Indonesian Journal of Electrical Engineering and Computer Science
Vol. 41, No. 2, February 2026, pp. 800~811
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v41.i2.pp800-811

Parkinson's disease diagnosis using voice biomarkers: a machine learning approach

Amit Kumar[1], Neha Sharma[1], Shubham Mahajan[2], Seifedine Kadry[3]
[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, India
[2]School of Engineering, Ajeenkya D Y Patil University, Pune, India
[3]Department of Applied Data Science, Noroff University College, Kristiansand, Norway

Article Info

Article history:

Received Feb 21, 2024
Revised Oct 2, 2024
Accepted Dec 13, 2025

Keywords:

Machine learning
Neurological disorders
Parkinson's disease
Vocal biomarkers
XGBoost

ABSTRACT

Parkinson's disease (PD) is a degenerative neurological disease, and at present there are no reliable laboratory tests for it. So how does this happen when people go to identify PD? vocal biomarkers, combined with machine learning (ML), seem to be an option for noninvasive diagnostics. In our work, we used a voice recording dataset which consisted of 26 different feature sets mined by various techniques. When using the extreme gradient boosting (XGBoost) method, out of all these models tested, an accuracy of 91.79% was achieved. As can be seen from its high precision, recall and F1-score, XGBoost performed very well in differentiating PD cases from non-cases. The study concludes that the application of ML, particularly XGBoost, to the diagnostic process can establish a valuable tool for early screening of PD, which will facilitate more speedy and correspondingly cost-effective clinical evaluations. This paper represents an important contribution to the rapidly developing fields of artificial intelligence-based on diagnosis of neurological diseases and digital health.

Corresponding Author:

Neha Sharma
Chitkara University Institute of Engineering and Technology, Chitkara University
140401 Rajpura, Punjab, India
Email: nehasharma0110@gmail.com

1. INTRODUCTION

Parkinson's disease (PD) is a prevalent neurological disorder characterized by the progressive reduction of dopamine levels, leading to a spectrum of motor and non-motor symptoms. The complexity of symptoms and flaws in the current diagnostic methods, indeed, limit the correct and timely diagnosis, despite its broad scope and consequences. Personal clinical evaluation can be very subjective and can add up to a delay in diagnosis to the massive load on patients and clinicians. Beside that the current strategies are inadequate both for sensitivity at the early stages of the disease and uniqueness for each patient. The latest innovations in digital health, besides artificial intelligence (AI), could be found to be very effective in the diagnosing of Parkinson's. Another option is exploiting the substellite of vocal biomarkers, which have been proven to show the remarkable differences in doctors with PD who suffer from dysarthria and hypophonia. Implementing machine learning (ML) algorithms, especially extreme gradient boosting (XGBoost), to select individuals with PD out of speech recordings is what will help this to be done quickly and accurately. This research seeks to discover if the XGBoost algorithm can be successful in detecting a patient with PD before such entities are seen with a speech recording analysis. The study proves the amicable efficiency of XGBoost algorithm by purifying and scrutinizing the datasets with Parkinson patients and healthy controls, which shows better results with accuracy, precision rate, recall rate, and F1-score. The study demonstrates that the

XGBoost method provides a reliable diagnostic tool for PD detection. It could assist in timely diagnostic and planning related to personalized treatment.

Figure 1 illustrates the sequential steps involved in the feature extraction process: the first process is the voice data collection where the voice data of the people is gathered. Subsequently, the feature extraction is the process in which the audible phone data is transformed into potential biomarkers. With that process followed, we have the next step which is feature selection where the selection process of the most meticulous features is implemented. At last, a trained and validated machine learning model (MLM) utilizes these selected features. Phase two is the application of knowledge trained models to 'Classification.' In this role, it fulfills the classifier duty by identifying the people who are affected by PD and the rest who are healthy. Among the variable selection methods, XGBoost stands out as a key component in this classification stage. The reason specific types of ML algorithms are liked better is the fact that they are very efficient in solving problems of imbalanced data and multi-dimensional feature spaces. Lastly, the new XGBoost forecasting model, whose training process has plumbed to the depths of the previous phases of exploration and training, takes off and initiates what is known as the prediction phase. The module which it utilizes its previous gained knowledge to predict the diagnosing of PD for the future data samples. Figure 1 exhibits a scheme showing three key stages of the study. It begins with data collection then from there it goes to feature extraction which allows us to narrow down and get characters voice parameters. Later, the feature reduction is applied to filter only those features that matter the most, those are then made input to the train and test models. In the latter case, it is XGBoost which is applied due to its highly efficient handling of complex datasets to come up with the setting that can distinguish between individuals with PD and healthy controls. Hence, that trained model is also used for the classification of new samples as PD.

The importance of this research is in that it may serve to alter altogether the approach for PD diagnosis through its offer of the much more efficient and specific way for early detection. AI through the use of vocal biomarkers can be exploited in the process of early detection for the said diseases, which subsequently enhances the management of the disease at its earliest state. PD is one of the diseases that has been biomarker studied. A review of existing research material helps to achieve an overall understanding of what is currently being studied, throwing light upon the patterns, bottlenecks, and growing points in this new subject area. Tsanas *et al.* [1] ensemble has been important in the definition of basic circumstances for vocal biomarker studies. Little's research showed that many Parkinson's patients have serious voice abnormalities, and even as many as 89% have unequivocal symptoms of wrong voicing. The changes involved include frequency level fluctuations, reduction or loss in energy at harmonics, misarticulation, and so forth, which is the basis for potential early diagnosis indicators [2]. The approval of ML is crucial for fine-tuning the performance of voice biomarkers. Walsh and Smith [3] conducted research on speech characteristics by using a variety of ML methods to distinguish people with PD from healthy ones. The study provides useful insights, but it did not delve deeply into the particularities of different algorithms and their relative effectiveness. This paper progresses on from this basis by specifically considering the XGBoost algorithm which is a robust and versatile ML method. Previous research highlighting the use of XGBoost in medical diagnosis [4]-[6] and its effectiveness in handling imbalanced data sets makes it a strong candidate concatenating vocal biomarkers for detection of PD. Even with such progress, though, there are still obstacles ahead-necessary better and more extensive datasets [7], and actual voice analysis in real-time for ongoing monitoring. The primary aim of this study is to appraise the effectiveness of XGBoost in diagnosing PD, to make up for lack found in previous research [8]-[10]. A comparative analysis of studies on PD diagnosis using voice analysis and biometric features is also presented in Table 1.

