iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80891

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Intelligent Framework for Early Parkinson's disease Detection

Ganta Bhanu Chandra[1], Dr. Dasari Haritha[2]
[1]*M. Tech,* [2]*Professor, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India*

Abstract: ***Early identification of Parkinson's Disease (PD) helps doctors provide better care and treatment. Analyzing vocal patterns serves as a straightforward, non-invasive method to recognize the initial symptoms of the condition. Despite this, many existing detection systems struggle with two persistent issues unequal distribution of class samples and high-dimensional feature spaces. This paper introduces an enhanced machine learning framework that applies Borderline-SMOTE to address data imbalance by synthesizing samples within challenging classification zones. A feature reduction step is incorporated to minimize redundancy, and a range of classifiers is evaluated against each other. Decision Tree classifier demonstrates superior performance on processed dataset. The model is tested using accuracy, precision, recall, and F1-score. The method achieved 98.67% accuracy, 98.69% precision, 98.66% recall, and 98.67% F1-score. These outcomes stem from improved handling of skewed data distributions and enhanced pattern recognition. The result is a transparent and dependable solution for speech-based early PD identification.***

Keywords: Parkinson's Disease, Machine Learning, Borderline-SMOTE, recursive feature elimination, Decision Tree, Explainable AI.

I. INTRODUCTION

The fast development of data-driven technologies has changed modern healthcare systems. Intelligent tools powered by machine learning now play a growing role in identifying diseases at their earliest stages and supporting ongoing patient monitoring. Among the many neurological conditions that affect people worldwide, Parkinson's Disease (PD) stands out as particularly challenging due to its progressive impact on both motor functions and vocal abilities. Timely identification of this condition is critical, as it allows medical professionals to begin appropriate treatment before the disease advances significantly and helps patients maintain a better standard of living. Despite this need, the current clinical approach relies heavily on specialist assessments and costly diagnostic procedures, which creates barriers to early and accessible detection. Vocal analysis has gained considerable attention as a practical alternative for identifying PD in its early stages. This is largely because changes in voice quality tend to appear before many other physical symptoms become noticeable. Collecting speech samples is simple, painless, and requires minimal resources, making it well-suited for routine health screening. However, applying machine learning to this type of data comes with notable difficulties. A major obstacle is the imbalanced nature of available datasets, where the number of healthy cases is much larger than PD cases, leading to unfair model results. Additionally, large volume of voice-related features often introduces redundancy and increases computational complexity, making it harder for models to identify truly meaningful signals. Conventional machine learning approaches have been widely explored for PD classification, yet they frequently fall short when dealing with skewed data distributions and intricate feature relationships. While these models may report strong overall accuracy figures, they tend to underperform on the minority class, which represents PD-positive cases, a critical shortcoming in any medical application. Furthermore, many of these systems operate as black boxes, offering little transparency into the reasoning behind each prediction, which limits their acceptability in clinical environments where accountability matters. To fix these drawbacks, this research presents a more effective machine learning pipeline that prioritizes both data quality and model clarity. The method employs Borderline-SMOTE to correct class imbalance by generating synthetic samples in the most informative and ambiguous regions of the feature space. This strategy directs the model's attention toward the cases where classification is most challenging. Alongside this, feature selection through Recursive Feature Elimination reduces the input space to attributes which meaningfully help make the forecast. Multiple classifiers are then trained and evaluated, with the Decision Tree emerging as the top-performing model due to its capacity to handle complex patterns while remaining straightforward to interpret. Beyond achieving strong classification metrics, this study emphasizes the importance of consistent and trustworthy model behavior across varying data conditions. Proper management of class imbalance combined with targeted feature selection results in a system that delivers stable and well-rounded predictions. The findings confirm that this approach is both efficient and interpretable, positioning it as a practical solution for early Parkinson's Disease screening through vocal data analysis.

