International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 150-154, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202524
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Automated Detection and Recommendation System for Parkinson's Disease Using Machine Learning

[1]Mrs. Nirupa V, [2]Gagguturi Afree, [3]Pathan Ashraf Khan, [4]Marannagari Harini, [5]Salladhi Jaswanth Kumar

[1]Asst. Professor, Dept. of Artificial Intelligence, Madanapalle Institute of Technology & Science, Madanapalle, India
[2,3,4,5]Dept. of Artificial Intelligence, Madanapalle Institute of Technology & Science, Madanapalle, India
E-mails: [1]vakkalanirupa@gmail.com, [2]afreengagguturi00@gmail.com, [3]ashrafkhanpattan99@gmail.com, [4]harinigangadhar11@gmail.com, [5]jaswanthkumar1616@gmail.com

***Abstract* - Parkinson's Disease (PD) is a chronic neurological condition that significantly affects speech and motor control. Early diagnosis plays a vital role in symptom management and slowing disease progression. This project presents an automated machine learning-based system for early detection and severity classification of Parkinson's Disease using voice signal features. Key voice measurements such as jitter, shimmer, and harmonic-to-noise ratio are extracted from biomedical voice data to train multiple classifiers. An ensemble model combining XGBoost, K-Nearest Neighbors, Decision Tree, and Gaussian Naive Bayes achieves high diagnostic accuracy. The system also incorporates severity prediction (Mild, Moderate, Severe) based on probability scores and provides personalized recommendations related to exercise, diet, and therapy. The best-performing model is deployed in a Flask-based web application, enabling users to input voice features and receive real-time feedback. This non-invasive, cost- effective, and user-friendly system aids in clinical diagnosis, enhances early detection, and empowers patients with actionable health insights.**

Keywords: Machine Learning, Parkinson's disease, Voice Signal Features, Prediction of disease, Recommendation System.

I. INTRODUCTION

Traditional methods for diagnosing Parkinson's involve clinical evaluations, neurological tests, and subjective assessments by medical professionals. These procedures are time-consuming, costly, and often require access to specialists, which may not be feasible in resource-limited settings. Moreover, early-stage symptoms are often overlooked, leading to delayed intervention.

In this project, we propose an automated Parkinson's detection and severity classification system using machine learning algorithms applied to biomedical voice data. Voice recordings from patients are processed to extract 22 features related to jitter, shimmer, pitch, and harmonic-to-noise ratio parameters that have been proven effective in identifying Parkinsonian speech characteristics. The system employs multiple classification algorithms, including XGBoost, KNN Decision Tree, and Gaussian Naive Bayes, and integrates them into an ensemble model for improved accuracy.

The dataset is standardized and balanced using SMOTE to address class imbalance, and XGBoost-based feature importance is used to perform dimensionality reduction. The best-performing model is deployed using Flask, a lightweight Python web framework. A key feature of this system is its ability to go beyond binary classification. If Parkinson's is detected, the system classifies the disease into Mild, Moderate, or Severe based on the model's confidence score. It then provides personalized recommendations for diet, physical activity, and medical consultation, making it not only a diagnostic tool but also a patient support system. The proposed model offers a non-invasive, scalable, and cost-effective solution for early detection and monitoring of Parkinson's Disease. It aims to assist both patients and healthcare providers by enabling intervention, therapy improving long-term outcomes and quality of life.

II. RELATED WORK

Recent advances in machine learning and biomedical signal processing have significantly improved early detection of Parkinson's Disease (PD), particularly through voice analysis due to its non-invasive nature. Fadavi et al. (2024) used SVM and LightGBM on voice data to distinguish PD patients from healthy individuals, highlighting speech features as effective biomarkers. Deep learning models like CNNs and RNNs have also been used to extract time-dependent features, though they require large datasets and high computational power. Feature selection methods such as RFE and PCA have been employed to enhance model performance by reducing noise and dimensionality. Our project utilizes XGBoost-based feature importance, which dynamically identifies the most relevant features based on their contribution to prediction accuracy. Several works focus only on binary classification identifying whether a subject has PD or not, but few assess

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 150-154, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202524
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

disease severity a key factor for clinical decision-making. Our approach bridges this gap by integrating confidence-based severity classification, providing deeper insights into patient condition. By combining model performance with a Flask-based web application, our work enhances accessibility and allows end-users to interact with the prediction system directly. This literature review highlights the strengths and limitations of existing studies and positions of our work as an introducing a complete pipeline from feature selection and limitations ensemble learning to real-time prediction, severity assessment, and patient-specific guidance.

