scientific reports

Check for updates

OPEN

Voice biomarkers as prognostic indicators for Parkinson's disease using machine learning techniques

Ifrah Naeem[1,9], Allah Ditta[1,9]✉, Tehseen Mazhar[2,6]✉, Muhammad Anwar[1], Mamoon M. Saeed[3]✉ & Habib Hamam[4,5,7,8]

Many people suffer from Parkinson's disease globally, a complicated neurological condition caused by the deficiency of dopamine, an organic chemical responsible for regulating movement in individuals. Patients with Parkinson face muscle stiffness or rigidity, tremors, vocal impairment, slow movement, loss of facial expressions, and problems with balance and coordination. As there is no cure for Parkinson, early diagnosis can help prevent the progression of this disease. The study explores the potential of vocal measures as significant indicators for early prediction of Parkinson. Different machine learning models such as Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Decision Tree (DT) are used to detect Parkinson using voice measures and differentiate between the healthy and Parkinson patients. The dataset contains 195 vocal recordings from 31 patients. The Synthetic Minority Over-Sampling Technique (SMOTE) is used for handling class imbalance to improve the performance of the models. The Principal Component Analysis (PCA) method was used for feature selection. The study uses different parameters to evaluate the model's classification results. The results highlight RF as the most effective model with an accuracy of 94% and a precision of 94%. In addition, SVM achieves an accuracy score of 92%, and precision of 91%. However, with the PCA method, SVM achieves an accuracy of 89%, 92%, and 87% for RF and DT respectively. This study highlights the significance of using vocal features along with advanced machine learning methods to reliably diagnose Parkinson's disease, considering the challenges associated with early detection.

Keywords PD, Machine learning, Early diagnosis, SMOTE, Voice signal

Parkinson's disease (PD) is a complex neurological disorder that impacts millions of people worldwide, progressively compromising motor function, cognition, and quality of life. PD is described as the gradual deterioration of dopaminergic neurons in the brain, which are essential in coordinating motor activities and message transmission between neurons. The progressive death of these neurons disturbs the balance of neurotransmitters, particularly dopamine, which causes tremors, stiffness, and abnormalities in gait, typical motor symptoms associated with PD[1,2]. PD symptoms arise when 60–80% of the neurons that produce dopamine are destroyed because there is insufficient dopamine to govern a person's movements. PD holds a significant global impact, standing as the second most prevalent neurodegenerative disorder worldwide, after Alzheimer's disease. It has a major impact on people, especially those over 50, and the expected rise in PD cases highlights the need for improved techniques for diagnosis[3]. Some soft sign cues could be identified before the progression of PD classical motor impairments. Some of these early characteristics are lost facial expressions, vocal changes, and walking patterns with fewer arm-related actions. However, these minor symptoms are often overlooked and considered normal aging factors which may defer the diagnosis of PD[4]. Several symptoms, including dyskinesia,

[1]Department of Information Sciences, Division of Science and Technology, University of Education, Lahore 54000, Pakistan. [2]School of Computer Science, National College of Business Administration and Economics, Lahore 54000, Pakistan. [3]Department of Communications and Electronics Engineering, Faculty of Engineering, University of Modern Sciences (UMS), Sana'a 00967, Yemen. [4]Faculty of Engineering, Uni de Moncton, Moncton, NB E1A3E9, Canada. [5]Department of Electrical and Electronic Engineering Science, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa. [6]Department of Computer Science, School Education Department, Government of Punjab, Layyah 31200, Pakistan. [7]International Institute of Technology and Management (IITG), Av. Grandes Ecoles, BP 1989, Libreville, Gabon. [8]Bridges for Academic Excellence, Spectrum, Tunis, Centerville, Tunisia. [9]These authors contributed equally: Ifrah Naeem and Allah Ditta. ✉email: allahditta@ue.edu.pk; tehseenmazhar719@gmail.com; dr.mamoon@ums-edu.com

dementia, tight muscles, static tremors, uncoordinated body movements, and unusual postures, are present in patients with the disease. Apart from the motor features, PD also presents with a range of changes in speech patterns (such as decreased speech volume and tempo, often pauses in speech, and abbreviated speech). Within the first five stages of PD, 90% of people with the disease show symptoms of vocal cord damage[5]. Two generic signs of voice modulation are dysphonia and dysarthria. Unfortunately, there is no cure for Parkinson to date due to its complex nature. Voice feature-based early diagnosis in combination with artificial intelligence is an important source of reliable, accurate, and non-invasive predictions for PD patients[6]. One of the promising and useful approaches for identifying PD is through vocal features[7,8]. In this approach, audio signals play a significant role in the early detection of the disease. Voice irregularities in the early stages of PD are not noticeable to listeners, but they can be identified by voice cue analysis. In addition to being simple to assess, vocal impairments come within the umbrella of remote medicine, or telemedicine. Despite physically visiting a doctor, individuals can conduct a quick test at home or capture sounds on their phones[9].

Physicians have long used the unified PD rating scale to track the disease's development and evaluate the results of treatments including oral medicine and surgery. Even though PD is still not curable, people with motor impairments now have some therapy choices. These choices include intensive (surgical) and non-invasive (drug) methods for diagnosis and treatment. Doctors may slow the development of PD by using deep brain stimulation or treatments that activate the brain's dopamine-producing neurons after an early diagnosis[10]. Invasive diagnostic techniques are very costly, and ineffective, and require incredibly complex equipment that has low accuracy. Therefore, less complex, more affordable, and effective approaches should be adopted for the diagnosis.

Traditionally, doctors have mostly relied on a patient's medical history to diagnose PD, especially when evaluating the patient's signs and symptoms. PD can now be diagnosed through neuroimaging methods like MRI scans, electroencephalogram (EEG), speech, and electromyography (EMG)[11]. Using FNIRS signals that track blood oxygenation and flow while measuring hemodynamic changes associated with brain stimulation is also another method being used for PD diagnosis. These signals collect data for different cases like rest, walking and finger tapping among control and PD patients, thus providing deep insights into the diagnosis[12].

Machine learning (ML) has become an invaluable tool in the realm of medical diagnosis. Machine learning (ML), which can examine complex patterns and correlations within massive datasets, is one potential method for increasing diagnosis accuracy. The detection of PD in the early stages can be enhanced by integrating behavioral changes and non-obvious indicators such as altered speech patterns and irregular walking patterns into the diagnostic procedure[13]. When it comes to identifying PD using medical image analysis, deep learning approaches have been shown to outperform standard machine learning techniques in many circumstances. The use of novel imaging modalities, such as deep learning algorithms for analyzing brain MRI scans, may provide a comprehensive understanding of the progression of the disease[14]. As an alternative to brain scanning and neurology testing, researchers have used a range of machine learning-based approaches and techniques to separate PD patients from healthy persons using handwriting, auditory, and speech samples[15]. The study relies on the concept that vocal measures when integrated with sophisticated machine learning models could function as potent predictive biomarkers for the early identification of PD. The research attempts to develop a strong prediction framework by looking at a wide range of voice factors, such as jitter, shimmer, basic frequency metrics, RPDE, NHR, PPE, harmonics parameters, and DFA. There are currently no validated biomarkers that can effectively give early PD detection. Therefore, artificial intelligence (AI) techniques must be used to support the healthcare industry in accurately and promptly diagnosing PD[16].

Different machine learning models have been used for classification between healthy and PD patients in previous studies; however, very little work focused on the use of vocal features for diagnosis. Since almost 90% of Parkinson's patients have vocal injuries, using voice measures as a diagnostic biomarker will give more accurate and faster results as compared to harmful methods such as neuroimaging testing and MRI scans. It will help in the early diagnosis and timely treatment of the disease which would help the patient to live a quality life.

By comparing the performance of machine learning models under varying pre-processing techniques and feature selection method, this study offers a novel way to improve the precision and reliability of detecting PD through voice biomarkers. Our research comprehensively assesses the influence of PCA and SMOTE on classification performance, in contrast to previous studies that frequently concentrate on a single classification technique and neglect the impact of class imbalance on model performance. A thorough grasp of model behavior is obtained by comparing four different machine learning models under various preprocessing circumstances. By tackling primary issues such as class imbalance, model's reliability and high dimensional datasets, the study provides a transparent and more practical AI-based framework for the identification of PD. This study also highlights an often-overlooked biological relevance of unique voice qualities in differentiating between healthy individuals and Parkinson patients.

The following is a summary of this paper's contributions:

- Investigate the effectiveness of using voice biomarkers as prognostic indicators for early-stage prediction of PD.
- Evaluate the performance of different machine learning models in classification between healthy and PD patients.
- Examine the efficacy of the Synthetic Minority Over-sampling Technique (SMOTE) for handling class imbalance.
- Explore using the Principal Component Analysis (PCA) method for feature extraction and its impact on the model's predictive performance.

The rest of the paper is divided into the following "Literature survey" section presents the literature work in the context of Parkinson's detection. "Proposed methodology" section describes the proposed methodology with subsections presenting the different preprocessing techniques, feature selection, and classification algorithms used for predicting PD patients. "Results" section presents the results of the classifiers evaluated with different metrics. "Comparative study and discussion" section is the discussion that analyses the results and compares them with some previous studies. "Limitations and future work" section gives the conclusion.