II. RELATED WORK

Many studies have focused on using speech and medical data for Parkinson's detection. In the early stages of this field, investigations focused primarily on traditional classification techniques paired with handcrafted acoustic descriptors. K. Shyamala and T. M. Navamani [1] proposed a classification pipeline combining feature selection with KMeans-SMOTE to correct class imbalance in PD speech data. The balanced feature subset fed into the classifier yielded significantly improved sensitivity toward PD-positive cases compared to unbalanced baselines. K. Shyamala and T. M. Navamani [2] extended their work to PD severity staging using a deep learning model that incorporates dimensionality reduction during preprocessing. The architecture reliably partitions disease progression into distinct stages, though limited interpretability restricts its direct clinical use. Q. Dao et al. [3] applied pre-trained speech foundation models to detect early-stage PD by fine-tuning transformer representations on disease-specific vocal recordings. The transfer-based approach captured subtle articulatory irregularities with strong detection performance, albeit at higher computational cost. M. Junaid et al. [4] introduced a multitask deep learning framework that jointly predicts PD prediction of progression and depression based on multimodal time-series information. Shared feature representations across tasks improved both predictions, though the model's complexity limits deployment in resource-constrained settings. M. Rey-Paredes et al. [5] used GANs to synthesize realistic PD voice waveforms for training data augmentation, improving minority-class classification. The approach demands substantial compute and may not generalize uniformly across patient demographics. M. Khan et al. [6] prototyped an FPGA-based PD detection system that executes machine learning inference on voice signals at the hardware level for real-time, low-power screening. The system matched software-level accuracy while offering significant speed advantages, though hardware updates after deployment are inflexible. M. Ullrich et al. [7] collected real-world inertial sensor gait data from wearable devices to predict fall risk in PD patients under naturalistic conditions. The study confirmed that accelerometer and gyroscope readings carry strong predictive value, though variability in real-world sensor conditions affects model consistency. E. Kumari et al. [8] proposed NeuroAid, an EEG-based system that analyzes emotion-linked brain activity patterns for PD identification. Multi-channel frequency features were extracted and classified, opening a complementary neurological diagnostic pathway, despite the practical barrier of costly EEG recording equipment. A. Rani Palakayala et al. [9] developed HAMF, a hierarchical attention-based multi-modal fusion model that dynamically weights each input modality at multiple feature levels for PD classification and severity prediction. The attention mechanism improves over simple concatenation, though simultaneous availability of all modalities is not always feasible clinically. A. Rezvani et al. [10] presented DiffuseGaitNet, which applies diffusion model-based generative learning to gait severity assessment in PD. The framework augments movement representations for robust severity prediction, though training diffusion models requires large annotated gait corpora and substantial computing resources. G. Amprimo et al. [11] performed a data-driven analysis of how deep brain stimulation alters gait parameters in PD patients using pre- and post-stimulation wearable sensor recordings. A predictive model was built to anticipate individual patient responses to DBS therapy, though the specialized cohort size limits broader generalizability. R. M. Al-Tam et al. [12] developed a stacking ensemble that aggregates multiple base classifiers through a meta-learner to strengthen PD diagnosis reliability. Although ensemble diversity reduced individual model bias, no specific mechanism targeted the class imbalance issue, leaving boundary-region detection insufficiently addressed. C. Dong et al. [13] Presented a combined static dynamic temporal network framework that fuses stable structural patterns with time-varying functional signal-based detection of Parkinson's disease and severity estimation. The dual-stream design generalized reasonably across datasets but increases architectural complexity and training cost. S. M. Abdullah et al. [14] applied deep transfer learning with optimized feature selection to improve PD detection across multiple datasets with limited labeled data. The strategy achieved better generalization than training from scratch, though the inherited opacity of deep representations restricts clinical interpretability. S. Gaba and H. Kaur [15] examined clinical voice recordings to validate jitter, shimmer, and pitch variability as acoustic biomarkers for PD diagnosis using classical classifiers. The study confirmed the relevance of these features but noted feature redundancy and class imbalance as unresolved obstacles. J. Jamuna and K. Kasturi [16] compared SVM, Random Forest, and KNN on acoustic PD speech features, finding that tree-based and kernel methods outperform distance-based approaches. However, the absence of any class-balancing step inflated overall accuracy while concealing poor recall on PD-positive samples. D. Kumar B. and K. France [17] benchmarked conventional machine learning techniques for PD prediction on clinical and voice feature sets under consistent experimental conditions. The study provides a useful interpretable baseline for evaluating performance gains offered by more advanced methods. A. Selvi S. and T. Kamalakannan [18] assessed filter and wrapper feature evaluation strategies alongside multiple classifiers for PD prediction from vocal data. The empirical comparison confirmed that preprocessing quality directly influences classification reliability and offers practical design guidance for detection systems. A. S. et al. [19] explored K-Nearest Neighbors for early PD prediction, examining the effect of distance metric and neighborhood size on imbalanced voice datasets.

While KNN achieved competitive accuracy with proper normalization, it suffers from scalability and dimensionality issues in larger clinical corpora. A. Akilandeswari [20] systematically compared filter, wrapper, and embedded feature selection methods on PD voice datasets, demonstrating that recursive and embedded strategies yield the most consistent accuracy gains. The study established feature selection as an essential preprocessing step, though class imbalance was not addressed.

III. PROPOSED METHOD

This study focuses on developing a system to identify Parkinson's Disease (PD) at an early stage using voice based data. Speech data is suitable for this purpose because vocal changes appear early in the disease and are easy to collect without any discomfort to the patient. The proposed pipeline is built in seven stages data loading, preprocessing, class balancing, feature selection, model training, classifier selection, and explainability. The complete workflow is shown in Figure 1. Where the dataset first goes through preprocessing and class balancing. Features are then selected using RFE, and the data is split for training and testing. Eight classifiers are compared, and the Decision Tree is identified as the best-performing model. SHAP analysis is finally applied to explain the model's predictions.

