

Speech Signal As A Suitable Biomarker For Early Prediction Of Parkinson's Disease Using Machine Learning

Shah Parveez Ahmed^{1*}, Dr. Siddhartha Baruah²

^{1*}Research Scholar, Assam Science and Technology University, Guwahati, India

Email: parveez82@gmail.com

²Professor, Dept. of Computer Application, Jorhat Engineering College, Jorhat, India

Email: sbaruain@gmail.com

***Corresponding Author**

Shah Parveez Ahmed

Research Scholar, Assam Science and Technology University, Guwahati, India

Email: parveez82@gmail.com

ABSTRACT

The prime objective of this study is to discover suitable biomarkers that may be helpful in the detection of Parkinson's disease (PD) at an early stage. Researchers used different biomarkers like imaging of brain, gait disorders and speech impairment to detect PD. Imaging biomarkers are useful in diagnosing at the clinical level and assessing effectiveness of new treatments; however, they are expensive, in some cases invasive and require specially trained medical personnel. Due to a high degree of variability across individuals, existing machine learning (ML) algorithms may fail to classify gait with the accuracy needed in clinical tests. In this study, speech-based biomarker is considered in the diagnosis of PD, because approximately 90% of PD patients suffer from speech disorder at an early stage. Moreover, speech impairment is a motor symptom that can be detected five years prior to the clinical diagnosis of PD. This study is conducted using voice samples collected from UCI machine learning repository, keeping an eye on the issues like class imbalance, data preprocessing and feature selection that most of the past studies did not resolve properly. ML techniques such as MLP, XGBoost and EasyEnsemble are applied on speech features to predict PD. Mutual Information is adopted to carry out feature selection. Finally, performance of the ML models is evaluated by using K-Fold and Stratified K-Fold cross validation methods. By addressing the issues faced by the previous studies we propose an XGBoost-based hybrid model named MIXSYN which shows better result with 94.6% of accuracy.

Keywords —ensemble learning, machine learning, Parkinson's disease, speech impairment, speech signals

How to cite this article: Shah Parveez Ahmed^{1*} Dr. Siddhartha Baruah². Speech Signal As A Suitable Biomarker For Early Prediction Of Parkinson's Disease Using Machine Learning. Int J Drug Deliv Technol. 2026;16(84s):834-846. DOI: 10.25258/ijddt.16.84s.92.

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disease after the Alzheimer [1][2]. It is a progressive disease which is caused by the death of dopamine-producing neurons in the substantia nigra of the brain [1][3]. Dopamine is a neurotransmitter involved in movement control of our body.

The substantia nigra is a vital part of the brain for the production of a neurochemical called dopamine. Dopamine affects the central nervous system by disrupting functions including movement control, emotional limbic activity and cognitive executive functions. Parkinson's disease is caused by the deficit of dopamine producing neurons in the substantia nigra [4]. Figure 1 depicts the substantia nigra which is located in the basal ganglia circuit of the brain.

Braak and colleagues [5] proposed the staging of brain pathology related to PD. In Stages 1 and 2 the involvement confined to the medulla oblongata. Stages 3 and 4 involve the midbrain including the substantia nigra. Stages 5 and 6 observe severe involvement of the brain, including neocortical areas. The degree of damage in the subcortical and mesocortical structures increases.

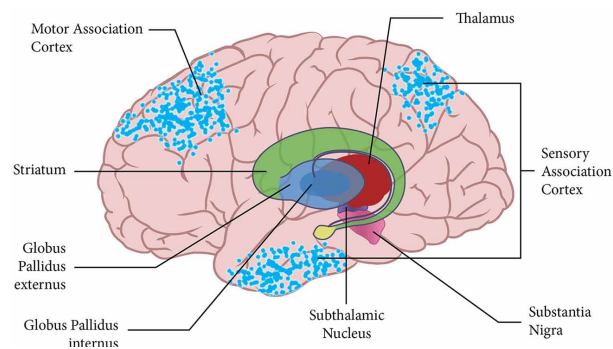


Figure1: The basal ganglia circuit and thalamus.

Clinical diagnosis of PD is mainly based on the motor symptoms. It includes shakiness in patients, tremors and stiffness when moving their limbs, bradykinesia (slowness of movement), FoG (Freeze of Gait), Micrographia (Small hand writing) and problems with balance and falls as the disease progresses [6]. As described by [3], [7], [8], [9], [10], [11], [12], [13], and [14] the major signs of PD are bradykinesia, rigidity, tremor and postural instability. Dopamine deficiency in the substantia nigra causes these symptoms to manifest. There is premotor stage of PD. The damage to specific neurons in the central nervous system (CNS) leads to

abnormalities in autonomic function that can serve as predictive biomarkers. These abnormalities are referred to as premotor parkinsonian phenotype. It includes constipation, urinary and sexual dysfunction, gastrointestinal dysfunction and cardiovascular dysfunction [15]. [16] mentioned non-dopaminergic and non-motor symptoms such as constipation, olfactory deficit, REM (Rapid eye movement) sleep behavior disorder, depression, fatigue, nausea, pain that appear with the progression of PD. These non-motor symptoms contribute to severe disability and diminished quality of life. Recognition of non-motor symptoms will form the basis of enhanced treatments that can slow down or prevent the progression of PD and its multicentric neurodegeneration and gives hope of better quality of life for the sufferer.

2. RELATED WORKS

Machine learning has a good history in medical field. Several researches have been done that demonstrated the application of machine learning techniques in diagnosis, prediction and identification of disease. In many recent studies, researchers applied machine learning techniques to classify the patients with PD and healthy subjects of control, based on imaging of brain, motion capabilities in the lower limbs (Gait) and speech signals. Overview of different biomarkers used in the diagnosis of PD is given in the subsequent sections.

2.1 Imaging

Structural and functional neuroimaging are used in neurological studies to illuminate neurodegenerative disease mechanisms, to assist in diagnosis and to assess the effectiveness of new treatments [17]. Diagnosis of neurodegenerative disorders using medical imaging is a difficult task due to a variety of reasons, such as different artifacts in the image acquisition process, errors involved in preprocessing, and inevitable inter-subject Table 1: Some imaging features used in previous experiments

	Feature	Imaging Type	Study
1	Striatum images	SPECT	[46]
2	Striatal uptake ratios	SPECT	[47]
3	Striatal uptake and uptake ratios	SPECT	[48]
4	Haralick features Computed over the image	SPECT	[49]
5	Voxel intensities, Striatal binding ratios	SPECT	[50]
6	Normalized intensity uptake values of putamen, thalamus, anterior cingulate gyrus, pars opercularis	PET	[51]

variabilities [18]. Imaging diagnostic techniques in the investigation of cerebral pathologies to confirm PD require specially trained dedicated personnel and are invasive, expensive, or not specific enough, such as the Transcranial Sonography [19]. Most of the Computer-Aided-Diagnosis (CAD) systems developed for automatic diagnosis of PD, collect information from medical images in the form of VAF (Voxels-As-Features), textural patterns or morphological features among others, and then use ML techniques to classify whether a patient is suffering the disease or not [20]. The dopamine transporters (DAT) measurements can be performed by using tomographic techniques, such as single-photon emission computed tomography (SPECT) or positron emission tomography (PET) [21]. NMR-DTI (nuclear magnetic resonance with diffusion tensor imaging) was relevant in studying the white matter integrity in PD pathophysiology [22]. Galantucci et al. [23] evaluated the structural NMR-DTI connectome of people with PD and cognitive impairment to find a structural network disruption in certain brain networks involving basal ganglia. Xu Nan et al. discovered a biomarker of PD detection. They used rs-fMRI (resting-state functional MRI) to identify the default functional network of the brain and then applied topological dimensional reduction to the data before using machine learning for classification and sensitivity analysis [44]. Mushta Illia et al. identified contralateral putamen SBR as the most predictive feature for the detection of PD [45].

