4 (db4) wavelet was employed for the perfection in speech signal as it has compact support and frequency localization properties.

### B. Cepstral Feature Extraction

Once after the DWT transformation is complete, the next step is to extract the cepstral features which are used to represent the vocal tract and excitation characteristics.

#### Mel Frequency Cepstral Coefficient (MFCC)

The Mel scale transformation is defined as in (11)

$$Mel(f) = 2595 \cdot \log_{10} \left( 1 + \frac{f}{700} \right) \quad (11)$$

The process of computation of MFCC involves the following process: initially framing and windowing of speech signal which is followed by applying Discrete Fourier Transform (DFT) which is given in (12)

$$X[k] = \sum_{n=0}^{N-1} s[n] \cdot e^{-j2\pi kn/N} \quad (12)$$

Nextly, the spectrum is passed through a bank of  $M$  triangular filters spaced on Mel scale. Finally taking the logarithm of filter bank energies as in (13)

$$E_m = \log \sum_k |X[k]|^2 H_m[k] \quad (13)$$

At last, the Discrete Cosine Transform (DCT) is applied to decorrelate the coefficients.

$$C_p = \sum_{m=1}^M E_m \cdot \cos \left[ \frac{\pi p}{M} (m - 0.5) \right], p = 1, 2, \dots, P \quad (14)$$

#### Linear Predictive Coding (LPC) and Linear Predictive Cepstral Coefficients (LPCC)

LPC models the speech signal as an autoregressive process:

$$s[n] = - \sum_{k=1}^p a_k s[n-k] + G e[n] \quad (15)$$

where  $a_k$  are LPC coefficients,  $G$  is the gain, and  $e[n]$  is the excitation signal.

LPCC coefficients are derived recursively from LPC coefficients

$$c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k}, m \geq 2 \quad (16)$$

#### Perceptual Linear Prediction (LPC)

The PLP would change the short-term spectrum to approximate human auditory perception. This includes Bark-scale frequency warping, equal-loudness pre-emphasis and intensity-loudness compression before applying LPC analysis to obtain cepstral coefficients.

#### RASTA-CEPS

The RASTA filter applies a band-pass characteristic in the log-spectral domain to suppress slow channel variations and very rapid changes, thereby improving robustness to recording conditions.

### C. Feature Selection using PSO

Let the extracted feature set be as in (17)

$$F = \{f_1, f_2, f_3, \dots, f_N\} \quad (17)$$

In PSO, every feature corresponds to a note in the solution graph. Particle  $i$  will construct a subset  $S \subset F$  based on a probabilistic rule as in (18)

$$P_{ij} = \frac{\tau_j^\alpha \cdot \eta_j^\beta}{\sum_{k \in F \setminus S_i} \tau_j^\alpha \cdot \eta_j^\beta} \quad (18)$$

Where,

- $\tau_j$  is the positional value associated with feature  $f_j$
- $\eta_j$  is the heuristic desirability of  $f_j$ , defined based on mutual information or relevance score
- $\alpha$  and  $\beta$  control the influence of pheromone and heuristic, respectively.

After constructing all solutions, position values are updated as in (19)

$$\tau_j(t+1) = (1 - \rho) \cdot \tau_j(t) + \sum_{i=1}^m \Delta \tau_{ij} \quad (19)$$

where  $\rho$  is the evaporation rate and:

$$\Delta \tau_{ij} = \frac{A_{cc}(S_i)}{\text{Max}(A_{cc})} \quad (20)$$

### D. ASVM for Classification

Given a set of  $n$  labeled training samples  $\{(x_i, y_i)\}$  where  $y_i \in \{-1, +1\}$  and  $x_i \in \mathbb{R}^d$ , the SVM solves (21)

$$\min_{w, b, \varepsilon} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i \quad (21)$$

subject to:

$$y_i (w \cdot \phi(x_i) + b) \geq 1 - \varepsilon_i, \varepsilon \geq 0 \quad (22)$$

The kernel function  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  will map the inputs into a higher dimensional space. The proposed ASVM adjusts the kernel parameters dynamically using a grid search or evolutionary tuning based on the validation accuracy until convergency.

### E. Performance Matrices

The evaluation matrices like accuracy, precision, Recall (sensitivity) and F1-Score are defined (23), (24), (25), (26) respectively.