Literature survey

In recent years, biomarkers such as Magnetic resonance imaging (MRI) scans, postures, handwriting, and genetic data have been used to predict PD (PD); however, little work has been done on aural impairment for early detection. The disease has been categorized by many researchers using diverse techniques. These findings offer a strong basis for applying machine learning to neurodegenerative disorders in light of the current difficulties in prognosis, risk evaluation, and subclassification of PD based on voice signal properties[6]. Ali et al.[17] propose a new ensemble strategy EOFSC for PD identification, by incorporating deep neural networks and feature selection for enhanced performance improving the detection accuracy of Parkinson's disease by 6.5% when compared to traditional techniques. The model was developed by exploiting the results of previous studies that indicate different forms of speech data require distinct optimal frameworks that are sensitive to varied samples and subsets of features. The results showed that the F-DNN technique enhances the PD detection accuracy by 6.5%, still, has limitations due to its dependence on particular datasets which may not represent the varied population.

Nijhawan et al.[18] suggested a novel method to predict PD in individuals using dysphonia readings (vocal characteristics) from a patient recording that outperforms the existing Gradient Boosting Decision Tree (GBDT) based solution with an Area under Curve (AUC) score of minimum 1%. The study highlights some limitations, as it might not be applicable to various phonation patterns, as demonstrated by MLP's superior performance on some datasets. Sayed et al.[2] investigate several machine learning methods such as XGBoost, LightGBM, Bagging, and SVM to analyze the potential of vocal features as early markers for PD. The dataset was collected from the UCI Repository website comprising 195 instances having 24 vocal features relevant to PD. Even though the other algorithms also show impressive results, LightGBM appears to be the most optimal choice for efficient diagnosis achieving an accuracy score of 96%. However, some limitations include the comparatively small size of the dataset and the imbalance for which various oversampling techniques were employed.

The research demonstrates that the ensemble of KNN and Gradient Boosting model named NNB classifier performs best when used for PD prediction using the acoustic features[19]. The obtained dataset was normalized using several techniques such as F-PER, F-MDI, and the correlation method (F-CORR) were employed to identify the most relevant features. The extracted features were fed to different machine-learning models, among which the proposed ensemble approach outperforms. The study presents several benefits, such as enhanced accuracy in the early detection of Parkinson's disease (PD) through speech data. However, limitations might emerge due to dependence on particular feature selection techniques, potentially impacting the model's performance on various datasets.

The authors[20] gathered audio recordings and extracted features such as mel-frequency cepstral coefficients (MFCC) to create two machine learning pipelines. With accuracy ranging from 65 to 75%, both machine learning pipelines showed reasonable performance. Using smartphone microphone recordings, the PD Predict system presents a promising method for early Parkinson's disease identification that could increase diagnostic availability and ease of use. The study[21] used Unified Parkinson's Disease Rating Scale in clinical tests and analyze the speech samples using machine learning models particularly SVM. With a high degree of accuracy in differentiating between healthy and PD patients, the findings showed that vocal anomalies are present in initial stage PD and deteriorate in middle stages. Even though the study used dataset with a well-defined range of participants, it still has certain limitations about the impact of L-Dopa on voice symptoms, which could influence how the results are understood.

The study[4] investigates how using an integrated approach comprising both vocal and facial expressions can effectively aid in the early prediction of PD. A total of 371 candidates' facial and audio recordings were obtained from a smartphone. The forward selection technique was used to extract relevant features for nine different machine-learning algorithms. The study performs well with an area under the receiver operating characteristic (AUROC) of 0.90, however, the study failed to group people with voice tremors that may impact the analysis. The study[22] proposes a methodology that uses a deep neural network model for the prediction of PD severity by using the telemonitoring vocal recordings of patients from the UCI repository. On the other hand, a different typing dataset of the participants was used for the diagnosis of the disease by employing the random forest classifier. With an accuracy range of 92–100%, and an AUC (Area Under Curve) of 0.97 and 1.00, the proposed solution was able to differentiate between people with early-stage PD.

Leung et al.[23] proposed a three-phase ensemble deep-learning approach for the prediction of patients with PD. In the first stage, DaTscan images were extracted using LSTM followed by the second stage where temporal MDS-UPDRS-III data was obtained. The final stage merges the extracted features from stages 1 and 2 with additional clinical measures and trains different neural networks to identify pertinent features for accurate prediction with a mean absolute percentage error of 18.36%. One of the study's drawbacks is that it requires more computing power than training a single model, which might limit its practical applicability. Neto[24] uses multi-source data to analyze the voice samples with four different ML models and ensemble techniques. ESM, SVM, and GB models outperformed the others in terms of accuracy and with 24 different voice relevant features exhibiting significant differences between the healthy and Parkinson's patients. The study encountered difficulties with choosing variables and data diversity, which could have an impact on the validity of the results.

Alkhatib et al.[25] employed a linear model based on the distribution of load during walking to differentiate between the balanced and unbalanced gait where unbalanced gait refers to PD obtaining an accuracy of 95%. Signals were gathered from 18 healthy individuals and 29 PD cases. Wang et al.[26] introduced a novel deep-learning method to differentiate between healthy and Parkinson's patients using premotor characteristics such as loss of smell, sleep disorder, and Rapid Eye Movement. The samples were obtained from the PPMI database comprising 401 PD and 183 healthy subjects. The proposed technique achieves an overall 96% accuracy when compared with other ensemble models and 12 machine-learning models. Despite the excellent results of the proposed method, the study still has some limitations in terms of large memory requirements and a relatively small dataset. Similar to the previous studies, Ahmed et al.[27] detected PD using six different machine learning models along with feature extraction methods. According to the results, the Random Forest classifier distinguished between Parkinson's patients and non-patients with the best accuracy score of 97%. One benefit of this strategy is the possibility of using voice analysis to diagnose Parkinson's disease effectively, which would lessen the need for in-depth physical testing.

Hema et al.[28] utilized different classification algorithms such as SVM, NB, KNN, and RF to classify between two classes, PD patients and healthy. Different feature selection approaches including LASSO feature selection, backward-forward selection, and wrapper methods were used to extract the relevant feature subset. The results of all the classifiers were compared with different feature selection techniques which concluded that Random Forest performs best with all the feature selection approaches achieving 96% accuracy. While the models with rough set feature selection for all the classifiers with an average of 97% accuracy.

Quan et al.[29] proposed a new approach by indicating the usefulness of dynamic speech patterns using a LSTM model with the conventional machine learning models. With a maximum and average accuracy of 90% and 74% the suggested strategy greatly increased PD detection accuracy and offer a more reliable method of early diagnosis. The study[30] shows enhanced classification performance of deep neural networks over traditional machine learning models in the diagnosis of PD patients using the voice patterns. Between the deep learning methods, the DNN2 model had the best accuracy of 95.41%, while the Extreme Gradient Boosting classifier had an overall accuracy of 92.18%.

Rana et al.[31] studied the impact of using machine learning models with a keen focus on feature selection methods such as filter and wrapper method. ANN achieved the highest accuracy at 96.7%, while both SVM and K-NN obtained an accuracy of 87.17%. The proposed approach minimizes the time and operational resources needed for screening Parkinson's disease by focusing on a limited number of clinical test features for diagnosis. The study's limitations include its dependence on particular speech traits and datasets, which might not fully represent the range of symptoms associated with Parkinson's disease.

The study[32] investigates the use of EEG signals by proposing a novel spatial-based method to identify PD in two different scenarios. The EEG signals are pre-processed to eliminate the significant distortions using a spatial pattern and different features are extracted from spatially filtered signals using metrics such as deviation, power, band spectrum, and entropy. The results indicate that alpha and beta bands gain the highest accuracy of 99% without medication for PD diagnosis, and 95 to 98% in case of medication. The limitations of the study include the difficulty in comparing outcomes arising from the use of diverse datasets, underscoring the call for uniform assessment techniques in EEG research.

Hossain et al.[33] examined voice recordings at various frequencies to differentiate PD patients from healthy individuals using machine learning models. A pipeline technique was created to aid in the selection of features. It led to the improvement in the accuracy scores of almost all the algorithms with Adaboost having the highest accuracy of 85% which is far higher than using it alone. The research successfully depicts that machine learning classifiers and pipelines can improve the classification accuracy for patients with Parkinson's disease. However, the high dimensionality in the data and possible biases in the dataset because of the small sample size could provide difficulties for the study. To categorize Parkinson's patients, Yuan et al.[8] employ auditory data; however, their proposed model is highly dependent on MATLAB. Therefore, this study aims to build models trained and tested in Python, that are simpler and efficient without any restriction of large memory and processing requirements.

The recent research shows several limitations, despite the increased interest in using vocal features for initial diagnosis of Parkinson. It has been observed that many earlier studies have focused on a small set of voice features and not thoroughly examined their overall importance. Also, the problem of class imbalance in the dataset that often results in biased model performance was neglected in several studies. Only a few studies address this problem. Furthermore, the investigation of feature selection methods is infrequent, leading to massive datasets that may introduce noise and other computational bottlenecks. Different machine learning models have been used in various studies, still there are very few comparative studies assessing how well these models perform in varying pre-processing conditions. Our research utilizes PCA for feature selection and SMOTE to handle the class imbalance problem in order to reduce this gap by thoroughly investigating a number of voice biomarkers. We offer a thorough analysis of the best method for Parkinson's disease diagnosis by comparing several classification models, such as SVM, LR, DT and RF. The results help in the development of a transparent AI-based framework for the identification of Parkinson. The study also indicates the potential of using vocal features as strong predictors for Parkinson's identification.