Different imaging techniques, such as SPECT, PET, Structural MRI and Functional MRI have been applied by researchers to collect data to be used in machine learning in the detection of PD. Some imaging features used in the experiments are presented in Table 1.

7	Normalized intensity uptake values of caudate, putamen, thalamus, olfactory, and SMA	PET	[52]
8	Striatal binding ratios, Gray matter, white matter and CSF volumes	SPECT, PET	[53]
9	Texture features of uptake data	PET	[54]
10	Whole-brain uptake data	PET	[55]
11	Gray and white matter volumes; Fractional Anisotropy and Mean Diffusivity from DTI	Structural MRI	[56]

12	Volumetric measures of ROIs in gray matter, white matter, brain lobes, cerebellum, midbrain etc.	Structural MRI	[57]
13	Gray and white matter volumes	Structural MRI	[58]
14	Discretized Voxel Intensity Changes	Structural MRI	[59]
15	Gray and white matter, Cerebrospinal Fluid volumes, cortical thickness etc.	Structural MRI	[60]

16	Network-based whole-brain Functional Connectivity	Functional MRI	[61]
17	Functional Connectivity obtained from ROI pairs	Functional MRI	[62]
18	Functional Connectivity from brain networks	Functional MRI	[63]
19	Amplitude of Low Frequency Fluctuations, Fractional Amplitude of Low Frequency Fluctuations	Functional MRI	[64]

2.2 Gait

Gait disorder is a cardinal sign used by clinicians to diagnose and assesses the progress of PD. Typically gait impairments are examined by measuring spatial and temporal parameters [24]. Gait is generally assessed subjectively by clinicians; hence it may be difficult to distinguish between different gait disorders [25]. Existing Machine Learning algorithms may fail to classify gait with the accuracy needed in clinical trials, because it possesses a high degree of variability across individuals, and is prone to a lot of annoying factors [26]. According to Beavers et al. and Hamer et al. as cited in [26], “Basic temporal (steptime/cadence) and spatial (stride-length and walking speed) parameters cannot be used as discriminative function, as they lack disease specificity”. Rovini E. et al. assessed the gait performance of the subjects by following an experimental protocol adopted by neurologist, which includes tasks described in the MDS-UPDRS (Movement Disorder Society-Unified Parkinson's Disease Rating Scale). In this experiment, four exercises related to the lower limbs, namely, tapping with heel pin (TTHP), heel tapping (HEHE) and rotation (ROTA) and gait (GAIT) were selected from the MDS-UPDRS III tasks and data were collected by an IMU unit based on gyroscope and geomagnetic module [19]. Yang et al. [24] used vertical ground reaction force (VGRF) to extract features from ten raw temporal variables by using four types of feature selection and construction methods Table 2: A sample of gait features used in previous experiments

	Feature/ Feature Set	Description
1	Right Stride	Period of time from right heel touching the ground to the same heel touching again the ground in a gait cycle

including principal component analysis (PCA) and SVM for classification. This experiment was based on gait cycle of human to extract ten temporal features. Tien et al. [25] used wireless inertial sensors to quantitatively characterize gait abnormalities in PD. The sensors they used include accelerometer and gyroscope to collect 67 features of pitch, roll, and yaw movements of the foot. By applying PCA, they selected 11 time-domain features. Mittra et al. [27] performed gait analysis of patients with PD by using data available in Phisonet's Gaitpdb database where, 8 sensors have been applied to each foot of the subjects to calculate VGRF. Each tuple in the database contains 19 features including 16 sensor values, two total force values exerted by the left and right foot and the rest one is time. Collecting gait measurements using Inertial Measurement Unit (IMU) is inexpensive, fast and easy in the sense that it does not require technical expertise [28]. Abujrida et al. [29] carried out smartphone-based gait assessment of patients with PD using crowdsourced data. They used accelerometer in the patients' smartphones to collect data while walking. PD severity was also assessed using noisy crowdsourced data.

Different instruments to collect gait data and units of measurements contribute to a divergent lot of features in predicting PD. A sample of gait features used in previous experiments [19], [24], [25], [29] applied with machine learning are presented in Table 2.

2	Right Stance	Period of time from right heel touching the ground to the right toe down in a gait cycle
3	Right Swing	Period time from right toe down to right heel touching the ground again in a gait cycle
4	Left Stride	Period of time from left heel touching the ground to

		the same heel touching again the ground in a gait cycle
5	Left Stance	Period of time from left heel touching the ground to the left toe down in a gait cycle
6	Left Swing	Period time from left toe down to left heel touching the ground again in a gait cycle
7	Double Support	Period of time while the two feet touching the ground in a gait cycle
8	GT_Time	Gait time
9	GT_Strd	Number of strides in 15meter walking
10	GT_Freq	Gait frequency
11	GT_StrdT	Stride time
12	GT_SWT	Swing time
13	GT_STT	Stance time
14	GT_Exc	Dorsiflexion angular excursion of the foot
15	RO_Time	Rotation time to cover 360° turn
16	RO_Strd	Number of strides in 360° turn
17	RO_Freq	Rotation frequency
18	TT_Freq	Tapping frequency
19	TT_Taps	Number of taps
20	TT_Exc	Toe angle
21	TT_CVfreq	Coefficient of variation of tapping frequency
22	TT_CVexc	Coefficient of variation of toe angle
23	HE_Power	Average power in power spectral density

24	HE_Freq	Fundamental frequency
25	HE_Peak	Maximum peak in power spectral density
26	Pitch features (Measured in Degree)	Range of motion, Maximum angle of dorsiflexion, Maximum angle of plantar flexion, Plantar flexion Standard deviation, Single-step maximum of maximum angle of plantar flexion
27	Roll features (Measured in Degree)	Range of motion, Maximum positive roll angle, Maximum negative roll angle
28	Yaw features (Measured in Degree)	Range of motion, Maximum positive yaw angle, Maximum negative yaw angle
29	Average Step Time	Average time passed for each step
30	Average Cadence	Total number of steps divided by total time
31	Skewness	Asymmetry of the signal distribution
32	Total Harmonic Distortion	Distortion of the whole signal compared to its harmonics
33	Gait Velocity	Total distance divided by total time
34	Signal Noise Ratio	Power of whole signal divided by power of its computed noise
35	Peak Frequency	Maximum spectral power

2.3 Speech

PD has few known symptoms; hence it is difficult for clinicians to discover it in its early stage. Speech disorder is one of the main symptoms that can help in predicting the disease as early as possible [30]. Vocal impairment is one of the early signs of PD [32], [33], [34], [35] and one of the main motor symptoms since it is seen in approximately 90% of the patients [32]. Vocal impairment can be detectable approximately five years prior to clinical diagnosis [31][35]. Classification of characteristic features of patients' voice is considered to be an early diagnosis of PD [66]. There is a growing interest in developing PD diagnostic systems based on vocal features [34], [36].