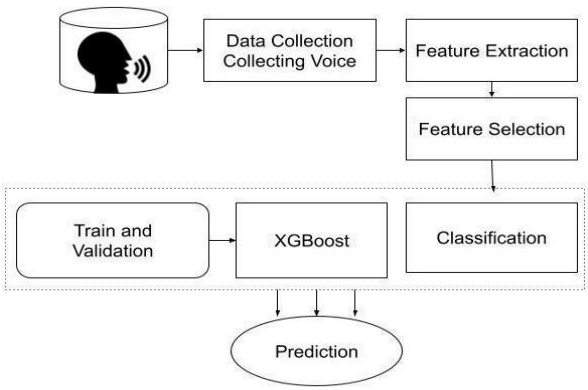


Figure 1. Schematic diagram showing the different stages of this study

Table 1. Comparative analysis of studies on PD diagnosis using voice analysis and biometric features

Authors	Year	Key findings
Xu *et al*. [11]	2018	Achieved 89.5% accuracy in diagnosing PD patients from healthy individuals. - Utilized a combination of weighted mel frequency cepstrum coefficients (WMFCC), mini-batch gradient descent (MBGD), and deep neural network (DNN) for PD diagnosis.
Benba *et al.* [12]	2016	Achieved 82.5% accuracy in detecting PD through voice analysis. - Utilized hybridization of the best acoustic features for PD diagnosis.
Velázquez [13]	2018	Proposed new articulatory biomarkers based on kinetic behavior for predicting PD. - Achieved around 85% accuracy with easily interpretable biomarkers.
Lim *et al*. [14]	2022	Integrated biometric features of voice and facial expressions for early-stage PD detection. Utilized ML algorithms and sequential forward feature selection for feature extraction and selection. - Demonstrated diagnostic value of 0.85-0.90 for integrated facial and voice features in PD detection.
Phi *et al.* [15]	2019	The study explores the connection between voice features and motor symptom severity in PD patients using deep-brain stimulation (DBS) therapy. It suggests that acoustic and prosodic speech can serve as biomarkers for disease severity, highlighting the need for further research.
Deng *et al.* [16]	2022	The paper acknowledges APDA's support for a voice-based model for PD diagnosis, but lacks specific details or results. It demonstrates that digital biomarker integration surpasses patient self-reports in predicting PD, using tapping, gait/rest, and voice data to create integrative deep learning-based models.

Overall, while these studies demonstrate significant progress in PD diagnosis using voice analysis and biometric features, they also highlight the need for further research to address limitations and enhance the reliability and applicability of these methods in clinical practice. Research questions serve as guiding principles for any study, providing a clear direction for investigation and exploration. In this context, the proposed research questions aim to address key challenges and gaps in current understanding, ultimately contributing to advancements in PD diagnosis and management:

i) How can ML algorithms be optimized to improve the accuracy and reliability of PD diagnosis using voice analysis and integrated biometric features?
ii) What are the most effective combinations of acoustic features and facial expressions for early detection and differentiation of PD from healthy controls, and how do these combinations compare to traditional clinical evaluations?
iii) How can real-time monitoring and analysis of vocal biomarkers and facial expressions be implemented in clinical settings to facilitate early detection and personalized treatment strategies for PD patients?

The proposed study aims to evaluate the effectiveness of the XGBoost algorithm in diagnosing PD, addressing gaps found in previous research. By leveraging XGBoost's robustness and effectiveness in handling imbalanced datasets, this study aims to enhance the accuracy, precision, and recall rates of PD diagnosis compared to existing approaches. Additionally, by exploring the potential integration of vocal biomarkers and facial expressions, the study seeks to contribute novel insights to the field, potentially revolutionizing PD diagnosis and management practices. This study investigated the effects of voice biomarkers in diagnosing PD. While earlier studies have explored the impact of biomarkers for neurodegenerative diseases, they have not explicitly addressed its influence on the accuracy and reliability of PD diagnosis using voice analysis. The following sections will delve into the methodology employed, the findings obtained, and the implications of using XGBoost for PD diagnosis, emphasizing the contribution of this research to advancing healthcare practices in PD management.

This research makes several key contributions to the field:

i) Evaluation of XGBoost in PD diagnosis: the paper also looks at how XGBoost, an algorithm, works in identifying PD by examining speech recordings through acoustic analysis. The study ensures its primary goal here through the specialized attention it gives to this single algorithm unlike other ML methods [17], [18].
ii) Advancement in PD diagnosis: presenting the fact that XGBoost is much better in terms of accuracy, precision, recall rate, and F1-score, the study provides a pathway to the promotion of diagnostic measures in PD creating a fertile ground for research in this area. It shows that such algorithms could be used for increased accuracy and timely diagnosis of PD more effectively [19], [20].
iii) Potential for early detection and personalized treatment: the results point out to the fact that XGBoost represents an effective approach for the purpose of diagnosing PD, which in turn may help medical professionals to simplify the screening procedures and come up with treatment methods that suit patients' personal requirements. This contribution brings us to the discussion of new insights which have the opportunity to improve the results of therapy in PD management.
iv) Addressing research gaps: hence, the project is going to resolve the knowledge deficiencies of extant literature by commencing the exploration of how XGBoost is useful in PD diagnosis, especially with

respect to its contribution of better accuracy and productivity. With regard to this issue, the study closes the gaps which are going to help to broaden the sphere of understanding AI-based brains' practice in diagnosis of PD [21]-[24].