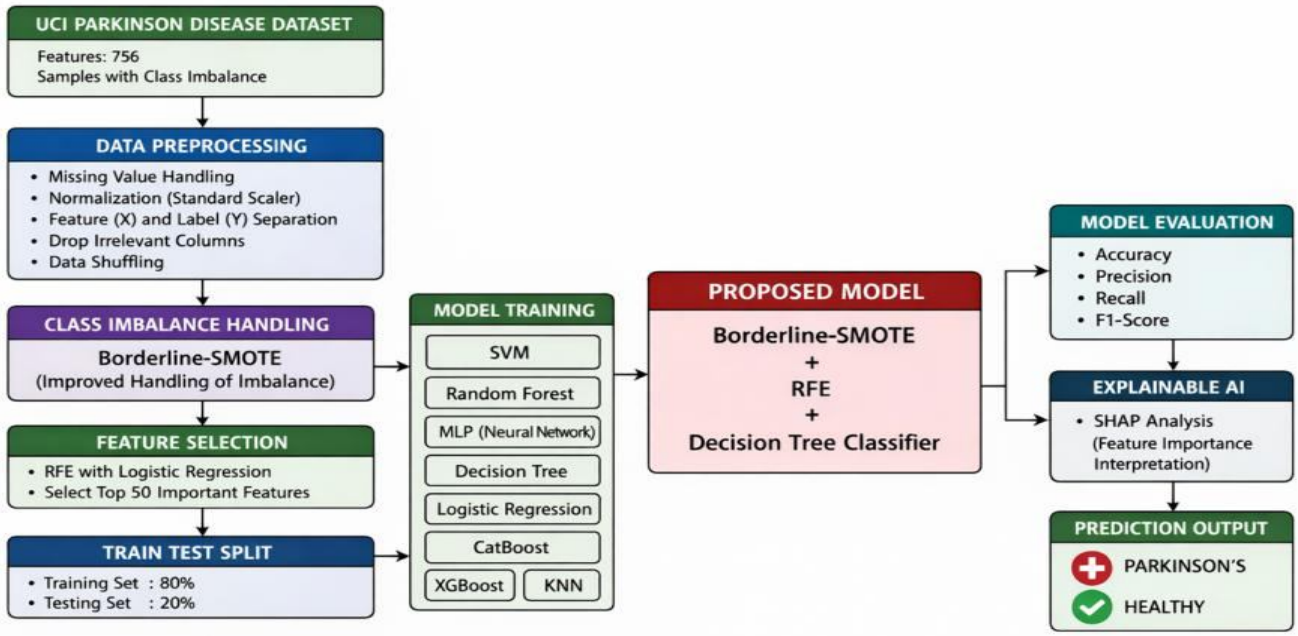


Figure 1. Working of Proposed System

A. *Data Preprocessing*

Before any learning can occur, the raw data must be cleaned and standardized. Any missing or corrupted entries in the dataset are replaced with the column mean so that no sample is entirely discarded due to incomplete information. Irrelevant columns, specifically the patient identifier and the class label, are removed before feature extraction, and the remaining attributes are treated as the input space while the class column is retained separately as the output. To eliminate the effect of differing measurement scales across features, all numerical values are transformed using standard normalization, which shifts each feature is adjusted to have a mean of zero and unit variance. The samples are then shuffled randomly to remove any ordering effects that might bias the training process. This sequence of operations produces a clean, uniformly scaled dataset that is ready for further processing.

B. *Class Imbalance Correction using Borderline-SMOTE*

Medical datasets frequently suffer from an unequal distribution of class samples, and the PD speech dataset is no exception. When a classifier is trained on data where one class dominates, it tends to favor the majority group and systematically miss cases belonging to the minority, which in this context means failing to detect actual PD patients. To counter this, the proposed framework applies Borderline-SMOTE, a targeted variant of the conventional synthetic oversampling technique. Unlike the basic approach, which generates new artificial samples at random positions within the minority region, Borderline-SMOTE concentrates its synthesis activity on samples located near the decision boundary the zone where the two classes are most difficult to distinguish.

By populating this critical region with additional training examples, the classifier receives stronger exposure to challenging and ambiguous cases, which directly improves its sensitivity toward the PD-positive class without introducing noise in well-separated areas of the feature space.

C. *Feature Selection using Recursive Feature Elimination (RFE)*

Even after balancing the class distribution, the dataset retains a very large number of input features, many of which carry little or no predictive signal. Retaining all of them increases training time and raises the likelihood of overfitting, model is just memorizing the answers instead of understanding the logic. To address this, Recursive Feature Elimination is applied as a structured dimensionality reduction technique. The method begins by training a Logistic Regression estimator on the full feature set and scoring the contribution of each attribute. In each subsequent cycle, the least informative features are dropped, and the process repeats until only the top 50 features remain. The importance of each retained feature is visualized through a ranking plot, giving a clear picture of which vocal characteristics contribute most to the classification task. This reduced feature set serves as the input to all subsequent model training stages.

D. *Train/Test Split*

After selecting the important Features, the data is divided into two groups one for learning and one for checking. We use an 80/20 approach, dedicating 80% of the information to build the model while saving the other 20% to see how it handles fresh challenges. By testing the model on unseen data, we get a much more accurate picture of how it will actually perform in the real world. The training set contains about 903 samples, while the test set has around 226 samples. Both sets maintain a balanced distribution of classes after applying Borderline-SMOTE in the earlier step.