The measurement of vocal features is noninvasive and easy to carry out. PD patients suffer from dysphonia (disability in producing vocal sounds) and dysarthria (difficulty in articulating words) [31]. To assess the severity of these symptoms mainly two types of vocal tests are performed, namely, sustained phonations and running speech tests. In the sustained phonations test, the patient is asked to pronounce a single vowel and hold the

pitch of it for as long as possible. In contrast, in the running speech test the patient is asked to speak scripted sentence representing a sample of linguistic units. Max Little et al. mentioned the fundamental frequency (F0), absolute sound pressure level, jitter, shimmer and noise-to-harmonics ratios as the prime traditional measurement processes. Max Little et al. introduced a novel measure that they called pitch period entropy (PPE) for efficiently characterizing dysphonia for segregating variations exhibited by both healthy and PD people, which the traditional measures are not capable of. Therefore, this new measure is robust against the confounding factors such as subject gender and highly variable acoustic environments. They also calculated some measures known as non-standard measures such as correlation dimension (D2), recurrence period density entropy (RPDE) and detrended fluctuation analysis (DFA) in the classification of people with PD. [37]. Tsanas et al. devised a rating scale based on speech signals that replicates the clinicians' Unified Parkinson's Disease Rating Scale (UPDRS). They achieved greater accuracy than the subjective UPDRS that requires the clinicians'

expertise and the guidelines for subjective assessment of the patient's ability to perform a set of empirical tests. In addition to this, apparent PD progression characteristics are found in male and female speech which suggests different pathological patterns in the two genders [38]. [35] Lists the extracted features with descriptions by grouping them into algorithmic families that share common characteristics, such as Glottis quotient (GQ), Glottal to noise excitation (GNE), Vocal fold excitation ratio (VFER), Empirical mode decomposition excitation ratio (EMD-ER), and Mel Frequency Cepstral Coefficients (MFCC). [65] applied various speech signal processing algorithms such as Time Frequency Features, MFCC, Wavelet Transform based Features, Vocal Fold Features and Tunable Q-factor Wavelet Transform (TQWT) to the speech recordings to extract useful information for PD prediction. MFCC carries speaker specific information which has been used for speaker recognition, and biomedical voice assessments for PD detection [38][68]. Wavelet transform method is used to differentiate voice samples for human speech recognition and also used in PD prediction [34][69]. Traditionally, in most of the studies, particular feature set is experimented with different classifiers by taking into Table 3: Some common vocal features used in previous experiments

	Feature	Description
1	Fo(Hz)	Average vocal fundamental frequency
2	Fhi(Hz)	Maximum vocal fundamental frequency
3	Flo(Hz)	Minimum vocal fundamental frequency
4	Jitter(%)	Average absolute difference between consecutive periods, divided by the average period
5	Jitter(Abs)	Average absolute difference between consecutive periods
6	Jitter:RAP	Relative Average Perturbation
7	Jitter:PPQ5	Five-point Period Perturbation Quotient
8	Jitter:DDP	Average absolute difference between consecutive differences between consecutive periods, divided by the average period
9	Shimmer	Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude

consideration the classifier as a variable and the feature set as a constant, e.g., studies [30],[36],[37],[40],[41],[42],[43],[67] used the Baseline features for the prediction of PD by experimenting with diverse classification algorithms such as Deep Belief Network, NNge, AdaBoostM1, KNN, SVM, Logistic Regression, Deep Neural Network, Random Tree, and Feedforward Back-propagation based Artificial Neural Network (FBANN) on the feature set. [31] used combination of diverse feature sets such as Baseline, MFCC, VFER, GQ and GNE, total of 132 dysphonia measures out of which only 10 features were selected to form a single feature set before applying different statistical classifiers such as random forest and SVM to predict PD. [34] considered in his study multiple feature sets such as Baseline, MFCC, Wavelet, Vocal Fold, TQWT and Time Frequency to experiment with deep learning based Convolutional Neural Network (CNN) with a detection rate of 86.9%.

Some common vocal features with high PD detection rates that have been used in previous experiments [30], [34], [35], [36], [39], [40], [41], [42], [43] are presented in Table 3.

10	Shimmer(dB)	Average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20
11	Shimmer:APQ3	Three-point Amplitude Perturbation Quotient; the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbors, divided by the average amplitude.
12	Shimmer:APQ5	Five-point Amplitude Perturbation Quotient; the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbors, divided by the average amplitude
13	Shimmer:DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods.
14	Noise to Harmonics Ratio (NHR)	Amplitude of noise relative to tonal components

15	Harmonics to Noise Ratio (HNR)	Amplitude of tonal relative to noise components
16	Recurrence period density entropy (RPDE)	Uncertainty in estimation of fundamental frequency
17	Detrended fluctuation analysis (DFA)	Quantifies the self-similarity of the noise present in the speech caused by the turbulent air flow
18	Pitch period entropy (PPE)	Quantifies the variability in Pitch period of a voice
19	Glottal to noise excitation Ratio(GNE)	Quantifies amount of turbulent noise in a voice compared to

20	Vocal fold excitation ratio (VFER)	signal from vocal fold vibrating Extent of noise in speech using energy, nonlinear energy, and entropy concepts
21	Empirical mode decomposition excitation ratio (EMD-ER)	Signal to noise ratios using EMD-based energy, nonlinear energy and entropy
22	Mel Frequency Cepstral Coefficients (MFCC)	Amplitude and spectral fluctuations

3. MATERIALS AND METHODS

MIXSYN- the proposed model in this study for detecting PD from speech signals has followed a sequence of procedural steps in the form of exploratory data analysis, feature selection, data preprocessing, data balancing, model training and testing, model evaluation and result analysis. The steps of building MIXSYN are discussed in the following sections.

3.1. Dataset

The data used in this study was originally gathered from 188 PD patients (107 males and 81 females) aged between 33 and 87, at the Department of Neurology, Istanbul University and is collected from UCI machine learning repository [86]. The control group comprises 64 healthy people (23 males and 41 females) aged between 41 and 82. After following physician's examination, three repeated sustained phonations of the vowel 'a' were gathered from each subject. During the recording, the microphone was set to 44.1 KHz.

3.2. Feature Selection

Feature selection is a technique of selecting a minimal subset of features to eliminate the irrelevant and redundant features which helps to improve performance and boost predictive power of a machine learning model.

$$I(X; Y) = \int_{\mathcal{Y}} \int_{\mathcal{X}} P(x, y) \log \left(\frac{P(x, y)}{P(x)P(y)} \right) dx dy \quad (1)$$

Table4 presents the Mutual Information Scores of TQWT speech features computed with the help of scikit-learn library in Python.

Table4: The Mutual Information Scores of the top 15 TQWT features of speech

ID	Features	MI Scores
1	tqwt_entropy_log_dec_35	0.107912
2	tqwt_TKEO_mean_dec_16	0.095862
3	tqwt_entropy_shannon_dec_35	0.094599
4	tqwt_TKEO_std_dec_12	0.093933
5	tqwt_maxValue_dec_12	0.092774
6	tqwt_entropy_log_dec_11	0.092258

The ideal approach to feature selection is often impractical because all the possible subsets of features are to be tried by providing as input to the pertaining machine learning algorithm and then to select the subset that gives the optimum result. This involves 2^X numbers of subsets with X numbers of features which is computationally expensive if x becomes too large. There are three approaches to feature selection: embedded, wrapper, and filter. In this study, a filter method known as Mutual Information is used.

3.2.1 Mutual Information (MI)

Mutual information (MI) in probability theory is defined as an estimate of mutual dependence between two variables. This quantification helps to increase the predictability or certainty about one variable by observing the other, that is, to quantify their mutual dependence statistically. MI is extended to machine learning for tasks like feature selection. In this case, MI selects features by filtering out the irrelevant features having low dependence with the target variable. A score of 0 indicates that there is no dependency between the variables, whereas higher scores imply stronger dependency. For two continuous variables X and Y , MI is computed by using Equation 1.

7	tqwt_stdValue_dec_15	0.091797
8	tqwt_TKEO_mean_dec_12	0.091018
9	tqwt_stdValue_dec_12	0.09025
1	tqwt_entropy_log_dec_12	0.090071
1	tqwt_entropy_shannon_dec_14	0.089607
1	tqwt_energy_dec_27	0.089097
1	tqwt_TKEO_std_dec_13	0.087579
1	tqwt_skewnessValue_dec_36	0.087065
1	tqwt_TKEO_mean_dec_11	0.087051

3.3. Data Balancing

The problem of class imbalance manifests when the data points of one class exceed the data points of another class in a dataset. Imbalanced datasets affect adversely and causes ML model to be biased towards the majority class. To address the problem of class imbalance, oversampling techniques SMOTE and ADASYN are used in this study. In oversampling, the samples of the dataset bearing the minority class are increased so that the number of both the minority and majority class samples are equal. Consequently, it reduces bias towards the majority class and improves the predictive power of ML model. The techniques employed in this study are outlined below:

- **SMOTE:** SMOTE, short for Synthetic Minority Over-sampling Technique, is a statistical technique used to increase the minority-class instances of the dataset by synthetically creating them in a uniform manner, i.e., by creating synthetic instances uniformly along the line joining a minority class instance to its k-nearest neighbors.
- **ADASYN:** ADASYN, short for Adaptive Synthetic Sampling, is a data balancing technique that uses oversampling to create synthetic minority class samples by calculating the density distribution of the minority class samples to target only those samples that are meagre in the distribution.