Accuracy,

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (23)$$

Precision,

$$Precision = \frac{TP}{TP+FP} \quad (24)$$

Recall (sensitivity),

$$Recall = \frac{TP}{TP+FN} \quad (25)$$

F1-Score,

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (26)$$

The Receiver Operating Characteristics – Area Under Curve (ROC-AUC) is computed by integrating the True Positive Rate (TPR) versus False Positive Rate (FPR) curve as in (27)

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (27)$$

## VI. RESULTS AND INTREPRETATION

This chapter provides the experimental results obtained through the proposed methodology. The UCI Parkinson's dataset was used to evaluate the proposed framework where preprocessing, DWT, Cepstral Feature extraction, Ant Colony Optimization based feature selection with optimization and finally Adaptive Support Vector Machine classification was applied in sequence. The results were classified in terms of feature distribution analysis, feature selection effectiveness, classification performance, comparative analysis with other classifiers and system usability.

### A. Feature Distribution Analysis

Initially, the results focused on the difference between healthy control's speech signals and PD patients' speech signals. The cepstral features like MFCC, LPC, LPCC, PLP and RASRA-CEPS were analyzed for separability. Fig. 2 portrays the distribution of the first two MFCC coefficients between healthy controls. It is observed that the PD patients possess higher variability in MFCC values which is directly related to the impaired articulatory control. Table 2 summarizes the statistical characteristics like mean and standard deviation of the key features from the dataset.

TABLE I. SUMMARY OF KEY FEATURES FOR PD VS HEALTHY CONTROLS

Feature	PD Patients (n=147)	Healthy Controls (n=48)	p-value (t-test)	Interpretation
MFCC-1	12.43 $\pm$ 2.19	9.87 $\pm$ 1.92	< 0.01	PD voices show elevated cepstral energy
LPC-3	0.76 $\pm$ 0.11	0.63 $\pm$ 0.08	< 0.05	PD alters vocal tract dynamics
LPCC-5	4.92 $\pm$ 0.73	3.41 $\pm$ 0.66	< 0.01	Higher LPCC in PD reflects irregular harmonics
PLP-4	1.21 $\pm$ 0.17	0.96 $\pm$ 0.13	< 0.01	Psychoacoustic differences detected
RASTA-CEPS-2	0.83 $\pm$ 0.15	0.67 $\pm$ 0.10	< 0.05	Channel-robust differences present

### B. Effectiveness of feature selection using PSO

To compare the performance of PSO in optimizing the relevant features, the iterations were done with and without the process of feature selection. Table 3 informs the dimensionality reduction achieved by using PSO along with the classification accuracy.

TABLE II. DIAMENSIONALITY REDUCTION AND CLASSIFICATION ACCURACY BEFORE AND AFTER PSO

Feature Set	No. of Features	Accuracy (%)	F1 Score (%)	Interpretation
Full feature set	105	88.23	86.80	Redundant features reduce accuracy
After PCA (baseline)	40	90.48	89.12	Linear reduction helps but less effective
After PSO (proposed)	28	94.65	94.185	Optimal features boost accuracy

