INTERNATIONAL JOURNAL OF ADVANCED INNOVATIVE TECHNOLOGY IN ENGINEERING

Published by Global Advanced Research Publication House Journal

Home page: www.ijaite.co.in

A Robust Ensemble Learning Framework for Early Detection and Classification of Parkinson's Disease Using Voice Features

[1]Komal P. Raut, [2]Dr. Vijaya Balpande

[1,2]Department of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India

[1]komalraut67@gmail.com, [2]vijaya.balpande@pcenagpur.edu.in

Article History

Received on: 15 April 2025

Revised on: 28 April 2025

Accepted on: 02 May 2025

Keywords: Parkinson's Disease, Ensemble Learning, Biomedical Voice Features, Soft Voting, Machine Learning,

e-ISSN: 2455-6491

DOI: 10.5281/zenodo.15332020

Production and hosted by
www.garph.org

ABSTRACT

Parkinson's disease is a progressive neurodegenerative disorder that significantly impacts motor and vocal functions, making early and accurate diagnosis essential for effective treatment. This study presents an ensemble learning framework for the detection and classification of PD using biomedical voice features, offering a non-invasive and data-driven diagnostic approach. The proposed model combines multiple base classifiers—including Logistic Regression, K-Nearest Neighbors, and Decision Tree—through a soft voting mechanism to improve classification performance and reduce prediction variance. Experimental results demonstrate that the ensemble model outperforms individual classifiers, achieving an accuracy of 95%, with balanced precision, recall, and F1-score values of 91.5%. These results highlight the model's robustness and generalizability, confirming its suitability for reliable Parkinson's disease detection. The findings suggest that ensemble learning, when applied to voice-based biomedical data, can serve as an effective tool in clinical decision support systems for early PD diagnosis.

1. INTRODUCTION

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder that affects millions of individuals globally. It primarily disrupts the central nervous system, leading to motor symptoms such as tremors, rigidity, bradykinesia (slowness of movement), and postural instability. In addition to these visible symptoms, PD also manifests through non-motor complications like speech impairment, sleep disturbances, mood disorders, and cognitive decline. These complexities make early and accurate diagnosis challenging yet critical for effective management and improved patient outcomes [1].

According to the Parkinson's Foundation, over 10 million people worldwide are living with PD, and the numbers are expected to rise with an aging global population [2]. Despite the increasing prevalence, early diagnosis of Parkinson's remains difficult due to the overlap of symptoms with other neurological disorders and the absence of a single definitive diagnostic test. Traditional diagnostic procedures often rely on clinical assessments and neurological examinations, which are inherently subjective and may result in delayed or

misdiagnosis, particularly in the early stages of the disease [3].

Recent advancements in biomedical engineering and data science have introduced new opportunities for leveraging computational models to assist in the early detection of PD. In particular, machine learning (ML) methods have gained traction for their ability to learn complex patterns from clinical and biomedical data such as voice signals, handwriting patterns, gait dynamics, and neuroimaging features [4]. Voice recordings, in particular, have emerged as a promising non-invasive biomarker for PD diagnosis, as the disease often affects vocal tract control and phonation early on.

While individual machine learning algorithms such as Decision Trees, Support Vector Machines (SVM), and Random Forests have shown encouraging results in PD detection, their performance is often limited by data imbalance, overfitting, and sensitivity to feature variations. These challenges necessitate a more robust, scalable approach that can generalize well across patient populations and diverse data modalities [5].

Despite the promising outcomes of machine learning in Parkinson's disease diagnosis, several challenges persist in deploying these models in clinical practice. Firstly, no single classifier performs optimally across all datasets and scenarios. For instance, while SVM may work well with high-dimensional data, it may not handle class imbalance or feature noise as effectively as ensemble-based models [6]. Secondly, most existing studies focus on individual models, overlooking the benefits of integrating multiple algorithms to compensate for individual weaknesses.