E. *Multi-Classifier Comparison*

To identify the most suitable classification algorithm for this problem, eight well-established machine learning models are trained and assessed under identical experimental conditions using the same preprocessed dataset. The models included in this comparison are Support Vector Machine, Random Forest, K-Nearest Neighbors, Multilayer Perceptron, XGBoost, CatBoost, Logistic Regression, and Decision Tree. Each algorithm is configured with tuned hyperparameters, also evaluated with use of four typical performance measures accuracy, precision, recall, and F1-score. The purpose of this evaluation is not merely to rank the classifiers but to understand how different algorithmic approaches respond to the specific data characteristics of this problem, including the handled imbalance and the reduced feature space produced by RFE.

F. *Proposed Classifier*

Among all the classifiers evaluated, the Decision Tree demonstrated superiority across the entire suite of evaluation criteria, making it the proposed model in this framework. The classifier operates by recursively splitting the input space into smaller regions based on feature thresholds, using the Gini impurity criterion to determine the most informative split at each node. This hierarchical partitioning continues until the tree reaches its configured maximum depth of 20 levels or until the leaves are sufficiently pure. The final output for any given input is determined by tracing a path from the root node through a series of binary decisions down to a leaf, where the predicted class label is assigned. A key advantage of this structure is its transparency every decision the model makes can be traced back to a specific feature value and threshold, which is especially useful in medical fields, where understanding the reason behind a prediction is as important as the result.

G. *Model Explainability using SHAP*

To complement the predictive capability of the Decision Tree with a deeper layer of interpretability, SHAP analysis is incorporated into the evaluation process. SHAP assigns a numerical contribution value to each input feature for every individual prediction made by the model, quantifying how much each attribute shifts the output toward or away from a PD diagnosis. A positive SHAP value for a feature indicates that its presence in the input pushed the model toward predicting Parkinson's Disease, while a negative value suggests the opposite. A summary plot is generated to display the distribution of these contribution values across the entire test set, immediately revealing which vocal characteristics exert the greatest influence on the model's decisions. This level of detail allows clinicians and researchers to connect model behavior directly to known biomarkers of the disease, fostering trust and supporting practical adoption of the system in a clinical setting.

IV. RESULTS AND DISCUSSION

This study uses the PD Speech Features dataset, that consists of voice features from both healthy individuals and patients confirmed to have Parkinson's Disease. The dataset holds 756 samples, each described by 753 acoustic measurements and one class label. A label of 0 denotes a healthy subject, while a label of 1 identifies a PD-positive case. Examining the distribution of these labels reveals a clear imbalance healthy samples are far more numerous than disease samples. This disproportion is a known obstacle in medical data analysis because models trained on such data tend to favor the majority group and overlook the minority, which in this case represents the very patients the system is meant to detect. The acoustic features span a wide range of voice properties, including pitch stability, frequency jitter, amplitude shimmer, and harmonic-to-noise ratios, all of these are clinically documented to deteriorate in PD patients before other physical symptoms become apparent. By applying the proposed methodology step by step, including preprocessing, data balancing, feature selection, and model training, following reliable results were obtained. This structured approach helped improve the model's ability to learn meaningful patterns from the data and produce accurate predictions.

A. Evaluation of Different Machine Learning Models

The effectiveness of each classifier was measured using accuracy, precision, recall, and F1-score. As shown in Table 1, the proposed system outclasses the other models in every evaluation area.

Algorithm Name	Accuracy	Precision	Recall	FSCORE
SVM	93.36	93.93	93.13	93.31
Random Forest	92.04	93.11	91.71	91.93
MLP	95.58	95.60	95.68	95.57
XGBoost	93.81	94.01	93.68	93.77
CatBoost	92.04	92.67	91.78	91.96
KNN	90.71	90.71	90.67	90.69
Logistic Regression	96.46	96.43	96.49	96.46
Proposed Model (RFE with Decision Tree)	98.67	98.69	98.66	98.67

Table 1. Outcome Analysis for All Models

B. Graphical Analysis of Model Performance

A comparative bar chart was utilized to visualize the success of each algorithm. This graphical representation confirms that the Extension model offers a significant improvement over baseline techniques. The consistent results across all metrics demonstrate stability and reliability of proposed framework.

Figure 2. Comparison Chart of Different Models

C. Analysis of Prediction Accuracy

The model's ability to guess correctly was tracked through a prediction matrix. High scores in the true-positive and true-negative sections confirm that the system is performing well. It clearly separates Parkinson's cases from healthy ones with a very low error rate. This also highlights the effectiveness of combining Borderline-SMOTE, Decision Tree, and RFE in improving the model's performance.