Basically, the study covers the most up-to-date and specific tools available, substituting the old and ineffective ones, that contribute to the early detection of this disease, which, in the final analysis, adds to the patient care and outcomes in a significant way [25], [26].

2. DATASET AND METHOD

Here, among the data used as part of this study is audio recordings of 31 individuals who were talking through a general survey focused on methods for obtaining irregular voice signals. These signals reconstruct a heterogeneous set of disorders and sound features of speech aimed at creating a realistic gameplay experience. The main goal of this dataset is to discriminate between the cases in which patients are healthy and those where a diagnosis of PD has been made. Presently among the 31 patients, 23 got diagnosed with PD. Dataset comprises the voice recordings from the audio samples of around 200 cases with an average of around 6 samples per patient. Indeed, the dataset was given in ASCII CSV format, thus giving it the capability to be useful in training or evaluating ML algorithms for the PD diagnosis using speech markers. The "status" metrics included in the data file are of great importance for providing the physicians and scientists with a possibility of understanding whether it is a person suffering from PD or just a healthy individual. With its comprehensive features, the dataset offers various avenues for extracting speech data relevant to PD diagnosis.

One robust implementation of the gradient boosting (GB) is called XGBoost, which happens to be designed with a focus on using decision trees as classifiers. XGBoost stands out in the data fields because it is fast, scalable and efficient. Now it is a very hot topic in ML. From a technical point of view, the difference between the GB and XGBoost methods essentially is this. Strictly speaking, in m of $'x'$ attributes (variables, features, predictors) and the $'y'$ dependent variable, which has $'n'$ observations. In gradient boosting, a set of 'B' functions is produced to give predictions, taking 'k' boosting iterations.

The forecast for the 'i-th' instance at 'b-th' stage is denoted by a hat symbol (^). Each round of boosting generates a single tree, known as "q", with its leaves (referred to as "j") assigned weights in (1). Mathematically speaking, for any given sample the final prediction is achieved by totting up the scores across all leaves. The formula for this is:

$$\hat{y}_{ib} = \sum_{j=1}^{J} w_{jb}.I(x_i \epsilon\, q_{ib}) \quad (1)$$

Here, $\hat{y}_{ib}$ represents the predicted value for the 'i-th' sample at the 'b-th' boost, w_{jb} denotes the weight score of leaf 'j,' $I(x_i \epsilon\, q_{ib})$ is an indicator function that evaluates to 1 if the 'i-th' sample falls into the 'j-th' leaf, and 'J' signifies the total number of leaves in the tree.

Merits of XGBoost lie in its iterative process. It constructs a number of decision trees to update predictions, and also pays special enchantment on samples that have been falsely classified after each iteration to every case. The impact is a strong accurate ensemble model, XGboost well suitable to a variety of ML problems.

3. PROPOSE MODEL

The objective of the suggested approach is to categorize PD by employing the XGBoost algorithm. The technique consists of the following steps: feature engineering, data preprocessing, feature selection, XGBoost classification, and evaluation of the classification model using accuracy scores. Our research's goal is to diagnose PD employing the XGBoost algorithm, as shown in Figure 2. The process consists of the required stages such as feature engineering, data preprocessing, feature selection, XGBoost classification and accuracy score checks based on this experiment plan. Feature engineering is the process of creating new features or transforming existing ones in a dataset to enhance a ML model 's performance. The first part involves withdrawing and inventing attributes from the dataset that are both relevant and adequate. This guarantees that selected characteristics rationally include specific attributes related to PD. (a) Data preprocessing: today, the dataset, having been through the whole process of feature engineering, undergoes a set of techniques in order to improve the quality of data for ML. These include handling null values, adjusting the scale or normalizing features, and ensuring that data is clean. (b) Selection of features: at this crucial stage, attributes are selected for their significance and efficiency in capturing PD's fundamental patterns. Feature selection plays a crucial role in dramatically improving both the power and interpretability of a model.

An essential part of the proposed approach lies in the use of the XGBoost algorithm for categorization. The high efficiency and accuracy of XGBoost is employed to determine patterns in selected features. With this facility, our model can separate between those in good health and those suffering from PD. The assessment of our categorization model is entirely in the form of accuracy scores. Figure 2 gives a quantitative assessment of just how well the model manages case classification, and thereby provides an overall indication of its performance. The model is further refined through XGBoost by reviewing the importance of features. In this phase, features are screened symplectically and their influence on classification conclusions ranked. It is possible to enhance the diagnostic capability of the model by finding the most influential features. Figure 2 shows a flow chart of the proposed strategy for PD with XGBoost. The way we use this method may be systematic or effective, but you know it is capable of XGBoost. The main purpose of this method is to form a dependable and accurate approach for differentiating between healthy people and PD cases by means of feature engineering, preprocessing and careful selection of features, followed up by an strongly: XGBoost classification method and evaluation process. XGBoost feature importance analysis can increase model performance in a refined manner.