Figure2 presents the proportion of the PD and Healthy people in the imbalanced dataset used in this study.

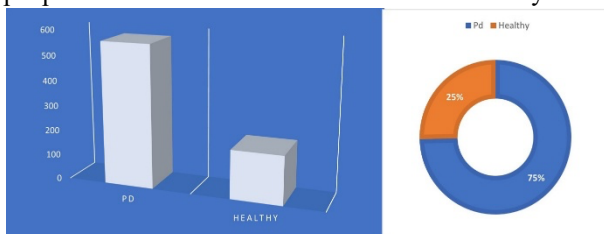


Figure2: Proportion of the PD and Healthy people in the imbalanced dataset

3.4. ML Techniques

In this section, a brief overview of the different ML techniques used in this study is given.

- **XGBoost-** XGBoost, short for eXtreme Gradient Boosting, is a ML library that implements supervised learning algorithm based on gradient boosted decision trees. The proposed ensemble model of this study is trained on XGBoost which is known for its efficiency, speed and scalability.
- **MLP-** Multi-Layer Perceptron (MLP) is an artificial neural network (ANN) which consists of one input layer, at least one hidden layer and one output layer of fully connected neurons. Neurons of each layer are fully connected to the neurons of the next layer. The MLP implemented in this study has 10 hidden layers with 5 neurons per layer. ReLU activation function is used. During model training, by calculating error at the output layer, backpropagation is used for adjusting the weights to improve the classification accuracy of the

model. Adam is used as an optimization algorithm to adjust the weights.

- **EasyEnsemble-It** is an advanced framework of ensemble learning developed specifically for handling imbalanced datasets (e.g., medical datasets) to better the classification accuracy. In this technique, the imbalanced dataset is divided into multiple subsets, keeping all the minority class samples in each subset. It then creates balanced datasets (i.e., majority to minority class samples in a 1:1 ratio) by applying random under sampling to the majority class samples and trains a model for each of the balanced datasets by using usually AdaBoost. Predictions from all the individual models are combined (using either voting or average) to get the final prediction.

3.5. Model Evaluation

Model evaluation in this study is done with robust and reliable 10-Fold cross validation method. The metrics for evaluating the performance of ML model used in this study- accuracy, precision, recall, and F1 score are based on 10-Fold cross validation method. In this method of model evaluation, data is divided into equal-sized 10 subsets (k folds with k=10). In each iteration, the ML model is trained with k-1, i.e., 9 folds and tested with the remaining fold. This process is repeated 10 times so that each fold may serve as a test set which can help in reducing bias and model overfitting. The mean of the evaluation metric from each fold is taken as the final performance of the model.

The evaluation metrics used to measure the performance of ML models are evaluated with the help of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) that are defined below:

TP- Total count of the positive cases predicted as positive by the ML model.

TN- Total count of the negative cases predicted as negative by the ML model.

FP- Total count of the negative cases predicted as positive by the ML model.

FN- Total count of the positive cases predicted as negative by the ML model.

Equations 2-5 give the different metrics used in this study to evaluate model performance.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Precision-recall (PR) curve is plotted for evaluating the performance of the ML models visually. With different thresholds, PR curve plots precision along the y-axis and recall along the x-axis. An ideal model with both high precision and recall, traverses through the top-right corner (1,1) in the xy plane, with an Area Under the Precision-Recall Curve (AUC-PR) of 1.0.

4. RESULTS

Choosing varying number of features based on the Mutual Information (MI) score gives varying accuracies by the classifiers. This is presented in Table5. While Table5: Accuracy of the classifiers with SMOTE against the number of features selected based on the MI Scores

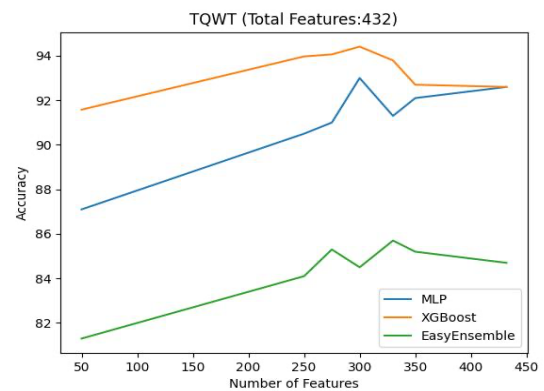
Classifiers	Number of Features							Accuracy (%)
	#5	#2	#2	#3	#3	#3	#4	
	0	50	75	00	30	50	32	
MLP	87	90	91	93	91	92	92	
	.1	.5	.0	.0	.3	.1	.6	
	0	0	0	0	0	0	0	

Figure3 shows the accuracies achieved by the three classifiers- EasyEnsemble, MLP and XGBoost against the number of selected features from the TQWT feature set based on the Mutual Information (MI) scores. The dataset used in this study contains 432 features of TQWT feature set. By varying the number of features from 50 onwards (as shown in Table5) the classifiers (MLP and XGBoost) show a steady PD detection rate until 330 numbers of features and then the accuracy falls, as depicted in Figure3. Hence, MI helps to eliminate irrelevant and redundant features by about 24%

increasing the number of features, accuracy steadily reaches the peak at a point and then it declines.

XGBo	91	93	94	94	93	92	92
ost	.5	.9	.0	.4	.7	.7	.6
	8	7	6	1	9	0	0
EasyE	81	84	85	84	85	85	84
nseml	.3	.1	.3	.5	.7	.2	.7
e	0	0	0	0	0	0	0

Figure3: Plotting Accuracy against the Number of



Features

Table6 presents a performance comparison of various models trained and tested in this study with respect to accuracy, precision, recall and F1 score.

Table6: Performance comparison of various models trained and tested in this study

Ensembl e Method	Data Balancing	Accur acy (%)	Preci sion (%)	Rec all (%)	F1 Score (%)
EasyEns emble	None	80.30	94.40	78.40	85.50
XGBoost	None	87.20	88.30	95.70	91.70
MLP	None	83.10	87.00	91.30	88.90

Figure4 visually presents the performance in terms of accuracy, precision, recall and F1 score of the various ML models trained and tested in this study using SMOTE and ADASYN. From Table6 and Figure4, it is observed that XGBoost with ADASYN performs the best in terms of accuracy (94.60%), precision (96.10%) and F1 score (94.40%) except for recall (92.90%) which is slightly lower than XGBoost without data balancing (95.70%). MLP with ADASYN may be the second candidate for predicting PD with impressive accuracy, precision, recall and F1 score of 93.00%, 95.10%, 90.60% and 92.80% respectively.