Another limitation is the tendency of some models to overfit small datasets, especially in medical domains where acquiring labeled data is expensive and time-consuming. Moreover, many models provide high accuracy on training data but lack the generalization power needed for real-world application, leading to unreliable performance on unseen cases [7]. This gap highlights the need for a method that can not only maintain high classification accuracy but also ensure robustness, interpretability, and consistency across various patient subgroups.

In this context, ensemble learning emerges as a promising solution. Ensemble models combine predictions from multiple base learners to improve overall performance and reduce the risk of misclassification. By leveraging diverse algorithms—each with its unique strength—ensemble methods can provide a more balanced and accurate classification of Parkinson's disease, especially when dealing with complex, high-dimensional, and imbalanced biomedical datasets.

2. Related Work

Numerous researchers have explored the classification of Parkinson's disease through different techniques. Their work offers a strong basis for understanding how machine learning can be leveraged to address ongoing challenges in the subclassification, risk evaluation, and prognosis of neurodegenerative disorders, particularly using vocal signal characteristics.

Srinivasan, S. et. al (2024) proposed the Machine Learning (ML) and Deep Learning (DL) techniques, specifically K-Nearest Neighbor (KNN) and Feed-forward Neural Network (FNN) models, to differentiate between individuals with PD and healthy individuals based on voice signal characteristics. The experimental results reveals that the FNN and KSVM models, trained on an 80–20 split of the dataset for training and testing respectively, yield the most promising results. The FNN model achieves an impressive overall accuracy of 99.11%, with 98.78% recall, 99.96% precision, and a 99.23% f1-score [8].

Alshammri R et. al. (2023) proposed to detect PD using different types of Machine Learning (ML) to differentiate between healthy and PD patients by voice signal features. The models were trained using different techniques such as Synthetic Minority Over-sampling Technique (SMOTE), Feature Selection, and hyperparameter tuning (GridSearchCV) to enhance their performance. At the end, we found that MLP and SVM with a ratio of 70:30 train/test split using GridSearchCV with SMOTE gave the best results [9].

Zhang J. et. al. (2023) proposed an eight commonly used machine learning models to allow for comprehensive assessment. Finally, the Shapley Additive Explanations (SHAP) method was used to investigate the contributions of each factor. Among the eight machine learning models considered, penalized logistic regression and XGBoost were the most accurate algorithms for assessing PD risk, with penalized logistic regression achieving an area under the curve of 0.94 and a Brier score of 0.08. Olfactory function and polygenic risk scores were the most important predictors for PD risk

Saeed, Faisal et. al. (2022) proposed a comprehensive approach to enhance the prediction of PD using several machine learning methods with different feature selection methods such as filter-based and wrapper-based. The dataset includes 240 recodes with 46 acoustic features extracted from 3 voice recording replications for 80 patients. The experimental results showed improvements when wrapper-based features selection method was used with K-NN classifier with accuracy of 88.33%.

Rahman A. et al. (2021) developed a PD diagnosis system using spectral features from voice samples. Data were collected from 160 people, 60 of whom had Parkinson's disease, by pronouncing /a/, /o/,

and /u/ three times each, and MFCC features were extracted from the voice samples. Local discriminant analysis was applied for feature reduction. The data were classified using ten distinct classifiers, with the SVM-RBF classifier obtaining an accuracy of 77.50%.

Solana-Lavalle G. et al. (2021) presented an assistance tool for PD detection based on voice signals. This study employed the UCI PD classification dataset. The methodology is broken down into four parts: dataset generation, feature selection, classification, and performance analysis. The feature selection was done using a wrapper-type method that operates in a stepwise manner for both forward and backward feature selection.

Zahid, L. et al. (2020) proposed a spectrogram-based deep feature extraction method for PD diagnosis. Spanish language PC-GITA data set was used in this study. Three different methods were proposed in this study. In the first approach, speech signals were converted into a spectrogram then pre-trained CNN with ALEXNET was used for feature extraction.

Solana-Lavalle, G. et al. (2020) proposed a pre-diagnosis tool for early diagnosis of PD by using fewer vocal features to increase classification accuracy. According to the author, this is the first time eight to twenty features have been selected by wrapper method. The performance of the selected four features subset was analyzed by four KNN, MLP, SVM, and RF classifiers. As shown in the results, 94.7% accuracy was obtained by using SVMRBF Classifier.