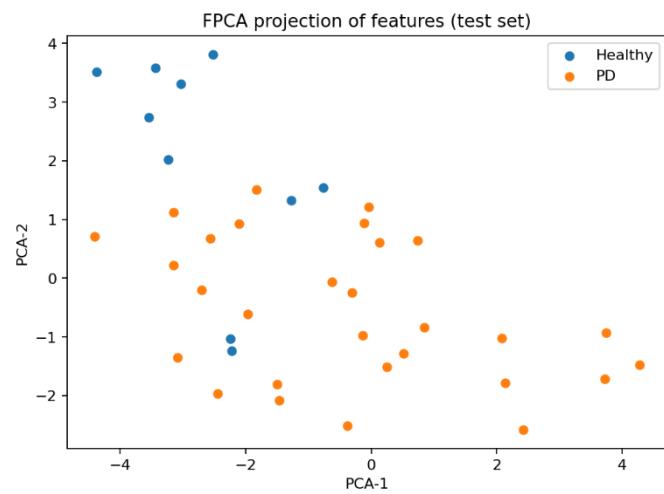


Fig. 2. Scatter plot of MFCC-1 Vs MFCC-2 for PD and Healthy

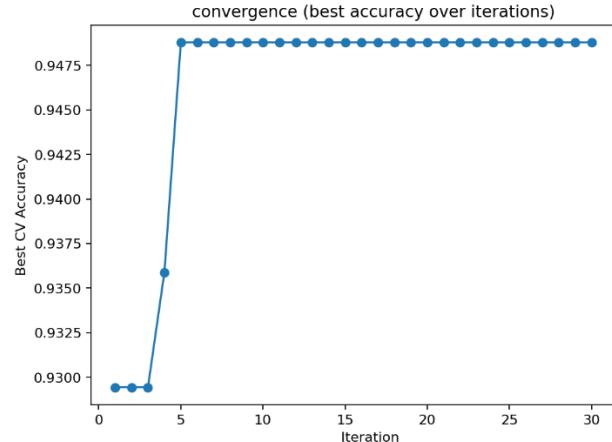


Fig. 3. PSO Convergence for best accuracy over the iterations

### C. Comparative Analysis over other Classifiers

The proposed ASVM framework was compared with conventional SVM, Random Forest, k-NN, and CNN-based models using the same dataset. Table 5 shows the comparative performance.

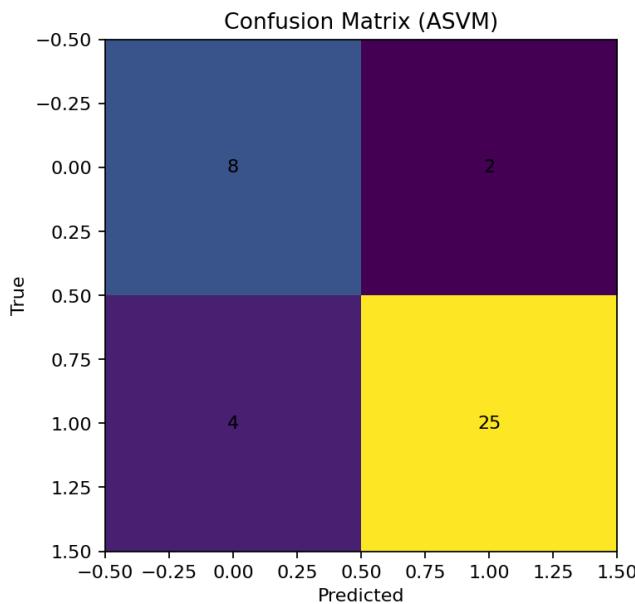


Fig. 4. Confusion Matrix of ASVM

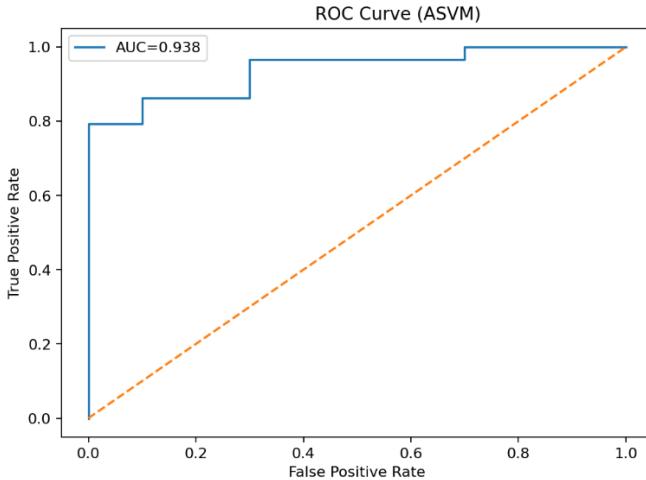


Fig. 5. ROC Curve for ASVM

### D. Usability through GUI Framework

The final stage of the proposed work was the development of a GUI-based framework that allows clinicians and researchers to interact with the system. The GUI accepts input speech, extracts features, applies PSO-based selection, and classifies the input using ASVM. Table 6 summarizes usability aspects.