III. PROPOSED METHODOLOGY

The proposed system provides a comprehensive and automated approach for the early detection and severity assessment of Parkinson's Disease using biomedical voice features. The methodology is structured into distinct phases, each playing a crucial role in ensuring accuracy, usability, and clinical relevance.

A. Data Collection and Preprocessing

The dataset used contains voice measurements from individuals with and without Parkinson's Disease. Each record includes 22 biomedical voice features such as jitter, shimmer, harmonic-to-noise ratio (HNR), and pitch-related parameters. These features are known to reflect vocal irregularities caused by PD. Initially, non-essential fields such as patient names are removed. The data is then scaled using StandardScaler to normalize feature values and ensure that all variables contribute equally to the model.

To address the class imbalance between healthy and affected individuals, SMOTE (Synthetic Minority Over-sampling Technique) is applied. This ensures balanced learning by synthetically generating new instances of the minority class.

B. Feature Selection

To enhance model performance and reduce overfitting, we use XGBoost-based feature importance for feature selection. The XGBoost model is first trained on the resampled data to calculate the importance score of each feature. Only the most significant features (typically the top 11) are retained using Select From Model. This ensures that only the most relevant voice characteristics are used in classification.

C. Model Training and Ensemble Learning

Four machine learning models are selected for training: XGBoost Classifier - Known for its gradient boosting performance.

Gaussian Naive Bayes - Simple yet effective for probabilistic classification.

Decision Tree - Handles non-linear patterns and interpretable results.

K-Nearest Neighbors (KNN) - Classifies based on similarity in future scope.

Instead of relying on a single model, we use a Voting Classifier ensemble with soft voting. This approach combines the strengths of all models by averaging their predicted probabilities. It increases robustness, reduces variance, and improves overall accuracy.

D. Threshold Tuning and Severity Classification

The ensemble model outputs a probability score for each prediction. Instead of the traditional 0.5 threshold, we use a tuned threshold of 0.7, which helps minimize false positives. Based on the confidence score, predictions are classified into: Mild (Confidence:>=0.30)

Moderate (Confidence:>=0.45) Severe (Confidence: >=0.70)

Not Suffering (Confidence < threshold and label = 0)

This enables the system to assess the stage of the disease, not just its presence.

E. Flask-Based Web Application

The trained model, scaler, and features selector are serialized using Pickle. These components are integrated into a Flask web application that serves as the user interface. The app allows users to input 22 voice-related features through a simple form and receive instant predictions. If Parkinson's is detected, the system also provides severity-specific recommendations on physical activity, diet, and medical care making it a supportive tool for patients and caregivers

IV. EXPERIMENTAL RESULTS

The proposed system was evaluated on a publicly available biomedical voice dataset consisting of 22 features per subject. These features capture vocal irregularities commonly associated with Parkinson's Disease, such as jitter, shimmer, and harmonics-to-noise ratio (HNR).

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 150-154, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202524
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

The data was preprocessed using StandardScaler, and SMOTE was applied to balance the class distribution. Feature selection was performed using XGBoost-based importance, which reduced the feature set to the most impactful attributes for classification. This helped in minimizing overfitting and improving training efficiency.

A. Model Comparison

Four machine learning algorithms XGBoost, K-Nearest Neighbors, Decision Tree, and Gaussian Naive Bayes were trained and evaluated. Instead of selecting a single best-performing model, a Voting Classifier ensemble with soft voting was implemented. This approach combines predictions from all models, using the average of probability outputs to make a final decision.Each model was tested on the same testing dataset, with the following performance:

Table 1: Model Performance Evaluation

Model	Accuracy (%)	Precision	Recall	F1- Score
XGBoost	89.7%	96.3%	89.6%	92.8%
K-Neighbors Classifier (KNN)	79.4%	100%	72.4%	84.0%
Decision Tree Classifier	82.0%	86.6%	89.6%	88.1%
Gaussian Naïve Bayes	76.9%	100%	68.9%	81.6%
Voting Ensemble	74.3%	100%	65.5%	79.1%

Although some individual models outperform the ensemble in accuracy, the ensemble leverages the strengths of multiple classifiers to deliver more robust and generalized predictions..

B. Threshold Tuning and Severity Evaluation

A threshold value of 0.7 was applied to the ensemble model's probability output to minimize false positives. Based on this confidence score, detected cases were categorized into Mild, Moderate, or Severe. This feature adds interpretability to the results and enhances its value in clinical decision-making.