EasyEns	SMO	84.50	87.70	80.50	83.90
emble	TE				
XGBoost	SMO	93.00	94.00	91.80	92.90
	TE				
MLP	SMO	93.00	94.30	91.40	92.90
	TE				
EasyEns	ADAS	84.30	86.70	81.40	83.90
emble	YN				
XGBoost	ADAS	94.60	96.10	92.90	94.40
	YN				
MLP	ADAS	93.00	95.10	90.60	92.80
	YN				

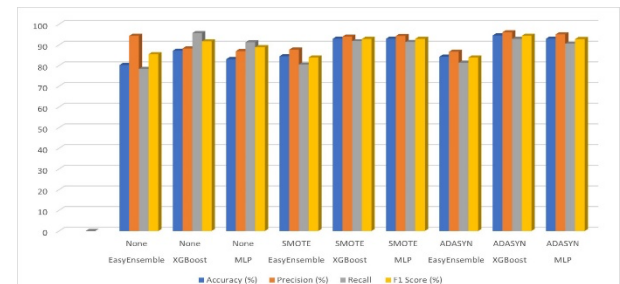


Figure4: Visual presentation of the performance of the ML models trained in this study

Figure5 presents precision-recall (PR) curve of the three classifiers-EasyEnsemble, MLP and XGBoost with both SMOTE and ADASYN each along with average-precision (AP) values. PR curve plots precision

and recall values at different thresholds along y and x axes respectively by giving insights about the performance of a classifier, i.e., for a good model, the curve appears near the top-right corner (1,1) in the xy plane, while a higher AP value ($0 \leq AP \leq 1$) shows better performance. In Figure5, it is seen that the performance of MLP is slightly better than XGBoost. With both SMOTE and ADASYN, the AP score of MLP is 2% better than XGBoost and 18-24% better than EasyEnsemble respectively. With this regard, EasyEnsemble is not performing up to the mark. For the three classifiers- EasyEnsemble, MLP and XGBoost, AP scores with SMOTE are 0.71, 0.95, and 0.93, while with ADASYN, AP scores are 0.78, 0.96, and 0.94 respectively.

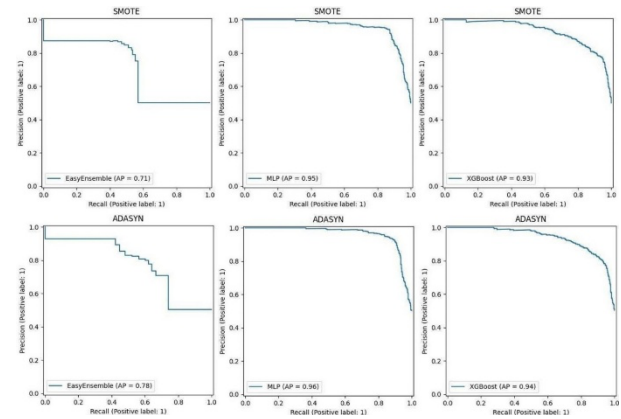


Figure5: Precision-recall curve of EasyEnsemble, MLP and XGBoost with both SMOTE and ADASYN

4.1 Ablation Study

To understand the significance of each component of the proposed MIXSYN model, we carried out an ablation study by replacing individual components as presented in Table7. The ablation study shows that the combination of XGBoost, ADASYN and MI achieved the highest accuracy, precision, and F1 score. Hence, by leveraging XGBoost, ADASYN and MI, MIXSYN model achieved an optimum accuracy of 94.6% in detecting PD.

Table7: Ablation study of the proposed *MIXSYN model

Experiment al Setup	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
XGBoost-ADASYN - MI*	94.60	96.10	92.90	94.40

XGBoost-SMOTE-MI	93.00	94.00	91.80	92.90
XGBoost-SMOTE	92.60	95.00	90.00	92.40
XGBoost-ADASYN	94.10	95.40	92.60	94.00
XGBoost-MI	87.20	88.30	95.70	91.70
XGBoost	84.70	86.30	94.60	90.20

4.2 Comparative Study

This section presents a comparison with the state-of-the-art studies that shows better performance of this study to Table8: Comparison of this study with the state-of-the-art studies using different biomarkers

Studies	Classifiers Used	Biomarker	Accuracy (%)
Aishwarya et al. [71]	Light GBM	Speech	81.35
Di Cesare et al. [72]	SVM, KNN, Neural Network	Speech	92.30
Sravanthi et al. [73]	Decision Tree	Speech	94.10
Alqam et al. [74]	KNN	Speech	89.70
Fouladi et al. [75]	Convolution Network	Speech	90.00
Cao et al. [76]	Deep learning	MRI	91.20

detect PD using machine learning. Table8 presents an outline by comparing this study with the state-of-the-art studies using different biomarkers in the detection of PD.

Vyas et al. [77]	3D Convolution Network	MRI	88.90
Jiaqiang Li et al. [78]	SVM	MRI	94.07
Haojie Song et al. [79]	SVM	MRI	83.79
Alicia et al. [80]	Random Forest	MRI	92.85
Kim et al. [81]	Random Forest	Gait	84.90
Yin W et al. [82]	Random Forest	Gait	91.43
Hwang et al. [83]	Naïve Bayes	Gait	67.30
Uchaipichat et al. [84]	Decision Tree	Gait	83.66

Park et al. [85]	CNN	Gait	90.35
------------------	-----	------	-------

Aishwarya et al. [71] applied Fisher score based recursive feature elimination and Light GBM on speech features to achieve 81.35% of classification accuracy. Di Cesare et al. [72] used speech signals to classify PD patients by using machine learning (ML) models that includes SVM, KNN, and Neural Network. Speech feature extraction techniques such as Mel frequency cepstral coefficient (MFCC) and Gammatone Frequency Cepstral Coefficient (GTCC) were used which produced an accuracy of 92.3%. Sravanthi et al. [73] used speech features to train various ML models including SVM, Random Forest, KNN, AdaBoost, Decision Tree, and Logistic Regression. They found Decision Tree to be 94.10% accurate in predicting PD. Alqam et al. [74] performed classification of speech disorder of PD patients by considering speech feature extraction methods such as discrete wavelet packet decomposition (DWPD), linear predictive coding (LPC), and MFCC to train and test three ML models including KNN, SVM and neural network. KNN with DWPD yielded the highest accuracy of 89.7%. Fouladi et al. [75] used voice samples by considering key feature sets such as MFCC, tunable Q-factor wavelet transform (TQWT) and wavelet transform (WT) and obtained a maximum classification accuracy of 90% by implementing one-dimensional convolution network with TQWT feature set.

Cao et al. [76] introduced a multimodal deep learning framework using single nucleotide polymorphism (SNP) and MRI data to classify PD with a classification accuracy of 91.2%. Vyas et al. [77] used deep learning-based techniques to train and test 2D and 3D convolution neural networks (CNN) to detect PD early. They obtained 88.9% accuracy with the 3D CNN model. Jiaqiang Li et al. [78] proposed a multimodal framework with neuroimaging (DTI and MRI) and genetic data to predict PD using SVM that yielded 94.07% accuracy. Haojie Song et al. [79] developed a deep learning-based PD diagnostic model by using multimodal, longitudinal data that includes neuroimaging. They obtained a classification accuracy of 83.79 % with SVM. Alicia et al. [80] developed an ML-based diagnostic model for prediction of PD at an early stage by using T1-weighted MRI radiomic features. Three ML models-random forest, SVM and logistic regression were trained and evaluated. Random Forest produced superior accuracy of 92.85% among all.

Kim et al. [81] studied gait-based biomarkers for early detection of PD using ML approach with random forest classifier that yielded 84.9% accuracy. Yin W et al. [82] performed gait analysis in the early stage of PD with an ML approach. Using random forest classifier, they obtained 91.43% accuracy in differentiating PD patients from healthy controls. Hwang et al. [83] studied accuracy of ML models for early detection of PD by using spatiotemporal gait features at varying walking speeds. With Naïve Bayes, they

This Study	XGBoost	Speech	94.60
------------	---------	--------	--------------

obtained 67.3% of accuracy of early PD detection. Uchaipichat et al. [84] investigated the application of ML in detecting PD using Vertical Ground Reaction Force (VGRF) and Stride data. They tested four models-Efficient Linear SVM, Efficient Logistic Regression, Decision Tree and Ensemble. Decision Tree with VGRF yielded the best accuracy of 83.66%. Park et al. [85] proposed a gait-based convolutional neural network (CNN) model that can detect PD patients with an accuracy rate of 90.35% which is claimed to outperform a conventional CNN by about 10%.

From the comparison of this section, it is seen that MIXSYN- the proposed model of this study shows superior performance over many state-of-the-art models in detecting PD.