Proposed methodology

The proposed methodology is aimed at classifying whether a patient has PD or not by analyzing the voice signals of a PD dataset using the Google Collaboratory platform and Python language. The dataset contains information on key vocal parameters of vowel phonation provided by Multidimensional Voice Program (MDVP). Preprocessing, analyzing, and visualization are done to ensure that attributes are well understood. The issue of class imbalance is also handled using SMOTE, a common oversampling method that creates new synthetic samples. To extract the most critical features a dimensionality reduction technique PCA was applied. Four machine learning models—LR, SVM, RF, and DT are used on 80% of the training data. Models are trained to distinguish the given audio signals into two groups: PD and healthy, based on the changes in frequency. 20%

of the data is used for testing the models and the performance is evaluated using different evaluation metrics. The proposed methodology of our study is shown in Fig. 1.

Ablation study and experimental setup

We performed an ablation study by training machine learning models under various settings to assess the effect of preprocessing methods on model performance. Our goal is to evaluate the impact of each technique on different evaluation metrics such as classification accuracy, F1-score, precision and recall.

We performed experiments under three conditions:

- Baseline Model: Models trained on all original features without SMOTE or PCA.
- With SMOTE: Before training, the dataset is balanced using SMOTE.
- With PCA: PCA is applied to minimize the dimensionality of features.

Dataset description

The dataset that has been used in the study for the experimental analysis to predict PD consists of different voice measures and is available on Kaggle. The dataset contains numerous biomedical voice metrics that record frequency, amplitude, noise, and nonlinear features of speech. The data is collected from 31 individuals, 23 of whom have PD. There are 195 biomedical voices in the speech signal dataset, 48 of which are phonetically classified for healthy individuals and 147 for PD patients. The information is in CSV ASCII format. Each row in the table presents one voice recording, with each field representing a specific voice measure. Each patient has about six recordings; the patient's name appears in the first column. The "status" column, which is set to 0 for healthy and 1 for PD, serves as the primary means of differentiating between individuals without Parkinson's and those with it. The attributes of the PD dataset are elaborated in the table below. Table 1 illustrates the description of different vocal features of the Parkinson Dataset used in this study for the diagnosis of PD.

The collective evaluation of the above-mentioned voice features offers a thorough understanding of the biological and physiological effects of Parkinson on an individual's speech. Tremors, breathiness, and diminished vocal strength are symptoms of Parkinson's disease, which alters the neuromuscular regulation of the vocal cords and causes abnormalities in pitch (jitter), loudness (shimmer), and harmonic structure (NHR, HNR). Nonlinear characteristics (RPDE, D2, DFA, Spread1, Spread2, PPE) capture the complexity and unpredictability of speech patterns, illustrating the ongoing deterioration of motor control in PD. These features help in capturing the minor details that might not be noticeable in clinical evaluations. In depth analysis of these features collectively improves the capacity to determine minimal voice impairments.

Preprocessing

Data preprocessing is an important step before analyzing the data and using it for the training of models. An important element in ensuring the effectiveness of subsequent initiatives is data analytics. The process consists of two phases (1) imputation, which refers to handling missing and null values, removal of duplicate rows, and dealing with outliers; and (2) making sure that data is consistent and complete[1]. The dataset was loaded as a CSV file into the Google Collab Platform using the Pandas library. The dataset was filtered for any null values to

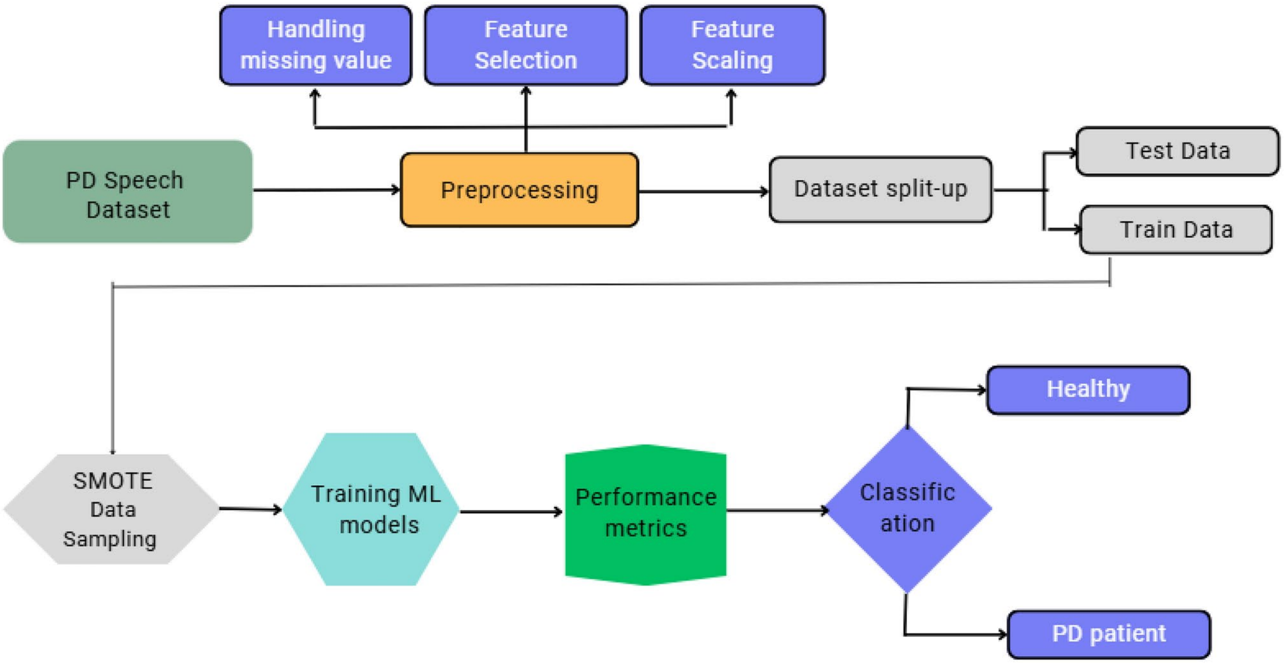

Fig. 1. Proposed architecture.

Voice measure	Description
Name	The name of the subject with its recording no
MDVP: Fo	Indicates the average vocal frequency
MDVP: Fhi	Maximum vocal frequency
MDVP: Flo	Minimum fundamental frequency
MDVP: Jitter, MDVP: Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP	Several variation measures in fundamental frequency
MDVP: shimmer, MDVP: shimmer (dB), shimmer: APQ3, shimmer: APQ5, MDVP: APQ, shimmer: DDA	Several variation parameters in amplitude
NHR, HNR	Two measurements of the voice's noise-to-tonal component ratio
Status	Health status, (one)-Parkinson's and (zero) healthy individual
RPDE, D2	Two metrics for nonlinear dynamical complexity
DFA	The exponent of signal fractal scaling
Spread1, Spread2, PPE	Measures of fundamental frequency that are nonlinear

Table 1. Parkinson dataset with description of vocal features.

name	0
MDVP:Fo(Hz)	0
MDVP:Fhi(Hz)	0
MDVP:Flo(Hz)	0
MDVP:Jitter(%)	0
MDVP:Jitter(Abs)	0
MDVP:RAP	0
MDVP:PPQ	0
Jitter:DDP	0
MDVP:Shimmer	0
MDVP:Shimmer(dB)	0
Shimmer:APQ3	0
Shimmer:APQ5	0
MDVP:APQ	0
Shimmer:DDA	0
NHR	0
HNR	0
status	0
RPDE	0
DFA	0
spread1	0
spread2	0
D2	0
PPE	0

dtype: int64

Fig. 2. Missing values in Parkinson's dataset.

maintain consistency. Figure 2 clearly shows that no missing value was found against each feature. Additionally, it was observed that every feature, aside from "status", which is of a binary type with two classes 0 and 1—is of the continuous "numerical variables" type. Using the status column, it was discovered that the dataset was imbalanced containing 147 records for PD and 48 records for healthy individuals, which is 25% for HC and 75% for PD. The dataset is divided into an 80:20 train/test split ratio to prevent both under- and overfitting.

There are 23 features of voice samples in the Parkinson dataset. To measure the symmetric distribution of the dataset, the skewness method is applied. It is observed that the features with minimum, maximum, and average vocal frequencies are positively skewed meaning that the mean value is higher than the median value. Similarly, the features with shimmer that are variation in amplitude also have skewness towards the right. Features such as HNR, and RPDE show negative skewness, while on the other hand features such as Spread1, and Spread2 have a value closer to 0 indicating a normal symmetric distribution. Depending on the precise nature and degree of inconsistencies, suitable measures are implemented if discrepancies occur during the pre-processing stage. The objective is to guarantee the data's quality and integrity for precise and accurate analysis.

Handling outliers

Outliers are the data points that differ noticeably from the majority in a dataset. In the context of machine learning, outliers have the potential to affect model performance by distorting results or introducing unwanted noise. To check for outliers, we use box plots (as shown in Figs. 3 and 4) an effective method for checking outliers that visualize the distribution of data using the five-number summary; minimum, first quartile(Q1), median, third quartile(Q3), and maximum. We used the interquartile method and capping to remove the outliers. Figures 3 and 4 show the dataset before and after the removal of outliers.