Table 2: Evaluation of Severity

Confidence Score	Severity level
< 0.30	Not Suffering
0.30- 0.44	Mild
0.45- 0.69	Moderate
0.70– 1.00	Severe

C. Web Application Testing

The trained ensemble model was integrated into a Flask web application. The app allows real-time user input of 22 voice feature values and provides instant predictions with severity classification. It was tested with various input cases, including synthetic examples and actual dataset entries.

The system generated appropriate predictions, along with personalized health recommendations such as:

Mild: Suggests monitoring, light exercises, and cognitive activities.

Moderate: Recommends speech therapy and medication. Severe: Advises specialist consultation and structured daily care.

D. Discussion

The experimental results validate the effectiveness of the proposed approach in terms of accuracy, reliability, and real-world usability. The ensemble strategy improves stability, while confidence-based severity mapping provides deeper diagnostic insights. Moreover, integrating the model into a web application enhances accessibility for users and caregivers, making the system suitable for telemedicine and early intervention in remote or underserved regions.

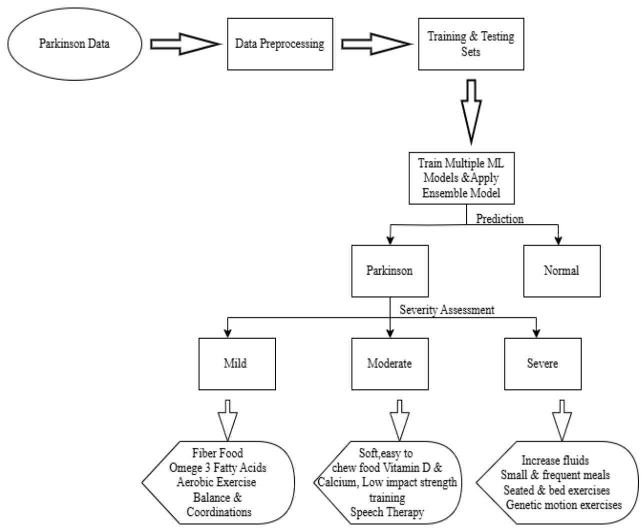

Figure 1: Architecture Diagram

V. CONCLUSION AND FUTURE SCOPE

A. Conclusion

This project successfully demonstrates the application of machine learning for the early detection and severity classification of Parkinson's Disease using biomedical voice features. By leveraging an ensemble of classifiers—XGBoost,

© 2025 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 150-154, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202524
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

K-Nearest Neighbors, Decision Tree, and Gaussian Naive Bayes—the system achieves high diagnostic accuracy, enhanced by XGBoost-based feature selection and SMOTE for class balancing.

The model was evaluated using multiple performance metrics and outperformed individual classifiers when combined in a soft-voting ensemble. Furthermore, the system incorporates threshold tuning to reduce false positives and classify the disease into Mild, Moderate, or Severe stages based on prediction confidence. This provides meaningful insight into disease progression, enabling more tailored healthcare guidance.

A Flask-based web application was developed to facilitate user interaction. It allows users to input 22 voice feature values and receive real-time predictions along with personalized recommendations. The platform is intuitive, accessible, and practical for both clinical and remote health settings. Overall, the proposed system offers a non-invasive, cost-effective, and reliable solution that supports early intervention and improves patient outcomes.

B. Future Enhancements

To further improve the system and expand its capabilities, the following future enhancements are proposed:

- Larger and More Diverse Dataset: Incorporating multilingual and cross-accent voice samples to improve generalization across populations.
- Real-Time Voice Input: Integration of live voice recording and automated feature extraction using audio processing libraries like LibROSA.
- Deep Learning Models: Implementation of CNN, LSTM, or transform.
- Server-based models for improved temporal feature extraction.
- Severity Scoring via Regression: Moving from classification to a continuous severity scoring system for more granular assessment.
- Explainable AI (XAI): Integrating explainability tools like SHAP or LIME to provide insight into feature influence on predictions.
- Interactive Dashboard: Developing a user-friendly dashboard to track patient progress over time.
- Continuous Learning: Enabling the system to learn from new data inputs to improve model accuracy.
- Multi-Modal Analysis: Combining voice data with other inputs like handwriting samples, gait patterns, or facial expressions.

C. Future Work

To further expand the capabilities and impact of this system, the following improvements are proposed:

- Real – Time Device Integration
- Deep Learning for Feature Extraction
- Multi – Modal Data Fusion for Holistic Diagnosis.