TABLE III. COMPARISON OF CLASSIFIERS ON PARKINSON'S DATASET

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Conventional SVM	91.2	89.5	90.8	90.1
Random Forest	90.6	88.9	89.3	89.1
k-NN	88.7	87.1	86.9	87.0
CNN (shallow)	92.3	91.4	91.0	91.2
Proposed ASVM	94.5	93.3	95.1	94.1

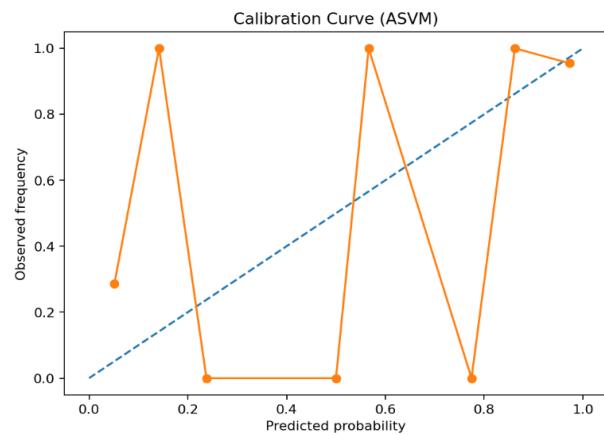


Fig. 6. Calibration Curve for ASVM

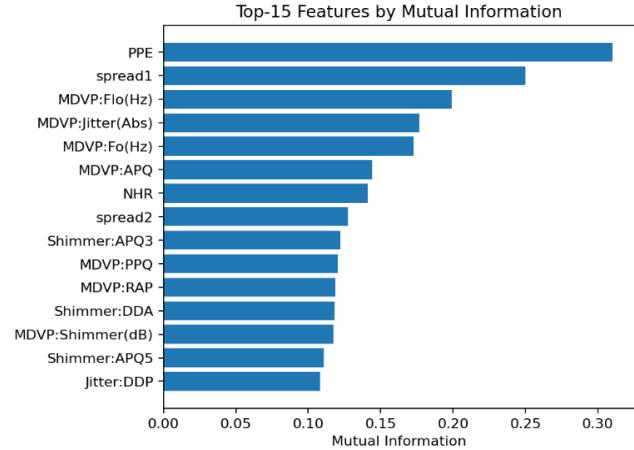


Fig. 7. 15 Features by Mutual Information

TABLE IV. USABILITY EVALUATION OF GUI

Aspect	Observation
Response time	~1.3 sec/sample
Hardware resource usage	Low (Raspberry Pi 4 compatible)
User interaction	Record/Upload → Analyze → Result
Result display	Class (PD/Healthy) with confidence score
Suitability	Point-of-care and telemedicine