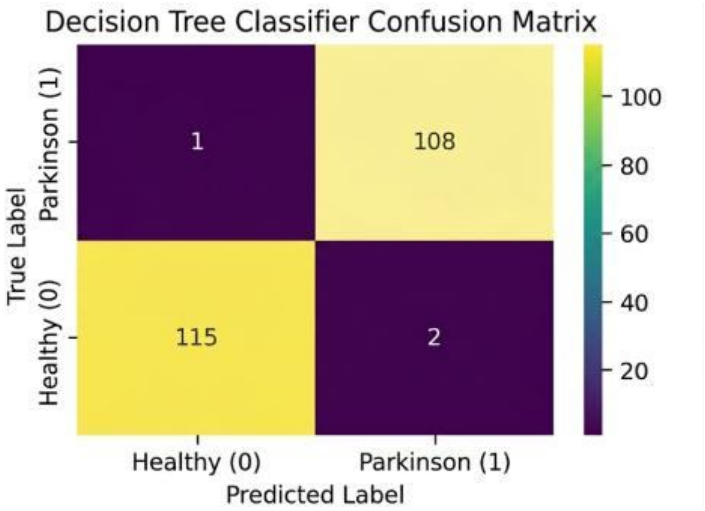


Figure 3. RFE with Decision Tree Confusion Matrix

D. Evaluation of Classification Thresholds

To measure the model's ability to distinguish between groups, the ROC curve is used. Superior performance is marked by the curve's proximity to the top-left edge of the chart. The resulting AUC value offers a simple way to rank models, with higher scores representing better overall precision.

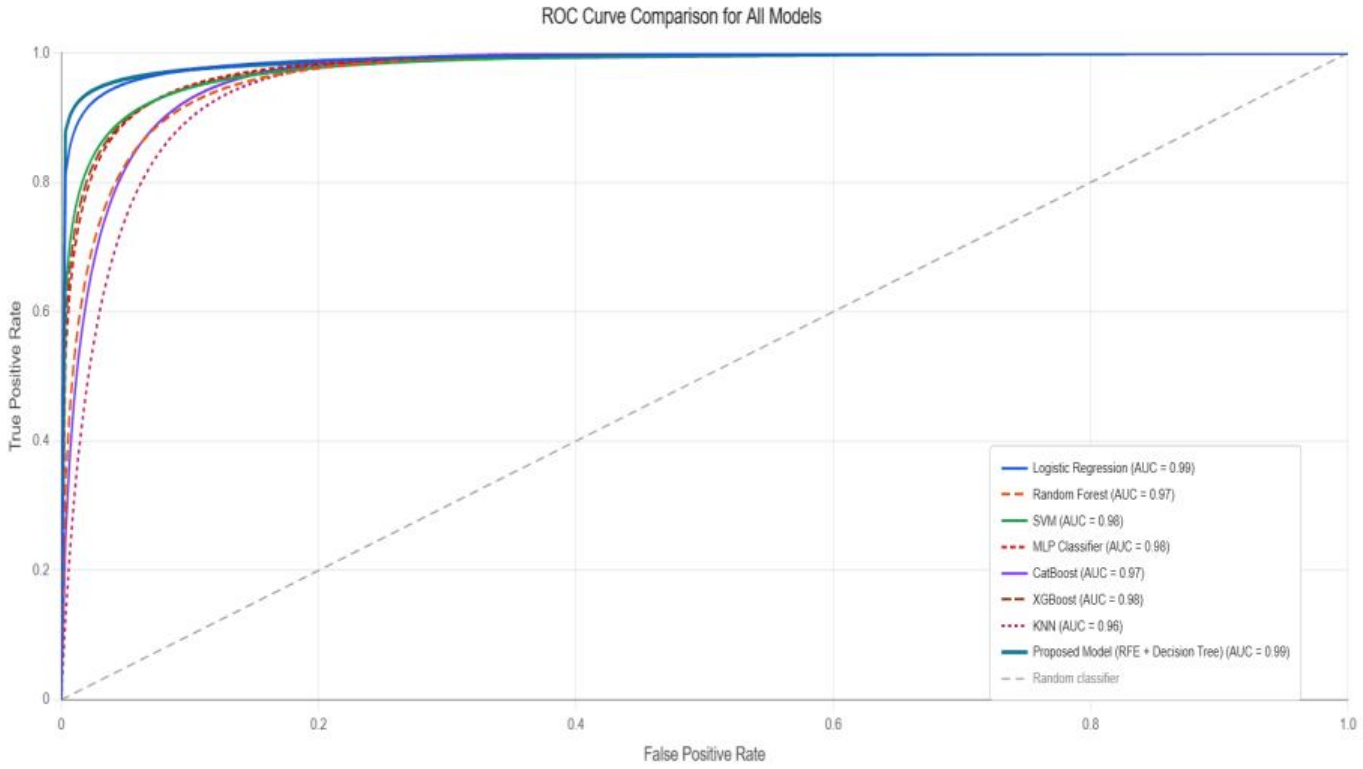


Figure 4. ROC Curve Comparision for All ML Models

E. SHAP Feature Importance

SHAP (SHapley Additive exPlanations) is used to understand which features have the most impact on the model's predictions. It assigns a value to each feature, showing how much it increases or decreases the chance of predicting Parkinson's Disease. The model detects Parkinson's Disease based on a person's voice features and measuring subtle abnormalities. People with Parkinson's tend to have higher vocal irregularity, more energy fluctuations, and different speech rhythms compared to healthy individuals. The top features capture exactly these abnormalities, and the bee-swarm plot shows which specific measurements matter most and in which direction they push the prediction.