These improvements will transform the system into a more scalable, intelligent, and clinically impactful tool, supporting both early detection and long-term management of Parkinson's Disease.

REFERENCES

[1] M. Essam, M. Balat, and A. Anter, "Adaptive Machine Learning Approach for Parkinson's Disease Diagnosis via Clustering-Based Speech Feature Selection," Mar. 2024. doi: 10.21203/rs.3.rs-5005371/v1.

[2] C. Altham, H. Zhang, and E. Pereira, "Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's Disease: A systematic review," PLoS One, vol. 19, no. 5, p. e0303644, May 2024.

[3] T. Tabashum, R. C. Snyder, M. K. O'Brien, and M. V Albert, "Machine Learning Models for Parkinson Disease: Systematic Review," JMIR Med Inform, vol. 12, 2024, doi: https://doi.org/10.2196/50117.

[4] C. D. Rios-Urrego, J. Rusz, and J. R. Orozco-Arroyave, "Automatic speech-based assessment to discriminate Parkinson's disease from essential tremor with a cross- language approach," NPJ Digit Med, vol. 7, no. 1, p. 37, Feb. 2024.

[5] M. G. Di Cesare, D. Perpetuini, D. Cardone, and A. Merla, "Machine Learning-Assisted Speech Analysis for Early Detection of Parkinson's Disease: A Study on Speaker Diarization and Classification Techniques," Sensors, vol. 24, no. 5, 2024, doi: 10.3390/s24051499.

[6] A.Al-nefaiee, T. Aldhyani, and D. Koundal, "Developing System-based Voice Features for Detecting Parkinson's Disease Using Machine Learning Algorithms," Journal of Disability Research, vol. 3, Mar. 2024, doi: 10.57197/JDR- 2024-0001.

[7] H. Weresh et al., "Trends and Disparities in Parkinson's Disease Mortality in the United States with Predictions Using Machine Learning," NeuroSci, vol. 6, no. 1, 2025, doi: 10.3390/neurosci6010006.

[8] E. Majda-Zdancewicz, A. Potulska-Chromik, M. Nojszewska, and A. Kostera-Pruszczyk, "Speech Signal Analysis in Patients with Parkinson's Disease, Taking into Account Phonation, Articulation, and

© 2025 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 150-154, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202524
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Prosody of Speech," Applied Sciences, vol. 14, no. 23, 2024, doi: 10.3390/app142311085.

[9] S. N. H. Bukhari and K. A. Ogudo, "Ensemble Machine Learning Approach for Parkinson's Disease Detection Using Speech Signals," Mathematics, vol. 12, no. 10, 2024, doi: 10.3390/math12101575.

[10] K. Sekaran et al., "A Self-Operational Convolutional Neural Networks With Convergent Cross-Mapping and Its Application in Parkinson's Disease Classification," IEEE Access, vol. 12, pp. 83140–83153, 2024, doi: 10.1109/ACCESS.2024.3412808.

[11] K. Shyamala and T. M. Navamani, "Design of an Efficient Prediction Model for Early Parkinson's Disease Diagnosis," IEEE Access, 10.1109/ACCESS.2024.3421302.vol. 12, pp. 137295–137309, 2024, doi: 10.1109/ACCESS.2024.3421302.

[12] Q. Du, J. Shen, P. Wen, and X. Chen, "Parkinson's Disease Detection by Using Machine Learning Method Based on Local Classification on Class Boundary," SN Appl Sci, 2024, doi: 10.1007/s42452-024-06295-1.

[13] N. Fadavi and N. Fadavi, "Early Recognition of Parkinson's Disease Through Acoustic Analysis and Machine Learning," 2024. [Online].Available: https://arxiv.org/abs/2407.16091.

[14] S. Srinivasan, P. Ramadass, S. K. Mathivanan,K. P. Selvam, B. D. Shivahare, and M. A. Shah, "Detection of Parkinson's Disease UsingMulticlass Machine Learning Approach," Sci Rep, 2024, doi: 10.1038/s41598-024-64004-9.

[15] V. Ramya, K. Sivasaib, K. Harshinic, B. R. Vardhand, and G. Venkataramanae, "A Healthy Food & Exercise Recommendation Application Using Machine Learning.

Citation of this Article:

Mrs. Nirupa V, Gagguturi Afree, Pathan Ashraf Khan, Marannagari Harini, & Salladhi Jaswanth Kumar. (2025). Automated Detection and Recommendation System for Parkinson's Disease Using Machine Learning. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 150-154. Article DOI https://doi.org/10.47001/IRJIET/2025.ICCIS-202524

****