3. MATERIAL AND METHODS

This Section presents complete workflow of the proposed model for predicting PD over the datasets based on the ensemble approach using several ML algorithms. Figure 1 shows the complete work flow of the system.

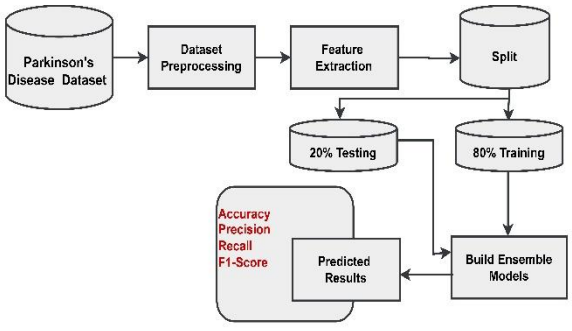

Figure 1: Workflow of the Proposed Model

A. Dataset Description

The dataset employed in this study was sourced from the University of Oxford (UO) repository in collaboration with the National Center for Voice, as developed in [15]-[16], and is publicly accessible through the UCI Machine Learning Repository [17]. The original research focused on extracting features relevant to general voice disorders. It comprised voice recordings from 31 individuals—23 diagnosed with Parkinson's Disease (16 males and 7 females) and 8 healthy control subjects (3 males and 5 females). In total, the dataset includes 195 instances and 24 attributes.

B. Data Preprocessing

Data preprocessing is a critical step in preparing raw biomedical data for effective machine learning model training, particularly in the context of Parkinson's disease detection. The dataset used in this research comprises voice-based features extracted from recordings of individuals diagnosed with Parkinson's Disease (PD) and healthy controls. To ensure data quality, consistency, and optimal model performance, the following preprocessing procedures were performed:

Data Cleaning and Validation: Initial data inspection was carried out to check for anomalies such as missing values, duplicate records, and inconsistencies in data format. Since no significant missing values were present in the dataset, further cleaning focused on removing any redundant entries and verifying that all feature values were within a plausible physiological range.

Label Encoding: The target variable, "status," was already labeled as binary—0 representing healthy individuals and 1 indicating PD patients. No additional label encoding was necessary for classification purposes.

Feature Scaling: To harmonize the numerical range of the features and prevent bias in distance-based classifiers like K-Nearest Neighbors (KNN), Min-Max Normalization was applied. This transformation scales each feature to a range of [0, 1] using the formula:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Feature Selection: To improve computational efficiency and reduce noise, feature selection was conducted using a logistic Regression algorithm, which ranks features based on their information gain. The most relevant features were retained to enhance model interpretability and performance, while low-importance features were removed.