5. CONCLUSION AND FUTURE WORK

In this study, three ML techniques- MLP, XGBoost and EasyEnsemble are used to train several ML models (with different configurations) on speech data to predict PD. Data balancing techniques- SMOTE and ADASYN are applied to the imbalanced dataset that contains voice samples of both the PD patients and healthy controls. Mutual Information (MI)- a statistical technique is used for feature selection from the speech feature set. Robust K-fold and stratified k-fold cross validation techniques for evaluating model performance are used. By addressing the issues like data preprocessing, class imbalance and feature selection that many earlier studies fell short of, this study proposes ensemble learning-based MIXSYN model which can detect PD with an accuracy of 94.60% which shows better result than many state-of-the-art studies including multimodal deep learning techniques. MLP with ADASYN also proved to be a second candidate for predicting PD with impressive accuracy, precision, and recall of 93.00%, 95.10%, and 90.60% respectively. Irrespective of the ML techniques to classify PD, it is important to incorporate robust data curation method that includes data cleaning and enhancing in order to arrive at a more accurate result. Moreover, studies with large sample size should be conducted to arrive at a more reliable result.

STATEMENTS AND DECLARATIONS

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflicts of Interest

The authors of this research paper have no conflict of interest and there is no copyright issue.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval

This article does not contain any studies with human or animal participants. No ethics approval was required.

Consent to participate

This article does not contain any studies with human. No consent to participate was required.

Consent to publish

This article does not contain any studies with human. No consent to publish was required.

Data Availability Statement

The data used in this research is taken from UCI machine learning repository. The URL of the data source is:

<https://archive.ics.uci.edu/dataset/174/parkinsons>

REFERENCES

- [1] R. A. Shirvan, E. Tahami, "Voice Analysis for Detecting Parkinson's Disease Using Genetic Algorithm and KNN Classification Method," 18th Iranian Conference on Biomedical Engineering, 14-16 December 2011, Tehran, Iran.
- [2] National Institute of Neurological Disorders and Stroke, 2009.06.21, http://www.ninds.nih.gov/disorders/parkinsons_disease/detail_parkinsons_disease.htm.
- [3] Ahlrichs, C., and Lawo, M., Parkinson's disease motor symptoms in machine learning: a review, arXiv preprint arXiv:1312.3825, 2013, doi: 10.5121/hij.2013.2401
- [4] Sonne J, Reddy V, Beato MR, Neuroanatomy, Substantia Nigra, StatPearls Publishing, 2025, <https://www.ncbi.nlm.nih.gov/books/NBK536995/>
- [5] Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiol Aging*, vol. 24, pp. 197–211, 2003 [PubMed: 12498954]
- [6] Abujrida, H., Agu, E., Pahlavan, K., "Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data," *IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)* (Bethesda, MD), pp. 208–211, 2017, doi: 10.1109/HIC.2017.8227621
- [7] Davie, C.A., "A review of Parkinson's disease," *British Medical Bulletin* 86(1), pp. 109– 127, 2008, <http://bmb.oxfordjournals.org/content/86/1/109.abstract>
- [8] Fahn, S., Concept and classification of dystonia, *Adv Neurol* 50, 1988
- [9] Hou, J.G.G., Lai, E.C., "Non-motor symptoms of Parkinson's disease," *International Journal of Gerontology*, vol. 1, no. 2, pp. 53 – 64, 2007, <http://www.sciencedirect.com/science/article/pii/S1873959808700243>
- [10] Jankovic, J., "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368– 376, 2008, <http://jnnp.bmj.com/content/79/4/368.abstract>
- [11] Korczyn, A.D., Parkinson's disease, *International Encyclopedia of Public Health*, Academic Press, Oxford, pp. 10 – 17, 2008, <http://www.sciencedirect.com/science/article/pii/B9780123739605000289>
- [12] Samii, A., Nutt, J.G., Ransom, B.R., "Parkinson's disease," *The Lancet*, vol. 363, no. 9423, pp. 1783 – 1793, 2004, <http://www.sciencedirect.com/science/article/pii/S0140673604163058>
- [13] Sian, J., Gerlach, M., Youdim, M.B.H., Riederer, P., "Parkinson's disease: a major hypokinetic basal ganglia disorder," *Journal of Neural Transmission*, vol. 106, pp. 443– 476, 1999, <http://dx.doi.org/10.1007/s007020050171>, 10.1007/s007020050171
- [14] Andlin-Sobocki, P., Jansson, B., Wittchen, H.U., Olesen, J., "Cost of disorders of the brain in Europe," *European Journal of Neurology* 12 Suppl 1, Jun 2005, <http://dx.doi.org/10.1111/j.1468-1331.2005.01202.x>
- [15] Palma, J.A., Kaufmann, H., "Autonomic disorders predicting Parkinson disease," *Parkinsonism Relat Disord*. 2014 January ; vol. 20, no. 1, S94–S98. doi:10.1016/S1353-8020(13)70024-5.
- [16] Chaudhuri, K.R., Healy, D.G., Schapira, A.H.V., "Non-motor symptoms of Parkinson's disease: diagnosis and management," *Lancet Neurol*, vol. 5, pp. 235–45, 2006
- [17] Niethammer, M., Feigin, A., Eidelberg, D., "Functional Neuroimaging in Parkinson's Disease," *Cold Spring Harb Perspect Med*, 2:a009274, 2012
- [18] Adeli, E., Shi, F., An, L., Wee, C.-Y., Wu, G., Wang, T., et al., "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data," *NeuroImage*, vol. 141, pp. 206–219, doi: 10.1016/j.neuroimage.2016.05.054
- [19] Rovini, E., Maremmani, C., Moschetti, A., Esposito, D., Cavallo, F., "Comparative motor pre-clinical assessment in parkinson's disease using supervised machine learning approaches," *Annals Biomed. Eng.*, vol. 46, pp. 2057–2068, 2018, doi: 10.1007/s10439-018-2104-9
- [20] Castillo-Barnes, D., Ramírez, J., Segovia, F., Martínez-Murcia, F. J., Salas-Gonzalez, D., Górriz, J. M., "Robust ensemble classification methodology for I123-Ioflupane SPECT images and multiple heterogeneous biomarkers in the diagnosis of Parkinson's disease," *Front. Neuroinf.*, vol. 12, no. 53, 2018, doi: 10.3389/fninf.2018.00053
- [21] Illan, I. A., Gorz, J. M., Ramirez, J., Segovia, F., Jimenez-Hoyuela, J. M., Ortega Lozano, S. J., "Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging," *Med. Phys.*, vol. 39, pp. 5971–5980, 2012, doi: 10.1118/1.4742055
- [22] Atkinson-Clement, C., S. Pinto, A. Eusebio, O. Coulon, "Diffusion tensor imaging in Parkinson's disease: review and meta-analysis," *NeuroImage Clin*, vol. 16, pp. 98–110, 2017
- [23] Galantucci, S., F. Agosta, E. Stefanova, S. Basaia, M. P. Van Den Heuvel, T. Stojkovic', E. Canu, I. Stankovic', V. Spica, M. Copetti, et al., "Structural brain connectome and cognitive impairment in Parkinson disease," *Radiology*, vol. 283, pp. 515–525, 2016.
- [24] Yang, M., Zheng, H., Wang, H., and McClean, S., "Feature selection and construction for the discrimination of neurodegenerative diseases based on gait analysis," 3rd International Conference on Pervasive