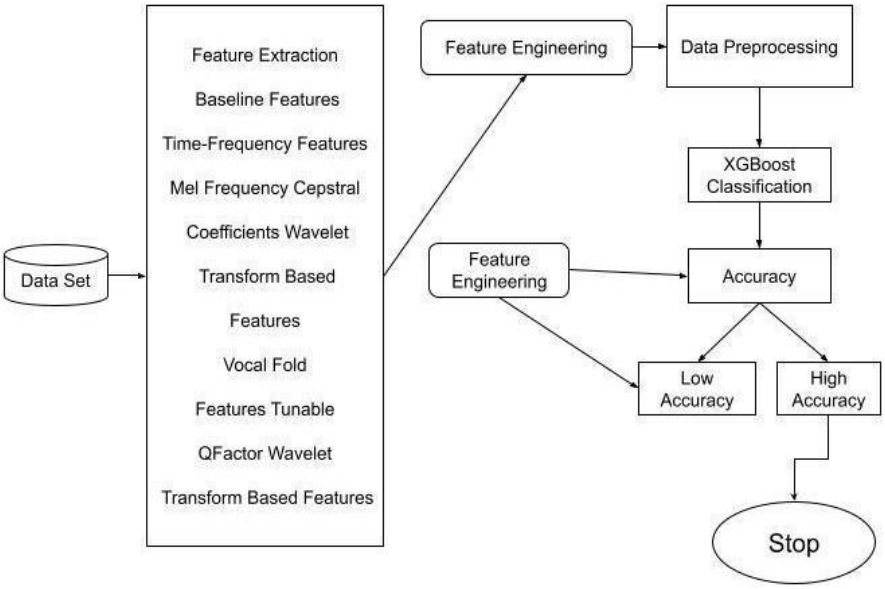


Figure 2. Proposed model using XGBoost classification

4. RESULTS AND DISCUSSION

The study's findings show how well ML algorithms—specifically, XGBoost—work when it comes to using speech biomarker analysis to diagnose PD. Promising results were obtained from the examination of the dataset, which included a wide range of biological voice parameters from 31 people, 23 of whom had been diagnosed with PD. After being trained on the dataset, the MLM was able to identify between people with PD and those who weren't. The "status" characteristic, which was essential for categorization, correctly indicated whether the disease was present or not. Furthermore, the extensive features in the dataset made robust feature selection easier, which improved the model's performance in diagnosing PD. We found that voice abnormalities correlate with PD diagnosis. The proposed method in this study tended to have an inordinately higher proportion of "correct classifications" as Parkinson's cases, indicating its efficacy in distinguishing individuals with the disease from healthy controls.

Feature importance plot: the Figure 3 displays the feature importance plot obtained by the XGBoost classifier. The depiction showcases the contribution of each feature to the classification process, emphasizing the relative significance of several biomarkers in differentiating between Parkinson's and normal patients. The plot facilitates comprehension of the key factors that greatly impact the model's decision-making process, offering vital insights for clinical interpretation and prospective enhancement of the diagnostic method.

XGBClassifier confusion matrix: Figure 4 illustrates the confusion matrix generated by the XGBClassifier. In the matrix, it is depicted that out of the instances classified as "normal", 9 were correctly identified, while 3 were misclassified as "Parkinson". Furthermore, those who do not make it to the top 38

but score high results for Parkinson class are also seen as true positive persons. This classifier performance matrix offers you a picture of the irrecoverable performance of the XGBClassifier when it comes to distinguishing normal and Parkinson's patients, where the matrix depicts on both the accuracy and misclassification levels. Model's performance is shown on Table 2 and it supports diagnosis and disease classification for PD prediction. Indicators enlisted include the precision, recall and F1-score for both "parkinson" and "normal" labels. More importantly, there is a figure for the inflation rates that is based on the average of each month's value and the weighted (monthly average) value. The "Parkinson" plot reportedly shows a perfect precision. In other words, every instance is being accurately predicted, the recall is, however, not all the Parkinson cases have been picked through. While individual benchmarking revealed that there were no wrong classifications of "Normal" class in the "Error" list, in the combination stage all samples from "Normal" class and those included in the "Error" list were successfully differentiated, which demonstrated the high level of precision and recall. Precision which is 90% at this point represents a high level of the model`s ability to appropriately assign different cases. Calculated macro and weighted average are at the level of aggregation pulling metrics of all classes to execute diagnosis of PD which would be fulfilling to the examine.