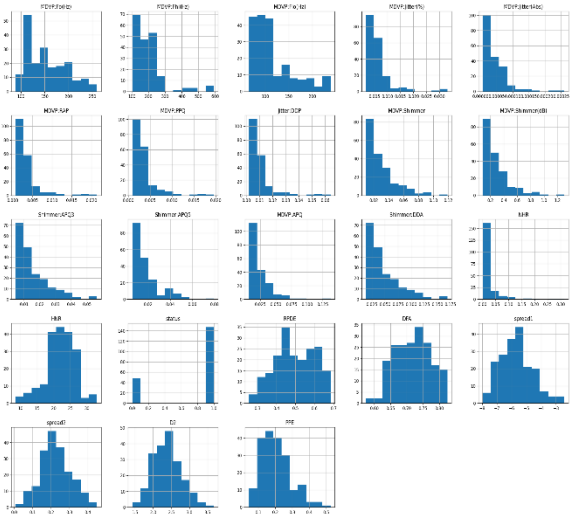

Figure 3: Pair plot of each Feature

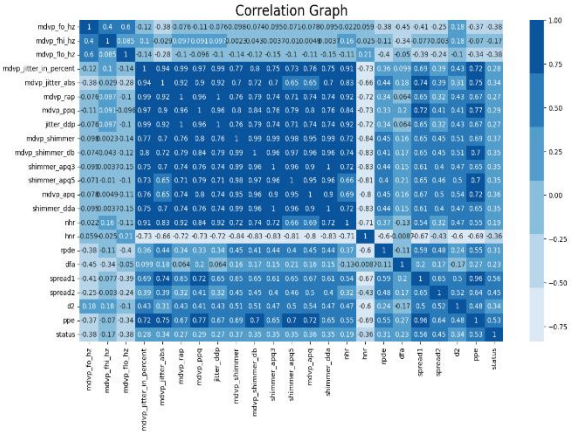

Figure 4: Heatmap of each Feature

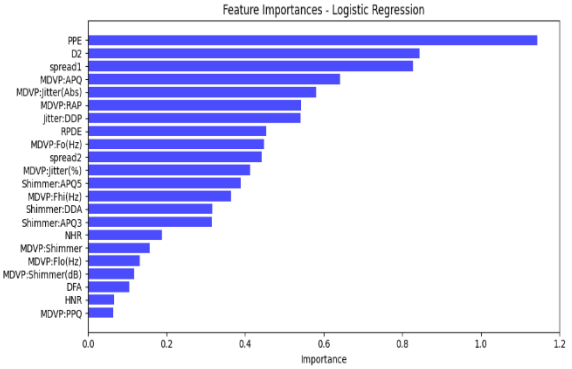

Figure 5: Feature Importance

C. Building Models

Logistic Regression (LR): Logistic Regression is a linear classification algorithm used to estimate the probability that a given input belongs to a particular class. In the case of Parkinson's disease detection, it models the relationship between voice features (e.g., jitter, shimmer, fundamental frequency) and the binary target variable (PD = 1, Healthy = 0).

K-Nearest Neighbors (KNN): KNN is a non-parametric, instance-based learning algorithm. It classifies a new sample based on the majority label among its K closest neighbors in the training dataset. In Parkinson's disease detection, KNN works by comparing a new patient's voice measurements with those of previously observed cases.

Decision Tree (DT): Decision Tree classifiers split the data into branches based on feature thresholds that maximize class separation, using measures like Gini Impurity or Information Gain. For Parkinson's detection, the model might split on key features such as MDVP:Fo(Hz) or HNR to separate healthy from PD patients.
Each internal node tests a feature, each branch represents an outcome, and each leaf node assigns a class label. Decision Trees are interpretable, capable of handling non-linear relationships, and contribute to ensemble diversity by capturing different patterns in the data.

Ensemble Model: The Ensemble model combines the predictions of multiple base classifiers—such as LR, KNN, Decision Tree, and others like Random Forest or XGBoost—through a soft voting mechanism.

RESULT ANALYSIS

The proposed ensemble learning approach was evaluated on a benchmark dataset of voice recordings from individuals diagnosed with Parkinson's disease and healthy controls. To assess the effectiveness of the model, multiple classification metrics were analyzed, including accuracy, precision, recall, F1-score, and ROC-AUC. These metrics were computed for each baseline classifier as well as the ensemble configurations to determine the relative improvements in performance.

Table 1: Performance Analysis of Proposed Models

Models	Accuracy	Precision	Recall	F1-Score
LR	90.00	85.00	77.00	80.5
KNN	91.00	87.00	80.00	85.5
DT	74.00	65.00	73.00	66.5
Ensemble	95.00	91.5	91.5	91.5

Table 1 shows the performance analysis that shows the comparative results of various models used for Parkinson's disease detection, showing that the ensemble model significantly outperforms individual classifiers in all evaluation metrics. The Logistic Regression (LR) model achieved a respectable accuracy of 90%, with precision and recall values of 85% and 77%, respectively. The K-

Nearest Neighbors (KNN) classifier slightly improved upon this, reaching 91% accuracy, but still showed limitations in recall (80%). The Decision Tree (DT) model exhibited the weakest performance with only 74% accuracy, indicating lower predictive reliability.