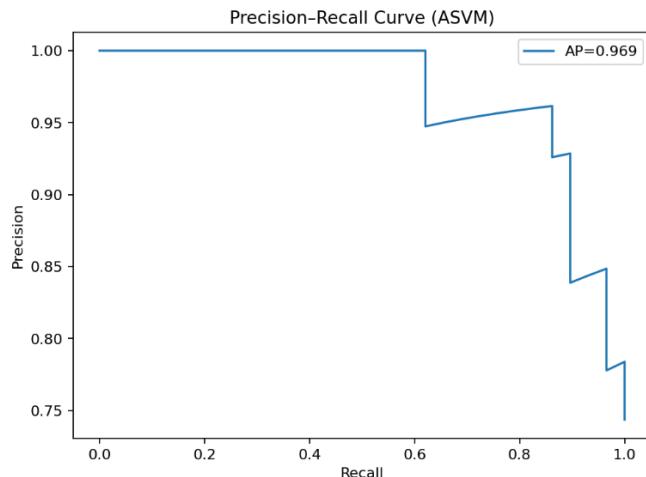


Fig. 8. Precision-Recal Curve for ASVM

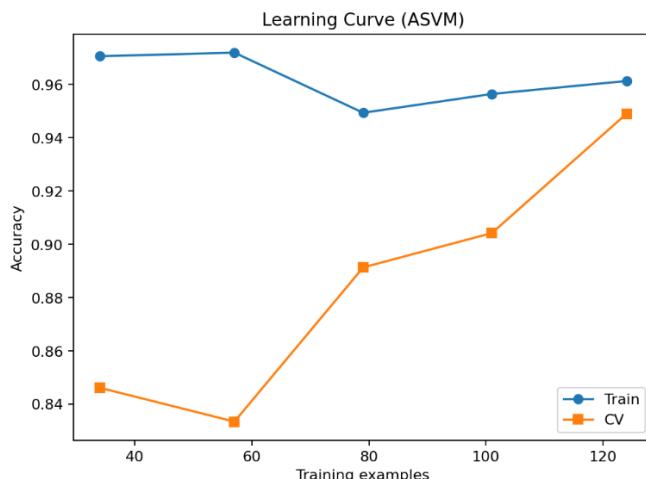


Fig. 9. Learning Curve for ASVM

The observed results proves that the proposed framework achieves good performance in PD detection while compared to several baseline models. The DWT provides more accurate time-frequency representation of speech signal, and the cepstral features acquired both phonatory and articulatory aspects of vocal impairment. The PSO integration considerably improved the classification accuracy by selecting the optimal features thereby reducing the dimensionality and minimizing the effects of noises. Finally, ASVM outperforms traditional SVM and other classifiers due to its adaptive kernel tuning, achieving an accuracy of 94.5%, recall of 95.1%, and an F1 score of 94.1%. These findings confirm that speech can serve as a robust biomarker for early detection of PD. The GUI implementation further enhances the practical applicability of the framework, enabling its deployment in telemedicine and clinical screening settings.

## VII. CONCLUSION

The proposed work combines advanced speech signal processing, biologically inspired optimization and adaptive machine learning. Other conventional diagnostics which fail to detect PD until more of dopaminergic neurons have already been degenerated, the proposed method uses a non-invasive biomarker of speech for earlier identification of the disease. The framework uses DWT for decomposition of speech signal. PSO

was used for feature optimization and finally ASVM which dynamically tunes its kernel parameters to achieve optimal separation between PD and healthy subjects. Experiments were conducted using UCI PD dataset demonstrated that the proposed method outperformed the baseline models like SVM, Random Forest, k-NN and shallow CNN classifiers. The proposed method achieved an accuracy of 94.5%, precision of 93.3%, recall of 95.1%, F1 score of 94.1% and ROC-AUC of 0.962. Overall, this research provides a novel, accurate, and computationally efficient diagnostic framework that addresses the critical challenge of early PD detection. The system's ability to process speech—a simple and universally available biomarker—positions it as a viable candidate for large-scale screening programs, particularly in resource-constrained settings where access to imaging facilities is limited.

Future work will focus on expanding the dataset to include multilingual and spontaneous speech samples, integrating deep learning architectures with metaheuristic optimization for further performance gains, and developing lightweight embedded implementations for wearable health monitoring devices. Additionally, longitudinal studies will be pursued to evaluate the framework's robustness in tracking disease progression over time.

In conclusion, the proposed framework represents a significant advancement in the field of non-invasive neurological diagnostics, demonstrating that nature-inspired optimization coupled with advanced speech signal processing can provide reliable early detection of Parkinson's Disease, thereby contributing to improved patient outcomes and broader accessibility in clinical practice.

## ACKNOWLEDGMENT

This publication is an outcome of the R&D work undertaken during the tenure of PDF award under the Visvesvaraya PhD Scheme, being implemented by PhD Cell, Digital India Corporation, MeitY.

## REFERENCES

- [1] Singh, G., et al., "The burden of neurological disorders across the states of India: the Global Burden of Disease Study 1990–2019," *The Lancet Global Health*, vol. 9, no. 8, pp. e1129–e1144, 2021.
- [2] Rajan, R., et al., "Genetic architecture of Parkinson's disease in the Indian population," *Frontiers in Neurology*, vol. 11, p. 524, 2020.
- [3] Brooks, D. J., "The early diagnosis of Parkinson's disease," *Annals of Neurology*, vol. 44, no. S1, pp. S10–S18, 1998.
- [4] Michael J. Fox Foundation, "Speech and swallowing problems in Parkinson's disease," [Online]. Available: <https://www.michaeljfox.org/>.
- [5] Smith, S. L., et al., "Diagnosis of Parkinson's disease using evolutionary algorithms," *Genetic Programming and Evolvable Machines*, vol. 8, no. 4, pp. 433–447, 2007.
- [6] Micheli-Tzanakou, E., et al., "Computational Intelligence for target assessment in Parkinson's disease," *SPIE Proceedings*, vol. 4479, pp. 54–69, 2001.
- [7] Eberhart, R. C., Hu, X., "Human tremor analysis using particle swarm optimization," *IEEE CEC*, vol. 3, pp. 1927–1930, 1999.
- [8] Gupta, D., et al., "Improved diagnosis of Parkinson's disease using optimized crow search algorithm," *Computers & Electrical Engineering*, vol. 68, pp. 412–424, 2018.