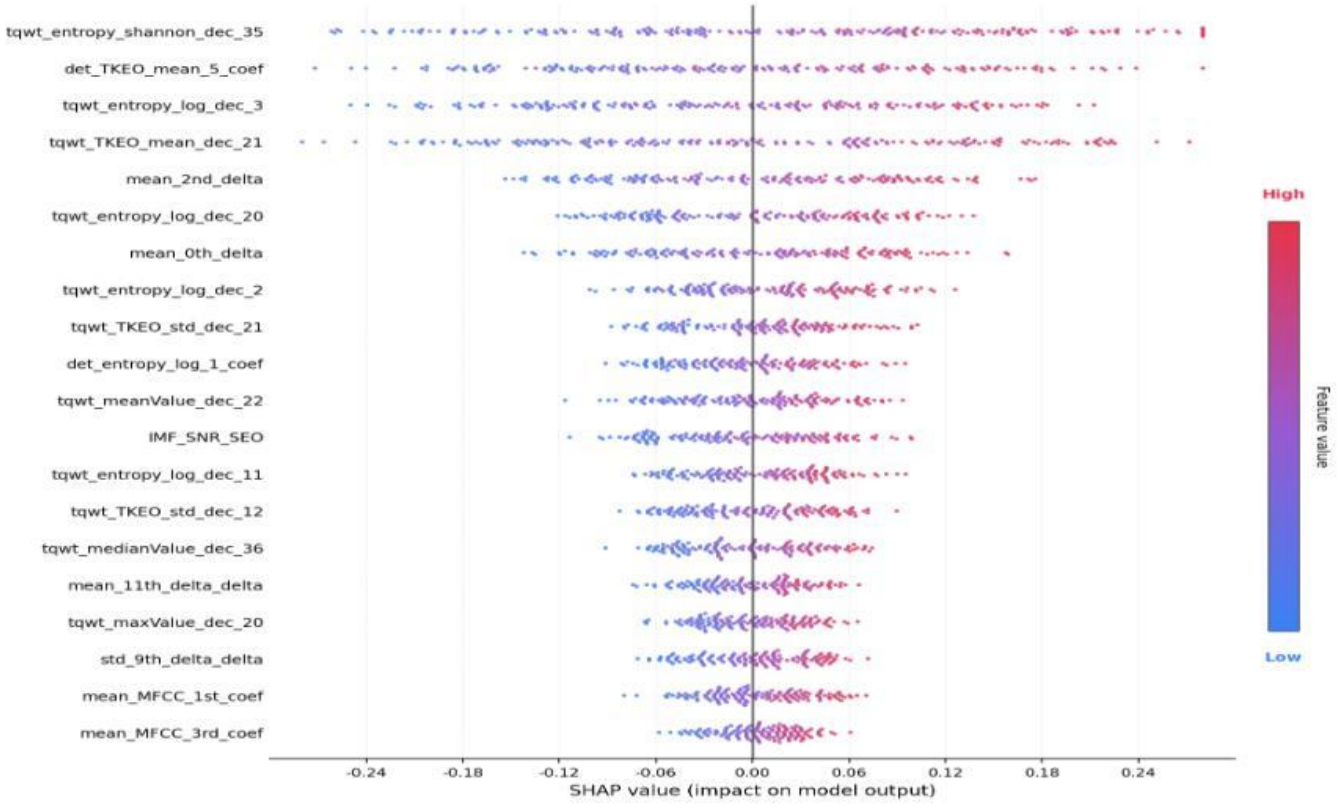


Figure 5. Analysis of Feature importance using SHAP

F. Discussion

The outcome suggests that the method performs reliably in detecting Parkinson's Disease using speech data. Decision Tree model gives the best performance with high accuracy and balanced precision and recall, which means it correctly identifies healthy as well as Parkinson's cases. This improvement is mainly due to Borderline-SMOTE, which helps the model learn from difficult samples, and feature selection, which removes unnecessary features and keeps only useful ones. Other models like Random Forest and XGBoost also give good results but are harder to interpret, while models such as KNN and Logistic Regression show lower performance, especially in identifying minority cases. Overall, the method provides accurate and stable predictions and is easy to understand, making it suitable for real-world healthcare use.

V. CONCLUSION

This work presents a reliable and easy-to-understand system for early PD detection by analyzing speech patterns. This framework was built to systematically address three fundamental limitations that are commonly encountered in medical machine learning applications. Class imbalance was resolved using Borderline-SMOTE, which directs synthetic sample generation toward the most critical regions of the feature space. The dimensionality of the input data was reduced through Recursive Feature Elimination, retaining only the attributes with the strongest predictive relevance. Model transparency was achieved by incorporating SHAP-based explainability, which maps the role of each feature in making a prediction. These three components work in concert to produce a system that is both technically sound and practically applicable.

A total of eight models were examined to select the most appropriate classifier for this task. Among them, the Decision Tree model showed the best results, obtaining better values in accuracy, precision, recall, and F1-score. This outcome demonstrates that targeting the most difficult classification cases and eliminating uninformative features produces tangible improvements in model learning. The addition of SHAP analysis further strengthens the system by giving clinicians a clear understanding of how individual vocal attributes influence each diagnostic outcome, thereby increasing trust and usability in real medical environments. The overall results of this study affirm that combining principled data balancing, targeted feature selection, and an inherently interpretable model can substantially elevate the quality of automated PD detection.