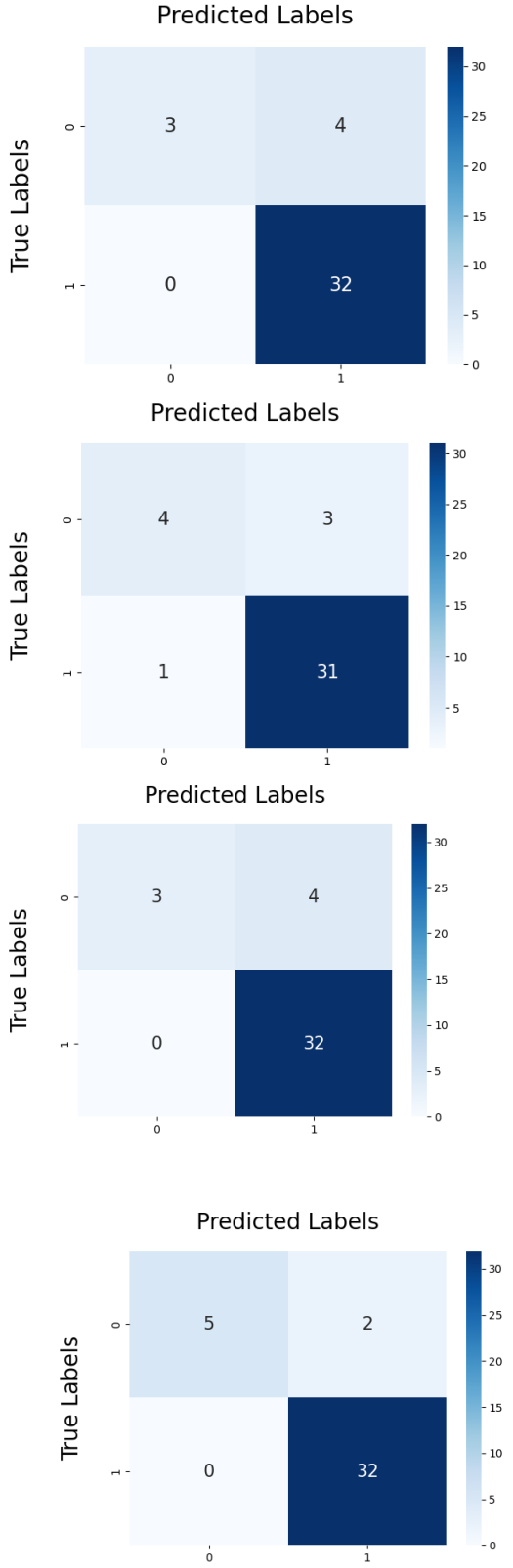

Figure 6: Confusion Matrix of LR, KNN, DT, Ensemble Models

Conclusion And Future Scope

In this study, we proposed a robust ensemble learning framework for the early detection and classification of Parkinson's disease (PD) using biomedical voice features. The ensemble approach integrated multiple base classifiers—including Logistic Regression, K-Nearest Neighbors, Decision Tree, that combined through soft voting to enhance classification accuracy, reduce variance, and ensure generalizability. The proposed ensemble model, which integrates multiple classifiers, demonstrated the best results with an impressive 95% accuracy, and balanced values of 91.5% for precision, recall, and F1-score. This underscores the ensemble model's strength in achieving both high accuracy and balanced classification performance, making it a more reliable choice for Parkinson's disease detection.

The proposed model has shown promising results, there are several avenues for further enhancement:

- Integration of multimodal data such as handwriting dynamics, gait analysis, or neuroimaging can provide a more comprehensive diagnostic tool and improve robustness.
- Expansion of the dataset with samples from diverse demographics and clinical environments would improve generalizability and reduce potential bias.
- Real-time deployment through mobile or wearable applications can allow for continuous monitoring and early symptom detection in remote or underserved areas.

Conflict Of Interest

The authors declare that they have no conflict of interest.

Funding Support

The author declare that they have no funding support for this study.