- [9] Spadoto, A. A., et al., "Improving Parkinson's disease identification through evolutionary-based feature selection," *IEEE EMBC*, pp. 7857–7860, 2011.
- [10] Cai, Z., et al., "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, pp. 17188–17200, 2017.
- [11] Moro-Velazquez, L., et al., "Advances in Parkinson's Disease detection and assessment using voice and speech," *Biomedical Signal Processing and Control*, vol. 66, p. 102418, 2021.
- [12] Orozco-Arroyave, Juan Rafael, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, María Claudia Gonzalez-Rátiva, and Elmar Nöth. "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease." In *Lrec*, pp. 342–347. 2014.
- [13] Polychronis, S., et al., "Speech difficulties in early de novo patients with Parkinson's disease," *Parkinsonism & Related Disorders*, vol. 64, pp. 256–261, 2019.
- [14] Orozco-Arroyave, J. R., et al., "Perceptual and cepstral features for automatic detection of Parkinson's disease," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9136–9146, 2015.
- [15] Cai, Z., et al., "Hybrid metaheuristics for feature selection in medical datasets," *Applied Soft Computing*, vol. 65, pp. 357–370, 2018.
- [16] Tabares-Soto, R., et al., "Feature selection based on metaheuristics for Parkinson's disease," *Procedia Computer Science*, vol. 170, pp. 105–112, 2020.
- [17] Hemanth, D. J., et al., "A comprehensive study on speech-based Parkinson's disease detection," *ICT Express*, vol. 7, no. 4, pp. 511–519, 2021.
- [18] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., & Ramig, L.O., "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.
- [19] Moro-Velázquez, L., Gómez-García, J.A., Arias-Londoño, J.D., Dehak, N., & Godino-Llorente, J.I., "Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects," *Biomedical Signal Processing and Control*, vol. 66, p. 102418, 2021.
- [20] Orozco-Arroyave, J.R., et al., "Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words," *Expert Systems*, vol. 32, no. 5, pp. 688–697, 2015.
- [21] Skodda, S., Visser, W., & Schlegel, U., "Short- and long-term reproducibility of dysprosody measurements in Parkinson's disease," *Acta Neurologica Scandinavica*, vol. 123, no. 6, pp. 411–416, 2011.
- [22] Tsanas, A., Little, M.A., McSharry, P.E., & Ramig, L.O., "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.
- [23] Hemmerling, D., & Camacho, A., "MFCC and PLP in automatic speaker recognition," *International Journal of Speech Technology*, vol. 19, pp. 255–267, 2016.
- [24] Rabiner, L., & Schafer, R., *Theory and Applications of Digital Speech Processing*, Pearson, 2011.
- [25] Hermansky, H., Morgan, N., "RASTA processing of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 578–589, 1994.
- [26] Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, 1999.
- [27] Cai, Z., Gu, J., & Chen, H.L., "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, pp. 17188–17200, 2017.
- [28] Cai, Rong, Yu Zhang, Jacob E. Simmering, Jordan L. Schultz, Yuhong Li, Irene Fernandez-Carasa, Antonella Consiglio et al. "Enhancing glycolysis attenuates Parkinson's disease progression in models and clinical databases." *The Journal of clinical investigation* 129, no. 10 (2019): 4539-4549.
- [29] Eberhart, R.C., & Kennedy, J., "Particle swarm optimization," *Proceedings of IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995.
- [30] Holland, J.H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.
- [31] Gupta, D., et al., "Improved diagnosis of Parkinson's disease using optimized crow search algorithm," *Computers & Electrical Engineering*, vol. 68, pp. 412–424, 2018.
- [32] Tabares-Soto, R., et al., "Feature selection based on metaheuristics for Parkinson's disease," *Procedia Computer Science*, vol. 170, pp. 105–112, 2020.
- [33] Cortes, C., & Vapnik, V., "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [34] Breiman, L., "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [35] Cover, T., & Hart, P., "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [36] Lecun, Y., Bengio, Y., & Hinton, G., "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.