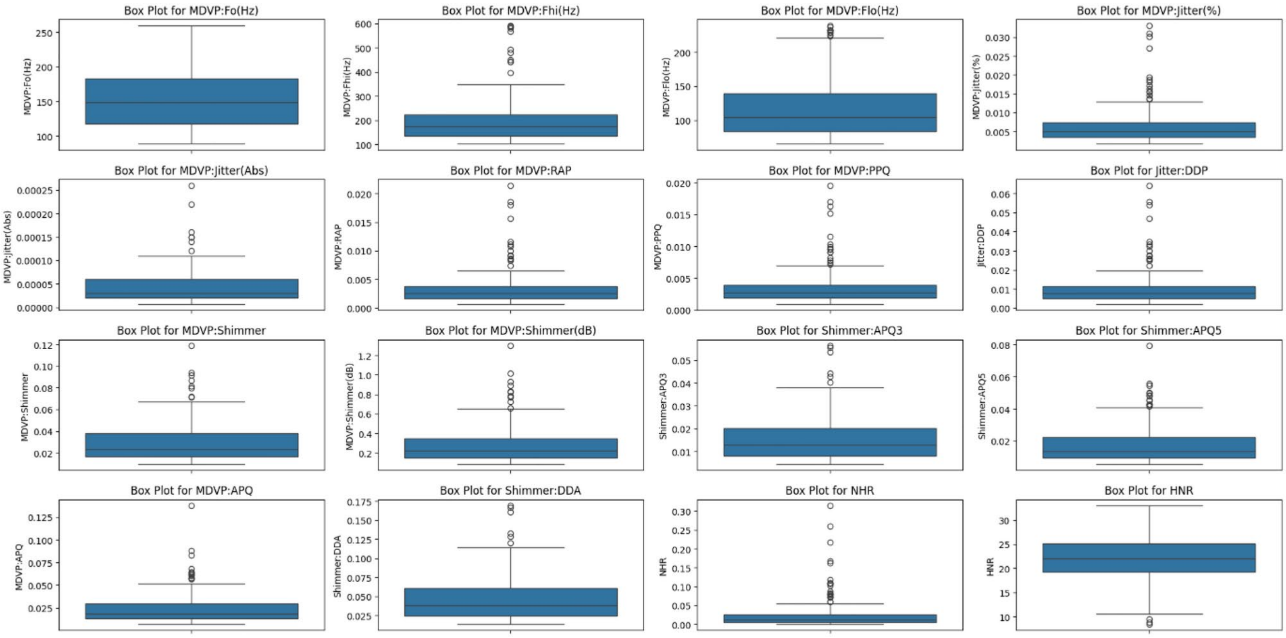

Fig. 3. Box plots with outliers.

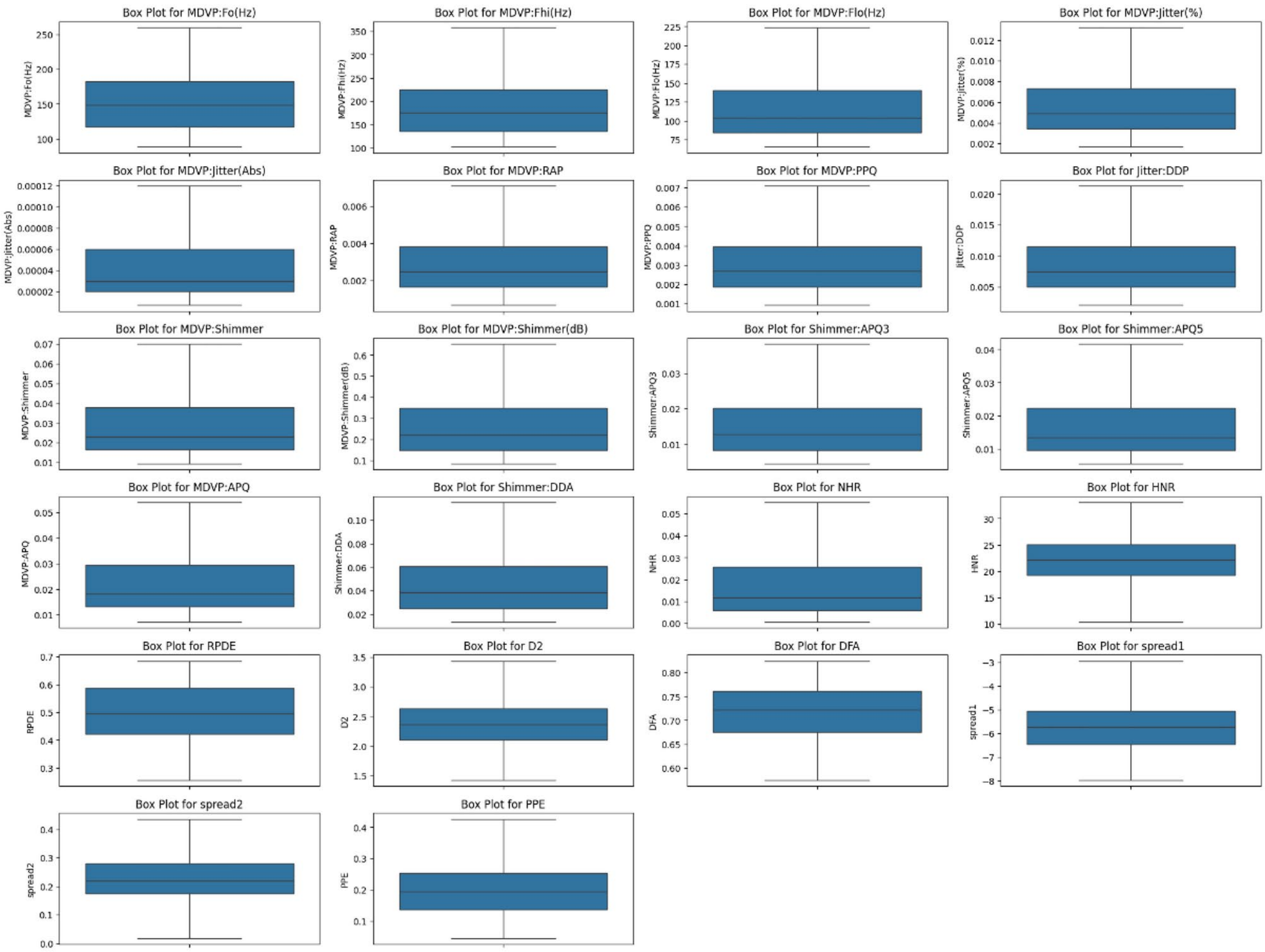

Fig. 4. Box plots after the removal of outliers.

Correlation features

The correlation between the different variables in a dataset can be visualized using a correlation matrix. The correlation matrix can be used to provide a thorough understanding of the various levels of relationship between various variables with the target variable[34]. A matrix showing correlations between the attributes is shown in Fig. 5. The figure shows how different attributes in the PD dataset are associated with each other. It has been observed that the dataset consists of some features with very high correlation, that should be eliminated before the classification. The high correlation is caused by the dataset's imbalance. Thus, the dataset was balanced by the use of the SMOTE technique.

Balance of dataset using SMOTE

The 195 records in the PD dataset are split into two imbalanced classes: Parkinson's patients (1) and healthy (0). The dataset is not balanced; there are 147 records for PD and 48 records for healthy individuals, which is 25% for HC and 75% for PD. Thus, the predictive models will neglect the minority class and the majority class will get attention. Therefore, balancing the dataset to get an equal distribution of all the classes is very important for the prediction performance. To solve this issue different techniques can be used such as up-sampling where samples of minority classes from the training dataset are duplicated[5]. This method may ensure the balance in the dataset, but it does not offer additional information. The duplication of samples by oversampling can result in identical values, leading to a higher risk of overfitting. Therefore, to address this issue, in this study, we are using SMOTE which has comparatively lower chances of overfitting.