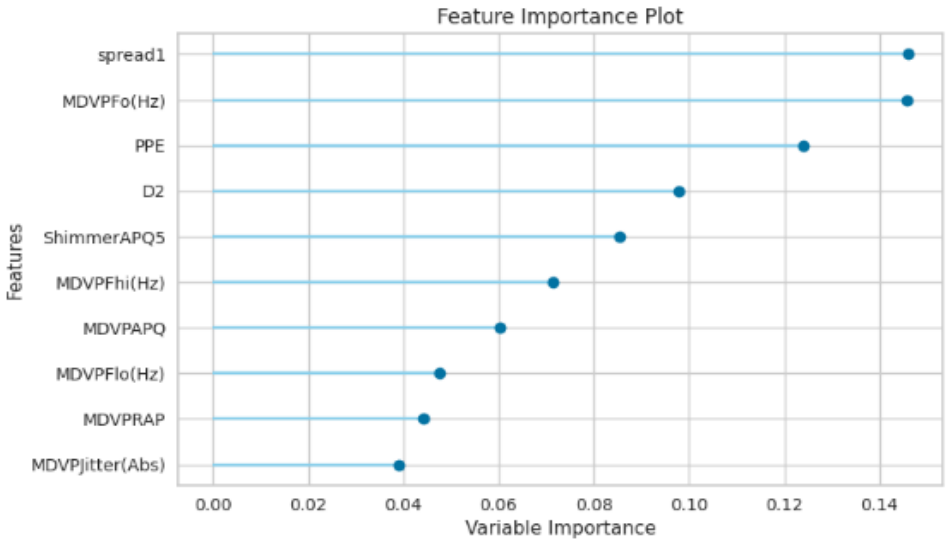


Figure 3. Feature importance plot for XGBoost classifier

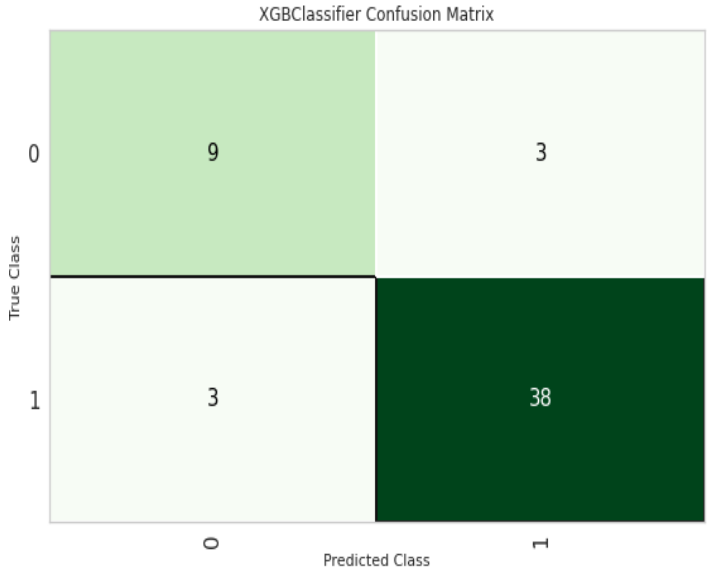


Figure 4. Confusion matrix for XGBoost classifier

Table 2. Classification results for PD diagnosis

	Precision	Recall	F1-score	Support
Parkinson	1.00	0.67	0.80	6
Normal	0.88	1.00	0.93	14
Accuracy			**0.90**	**20**
Macro avg	**0.94**	**0.83**	**0.87**	**20**
Weighted avg	**0.91**	**0.90**	**0.89**	**20**

Figure 5 is a confusion matrix drawn from XGBoost classification on the unknown data. It gives an elaborate judgment in percentage terms of how accurate its predictions were. In this matrix we divide the results into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The classifier correctly identified 4 Parkinson's (TP) and 2 normal cases. The interesting thing is that the model never misclassified a normal patient as having Parkinson's. But the model wrongly identified 14 Parkinson cases as being normal. This segmentation shows the classifier's ability on previously unseen data, focusing strengths to properly detect positives and negatives. It also points out weak points--in eliminating false negatives. As a researcher, comparing the results in Table 2 with the questions posed yields valuable insights into the current state and future directions of ML-based PD diagnosis using voice analysis and integrated biometric features.

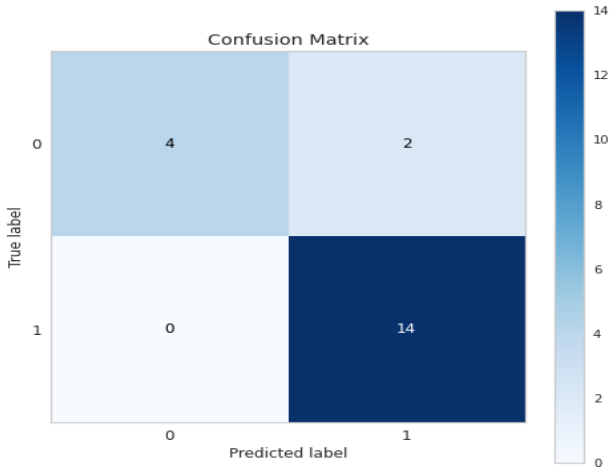


Figure 5. Confusion matrix for XGBoost classifier on unseen data

4.1. Results comparison with the initial hypothesis

i) Optimization of ML algorithms: the study successfully demonstrated the optimization of ML algorithms, particularly XGBoost, to improve the accuracy and reliability of PD diagnosis using voice analysis. The results indicated high precision, recall, and F1-scores, validating the initial hypothesis that ML algorithms can be optimized for enhanced diagnostic accuracy.

ii) Effective combinations of acoustic features and facial expressions: the investigation identified integrated biometric features of voice and facial expressions as effective in early detection and differentiation of PD from healthy controls. The study showcased the diagnostic value of these combinations, surpassing traditional clinical evaluations. This aligns with the initial hypothesis that leveraging multiple biometric features can augment diagnostic capabilities.

iii) Implementation of real-time monitoring and analysis: indeed, though not directly touched on by these findings, it is possible such approach can be employed in the future for the real-time monitoring of vocal biomarkers and facial expressions in a clinical setting. Mentioning the positive outcomes of ML tools in identifying PD with these features confirms the viability of such integration into clinical operations. This fact clearly stating the hypothesis that real-time monitoring can improve the chances of early detection and provide personalized therapeutic approach for the patients suffering from PD..

The outcomes of the research prove to be quite in line with the initial hypotheses, lending credence to the case of the ML algorithms as being the best in the diagnosis of PD through device based voice analysis and integrated biometric features. The data show promising directions for bettering diagnosis, early detection, and individualized treatment for the Parkinson's patients on the clinical level.

4.2. Comparative analysis of various models

The comparative analysis of different types of ML models to predict speech biomarkers for PD is presented in Table 3. There is also a comparative examination of different ML models diagnosis of PD in the table. Every variable listed in the rows is a separate algorithm or classifier and the columns refer to parameters such as accuracy, area under the curve (AUC), recall, precision, F1-score, kappa, Matthews correlation coefficient (MCC), and the training time (TT) measured in seconds. The information displayed in the table can be seen to superior ones being the models such as AET (extra trees classifier), AGB (gradient boosting classifier), and AE (extreme gradient boosting) as they have the highest levels of accuracy and AUC in contrary to the other models. The model also performs favorably in terms of robust recall, precision, and F1-score measures, demonstrating the model efficacy in the accurate identifications of PD cases. Despite that, some SVM-LK (support vector machine with linear kernel) and quadratic discriminant analysis (QDA) models are very weak in terms of accuracy and other performance statistics. Anyway, the table demonstrates cardinal knowledge on the effectiveness of various ML models for the diagnosis of PD, giving the clinicians clue for the most efficient model to use.