- Computing Technologies for Healthcare (London), pp. 1–7, 2009, doi: 10.4108/ICST.PERVASIVEHEALTH2009.6053
- [25] Tien, I., Glaser, S. D., Aminoff, M. J., “Characterization of gait abnormalities in Parkinson's disease using a wireless inertial sensor system,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 3353–3356, 2010, doi: 10.1109/IEMBS.2010.5627904
- [26] Cuzzolin, F., Sapienza, M., Esser, P., Saha, S., Franssen, M. M., Collett, J., et al., “Metric learning for Parkinsonian identification from IMU gait measurements,” *Gait Posture*, vol. 54, pp. 127–132, doi: 10.1016/j.gaitpost.2017.02.012
- [27] Mittra, Y., and Rustagi, V., “Classification of subjects with Parkinson's disease using gait data analysis,” *International Conference on Automation and Computational Engineering (ICACE)* (Greater Noida), pp. 84–89, 2018, doi: 10.1109/ICACE.2018.8687022
- [28] Esser P, Dawes H, Collett J, Feltham MG, Howells K., “Assessment of spatio-temporal gait parameters using inertial measurement units in neurological populations,” *Gait Posture*, vol. 34, no. 4, pp. 558–560, 2011
- [29] Abujrida, H., Agu, E., and Pahlavan, K. “Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data,” *IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)* (Bethesda, MD), pp. 208–211, 2017, doi: 10.1109/HIC.2017.8227621
- [30] Alqahtani, E. J., Alshamrani, F. H., Syed, H. F., Olatunji, S. O. “Classification of Parkinson's disease using NNge classification algorithm,” *21st Saudi Computer Society National Computer Conference (NCC)* (Riyadh), pp. 1–7, 2018, doi: 10.1109/NCG.2018.8592989
- [31] Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., Ramig, L.O. “Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease,” *IEEE Trans. Bio. Med. Eng.*, vol. 59, pp. 1264–1271, 2012, doi: 10.1109/TBME.2012.2183367
- [32] Tsanas A, Little MA, McSharry PE, Ramig LO, “Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests,” *IEEE transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–93, 2010, <https://doi.org/10.1109/TBME.2009.2036000> PMID: 19932995
- [33] Duffy, J., *Motor speech disorders: Substrates, differential diagnosis, and management*, St. Louis, MO, Mosby-Year Book, 2005
- [34] Gunduz, H., “Deep learning-based parkinson's disease classification using vocal feature sets,” *IEEE Access*, vol. 7, pp. 115540–115551, 2019, doi: 10.1109/ACCESS.2019.2936564
- [35] Harel B, Cannizzaro M, Snyder PJ, “Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study,” *Brain and cognition*, vol. 56, no. 1, pp. 24–9, 2004, <https://doi.org/10.1016/j.bandc.2004.05.002> PMID: 15380872
- [36] Erdogdu Sakar, B., Serbes, G., Sakar, C. O., “Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease,” *PLoS ONE*, vol. 12, no. e0182428, 2017, doi: 10.1371/journal.pone.0182428
- [37] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, L.O. Ramig, “Suitability of dysphonia measurements for telemonitoring of Parkinson's disease,” *IEEE Trans. Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009
- [38] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, “Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity,” *Journal of the Royal Society Interface*, Vol. 8, pp. 842–855, 2011
- [39] Bakar, Z. A., Ispawi, D. I., Ibrahim, N. F., Tahir, N. M., “Classification of Parkinson's disease based on Multilayer Perceptrons (MLPs) neural network and ANOVA as a feature extraction,” *8th International Colloquium on Signal Processing and its Applications* (Malacca), pp. 63–67, 2012, doi: 10.1109/CSPA.2012.6194692
- [40] Ma, H., Tan, T., Zhou, H., Gao, T., “Support Vector Machine-recursive feature elimination for the diagnosis of Parkinson disease based on speech analysis,” *Seventh International Conference on Intelligent Control and Information Processing (ICICIP)* (Siem Reap), pp. 34–40, 2016, doi: 10.1109/ICICIP.2016.7885912
- [41] Haq, A. U., Li, J., Memon, M. H., Khan, J., Din, S. U., Ahad, I., et al., “Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease,” *15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)* (Chengdu), pp. 101–106, 2018
- [42] Haq, A. U., Li, J. P., Memon, M. H., Khan, J., Malik, A., Ahmad, T., et al., “Feature selection based on L1-Norm support vector machine and effective recognition system for Parkinson's disease using voice recordings,” *IEEE Access* 7, pp. 37718–37734, doi: 10.1109/ACCESS.2019.2906350, 2019
- [43] Islam, M. S., Parvez, I., Hai, D., and Goswami, P. (2014). “Performance comparison of heterogeneous classifiers for detection of Parkinson's disease using voice disorder (dysphonia),” *International Conference on Informatics, Electronics & Vision (ICIEV)* (Dhaka), pp. 1–7, doi: 10.1109/ICIEV.2014.6850849, 2014
- [44] Xu, N. , Zhou, Y., Patel, A., Zhang, N., Liu, Y., *Parkinson's Disease Diagnosis beyond Clinical Features: A Bio-marker using Topological Machine Learning of rs-fMRI*, <https://ssrn.com/abstract=4165341>
- [45] Mushta, I.; Koks, S.; Popov, A.; Lysenko, O., “Exploring the Potential Imaging Biomarkers for Parkinson's Disease Using Machine Learning

- Approach," *Bioengineering*, vol. 12, no. 11, <https://doi.org/10.3390/bioengineering12010011>, 2025
- [46] Acton, P. D. & Newberg, A., "Artificial neural network classifier for the diagnosis of Parkinson's disease using [99mTc]TRODAT-1 and SPECT," *Phys. Med Biol.*, vol. 51, pp. 3057–3066, 2006.
- [47] Hamilton, D., List, A., Butler, T., Hogg, S., Cawley, M., "Discrimination between parkinsonian syndrome and essential tremor using artificial neural network classification of quantified DaTSCAN data," *Nucl. Med. Commun.*, vol. 27, pp. 939–944, 2006
- [48] Palumbo, B. et al., "Comparison of two neural network classifiers in the differential diagnosis of essential tremor and Parkinson's disease by 123I-FP-CIT brain SPECT," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 37, pp. 2146–2153, 2010
- [49] Martínez-Murcia, F., G'orriz, J., Ram'irez, J., Moreno-Caballero, M., G'omez-R'io, M., "Parametrization of textural patterns in 123I-ioflupane imaging for the automatic detection of Parkinsonism," *Med. Phys.*, vol. 41, no. 012502, 2014
- [50] Taylor, J. C., Fenner, J. W., "Comparison of machine learning and semiquantification algorithms for (I123)FP-CIT classification: the beginning of the end for semi-quantification," *EJNMMI Phys.*, vol. 4, no. 29, 2017
- [51] Segovia, F. et al., "Distinguishing Parkinson's disease from atypical parkinsonian syndromes using PET data and a computer system based on support vector machines and Bayesian networks," *Front Comput Neurosci.*, vol. 9, no. 137, 2015
- [52] Segovia, F. et al., "Multivariate analysis of 18F-DMFP PET data to assist the diagnosis of Parkinsonism," *Front Neuroinform.*, vol. 11, no. 23, 2017a
- [53] Wang, Z. et al., "ADNI and PPMI. Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning," *Med. Image Anal.*, vol. 39, pp. 218–230, 2017
- [54] Wu, Y. et al., "Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls," *Ann. Transl. Med.*, vol. 7, no. 773, 2019
- [55] Glaab, E. et al., "Integrative analysis of blood metabolomics and PET brain neuroimaging data for Parkinson's disease," *Neurobiol. Dis.*, vol. 124, pp. 555–562, 2019
- [56] Cherubini, A. et al., "Magnetic resonance support vector machine discriminates between Parkinson disease and progressive supranuclear palsy," *Mov. Disord.*, vol. 29, pp. 266–269, 2014
- [57] Huppertz, H. J. et al., "Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification," *Mov. Disord.*, vol. 31, pp. 1506–1517, 2016
- [58] Adeli, E. et al., "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data," *Neuroimage*, vol. 141, pp. 206–219, 2016
- [59] Singh, G., Samavedham, L., Lim, E. C., "Alzheimer's disease neuroimaging initiative; Parkinson progression marker initiative. Determination of imaging biomarkers to decipher disease trajectories and differential diagnosis of neurodegenerative diseases (Disease TreND). *J. Neurosci.*," *Methods*, vol. 305, pp. 105–116, 2018
- [60] Peng, B. et al., "A multilevel-ROI-features-based machine learning method for detection of morphometric biomarkers in Parkinson's disease," *Neurosci. Lett.*, vol. 651, pp. 88–94, 2017
- [61] Chen, Y. et al., "Discriminative analysis of Parkinson's disease based on whole brain functional connectivity," *PLoS ONE*, vol. 10, pp. 1–16, 2015
- [62] Badea, L., Onu, M., Wu, T., Roceanu, A., Bajenaru, O., "Exploring the reproducibility of functional connectivity alterations in Parkinson's disease," *PLoS ONE*, vol. 12, no. e0188196, 2017
- [63] Pläschke, R. N. et al., "On the integrity of functional brain networks in schizophrenia, Parkinson's disease, and advanced age: Evidence from connectivity based single-subject classification," *Hum. Brain Mapp.*, vol. 38, pp. 5845–5858, 2017
- [64] Tang, Y. et al., "Identifying the presence of Parkinson's disease using lowfrequency fluctuations in BOLD signals," *Neurosci. Lett.*, vol. 645, pp. 1–6, 2017
- [65] Sakar, C., et al., "Parkinson's Disease Classification." *UCI Machine Learning Repository*, 2018, <https://doi.org/10.24432/C5MS4X>.
- [66] Froelich, W., Wróbel, K., Porwik, P., "Diagnosing Parkinson's disease using the classification of speech signals," *Journal of Medical Informatics & Technologies*, vol. 23, pp. 187–194, 2014
- [67] Al-Fatlawi, A. H., Jabardi, M. H., Ling, S. H., "An Efficient diagnosis system for Parkinson's disease using deep belief network," *IEEE Congress on Evolutionary Computation (CEC) (Vancouver, BC)*, pp. 1324–1330, 2016, doi: 10.1109/CEC.2016.7743941
- [68] K.S.R.Murty, B.Yegnanarayana, "Combining evidence from residual phase and MFCC features for speaker recognition," *IEEE Signal Processing Letters*, vol. 13, no. 1, pp. 5255, 2006
- [69] Putraa, D.S., Werub, Y. U. W., "Feature Extraction of Facial Electromyography (EMG) Signal for Aceh Languages Speech using Discrete Wavelet Transform (DWT)," *Jurnal Inovasi Teknologi dan Rekayasa* Vol. 4, no 1, pp. 31-31, 2019
- [70] Colder B., "The basal ganglia select the expected sensory input used for predictive coding," *Front. Comput. Neurosci.*, vol. 9, no. 119, 2015, doi: 10.3389/fncom.2015.00119 <http://journal.frontiersin.org/article/10.3389/fncom.2015.00119/full>
- [71] R. Aishwarya, K. Pavitra, P. V. Miranda, K. Keerthana, L. Kamatchi Priya, "Parkinson's disease prediction using Fisher score based recursive feature elimination," *Proc. Int. Conf. Advancement Comput. Technol. (InCACCT)*, pp. 1–8, 2023

- [72] Di Cesare, M.G., Perpetuini, D., Cardone, D., Merla, A., "Machine Learning-Assisted Speech Analysis for Early Detection of Parkinson's Disease: A Study on Speaker Diarization and Classification Techniques," *Sensors*, vol. 24, no. 1499, 2024, <https://doi.org/10.3390/s24051499>
- [73] Sravanthi, Y., Soundarya, P., Babu, S. R., Devender, V. "Parkinson's disease detection by using machine learning," *Advances in Engineering Research*, pp. 60–70, 2026, https://doi.org/10.2991/978-94-6239-628-9_7
- [74] Alqam, M., Hossen, A., Al-Abri, S. et al., "Classification of the Severity of Speech Disorders in Patients with Parkinson's Disease Using Machine Learning Techniques," *Int J Comput Intell Syst*, vol. 19, no. 31, 2026 <https://doi.org/10.1007/s44196-025-01109-7>
- [75] Fouladi, S., Safaei, A.A., Alayari, D. et al., "Advanced Parkinson's disease diagnosis: leveraging deep learning and machine learning with vowel analysis," *Neural Comput & Applic*, vol. 38, no. 76, 2026 <https://doi.org/10.1007/s00521-025-11727-z>
- [76] Cao, Jiangbo, Xiaojing Long., "RACF: A Multimodal Deep Learning Framework for Parkinson's Disease Diagnosis Using SNP and MRI Data," *Applied Sciences*, vol. 15, no. 8, p. 4513, 2025, <https://doi.org/10.3390/app15084513>
- [77] Vyas, T., Yadav, R., Solanki, C., Darji, R., Desai, S., Tanwar, S., "Deep learning-based scheme to diagnose Parkinson's disease," *Expert Systems*, vol. 39, no. 3, e12739, 2022, <https://doi.org/10.1111/exsy.12739>
- [78] Jiaqiang Li, Yukang Lei, Zhenghua Guan et al., "Multimodal adaptive subspace learning for Parkinson's disease diagnosis and prediction," *Applied Soft Computing*, Volume 192, no. 114691, 2026, <https://doi.org/10.1016/j.asoc.2026.114691>, <https://www.sciencedirect.com/science/article/pii/S1568494626001390>
- [79] Haojie Song, Haijun Lei, Yukang Lei et al., "Multimodal joint subspace model for Parkinson's disease diagnosis," *Expert Systems with Applications*, Volume 298, no. 129556, 2026, <https://doi.org/10.1016/j.eswa.2025.129556>
- [80] Alicia R.M. Accioly, Vinícius O. Menezes et al., "Machine Learning-Based Diagnostic Prediction Model Using T1-Weighted Striatal Magnetic Resonance Imaging for Early-Stage Parkinson's Disease Detection," *Academic Radiology*, Volume 32, no. 7, 2025, pp. 4177-4187, 2026, <https://doi.org/10.1016/j.acra.2025.04.001>
- [81] Kim, M., Youm, C., Park, H. et al., "Machine learning classification of early-stage Parkinson's disease using sit-to-walk biomechanical features," *Sci Rep*, vol. 16, no. 10559, 2026, <https://doi.org/10.1038/s41598-026-45122-y>
- [82] Yin W, Zhu W, Gao H, Niu X, Shen C, Fan X, Wang C, "Gait analysis in the early stage of Parkinson's disease with a machine learning approach," *Front. Neurol.*, vol. 15, no. 1472956, 2024, doi: 10.3389/fneur.2024.1472956
- [83] Hwang, J., Youm, C., Park, H. et al., "Machine learning for early detection and severity classification in people with Parkinson's disease," *Sci Rep*, vol. 15, no. 234, 2025, <https://doi.org/10.1038/s41598-024-83975-3>
- [84] Nopadol Uchaipichat, Annop Tangjirachaiyapoka, Chisanupong Chamminavakul, "Machine Learning-Based Detection of Parkinson's Disease Using Vertical Ground Reaction Force and Gait Analysis," *Proceedings of the 2024 9th International Conference on Biomedical Imaging, Signal Processing (ICBSP '24)*, Association for Computing Machinery, New York, NY, USA, pp. 153–157, 2024, <https://doi.org/10.1145/3707172.3707195>
- [85] Park, E.-S., Liu, X., Hwang, H.-J., Han, C.H., "Deep Convolutional Neural Network-Based Detection of Gait Abnormalities in Parkinson's Disease Using Fewer Plantar Sensors in a Smart Insole," *Biosensors*, vol. 16, no. 1, <https://doi.org/10.3390/bios16010040>
- [86] Sakar, C., Serbes, G., Gunduz, A., Nizam, H., Sakar, B., "Parkinson's Disease Classification [Dataset]. UCI Machine Learning Repository," 2018, <https://doi.org/10.24432/C5MS4X>.