SMOTE is a common oversampling method that eliminates data duplication by using synthetic samples. By selecting the k-nearest neighbors of the minority class samples, new samples are created. The count of

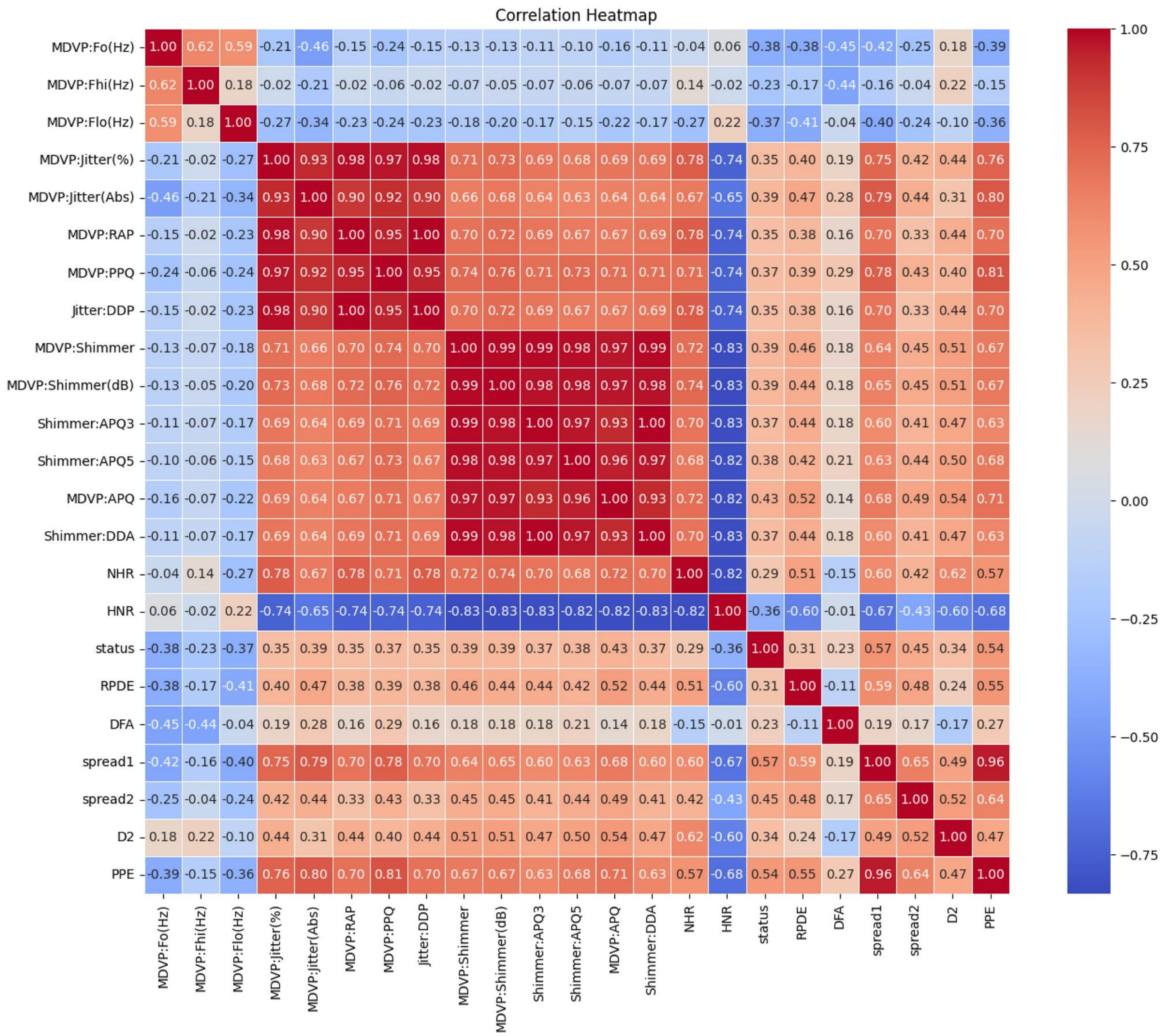

Fig. 5. Correlation matrix of different features of PD.

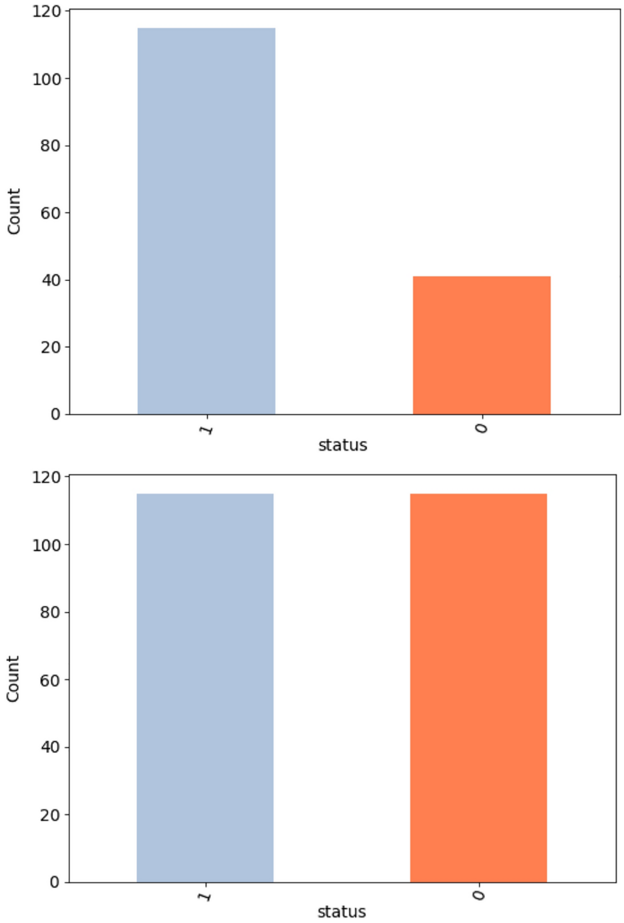

Fig. 6. Distribution of the dataset before and after the SMOTE.

Phase	Training 80%		Testing 20%	
Classes	Healthy	Parkinson	Healthy	Parkinson
Before SMOTE	41	115	7	32
After SMOTE	115	115	7	32

Table 2. Balanced dataset with SMOTE.

new samples that must be generated determines the value that should be entered for "k." After calculating the difference between the highlighted vector and the nearby sample, this value is multiplied by a random integer between 0 and 1. To construct a new sample of the minority class, this new value is appended to the chosen featured vector[35]. Figure 6 shows the distribution of the dataset before and after the SMOTE.

The distribution of samples before and after the use of SMOTE in the training and testing data has been shown in Table 2 below.

Standardization of data

Standardization is a widely used technique for scaling the data before loading it into machine learning models. Even when comparing values on various scales, standardization helps the comparison appear simpler. To attain uniformity, the data must be adjusted to a standardized range. The study uses Standard Scaler for the purpose of feature scaling. A standardized distribution with a zero mean and one standard deviation (unit variance) can be obtained with the aid of a standard scaler. It does so by subtracting the mean from each data point and then dividing them by the Standard Variation[36]. The PD dataset contains a wide range of voice quality. The standardization process has contributed to the values being more comparable, which will facilitate data analysis.

PCA for feature selection

Principal Component Analysis (PCA) is a technique used to reduce the dimensions of the dataset while getting the maximum information and patterns. The PCA method aims to minimize the dataset's dimensions first and then it seeks to discover new, significant features. An orthogonal conversion is used in the PCA method to change a set of correlated variables into a set of independent variables[37]. The principal component analysis (PCA) finds the principal components with the maximum variance in the dataset. These principal components contain variance that is equal to the entire variance of the dataset[38]. New variables known as principal components are created by mixing or combining the original variables linearly. It is done in a way that the newly created variables also called the principal components are independent and the majority of information from the actual features is reduced into these first components. Principal components are built so that the first principal component may account for the greatest amount of variance in the data set. There are as many principal components in the data as there are variables. Then the second principal component is calculated as having the next higher variance and orthogonal to the first component and so on[39]. We need to standardize the data so that each feature has a zero mean and one standard deviation to ensure that each of the features plays its part equally in the analysis. This study uses the PCA method to extract the features preserving the most critical information.

Classification algorithms

Different classification algorithms have been used in this study for the early diagnosis of Parkinson's patients. The performance of all the machine learning models is compared with the help of different evaluation metrics to find out which model performs the best among all.

Support vector machine

SVM, or support vector machines, are widely used in classification. It is a machine-learning technique, that belongs to supervised algorithms and is applied to problems involving regression or classification. The use of SVM for classification is appreciated by many as it generates notable accuracy with minimal processing power[40]. SVM maps N features to a multidimensional space and uses this information to generate a hyperplane that divides the features. Using SVM, the dataset is divided into classes to determine the maximum marginal hyperplane (MMH)[5].

Since Parkinson's audio data cannot be separated linearly, we translate it into a high-dimensional space using an SVM kernel. Support vectors created from a subset of training points and memory utilization are the reasons SVM functions effectively with PD data.

Random forest

The random forest classifier is a supervised machine-learning technique that addresses regression and classification tasks. This technique is also highly accurate because it uses numerous decision trees to produce its output. The random forest classifier does not encounter the overfitting problem since it aggregates all predictions, removing biases and fixing the overfitting issue[41]. The RF algorithm uses different subsets of the database's features to generate many decision trees. To forecast the dataset's correctness, it finally combines the performances of each subset. Random forest does not consider a single decision tree model prediction, instead, the results from all the decision trees are considered to give an average prediction. It can be used to solve complex problems avoiding the issue of overfitting[28].

Logistic regression

A popular supervised machine learning approach called logistic regression uses many independent variables to predict the dependent variables. It predicts a possible value within the range of 0–1 as the result of a continuous or categorical input using the curve-fitting approach[42]. A mathematical function known as the sigmoid function associates a value with the anticipated real values that fall between 0 and 1. Thus logistic regression models form a S-like curve[43]. This is perfect for audio data because the characteristics that determine PD categorization follow an exponential trend rather than a linear one.

Decision trees

Decision Tree model falls under the category of supervised algorithms, majorly used for classification but can also be used for regression tasks. It is comprised of leaf nodes that each indicate a result and inner nodes that define the branch structures. Decision nodes are used to make choices and have multiple branches. On the other hand, leaf nodes are the outcome of decision nodes with no further branch. A decision tree typically splits the tree into smaller trees based on whether the answer to the query is yes or no[40].

Model evaluation

To measure the efficiency of the trained models, different evaluation metrics have been used in this study. The detail is given below:

Accuracy

Accuracy is a crucial performance metric. It gives the proportion of correctly predicted data points with the actual data points. The overall performance of the classification system is shown by accuracy.

It can be calculated using the equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \tag{1}$$

Precision

Precision can be classified as an indicator of a classifier's accuracy. It is defined as a ratio of true positive values to the addition of both true and false positives for every class. It can be calculated using the formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall
Sensitivity also called recall is a metric used to assess the comprehensiveness of a classifier; it quantifies its ability to find every positive occurrence. It can be calculated using the equation:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1-score
By computing the weighted average of Precision and Recall, the F1-Score is utilized to find the balance between these two metrics. It considers both false negatives and false positives. It can be calculated by using the equation:

$$\text{F1\_Score} = \frac{2^{*}\ \text{Precision}\ ^{*}\ \text{Recall}}{\text{Precision}\ +\ \text{Recall}} \tag{4}$$

Results

Baseline model (without smote or PCA)

In the first approach, all the given attributes of the PD dataset have been used. Data preprocessing is performed to handle any duplicate or null values. The data has been split into 80% for testing the models and 20% for training. Data Standardization has been done using the Standard Scaler function. In this approach, the original dataset is used with an imbalanced distribution of both classes. All the attributes for the original dataset have been used for classification. Then, different models such as SVM, RF, LR, and DT are trained on the testing dataset. Training on the complete dataset with all the attributes serves as a basic test for Parkinson's classification. Table 3 depicts the results of the model's performance using this approach, that is models are trained on the original dataset with complete features.

As we can see from the table, the RF model performs the best in this case achieving an accuracy of 92%. Since the RF is an ensemble model, it is perfect for entire datasets. Before a prediction is made, the average of 100 decision trees is assessed. All the attributes are given the same weight during the classification procedure. RF achieved a precision of 93%, and a recall score of 96%, while the F1 score was 95%. On the testing data, SVM achieved an accuracy of 89%. The precision score was 88% and the f1 score was 94%. LR model obtained the accuracy, precision, recall, and F1-score of 89%, 88%, 95%, and 93% respectively. Finally, the DT achieved an accuracy of 89% with a precision of 93% and a recall and f1-score of 93%. It has been observed from the results that the random forest obtained a greater accuracy among all the models for the complete dataset with 22 attributes. Figure 7 shows the confusion matrix for the model where the model classifies 5 true positives, 2 false negatives, 1 false positive, and 31 true negatives.

Effect of smote (class balancing applied)

The dataset is divided into an 80–20% ratio for training and testing. The dataset is not balanced as there are 147 records for PD and 48 records for healthy individuals, which is 25% for HC and 75% for PD. Therefore, balancing the dataset is an essential step for accurate predictions so that both classes contribute equally to the classification. Otherwise, the majority class will be given attention, and the minority class with fewer samples will be neglected. To solve this issue SMOTE technique is used to balance the dataset as shown in Table 2. SMOTE is a common oversampling method that eliminates data duplication by using synthetic samples. All the other preprocessing steps are the same for this approach as well as carried out with the complete dataset. Using this approach will help us to understand the effect of imbalance class distribution in the predictive performance of the models. Table 4 describes the results obtained by different models with the balanced dataset.

It is visible from the above table that the use of the SMOTE technique for handling the class imbalance proves beneficial. Most models showed an increase in accuracy when SMOTE was applied; Random Forest showed the largest increase, going from 92 to 94.87%, while SVM improved from 89.74 to 92%. Addressing class imbalance helps improve prediction stability, as shown by improvements in precision and F1-score, especially in Random Forest and Decision Tree.

Figure 8 illustrates the Classification performance of models using SMOTE and Fig. 9 show the Classification Performance of models without SMOTE.

Metric	SVM	Random forest	Logistic regression	Decision tree
Accuracy	89.74%	92%	89.73%	89.74%
Precision	0.88	0.93	0.88	0.93
Recall	1.0	0.96	0.95	0.93
F1-score	0.94	0.95	0.93	0.93

Table 3. Results with approach 1: using 22 attributes.

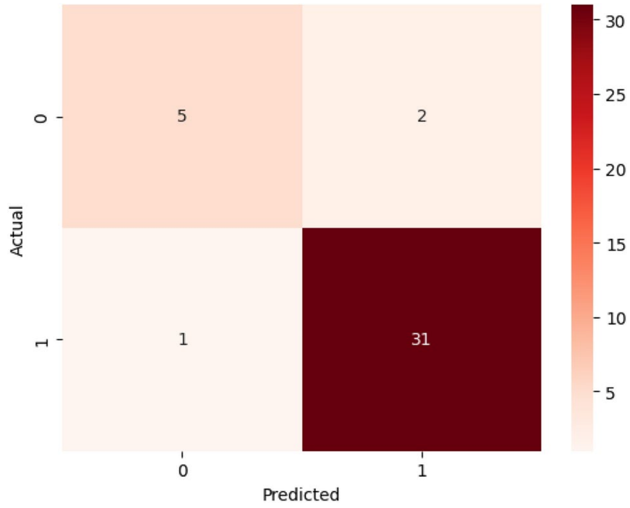

Fig. 7. Confusion matrix for random forest.

Metric	SVM	Random forest	Logistic regression	Decision tree
Accuracy	92%	94.87%	87.17%	89.74
Precision	0.91	0.94	0.90	0.96
Recall	1.0	1.0	0.93	0.90
F1-score	0.95	0.96	0.92	0.93

Table 4. Results with a balanced dataset.

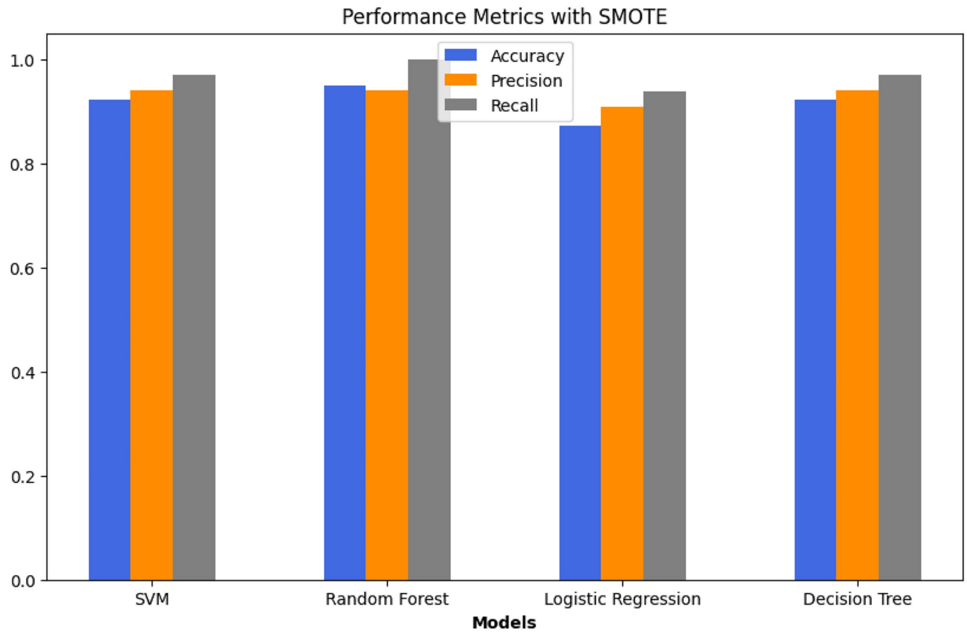

Fig. 8. Classification performance of models using SMOTE.

Effect of PCA (dimensionality reduction applied)

In this experiment, the dataset goes through the same preprocessing. However, an additional step is performed which is the use of the Principal Component Analysis (PCA) method for feature selection. The features were reduced to twelve critical features for diagnosis by the PCA technique. It is used to reduce large data with high dimensions into a smaller set while preserving the most important patterns and information. The results of all the

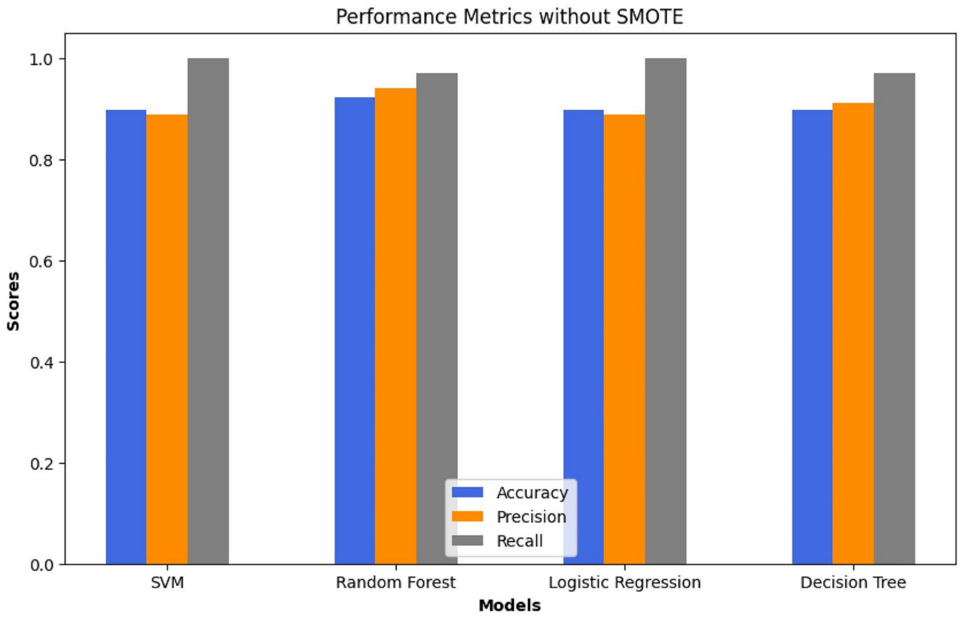

Fig. 9. Classification performance of models without SMOTE.

Metric	SVM	Random forest	Logistic regression	Decision tree
Accuracy	92.30	89.76%	82.05%	87.17%
Precision	0.91	0.91	0.90	0.90
Recall	1.0	0.96	0.87	0.93
F1-score	0.95	0.93	0.88	0.92

Table 5. Results using the PCA method.

models after the application of PCA are shown in Table 5. Even after PCA feature reduction, SVM demonstrated the highest accuracy (92.3%), demonstrating its resilience in classification with a smaller feature set. Following PCA, RF accuracy decreased somewhat from 92 (the baseline) to 89.76%, indicating that some valuable features might have been lost due to dimensionality reduction. Following PCA, the accuracy of LR decreased the highest, from 89.73 to 82.05%. This suggests that PCA might not have been able to capture important feature correlations.

Overall findings

There is a visible improvement in performance through the use of SMOTE, that highlights the importance of handling the problem of class imbalance for accurate classification. Even though the dimensionality reduction technique improves the efficiency of the models, it can also eliminate certain critical discriminative features. As we can see some slight changes are observed in the accuracy scores after the PCA method is applied. These findings suggest that the combination of PCA for feature selection and SMOTE for balancing classes leads to better model performance, enhancing the reliability of machine learning models for detecting PD using voice biomarkers.

Comparative study and discussion

This study aims to develop an effective method for detecting PD in patients using audio data. The dataset used contains 195 voice signal feature recordings gathered from 48 healthy and 147 PD patients. Different approaches have been used to understand their impact on classification performance. Machine learning classification models are used for classification between healthy and PD patients. The dataset is imbalanced as the number of samples for PD patients is more than the number of samples for healthy individuals. SMOTE is used to overcome this issue resulting in a balanced dataset with equal distribution of both classes. Different approaches are used in this paper. First of all, the models were trained on the complete dataset with all the attributes. This baseline test helps to understand how all the features are equally important in the classification. Then in the second approach, the imbalance problem was solved using SMOTE, and models were trained on the balanced dataset. In the third approach, the PCA method was used to reduce the dimensionality and feature selection to select the key features while preserving the most important information.

Parkinson's classification using the audio data with all the attributes gives an accuracy of 92% with Random Forest. Because each of the 22 attributes in the MDVP dataset is given equal weight, the RF model produces perfect results. On the other when using the balanced dataset, the accuracy score rises from 89 to 92% for SVM, and 92 to 94% for RF. The precision for SVM also rise to 91% from 88%. DT also performs best with a balanced dataset. However, the accuracy score of logistic regression dropped to 87% from 89%, and the precision score was improved from 88 to 90%. Thus, it can be seen that the models perform best when trained with the balanced dataset as shown in Fig. 5. Using SMOTE proves to be beneficial and demonstrates that a balanced dataset improves the classification performance of the models. The study also uses the PCA method which selects the critical features with the maximum variance while preserving most of the critical information. However, it is observed that using the PCA method does not yield as much satisfactory results. As can be observed the RF model, LR, and DT obtained an average accuracy score of 89%, 82%, and 87% respectively. Therefore, it is observed that feature selection did not perform well in this case, due to the importance of all the features in the training. It is suggested to use all the features for the training of models.

This study uses acoustic measures for the diagnosis rather than the widely accepted diagnostic biomarkers such as MRI scans, PET, or DaT scans. Since almost 90% of Parkinson's patients have vocal injuries, using voice measures as a diagnostic biomarker will give more accurate and faster results as compared to harmful methods such as neuroimaging testing and MRI scans. Table 4 shows the result analysis of the suggested approach, and the best results were achieved with SMOTE using all the features. Machine learning models perform best with Random Forests with an excellent accuracy of 94%.

Various researchers have used the voice measures for Parkinson's classification and regression tasks. A comparison of our study with the previous studies has been shown (Table 6) which used the same data. It can be seen that the proposed work performs best in this case. Alshammri et al.[10], they used SMOTE and hyperparameter tuning; the highest accuracy was 95% for SVM. However other models such as KNN, DT, and RF obtained lower accuracies as compared to our proposed approach. In[5], the study uses up-sampling to handle the class imbalance and trained four models with 23 attributes, all the models such as RF, SVM, and LR did not achieve accuracy above 85%. However, the study has limitations as there are more chances of overfitting due to similar values created by up-sampling. In[40], LR, DT, SVM, and Bagging to classify whether a person is having PD or not based on the acoustic features. This paper did not mention the use of SMOTE and the best results were obtained by SVM with an accuracy of 92%. The study[44] explores algorithms such as Logistic Regression, Support Vector, and Random Forest. SMOTE was used to balance the dataset. The machine learning models showed satisfactory results, however, the accuracy is low compared to this study. The study is compared with some previous studies in terms of accuracy (as shown in Fig. 10) and it can be seen that the best results were obtained in our proposed approach. The results show that SMOTE was the significant factor that contributed to the best performance of machine learning models. Also, the results shown by all three proposed approaches (as shown in Tables 3, 4, and 5) indicate that feature selection did not perform the best in this case as it did not yield satisfactory results. It suggests that each feature in the Parkinson's dataset is equally important for the classification of people as healthy or PD patients.

Our findings align with the existing literature by offering a more thorough assessment by the use of SMOTE and PCA in contrast with the previous studies. Previous studies frequently neglect the importance of analyzing dataset class imbalance and feature reduction techniques, which we have thoroughly examined. The increase in recall and precision, especially once the dataset has been balanced, shows that our method addresses frequent problems in medical datasets, where incorrect classifications can have serious repercussions.

References	Method	Accuracy (%)	Precision (%)	Recall (%)
10	SVM	95	98	96
	DT	90	96	92
	RF	92	98	92
5	LR	85	89	92
	RF	85	89	92
	SVM	81	82	94
40	LR	79	–	–
	DT	90	–	–
	SVM	92	–	–
44	SVM	82	–	70
	DT	87	–	80
	RF	92	–	100
	LR	84	–	85
16	RF	84	–	–
	LR	82	–	–
	KNN	91	–	–
Proposed work	SVM	92	91	100
	RF	94	94	100
	LR	87	90	93
	DT	89	96	90

Table 6. A comparison of the proposed study with some previous studies.

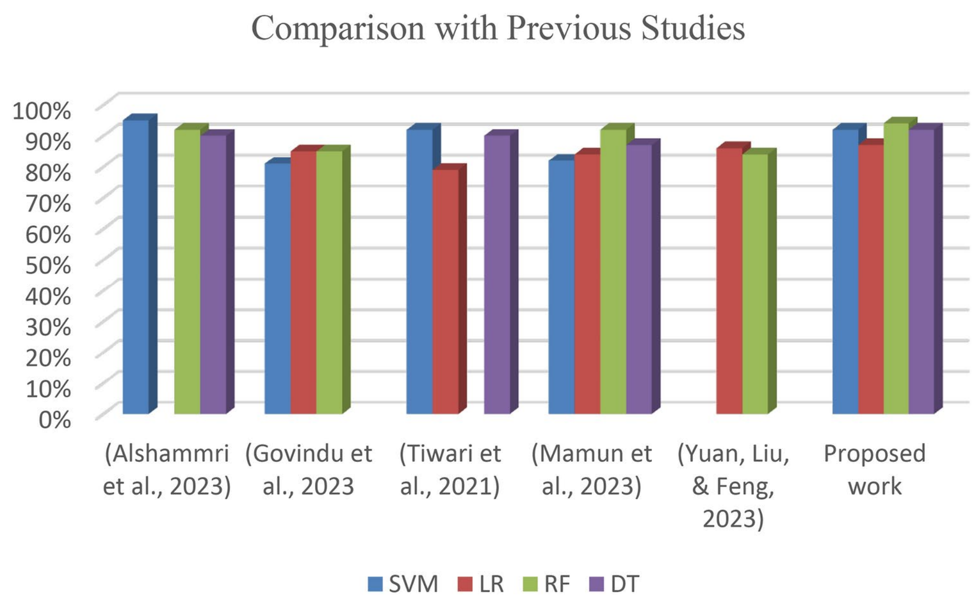

Fig. 10. Comparison of accuracy between the proposed work and previous studies.

Limitations and future work

Some limitations need to be highlighted even though the study explores the potential of vocal measures as strong predictors for the initial diagnosis of PD using machine learning. As the dataset contain biomedical voice measurements from 31 individuals, the size is quite small and may not properly represent the diversity among Parkinson's patients, thereby affecting the applicability of the results. Furthermore, even though the models showed strong performance on this dataset, their effectiveness when scaled to larger datasets has not been investigated. Additionally, it may face issues like distracting background sounds, differing recording settings, and speech variability when implement in real world applications. Future studies should emphasize broadening datasets to involve a more varied population, using the models into clinical decision-support tools. Other biomarkers such as MRI scans, gaits and handwriting can also be used in combination with voice for better results. Deep learning methodologies can be investigated to enhance feature extraction and classification precision. Additionally, such a voice analysis system could promote practical use in telemedicine by improving the model's performance. Focusing on these elements will address the gap between experimental findings and real-world implementation, enhancing the reliability of AI-based detection of PD.

Conclusion

Parkinson which affects the lives of the elderly is a disease caused due to deficiency of dopamine in the brain that controls the person's motor movements. Numerous psychological and physical examinations and specialized scans of the patient's nervous system are required for the diagnosis, presenting several problems. The study explores the potential of vocal measures as strong predictors for the initial diagnosis of PD. The study investigates several machine-learning methods, using an audio dataset with an extensive range of 23 voice parameters. SMOTE was used to balance the dataset with an equal distribution of classes. The dimensionality of the dataset was reduced using the PCA method to choose the relevant features. However, the results show that the models perform best using SMOTE. All the classifiers perform best in predicting healthy and PD patients. RF achieves an accuracy score of 94%, precision of 94%, and f1-score of 96%. While SVM and DT achieve an accuracy of 92%. With feature selection, the models didn't perform well and achieved an accuracy of 87% for DT, and 89% for RF. The study also investigates the effect of the imbalance class on the predictive performance of models. The results highlight the capability of vocal features as a significant indicator for the early diagnosis of disease which could improve the way of living of Parkinson's patients.

Data availability

The data is available on the https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection.

Received: 10 October 2024; Accepted: 1 April 2025
Published online: 09 April 2025

References

1. Alalayah, K. M. et al. Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method. *Diagnostics* **13**(11), 1924 (2023).
2. Sayed, M. A. et al. Parkinson's disease detection through vocal biomarkers and advanced machine learning algorithms. *J. Comput. Sci. Technol. Stud.* **5**(4), 142–149 (2023).

3. Ayaz, Z. et al. Automated methods for diagnosis of Parkinson's disease and predicting severity level. *Neural Comput. Appl.* **35**(20), 14499–14534 (2023).
4. Lim, W. S. et al. An integrated biometric voice and facial features for early detection of Parkinson's disease. *NPJ Parkinson's Dis.* **8**(1), 145 (2022).
5. Govindu, A. & Palwe, S. Early detection of Parkinson's disease using machine learning. *Procedia Comput. Sci.* **218**, 249–261 (2023).
6. Saravanan, S. et al. A systematic review of artificial intelligence (AI) based approaches for the diagnosis of Parkinson's disease. *Arch. Comput. Methods Eng.* **29**(6), 3639–3653 (2022).
7. Iyer, A. et al. A machine learning method to process voice samples for identification of Parkinson's disease. *Sci. Rep.* **13**(1), 20615 (2023).
8. Yuan, L., Liu, Y. & Feng, H.-M. Parkinson disease prediction using machine learning-based features from speech signal. *Serv. Oriented Comput. Appl.* 1–7 (2023).
9. Khaskhoussy, R. & Ayed, Y. B. An I-vector-based approach for discriminating between patients with Parkinson's disease and healthy people. In *Fourteenth International Conference on Machine Vision (ICMV 2021)*, (SPIE, 2022).
10. Alshammri, R. et al. Machine learning approaches to identify Parkinson's disease using voice signal features. *Front. Artif. Intell.* **6**, 1084001 (2023).
11. Bougea, A. New markers in Parkinson's disease. *Adv. Clin. Chem.* **96**, 137–178 (2020).
12. Tuncer, T. et al. TPat: Transition pattern feature extraction based Parkinson's disorder detection using FNIRS signals. *Appl. Acoust.* **228**, 110307 (2025).
13. Kaur, S. et al. Voice analysis and gait patterns: Machine learning approaches for Parkinson's disease detection. In *AIP Conference Proceedings*, (AIP Publishing, 2024).
14. Balnarsaiah, B. et al. Parkinson's disease detection using modified ResNeXt deep learning model from brain MRI images. *Soft. Comput.* **27**(16), 11905–11914 (2023).
15. Salari, N. et al. The performance of various machine learning methods for Parkinson's disease recognition: A systematic review. *Curr. Psychol.* **42**(20), 16637–16660 (2023).
16. Yuan, L., Liu, Y. & Feng, H.-M. Parkinson's disease prediction using machine learning-based features from the speech signal. *SOCA* **18**(1), 101–107 (2024).
17. Ali, L. et al. A novel sample and feature dependent ensemble approach for Parkinson's disease detection. *Neural Comput. Appl.* **35**(22), 15997–16010 (2023).
18. Nijhawan, R. et al. A novel artificial-intelligence-based approach for classification of Parkinson's disease using complex and large vocal features. *Biomimetics* **8**(4), 351 (2023).
19. Shastry, K. A. An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease. *Healthc. Anal.* **3**, 100181 (2023).
20. Tougui, I., Jilbab, A. & El Mhamdi, J. Machine learning smart system for Parkinson disease classification using the voice as a biomarker. *Healthc. Inf. Res.* **28**(3), 210–221 (2022).
21. Suppa, A. et al. Voice in Parkinson's disease: A machine learning study. *Front. Neurol.* **13**, 831428 (2022).
22. Raundale, P., Thosar, C. & Rane, S. Prediction of Parkinson's disease and severity of the disease using machine learning and deep learning algorithm. In *2021 2nd International Conference for Emerging Technology (INCET)*, (IEEE, 2021).
23. Leung, K. H. et al. A three-stage, deep learning, ensemble approach for prognosis in patients with Parkinson's disease. *EJNMMI Res.* **11**, 1–14 (2021).
24. Neto, O. P. Harnessing voice analysis and machine learning for early diagnosis of Parkinson's disease: A comparative study across three datasets. *J. Voice* (2024).
25. Alkhatib, R. et al. Machine learning algorithm for gait analysis and classification on early detection of Parkinson. *IEEE Sens. Lett.* **4**(6), 1–4 (2020).
26. Wang, W. et al. Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access* **8**, 147635–147646 (2020).
27. Ahmed, I. et al. Classification of Parkinson disease based on patient's voice signal using machine learning. *Intell. Autom. Soft Comput.* **32**(2), 705 (2022).
28. Hema, M. et al. Prediction analysis for Parkinson disease using multiple feature selection & classification methods. *Multimed. Tools Appl.* **82**(27), 42995–43012 (2023).
29. Quan, C., Ren, K. & Luo, Z. A deep learning based method for Parkinson's disease detection using dynamic features of speech. *IEEE Access* **9**, 10239–10252 (2021).
30. Rahman, S. et al. Classification of Parkinson's disease using speech signal with machine learning and deep learning approaches. *Eur. J. Electr. Eng. Comput. Sci.* **7**(2), 20–27 (2023).
31. Rana, A. et al. An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features. *Electronics* **11**(22), 3782 (2022).
32. Aljalal, M. et al. Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques. *Diagnostics* **12**(5), 1033 (2022).
33. Hossain, M. A. & Amenta, F. Machine learning-based classification of Parkinson's disease patients using speech biomarkers. *J. Parkinson's Dis.* **14**, 1–15 (2024).
34. Yadav, S., Singh, M. K. & Pal, S. Artificial intelligence model for parkinson disease detection using machine learning algorithms. *Biomed. Mater. Devices* **1**(2), 899–911 (2023).
35. Prakash, D. B., Narendar, C. & Sumalatha, D. Handling class imbalance problem in machine learning using synthetic minority oversampling technique (SMOTE).
36. Raju, V. G., et al. Study the influence of normalization/transformation process on the accuracy of supervised classification. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (IEEE, 2020).
37. Hasan, B. M. S. & Abdulazeez, A. M. A review of principal component analysis algorithm for dimensionality reduction. *J. Soft Comput. Data Min.* **2**(1), 20–30 (2021).
38. Bharadiya, J. P. A tutorial on principal component analysis for dimensionality reduction in machine learning. *Int. J. Innov. Sci. Res. Technol.* **8**(5), 2028–2032 (2023).
39. Saeed, F. et al. Enhancing Parkinson's disease prediction using machine learning and feature selection methods. *Comput., Mater. Contin.* **71**(3), 5639–5658 (2022).
40. Tiwari, H. et al. Early prediction of Parkinson disease using machine learning and deep learning approaches. *EasyChair Prepr.* **4889**, 1–14 (2021).
41. Mall, P. K. et al. Early warning signs of Parkinson's disease prediction using machine learning technique. *J. Pharm. Negat. Results* 4784–4792 (2022).
42. Guan, Y. Application of logistic regression algorithm in the diagnosis of expression disorder in Parkinson's disease. In *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, (IEEE, 2021).
43. Sharanyaa, S., Renjith, P. N. & Ramesh, K. Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, (IEEE, 2020).
44. Mamun, M. et al. Vocal feature guided detection of parkinson's disease using machine learning algorithms. In *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, (IEEE, 2022).

Author contributions
All authors have equally contributed.

Funding
Not applicable.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.D., T.M. or M.M.S.