The proposed system is not only high-performing but also straightforward to validate and explain qualities that are particularly important when deploying AI-based tools in healthcare contexts where accountability and transparency are non-negotiable.

Looking ahead, several directions exist for extending this work. Testing the framework on larger and more demographically diverse datasets would strengthen confidence in its generalizability. Incorporating complementary data modalities such as gait analysis, neuroimaging, or handwriting patterns alongside speech features could provide a richer diagnostic signal and further boost performance. Ultimately, translating this pipeline into real-time clinical system would be a major step, allowing quick, low-cost, and Conservative screening of PD directly in healthcare settings.

REFERENCES

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

[2] K. Shyamala and T. M. Navamani, “Design of an Optimized Feature Driven Severity Stage Classifier for Parkinson’s Disease Prediction Using Deep Learning,” *IEEE Access*, vol. 13, pp. 142140–142160, 2025, doi: 10.1109/ACCESS.2025.3597851.

[3] Q. Dao *et al*., “Detection of Early Parkinson's Disease by Leveraging Speech Foundation Models,” *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 7, pp. 5181–5190, Jul. 2025, doi: 10.1109/JBHI.2025.3548917.

[4] M. Junaid, M. Ghergherehchi, and S. Lee, “Multitask Deep Learning for Predicting Parkinson’s Progression and Depression From Multimodal Time Series Data,” *IEEE Access*, vol. 13, pp. 147818–147841, 2025, doi: 10.1109/ACCESS.2025.3593254.

[5] M. Rey-Paredes, C. J. Pérez, and A. Mateos-Caballero, “Time Series Classification of Raw Voice Waveforms for Parkinson's Disease Detection Using Generative Adversarial Network-Driven Data Augmentation,” *IEEE Open Journal of the Computer Society*, vol. 6, pp. 72–84, 2025, doi: 10.1109/OJCS.2024.3504864.

[6] M. Khan, A. Moiz, G. Nawaz Khan, M. Wajid, M. Usman, and J. Ali, “An FPGA Prototype for Parkinson’s Disease Detection Using Machine Learning on Voice Signal,” *IEEE Access*, vol. 13, pp. 91113–91128, 2025, doi: 10.1109/ACCESS.2025.3572092.

[7] M. Ullrich *et al*., “Fall Risk Prediction in Parkinson's Disease Using Real-World Inertial Sensor Gait Data,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 319–328, Jan. 2023, doi: 10.1109/JBHI.2022.3215921.

[8] E. Kumari, M. K. Shukla, O. J. Pandey, and S. Yadav, “NeuroAid: Emotion-Based EEG Analysis for Parkinson's Disease Identification,” *IEEE Sensors Letters*, vol. 7, no. 12, pp. 1–4, Dec. 2023, doi: 10.1109/LSENS.2023.3335226.

[9] A. Rani Palakayala, P. Kuppusamy, D. Kothandaraman, G. Archana, and J. Gera, “HAMF: A Novel Hierarchical Attention-Based Multi-Modal Fusion Model for Parkinson’s Disease Classification and Severity Prediction,” *IEEE Access*, vol. 13, pp. 81252–81278, 2025.

[10] A. Rezvani *et al*., “DiffuseGaitNet: Improving Parkinson’s Disease Gait Severity Assessment With a Diffusion Model Framework,” *IEEE Journal of Biomedical and Health Informatics*, 2024.

[11] G. Amprimo, Z. Mei, C. Ferraris, G. Olmo, and D. Ravi, “A Data-Driven Exploration and Prediction of Deep Brain Stimulation Effects on Gait in Parkinson's Disease,” *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 7, pp. 4647–4658, Jul. 2025, doi: 10.1109/JBHI.2024.3446548.

[12] R. M. Al-Tam, F. A. Hashim, S. Maqsood, L. Abualigah, and R. M. Alwhaibi, “Enhancing Parkinson’s Disease Diagnosis Through Stacking Ensemble-Based Machine Learning Approach,” *IEEE Access*, vol. 12, pp. 79549–79567, 2024, doi: 10.1109/ACCESS.2024.3408680.

[13] C. Dong *et al*., “Static-Dynamic Temporal Networks for Parkinson’s Disease Detection and Severity Prediction,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2205–2213, 2023, doi: 10.1109/TNSRE.2023.3269569.

[14] S. M. Abdullah *et al*., “Deep Transfer Learning Based Parkinson’s Disease Detection Using Optimized Feature Selection,” *IEEE Access*, vol. 11, pp. 3511–3524, 2023, doi: 10.1109/ACCESS.2023.3233969.

[15] S. Gaba and H. Kaur, “Clinical Voice Data Collection and Analysis for Parkinson’s Disease Diagnosis,” in *Proc. 3rd World Conf. Communication & Computing (WCONF)*, Raipur, India, 2025, pp. 1–6, doi: 10.1109/WCONF64849.2025.11233316.