References

[1] F. N. Emamzadeh and A. Surguchov, "Parkinson's disease: Biomarkers, treatment, and risk factors," Frontiers in Neuroscience, vol. 12, no. AUG. 2018. doi:10.3389/fnins.2018.00612

[2] A. Draoui, O. El Hiba, A. Aimrane, A. El Khiat, and H. Gamrani, "Parkinson's disease: From bench to bedside," Revue Neurologique, vol. 176, no. 7–8. 2020. doi: 10.1016/j.neurol.2019.11.002

[3] F. N. Emamzadeh and A. Surguchov, "Parkinson's disease: Biomarkers, treatment, and risk factors," Frontiers in Neuroscience, vol. 12, no. AUG. 2018. doi:10.3389/fnins.2018.00612.

[4] [10] S. Przedborski, "The two-century journey of Parkinson disease research," Nature Reviews Neuroscience, vol. 18, no. 4. 2017. doi: 10.1038/nrn.2017.25

[5] R. Yuvaraj, U. Rajendra Acharya, and Y. Hagiwara, "A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals," Neural Comput. Appl., vol. 30, no. 4, 2018, doi: 10.1007/s00521-016-2756-z.

[6] C. R. Pereira, D. R. Pereira, S. A. T. Weber, C. Hook, V. H. C. de Albuquerque, and J. P. Papa, "A survey on computer-assisted Parkinson's Disease diagnosis," Artificial Intelligence in Medicine, vol. 95. 2019. doi:10.1016/j.artmed.2018.08.007

[7] Nour M, Senturk U, Polat K. Diagnosis and classification of Parkinson's disease using ensemble learning and 1D-PDCovNN. Comput Biol Med. 2023 Jul;161:107031. doi: 10.1016/j.compbiomed.2023.107031. Epub 2023 May 17. PMID: 37211002.

[8] Bukhari, S. N. H., & Ogudo, K. A. (2024). Ensemble Machine Learning Approach for Parkinson's Disease Detection Using Speech Signals. Mathematics, 12(10), 1575. https://doi.org/10.3390/math12101575

[9] S. Srinivasan, P. Ramadass, S. Mathivanan, et al., "Detection of Parkinson disease using multiclass machine learning approach," Science Report, vol. 14, p. 13813, 2024, doi: 10.1038/s41598-024-64004-9.

[10] R. Alshammri, G. Alharbi, E. Alharbi, and I. Almubark, "Machine learning approaches to identify Parkinson's disease using voice signal features," Front. Artif. Intell., vol. 6, p. 1084001, Mar. 2023, doi: 10.3389/frai.2023.1084001.

[11] J. Zhang, W. Zhou, H. Yu, T. Wang, X. Wang, L. Liu, and Y. Wen, "Prediction of Parkinson's Disease Using Machine Learning Methods," Biomolecules, vol. 13, no. 12, p. 1761, Dec. 2023, doi: 10.3390/biom13121761.

[12] F. Saeed, M. Al-Sarem, M. Al-Mohaimeed, A. Emara, W. Boulila, M. Alasli, and D. Ghabban, "Enhancing Parkinson's Disease Prediction Using Machine Learning and Feature Selection Methods," Computers, Materials & Continua, vol. 71, pp. 5639-5658, 2022, doi: 10.32604/cmc.2022.023124.

[13] A. Rahman, S. S. Rizvi, A. Khan, A. A. Abbasi, S. U. Khan, and T. S. Chung, "Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier," Mob. Inf. Syst., vol. 2021, 2021, doi: 10.1155/2021/8822069.

[14] G. Solana-Lavalle and R. Rosas-Romero, "Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation," Biomed. Signal Process. Control, vol. 66, 2021, doi: 10.1016/j.bspc.2021.102415.

[15] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D., and Moroz, I. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. BioMed. Eng. 6, 23. doi: 10.1186/1475-925X-6-23

[16] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., and Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Tran. Biomed. Eng. 56, 1015-1022. doi: 10.1109/TBME.2008.2005954

[17] Little, M. (2008). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. Available online at: https://archive.ics.uci.edu/ml/datasets/parkinsons (accessed March 17, 2023).