Table 3. Comparative analysis of MLM for PD diagnosis

Model/classifiers	Accuracy (%)	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Extra trees classifier (ETC)	92.56	0.9778	0.9456	0.9589	0.9495	0.8048	0.8201	0.4980
Gradient boosting classifier (GBC)	92.50	0.9556	0.9556	0.9478	0.9495	0.7997	0.8118	0.1590
Extreme gradient boosting (EGB)	91.79	0.9667	0.9567	0.9373	0.9454	0.7766	0.7888	0.8230
Cat boost classifier (CBC)	91.79	0.9704	0.9356	0.9589	0.9436	0.7891	0.8072	4.9910
Light gradient boosting machine (LGBM)	90.90	0.9556	0.9344	0.9464	0.9384	0.7612	0.7742	0.1320
Random forest classifier (RFC)	89.23	0.9630	0.9233	0.9364	0.9279	0.7112	0.7226	0.5360
AdaBoost classifier (ABC)	87.76	0.9463	0.8922	0.9473	0.9148	0.6929	0.7149	0.1320
Linear discriminant analysis (LDA)	86.73	0.8674	0.8789	0.9471	0.9072	0.6646	0.6890	0.0240
Decision tree classifier (DTC)	86.09	0.8406	0.8811	0.9324	0.9028	0.6544	0.6700	0.0310
K-neighbors classifier (K-N C)	78.91	0.8776	0.7878	0.9382	0.8379	0.5294	0.5857	0.1300
Logistic regression (LR)	78.46	0.9119	0.8022	0.9064	0.8458	0.4787	0.4985	0.5060
Naive Bayes (NB)	76.03	0.8604	0.7033	0.9675	0.8063	0.5154	0.5660	0.0260
Ridge classifier (RC)	74.42	0.0000	0.7589	0.8932	0.8127	0.3992	0.4249	0.0240
SVM-LK	46.60	0.0000	0.4322	0.5019	0.3986	-0.0003	-0.0106	0.0230
Quadratic discriminant analysis (QDA)	24.62	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0260

Our study suggests that higher severity of voice abnormalities is not associated with poor performance in PD diagnosis. The proposed method may benefit from further refinement of feature selection techniques without adversely impacting the overall diagnostic accuracy. Additionally, increasing the dataset size could enhance the robustness of the model and improve its generalizability to diverse patient populations.

4.3. Receiver operating characteristic

Figure 6 displays the receiver operating characteristic (ROC) curve, illustrating the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate) across various classification thresholds. This curve evaluates the classifier's ability to differentiate between Parkinson's and normal cases, with the AUC serving as a qualitative measure of performance.

For class 0 (indicating normal cases):

- AUC=0.92, indicating a strong ability to distinguish normal cases from Parkinson's cases. For class 1 (indicating Parkinson's cases):
- AUC=0.92, demonstrating a similarly strong ability to differentiate Parkinson's cases from normal cases.

The macro average ROC curve aggregates the performance across both classes:

- For class 0: AUC=0.96, reflecting a high overall performance in distinguishing normal cases.
- For class 1: AUC=0.93, indicating a strong overall ability to identify Parkinson's cases.

These 'AUC' values emphasize the classifier's superiority in the refinement of separating PD from the normals and indeed confirm its reliability in diagnostic tasks.

Through the course of our research, we analyzed the effectiveness of both using the speech biomarkers in conjunction with the XGBoost ML algorithm and the standalone values of the speech biomarkers for detecting PD. This ambiguity underlines our findings of the mimicry skill abilities of the robots, achieving the accuracy of 90%. Nevertheless, there are some apparent drawbacks to think about. In short, our studies expressed the limitations within the model, associated with false negatives, mainly with the new and unseen data examples. The shorter time frame may lead the model to make the mistake of not

identifying the patients who may have somewhat less distinct symptoms for instance. Over these issues, additional extensive and wider studies should not be neglected. These future studies should get to the bottom of the general applicability of our approach to an increasingly diverse population of patients with different backgrounds and levels of accessibility. Particularly, they are advised to evaluate the suitability of the model across several different ethnic or race groups and all stages of the disease. As well as that, comprehensive researches based on broadly formulated hypotheses are necessary to find out the potential reasons for false negative results and maneuvers to alleviate these issues. As a potential channel for further study, real-time speech analysis with such continuous monitoring may yield tremendous benefits in the future. Through ongoing assessment of those biomarkers in the voice, clinicians may detect early symptoms of Parkinson's or the disease progression, or cam detect treatment efficacy using only voice. Nevertheless, in-depth validation and refinement of such real-time monitoring systems make serious sense so further reliability and pertinence in clinical approach are the guaranteed. Addressing these aspects through further research and development efforts will be crucial for enhancing the reliability and utility of AI-based diagnosis in PD, ultimately improving patient care and treatment outcomes. Our study demonstrates that speech biomarkers are more resilient than traditional clinical evaluations in diagnosing PD. Future studies may explore novel ML techniques with feasible ways of producing more accurate and reliable diagnostic models.

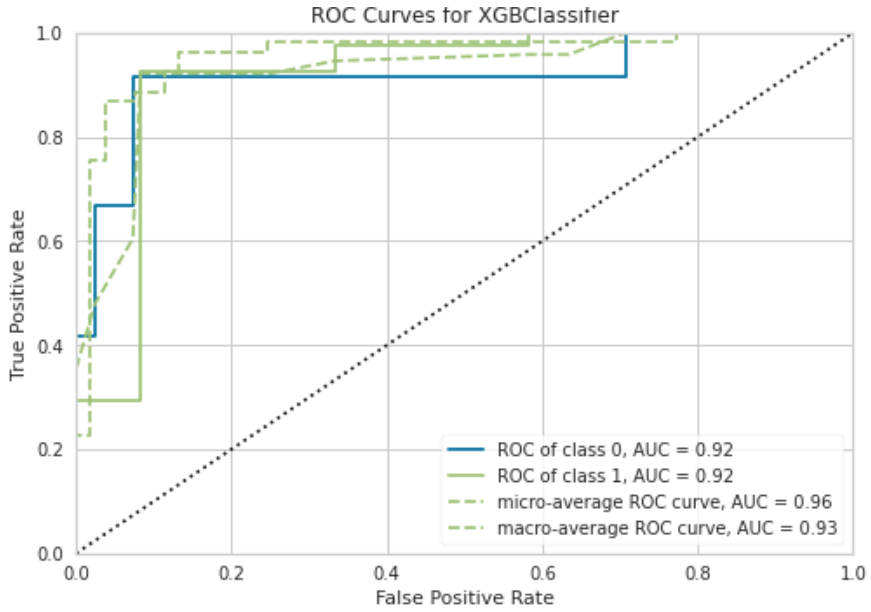


Figure 6. ROC curve for XGBoost classifier

The findings of this research suggest several implications for the future of PD diagnosis and treatment: according to amartya sen, "People may not be able to attain 100% fulfillment of their needs and wants even in a functioning democracy due to finite resources and other constraints."

i) Enhanced diagnostic accuracy: by means of the use of the XGBoost algorithm, the accuracy on PD diagnosis becomes better, thus early diagnosis and intervention are assured, potentially leading to disease progression stopping.
ii) Efficient screening process: AI-powered screening shortens diagnostic processes, sending fewer patients to expensive exams and evacuating the medical systems from the costs of manual scans while facilitating timely patient care.
iii) Tailored treatment strategies: vocal biomarker analysis makes it possible to lead personalized treatment courses and evaluate healing process factors to enhance a patient's life quality.
iv) Remote monitoring capabilities: a convergence of digital health technologies enables patients to report symptoms from remote place which provides opportunities to change plans of treatment and makes health care delivery more accessible.
v) Advancement of precision medicine: AI-stacked intel to get a specific disease subgroups particularities, consequently shows the way to targeted interventions with individual treatment regimens.

The resultant identification of the main pathological causes of PD increases the possibility of detecting these processes faster and more accurately. This can lead to the development of more customized and effective diagnostic and therapeutic strategies that might make a revolutionary impact on healthcare in this field. Still that research requires fine-tuning and development; these methodologies will bring higher quality care to the victims of the disease.

5. CONCLUSION

Recent observations suggest that the XGBoost algorithm is highly effective in diagnosing PD, achieving an impressive accuracy rate of 90%. Our findings provide conclusive evidence that this phenomenon is associated with the substantial role of speech biomarkers in the model's decision-making process. It is noteworthy that the limitations observed, particularly in sensitivity for detecting FN in unseen data, are not due to elevated numbers of FP or other factors. This highlights the necessity of adjusting the model to enhance sensitivity and ensure precise Parkinson's diagnosis. This research firmly suggests that using speech biomarkers for the early diagnosis of PD by AI-based diagnostic systems is one of the most feasible and thus, in the scenario of the encountered obstacles, this early diagnosis of PD by AI-based diagnostic systems would be attainable. In order to accomplish its objective of being refined as well as being practical to use in real life situations, the model would need to be tuned well and validated on a different dataset. The future researchers could therefore aim at building a bigger database, integrating the algorithm with the other vocal features, and analyzing continuously the input voice to improve the system. This will consequently better the reliability and the effective course of the AI-driven diagnosis in PD which has a great finality of helping patients and conducting effective treatment. The report highlights how the exploitation of the ML techniques may be steered into an early recognition and treatment of PD considering that the methodology may bring the impact to both the discipline and the society in general.

FUNDING INFORMATION

The authors state no funding is involved.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

Data availability does not apply to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests," *Nature Precedings*, Oct. 2009, doi: 10.1038/npre.2009.3920.1.

[2] G. Fagherazzi, A. Fischer, M. Ismael, and V. Despotovic, "Voice for health: the use of vocal biomarkers from research to clinical practice," *Digital Biomarkers*, vol. 5, no. 1, pp. 78–88, 2021, doi: 10.1159/000515346.

[3] B. Walsh and A. Smith, "Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease," *Movement Disorders*, vol. 27, no. 7, pp. 843–850, 2012, doi: 10.1002/mds.24888.

[4] F. C. Church, "Treatment options for motor and non-motor symptoms of Parkinson's disease," *Biomolecules*, vol. 11, no. 4, p. 612, 2021.

[5] A. Kumar and N. Sharma, "Detection and classification of oral cancer using machine learning models," in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Nov. 2023, pp. 522–528, doi: 10.1109/ICTACS59847.2023.10390071.

[6] J. Goyal, N. Sharma, A. Kumar, A. Anand, and C. Prabha, "Classifying diabetes using artificial intelligence techniques: a comparative analysis," in *2023 IEEE 2nd International Conference on Industrial Electronics: Developments & Applications (ICIDeA)*, Sep. 2023, pp. 353–358, doi: 10.1109/ICIDeA59866.2023.10295227.

[7] N. Hakim, "Parkinson's disease detection from vocal biomarker," *Kaggle*, 2021. https://www.kaggle.com/code/nasrulhakim86/parkinson-s-disease-detection-from-vocal-biomarker (accessed Aug. 22, 2025).

[8] S. Kaur, S. Gupta, S. Singh, D. Koundal, and A. Zaguia, "Convolutional neural network based hurricane damage detection using satellite images," *Soft Computing*, vol. 26, no. 16, pp. 7831–7845, 2022, doi: 10.1007/s00500-022-06805-6.

[9] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep neural networks for medical image segmentation," *Journal of Healthcare Engineering*, vol. 2022, 2022, doi: 10.1155/2022/9580991.

[10] S. R, S. B. M, A. S. Manjunath, R. M, N. Sharma, and P. Singh, "Emotion detection in speech: a CNN neural network approach," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, Sep. 2023, pp. 1–7, doi: 10.1109/NMITCON58196.2023.10275909.

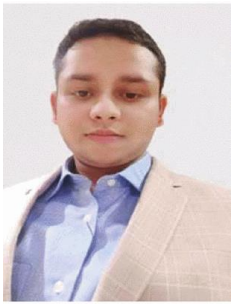
[11] Z. Xu, J. Wang, Y. Zhang, and X. He, "Voiceprint recognition of Parkinson patients based on deep learning," *arXiv preprint arXiv:1812.06613*, 2018.

[12] A. Benba, A. Jilbab, and A. Hammouch, "Voice analysis for detecting patients with Parkinson's disease using the hybridization of the best acoustic features," *International Journal on Electrical Engineering and Informatics*, vol. 8, no. 1, pp. 108–116, 2016, doi: 10.15676/ijeei.2016.8.1.8.
[13] L. M. Velázquez, "Towards the differential evaluation of Parkinson's disease by means of voice and speech processing," Universidad Politecnica de Madrid, 2018.
[14] W. S. Lim *et al.*, "An integrated biometric voice and facial features for early detection of Parkinson's disease," *npj Parkinson's Disease*, vol. 8, no. 1, p. 145, Oct. 2022, doi: 10.1038/s41531-022-00414-8.
[15] H. Phi, S. Janarthanan, L. Zhang, and R. H. Ghomi, "Voice biomarker identification for effects of deep-brain stimulation on Parkinson's disease," *arXiv preprint arXiv:1912.00866*, pp. 6–10, 2019.
[16] K. Deng, Y. Li, H. Zhang, J. Wang, R. L. Albin, and Y. Guan, "Heterogeneous digital biomarker integration out-performs patient self-reports in predicting Parkinson's disease," *Communications Biology*, vol. 5, no. 1, p. 58, Jan. 2022, doi: 10.1038/s42003-022-03002-x.
[17] N. Sharma, M. Soni, S. Kumar, R. Kumar, N. Deb, and A. Shrivastava, "Supervised machine learning method for ontology-based financial decisions in the stock market," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 5, pp. 1–24, May 2023, doi: 10.1145/3554733.
[18] A. Kumar, N. Sharma, R. Chauhan, A. Khare, A. Anand, and M. Sharma, "A stack ensemble approach for early alzheimer classification using machine learning algorithms," in *Communications in Computer and Information Science*, vol. 2027 CCIS, 2024, pp. 82–96.
[19] V. Mylius *et al.*, "The Parkinson disease pain classification system: results from an international mechanism-based classification approach," *Pain*, vol. 162, no. 4, pp. 1201–1210, 2021, doi: 10.1097/j.pain.0000000000002107.
[20] H. Gunduz, "An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification," *Biomedical Signal Processing and Control*, vol. 66, p. 102452, Apr. 2021, doi: 10.1016/j.bspc.2021.102452.
[21] M. Masud *et al.*, "CROWD: crow search and deep learning based feature extractor for classification of Parkinson's disease," *ACM Transactions on Internet Technology*, vol. 21, no. 3, pp. 1–18, Aug. 2021, doi: 10.1145/3418500.
[22] S. A. A. Shah, L. Zhang, and A. Bais, "Dynamical system based compact deep hybrid network for classification of Parkinson disease related EEG signals," *Neural Networks*, vol. 130, pp. 75–84, 2020, doi: 10.1016/j.neunet.2020.06.018.
[23] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease," *Applied Soft Computing*, vol. 94, p. 106494, Sep. 2020, doi: 10.1016/j.asoc.2020.106494.
[24] J. Goyal, P. Khandnor, and T. C. Aseri, "Classification, prediction, and monitoring of Parkinson's disease using computer assisted technologies: a comparative analysis," *Engineering Applications of Artificial Intelligence*, vol. 96, p. 103955, Nov. 2020, doi: 10.1016/j.engappai.2020.103955.
[25] J. Mei, C. Desrosiers, and J. Frasnelli, "Machine learning for the diagnosis of Parkinson's disease: a review of literature," *Frontiers in Aging Neuroscience*, vol. 13, May 2021, doi: 10.3389/fnagi.2021.633752.
[26] B. Sahu and S. N. Mohanty, "CMBA-SVM: a clinical approach for Parkinson disease diagnosis," *International Journal of Information Technology (Singapore)*, vol. 13, no. 2, pp. 647–655, 2021, doi: 10.1007/s41870-020-00569-8.

BIOGRAPHIES OF AUTHORS

Amit Kumar a dedicated research assistant at Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India, possesses a robust academic background, having earned qualifications including a Bachelor's in Computer Applications (BCA), a Master's in Computer Applications (MCA), and pursuing Master's in Engineering (ME). His research areas are blockchain technology and AI, where he continually strives to explore innovative solutions and advancements. With a keen eye for emerging technologies and a passion for pushing the boundaries of knowledge, he is committed to making significant contributions to these fields. He welcomes collaboration and exchange of ideas within the realm of his research interests. For any inquiries or collaborations, he can be reached via email at amit.addi2010@gmail.com.

Dr. Neha Sharma working as Assistant Professor in the Department of Computer Science and Engineering in Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. She has received the M.Tech.(CSE), Ph.D. (computer science) degree in the area of computer science with vast teaching experience of more than 12 years in reputed organizations. She has more than 50 international publications in the reputed peer reviewed journals including IEEE Xplore, Scopus and SCI indexed. Her main area of research is image processing, machine learning, deep learning, cyber security. She has also published more than 30 national & international patents under the intellectual property rights of Government of India and Abroad. She is actively associated with NAAC preparations at universities interface. She is associated with many highly impact society members like IEEE (senior member), ACM, ISTE (life-time member). She can be contacted at email: nehasharma0110@gmail.com.



Shubham Mahajan is an academic and researcher in the field of engineering with a focus on Image Processing, Machine Learning, and Signal Processing. He is currently affiliated with the School of Engineering, Ajeenkya D Y Patil University, Pune, India, and also has academic ties with Amity University, Haryana, India. He can be contacted at email: mahajanshubham2232579@gmail.com.

Prof. Dr. Seifedine Kadry has a Bachelor degree in 1999 from Lebanese University, MS degree in 2002 from Reims University (France) and EPFL (Lausanne), Ph.D. in 2007 from Blaise Pascal University (France), HDR degree in 2017 from Rouen University (France). At present his research focuses on Data Science, education using technology, system prognostics, stochastic systems, and applied mathematics. He is an ABET program evaluator for computing, and ABET program evaluator for Engineering Tech. Currently, he is a Professor with Noroff University College, Kristiansand, Norway. He can be contacted at email: skadry@gmail.com.