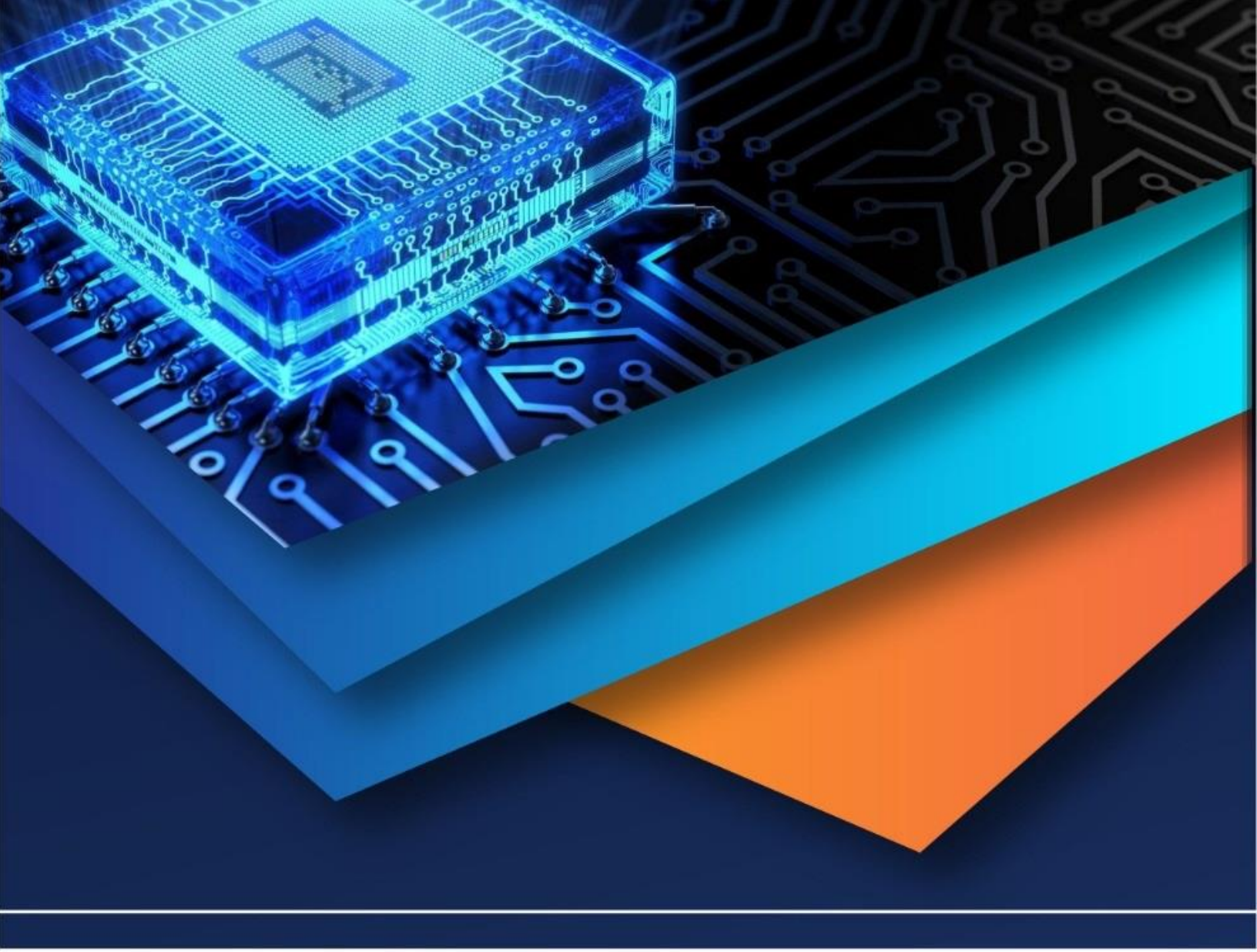
[16] J. Jamuna and K. Kasturi, “Enhancing Parkinson's Disease Prediction Using Machine Learning Techniques,” in *Proc. 9th Int. Conf. Inventive Systems and Control (ICISC)*, Coimbatore, India, 2025, pp. 958–964, doi: 10.1109/ICISC65841.2025.11188216.

[17] D. Kumar B. and K. France, “Prediction of Parkinson's Disease Using Machine Learning,” in *Proc. 9th Int. Conf. Inventive Systems and Control (ICISC)*, Coimbatore, India, 2025, pp. 147–152, doi: 10.1109/ICISC65841.2025.11187909.

[18] A. Selvi S. and T. Kamalakannan, “Machine Learning Based Prediction of Parkinson's Diseases,” in *Proc. 4th Int. Conf. Sentiment Analysis and Deep Learning (ICSADL)*, Nepal, 2025, pp. 1499–1502, doi: 10.1109/ICSADL65848.2025.10933153.

[19] A. S. *et al*., “Early Prediction of Parkinson's Disease with Machine Learning: A KNN Approach,” in *Proc. 5th Int. Conf. Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, 2025, pp. 1003–1007, doi: 10.1109/ICPCSN65854.2025.11035935.

[20] A. Akilandeswari, “Evaluation of Feature Selection Techniques for Predicting Parkinson's Disease using Machine Learning Models,” in *Proc. Int. Conf. Electronics and Renewable Systems (ICEARS)*, Tuticorin, India, 2025, pp. 1431–1435, doi: 10.1109/ICEARS64219.2025.10940164.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY