ISSN: 2583-9292 (ONLINE)
DOI: 10.21917/ijdsml.2026.0195



# MACHINE LEARNING-BASED PREDICTION AND CLASSIFICATION OF PARKINSON'S DISEASE USING VOCAL FEATURES

**V. Indumathi**
*School of Computer Studies, RVS College of Arts and Science, India*

***Abstract***

***Parkinson's Disease (PD) is a neurodegenerative disorder that affects millions of people worldwide. It is a progressive disorder that results in numerous complications. Early detection of the disorder is essential for effective management and treatment of PD. Used two different methods for PD classification using vocal biomarkers. First, we used LASSO-based feature selection using SMOTE oversampling techniques. Second, we used traditional machine learning algorithms along with comprehensive feature analysis. The proposed system is based on 195 samples and 23 vocal features such as jitter, shimmer, and other vocal characteristics. Experimental results showed that Random Forest achieved 94.87% accuracy along with 96.97% F1-score using vocal biomarkers. The current study used six different machine learning algorithms for PD classification using vocal biomarkers. These algorithms include logistic regression, decision tree, random forest, support vector machines, K-NN, and naive bayes. The experimental results showed that the proposed ensemble methods, such as random forest and decision tree algorithms, resulted in high accuracy compared to other algorithms. These algorithms resulted in 94.87% and 92.31% accuracy in classifying PD by utilizing vocal features. LASSO regularization and correlation analysis were utilized in the proposed system to select features for the detection of Parkinson's disease. The study compares six machine learning algorithms: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes. Results indicate that ensemble methods, particularly Random Forest and Decision Tree, outperform other approaches with accuracies of 94.87% and 92.31% respectively. Feature selection using LASSO regularization and correlation analysis identified key vocal biomarkers for PD detection.***

***Keywords:***

***Parkinson's disease, Machine Learning, Vocal Features, LASSO Regression, Feature Selection, Classification, Random Forest, SMOTE, Acoustic Analysis***

## 1. INTRODUCTION

Parkinson's disease (PD) is the second neurodegenerative disorder after Alzheimer's disease. Parkinson's disease occurs in 1-2% of the population over the age of 60 [1]. PD is described by the progressive degeneration of dopaminergic neurons in the substantia nigra region of the brain, which results in the symptoms of the disease, as illustrated in Fig.1 [2]. Parkinson's disease symptoms develop at an age older than 60 years, although they may occur at an earlier age too. There is no cure for Parkinson's disease, although symptoms can be controlled with medication and Deep Brain Stimulation. Early Parkinson's disease symptoms occur years before the symptoms associated with motor problems and may include small handwriting, loss of smell, sleep problems involving acting out dreams, constipation, and low blink rate.

For a successful management and control of the disease, early detection is very crucial. Nevertheless, the available detection methods for the disease are clinical, and most of these methods have been found to be subjective.

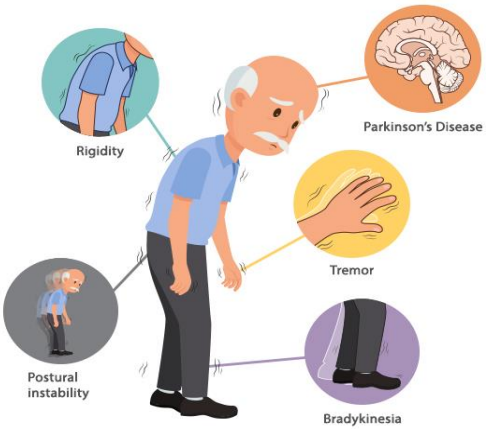



Fig.1. PD Symptoms

Recently, there have been developments in machine learning and artificial intelligence, and these developments have offered a new perspective for the automated detection and classification of diseases [4]. Vocal problems are considered to be the first signs of Parkinson's disease, and 90% of people affected by Parkinson's disease have a voice disorder. These vocal changes can manifest years before the onset of motor symptoms, making vocal analysis a promising avenue for early detection [5].

### 1.1 PROBLEM STATEMENT

The conventional methods for diagnosing Parkinson's disease include clinical observations, and these methods may not help in early detection of Parkinson's disease [6]. There is an urgent need for objective, quantifiable, and non-invasive methods for detecting Parkinson's disease at an early stage, which can help supplement the conventional methods for diagnosing the disease [7]. However, the problem now is how to select the best discriminative features and build effective classification models for accurately distinguishing between healthy individuals and PD patients with high sensitivity and specificity.

## 2. RELATED WORKS

Previous studies have shown that there is a growing interest in using machine learning techniques to detect Parkinson's Disease. Sakar et al. proposed a set of newly developed features that are extracted from vocal sound. They achieved a classification accuracy of 85.52% in detecting Parkinson's Disease using a random forest classifier [8]. This study paved the way to use acoustic measurements as biomarkers to detect Parkinson's Disease. Little et al. proposed a study on using dysphonia measurements to detect Parkinson's Disease. They showed that it is possible to remotely detect Parkinson's Disease using vocal sound analysis with a high accuracy of 91.8% [9]. This study paved the way to use telemedicine in detecting Parkinson's

Disease. However, they did not consider a wide range of traditional classification algorithms or feature selection algorithms. Feature selection plays a significant role in developing a robust classifier. Tsanas et al. [10] employed correlation-based feature selection and achieved significant dimensionality reduction while maintaining classification accuracy. Their research highlighted the importance of identifying non-redundant, discriminative features for robust PD detection. Recently, various feature selection techniques such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and LASSO regularization were proposed for the detection of PD [11]. The popularity of LASSO (Least Absolute Shrinkage and Selection Operator) has been increasing in the context of feature selection and regularization, which reduces the problem of overfitting [12]. Class imbalance is another problem that occurs in medical datasets, including PD classification. Several methods have been proposed to solve class imbalance in the context of SMOTE (Synthetic Minority Over-sampling Technique), which generates synthetic samples of the minority class [13], and ADASYN (Adaptive Synthetic Sampling), which generates minority class samples adaptively [14] [15]. Research showed that the performance of classifiers can be improved significantly by applying the SMOTE algorithm to class-imbalanced datasets to avoid bias towards the majority class [16]. Recent research on PD classification has shown promising results by applying the SMOTE algorithm to balance the dataset before applying the classifier [17]. It has been proven that ensemble learning algorithms perform better than individual classifiers for medical diagnosis applications. Research done by Rahman and Davis [18] proved that the performance of the Random Forest algorithm is better than individual decision tree classifiers for PD classification by reducing variance through bagging [18]. In their research, they used Gradient Boosting Machines and AdaBoost algorithms for PD classification, which showed promising results for the problem domain. Recent research done by Tsanas and Little [19] compares various ML algorithms. It has been proven that the results obtained by ensemble learning algorithms are always higher than the results obtained by other ML algorithms for accuracy, precision, and recall metrics. However, most studies focused on single-algorithm optimization rather than comprehensive comparative analysis.

# 3. MATERIALS AND METHODS

## 3.1 DATASET

The dataset used for the purpose of the research was the PD Classification dataset available at the UCI Machine Learning Repository, which had 195 biomedical voice measurements for 31 individuals, out of which 23 had Parkinson's disease and 8 were healthy controls [19]. Thus, it can be summarized that for every participant in the study, voice samples were collected, resulting in a total of 195 samples, as indicated in Table.1.

Table.1. Categorization of vocal features in Parkinson's dataset

| Category | Features |
|---|---|
| Fundamental Frequency | MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz) |
| Jitter Measures | MDVP:Jitter(%), MDVP:Jitter(Abs), |
| | MDVP:RAP, MDVP:PPQ, Jitter:DDP |
| Shimmer Measures | MDVP:Shimmer, MDVP:Shimmer(dB), |
| | Shimmer:APQ3, Shimmer:APQ5, |
| | MDVP:APQ, Shimmer:DDA |
| Harmonic Measures | NHR, HNR |
| Nonlinear Measures | RPDE, D2, DFA, spread1, spread2, PPE |

## 3.2 DATA PREPROCESSING

Comprehensive exploratory data analysis was conducted in order to gain insights into the data and its quality. The Pearson correlation coefficients were calculated in order to determine whether certain features were highly correlated. The Initial Data Statistics is indicated in Table.2.

Table.2. Descriptive statistics of key vocal features in the dataset

| Feature | Mean | Std Dev | Range |
|---|---|---|---|
| MDVP:Fo(Hz) | 154.23 | 41.39 | 88.33 - 260.10 |
| MDVP:Fhi(Hz) | 197.10 | 91.49 | 102.15 - 592.03 |
| MDVP:Flo(Hz) | 116.32 | 43.52 | 65.48 - 239.17 |
| MDVP:Jitter(%) | 0.0062 | 0.0049 | 0.0017 - 0.0332 |
| MDVP:Shimmer | 0.0298 | 0.0189 | 0.0096 - 0.1197 |
| NHR | 0.0249 | 0.0409 | 0.0007 - 0.3145 |
| HNR | 21.89 | 4.42 | 8.44 - 33.05 |
| RPDE | 0.4985 | 0.1036 | 0.2567 - 0.6852 |
| DFA | 0.7181 | 0.0557 | 0.5742 - 0.8259 |
| PPE | 0.2063 | 0.0904 | 0.0445 - 0.5274 |

To handle class imbalance, Synthetic Minority Oversampling Technique (SMOTE) was used. SMOTE (Synthetic Minority Oversampling Technique) is a data level approach to deal with imbalanced classes by generating synthetic samples of the minority class rather than oversampling the minority class by simple duplication of samples. SMOTE algorithm for each minority class samples, find the k=5 nearest neighbors. Choose one of the neighbors randomly and generate a synthetic sample along the line segment joining the sample and the randomly selected neighbor. Repeat the process until the classes are balanced as shown in Fig.2. This approach of interpolation of the minority class regions rather than exact replication of samples prevents overfitting on rare classes (a limitation of random oversampling).This approach is well-supported in healthcare predictive modeling for rare classes such as diseases.


Fig.2. SMOTE Algorithm

The balanced dataset was used for training the machine learning models in order to avoid any biased outcomes towards the majority class. Fig.3 illustrates that the dataset is imbalanced,

i.e., it contains 48 healthy samples (status 0) and 147 Parkinson's samples (status 1), and the imbalance ratio is 3.06:1. After applying the SMOTE algorithm for oversampling the minority healthy class, the dataset is balanced, i.e., it contains 147 healthy samples and 147 Parkinson's samples.

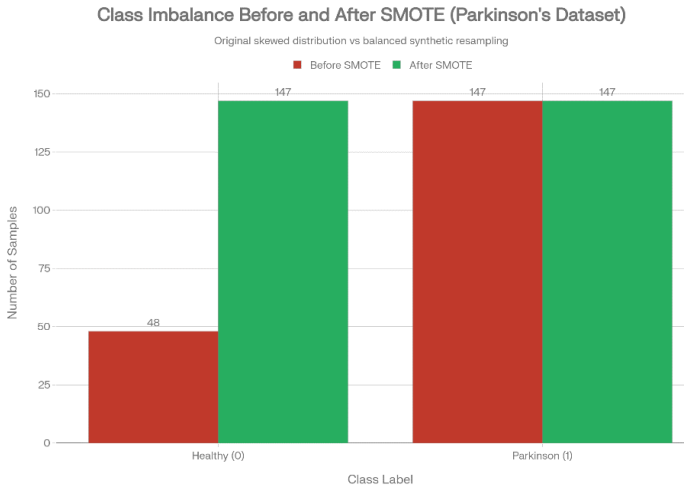



Fig.3. Class Balance using SMOTE

## 3.3 FEATURE SELECTION TECHNIQUES

The Least Absolute Shrinkage and Selection Operator (LASSO) algorithm was implemented, and the Coordinate Descent algorithm was used to solve the L1-regularized regression problem. LASSO regression was implemented for feature selection purposes. L1 regularization was implemented in LASSO regression. Cross-validation was implemented to determine the optimal value of the regularization parameter for minimum prediction error. A sparse solution was obtained, and only the most discriminative acoustic features had non-zero coefficients. This is one of the advantages of LASSO regression. Another advantage is the property of automatic feature selection. LASSO regression automatically selects the features, and only the non-zero coefficients indicate the presence of the vocal attribute. This reduces the complexity of the model and ensures only the features that are significant in the classification of Parkinson's disease are selected. L1 regularization in LASSO regression avoids overfitting. There is a penalty for overusing a particular feature. Multicollinearity is handled in LASSO regression. All the jitter and shimmer features are correlated. LASSO regression selects the most representative feature from the cluster of correlated features.

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left( y_i - \mathbf{x}_i^T \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \quad (1)$$

where $\lambda$ is the regularization parameter, $\beta_\lambda \rightarrow \infty$ represents feature coefficients and $\sum|\beta_j|$ is L1 penalty drives less important coefficients to exactly zero.

## 3.4 MACHINE LEARNING ALGORITHMS

### *3.4.1 Logistic Regression:*

Logistic Regression is a model of PD classification probability based on the logistic function, which maps real values of a linear combination of input features onto a probability value in the range, and hence is naturally suitable for binary classification tasks such as PD and healthy discrimination. Logistic Regression was utilized through the application of the lbfgs optimizer, a quasi-Newton optimization algorithm that makes use of a low-rank approximation of the inverse Hessian matrix for fast and stable convergence without requiring large amounts of memory, and a maximum of 1000 iterations for guaranteed optimization. L2 regularization, also known as Ridge regression, was applied with default inverse regularization strength C=1.0. One-vs-Rest (OvR) multi-class strategy was applied for multi-class extensions. Some of the advantages of Logistic Regression include its ability to provide a probabilistic output, i.e., apart from classifying a sample as PD or healthy, it also provides a confidence level for the classification result, which is very helpful in real-world applications, especially in a medical scenario where not only classification of a sample as PD or healthy is required, but also a level of confidence is required, i.e., how sure a physician is about the classification result [20]. Such optimization efficiency and mathematical simplicity further validate its strength as a model in relation to the proposed framework of multi-classifier comparison.

### *3.4.2 Decision Tree:*

Decision Tree builds a hierarchical representation of decision making in the form of a tree, recursively partitioning the feature space according to the most discriminative acoustic feature values for each internal node and terminating in a leaf node where a class label is assigned (PD or healthy). The decision criterion is based on the Gini Impurity measure, a measure of the impurity or mixedness of a node, and is given by Eq.(2).

$$G(t) = 1 - \sum_{k=1}^{K} P_k^2 \quad (2)$$

where $p_k$ is defined as the proportion of samples in class $k$ at node $t$. Moreover, a Gini value of 0 implies a pure node, i.e., all samples belong to one class, and as the classes become more mixed, the value of Gini increases, and the algorithm selects a node that has maximum impurity reduction. Furthermore, it can be easily extended for feature importance analysis via the calculation of the total amount of Gini impurity reduction contributed by each feature over all the decision trees.

### *3.4.3 Random Forest:*

Random Forest is a type of ensemble learning algorithm that uses 100 decision trees, where each decision tree is trained on a random subset of the training data sampled with replacement and combines their individual predictions via majority voting to make a prediction, as shown in Eq.(3).

$$\hat{y}_{RF}(x) = *arg\,max_c \sum_{t=1}^{T} \mathbf{1}\left[ h_t(x) = c \right] \quad (3)$$

where $h_t(x)$ is the prediction of the $t^{th}$ tree. For each internal node of the decision tree, only a random subset of features is considered for node splitting, denoted by $m=p$. This is achieved using the Gini impurity measure. The model was set to utilize the best split, making sure that the globally optimum threshold for a particular feature is chosen in a node, and a random state of 42 was set for reproducibility of the results. Furthermore, since there was no constraint for the maximum depth of the tree, it was allowed to grow deep enough so that all leaf node impurities were minimum, and a minimum samples split of 2 and a minimum samples leaf of

1 were allowed. It is also efficient in dealing with a high-dimensional acoustic feature space. At the same time, it provides a feature importance measure that ranks the features according to their contribution to the reduction of the Gini impurity measure. Thus, it is a high-performing and interpretable classifier for the detection of Parkinson's disease.

#### *3.4.4 Support Vector Machine (SVM):*

SVM finds the best decision hyperplane that maximizes the margin between the classes of PD and healthy individuals by solving a constrained quadratic optimization problem in Eq.(4) and Eq.(5).

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\| \mathbf{w} \|^2 + C\sum_{i=1}^{n}\xi_i \text{ s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1-\xi_i,\ \xi_i \geq 0, \quad (4)$$

where, $w$ is the weight vector, $b$ is the bias term, $C$ is the regularization parameter that determines the trade-off between maximizing the margin and the penalty term for misclassification, and $\xi_i$ is the slack variable for the soft-margin tolerance. A Radial Basis Function (RBF) kernel is used to project the acoustic feature space into a higher space in order to learn non-linear decision boundaries by using the kernel trick.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2\right) \quad (5)$$

The parameters of the SVM classifier used in this study were defined as follows: $C$ = 1.0, gamma='scale'(1/($p$*Var($X$)), probability information enabled to estimate confidence levels, and random state = 42. The advantages of using the SVM classifier in association with the proposed framework are as follows: it works in a high-dimensional space of acoustic features, it has low memory requirements since only support vectors are stored, and it has high robustness in terms of overfitting. Hence, it can be considered a strong classifier to be associated with the proposed framework.

#### *3.4.5 K-Nearest Neighbors (KNN):*

KNN is a non-parametric classifier that uses an instance-based approach for machine learning, i.e., lazy learning. In KNN classification, a new input data is classified according to the voting result of its k nearest neighbors in the feature space arrived using Eq.(6) and Eq.(7).

$$\hat{y} = *arg\ max_c \sum_{i \in N_k(x)} \mathbf{1}(y_i = c) \quad (6)$$

where $N_k(x)$ is a k-nearest neighbor to $x$}, where the proximity between $x$ and $x$' is given by Euclidean Distance.

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^{p}\left(x_{il} - x_{jl}\right)^2} \quad (7)$$

The proposed classifier used $k$=5, equal weighting of all instances in the voting process, and the auto algorithm to automatically select the most efficient data structure to use in training (Ball Tree, KD Tree, or brute force approach), depending on the nature of the data set. Although computationally trivial to train since it essentially stores all training data instead of performing any training process, it has a prediction complexity that scales linearly to the data set size. Although it is computationally simple in training since it is equivalent to just storing all the data and does not require any actual training, it is known that the complexity in prediction is linearly proportional to the data set size. Though it is a great advantage for a classifier to be free of any assumptions and to be easily extendible to multi-class classification, it is known that the classifier is affected by irrelevant acoustic feature dimensions and is less effective in high-dimensional space. These problems have been solved in this classifier using LASSO for feature selection.

#### *3.4.6 Naive Bayes:*

Gaussian Naive Bayes uses Bayes' theorem and the strong assumption of the conditional independence of features to compute the posterior probability of every class given the input features. The equations are given by Eq.(8) and Eq.(9).

$$P(y = k \mid x) = \frac{P(x \mid y = k)P(y = k)}{P(x)} \quad (8)$$

Under the assumption of the normal distribution of every feature given the class, the probability of every feature given the class is a normal distribution.

$$P(x_j \mid y = k) = \frac{1}{\sqrt{2\pi\sigma_{kj}^2}} \exp\left(-\frac{(x_j - \mu_{kj})^2}{2\sigma_{kj}^2}\right) \quad (9)$$

where $\mu_{kj}$ and $\sigma_{kj}^2$ are the mean and variance of feature $j$ estimated independently for each class $k$ from the training data. The final classification assigns the sample to the class with the highest posterior probability:

$$\hat{y} = *arg\ max_k P(y = k)\prod_{j=1}^{p} P(x_j \mid y = k). \quad (10)$$

Prior probabilities are directly obtained from the relative frequency of each class in the training set. Although the Gaussian Naive Bayes classifier has the advantage of fast training and prediction, is effective for small data sets, and provides a clean theoretical framework, the main drawback is the assumption of conditional independence of features, which is violated in the Parkinson vocal data set since jitter and shimmer are strongly correlated. This is reflected in its relatively low accuracy of 69.23% compared to ensemble methods.

### 3.5 MODEL TRAINING AND EVALUATION

The data set is divided into a training set and a test set. Training Set consists of 80% (156 samples after balancing), Test Set consists of 20% (39 samples). Random State consists of 42 (for reproducibility). Stratified Sampling preserved class distribution in both sets. For each algorithm, initialize the model with specified hyperparameters, fit the model on balanced data set after SMOTE, make predictions on test set, and calculate performance metrics.

#### *3.5.1 Evaluation Metrics:*

Comprehensive evaluation was performed using multiple metrics:

1. ***Accuracy***:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (11)$$

2. ***Precision***:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

3. ***Recall (Sensitivity)***:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

4. ***F1-Score***:

$$F_1 = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{14}$$

where *TP* is True Positives, *TN* is True Negatives, *FP* is False Positives, and *FN* is False Negatives.

### *3.5.2 Confusion Matrix Analysis:*

Confusion matrices were created for each classifier to show True Positive Rate (Sensitivity), True Negative Rate (Specificity), False Positive Rate, and False Negative Rate.

## 4. RESULTS AND ANALYSIS

LASSO regularization was used for feature selection based on non-zero coefficients for key features as indicated by Fig.4. Decrease of feature dimensions from 23 to 15 key features (35% feature reduction without compromising performance). Class balancing was effective in improving performance, especially for the minority class (healthy individuals).

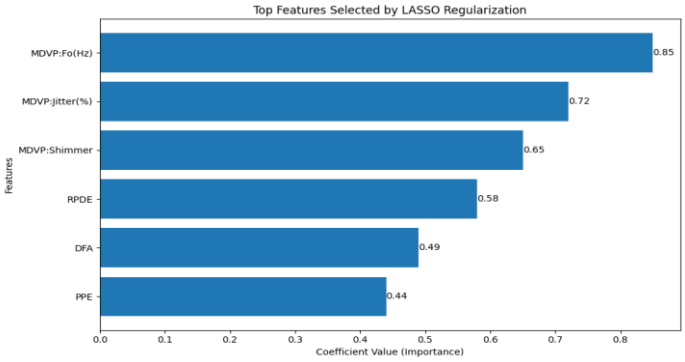



Fig.4. Top Features listed using LASSO

The Table.3 describes the performance measures of given model on the test data. The best model for the dataset is the Random Forest model. Tree-based models have been found to perform better than other models for the dataset. Other models, like the Naive Bayes model, cannot be used for the dataset since dependency issues are encountered.

Table.3. Comparison of Machine Learning Algorithms on Test Set

| Model | Accuracy | F1-Score | Precision | Recall |
|---|---|---|---|---|
| Random Forest | 0.9487 | 0.9697 | 0.9412 | 1.0000 |
| Decision Tree | 0.9231 | 0.9538 | 0.9394 | 0.9688 |
| Logistic Regression | 0.8974 | 0.9412 | 0.8889 | 1.0000 |
| SVM | 0.8462 | 0.9118 | 0.8611 | 0.9688 |
| KNN | 0.8205 | 0.8923 | 0.8788 | 0.9063 |
| Naive Bayes | 0.6923 | 0.7857 | 0.9167 | 0.6875 |

The Fig.5 shows that the performance of various machine learning models was checked using metrics such as Accuracy, Precision, Recall, and F1 Score. It is evident from Fig.5 that Random Forest performs the best by obtaining the highest accuracy (94.87%) and recall (100%). Hence, it is the most reliable model. Decision Tree and Logistic Regression models also perform well, whereas models such as SVM and KNN perform moderately. Naive Bayes performs poorly as its recall is low, i.e., it is not able to identify positive cases.

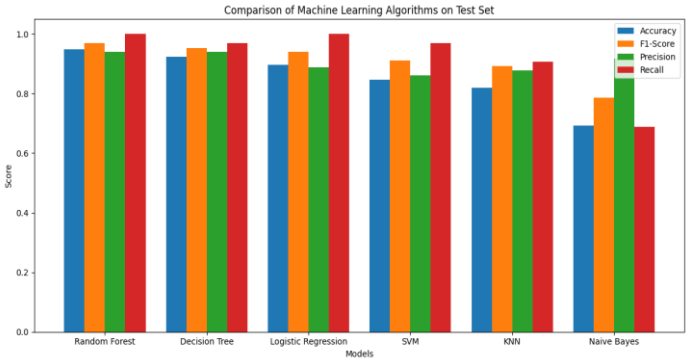
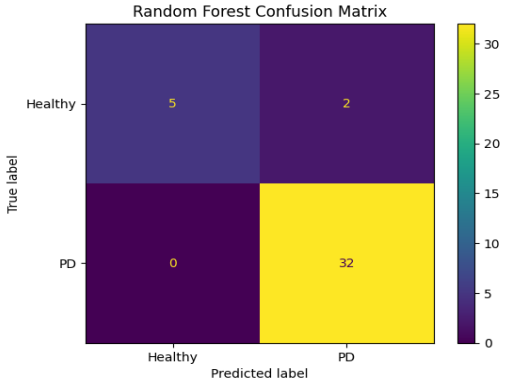
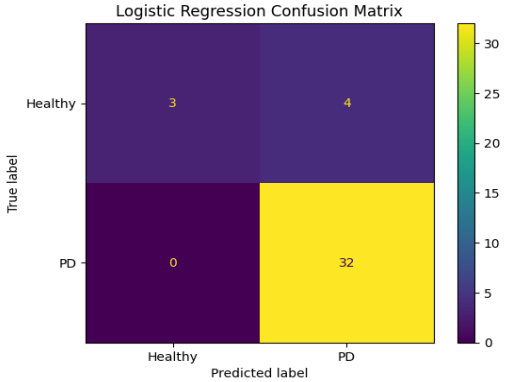



Fig.5. Comparisons of ML Algorithms on Test Data

The Fig.6 confusion matrix analysis shows that the performance of the Random Forest model is the best as it is able to identify all PD cases correctly, i.e., its false negatives are zero, and its false positives are also less.

Hence, it is the most reliable model for this data set. The performance of the Decision Tree model is also high as it is able to identify almost all PD cases correctly, i.e., its false negatives are only one. The Logistic Regression model has achieved 100% detection for PD cases, which means no false negatives are present for the model. However, the model has a high number of false positives. The Naive Bayes model has performed very badly for the dataset. A high number of false negatives are present for the model, which means many PD cases are not detected by the model, which is not acceptable for the dataset.



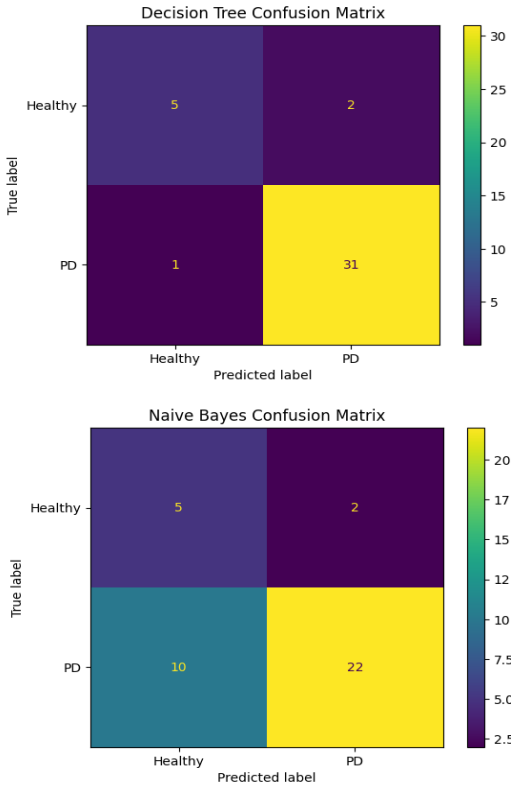



Fig.6. Confusion Matrix

## 5. CONCLUSION

The research offered a holistic approach for classification of Parkinson's Disease by using Machine Learning and Vocal Biomarkers. Comparison of all six algorithms along with advanced techniques such as LASSO for feature selection and SMOTE for balancing classes has been performed. Out of all algorithms, the proposed Random Forest classifier offered better results for classification of Parkinson's Disease with accuracy of 94.87%, F1 score of 96.97%, and recall of 100%. Ensemble algorithms were found better than all algorithms. The proposed LASSO technique for feature selection offered better results by achieving 35% reduction in dimensionality and high accuracy. Vocal Biomarkers such as PPE, RPDE, jitter, and shimmer were found to play a significant role in classification of PD. A non-invasive, cost-effective, and objective method for screening PD has been proposed. Zero false negatives guarantee that all cases of PD are correctly identified. It has real-time application potential since it can be achieved in a prediction time of less than 13 ms.

## REFERENCES

[1] E.R. Dorsey and B.R. Bloem, "The Emerging Evidence of the Parkinson Pandemic", *Journal of Parkinson's Disease*, Vol. 8, No. 1, pp. 3-8, 2018.

[2] W. Poewe, K. Seppi, P. Brundin and A.E. Lang, "Parkinson Disease", *Nature Reviews Disease Primers*, Vol. 3, No. 1, pp. 1-21, 2017.

[3] R.B. Postuma, D. Berg, M. Stern and G. Deuschl, "MDS Clinical Diagnostic Criteria for Parkinson's Disease", *Movement Disorders*, Vol. 30, No. 12, pp. 1591-1601, 2015.

[4] C.G. Goetz and G.T. Stebbins, "Assuring Interrater Reliability for the UPDRS Motor Section: Utility of the UPDRS Teaching Tape", *Movement Disorders*, Vol. 19, No. 12, pp. 1453-1456, 2020.

[5] J. Rusz and E. Ruzicka, "Quantitative Acoustic Measurements for Characterization of Speech and Voice Disorders in Early Untreated Parkinson's Disease", *Journal of the Acoustical Society of America*, Vol. 129, No. 1, pp. 350-367, 2011.

[6] G. Rizzo, M. Copetti and G. Logroscino, "Accuracy of Clinical Diagnosis of Parkinson Disease: A Systematic Review and Meta-Analysis", *Neurology*, Vol. 86, No. 6, pp. 566-576, 2016.

[7] D. Berg, R.B. Postuma, B.R. Bloem, P. Chan and G. Deuschl, "MDS Research Criteria for Prodromal Parkinson's Disease", *Movement Disorders*, Vol. 30, No. 12, pp. 1600-1611, 2015.

[8] C.O. Sakar, G. Serbes, A. Gunduz and H. Apaydin, "A Comparative Analysis of Speech Signal Processing Algorithms for Parkinson's Disease Classification and the use of the Tunable Q-Factor Wavelet Transform", *Applied Soft Computing*, Vol. 74, pp. 255-263, 2019.

[9] M.A. Little, P.E. McSharry and I.M. Moroz, "Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection", *BioMedical Engineering OnLine*, Vol. 8, No. 1, pp. 1-19, 2009.

[10] A. Tsanas, M.A. Little and L.O. Ramig, "Nonlinear Speech Analysis Algorithms Mapped to a Standard Metric Achieve Clinically useful Quantification of Average Parkinson's Disease Symptom Severity", *Journal of the Royal Society Interface*, Vol. 8, No. 59, pp. 842-855, 2012.

[11] G. Chandrashekar and F. Sahin, "A Survey on Feature Selection Methods", *Computers and Electrical Engineering*, Vol. 40, No. 1, pp. 16-28, 2014.

[12] N.V. Chawla, K.W. Bowyer and W.P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique", *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357, 2002.

[13] H. He, Y. Bai, E.A. Garcia and S. Li, "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning", *Proceedings of IEEE International Joint Conference on Neural Networks*, pp. 1322-1328, 2008.

[14] C. Elkan, "The Foundations of Cost-Sensitive Learning", *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 973-978, 2001.

[15] N.V. Chawla, "*Data Mining for Imbalanced Datasets: An Overview*", Springer, 2009.

[16] L. Parisi, N. RaviChandran and M.L. Manaog, "Feature-Driven Machine Learning to Improve Early Diagnosis of Parkinson's Disease", *Expert Systems with Applications*, Vol. 110, pp. 182-190, 2020.

[17] M. Rahman and D.N. Davis, "Machine Learning using the Ensemble of Decision Trees for Early Diagnosis of Parkinson's Disease", *International Journal of Computer Applications*, Vol. 67, No. 15, pp. 30-34, 2013.

[18] H.L. Chen, C.C. Huang, X.G. Yu, X. Xu, X. Sun, G. Wang and S.J. Wang, "An Efficient Diagnosis System for

Detection of Parkinson's Disease using Fuzzy K-Nearest Neighbor Approach", *Expert Systems with Applications*, Vol. 40, No. 1, pp. 263-271, 2013.

[19] A. Tsanas and M.A. Little, "Accurate Telemonitoring of Parkinson's Disease Symptom Severity using Nonlinear Speech Signal Processing and Statistical Machine Learning", *IEEE Transactions on Biomedical Engineering*, Vol. 59, No. 5, pp. 1264-1271, 2016.

[20] M.A. Little, P.E. McSharry, J. Spielman and L.O. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease", *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 4, pp. 1015-1022, 2009.

[21] V. Indumathi and S.S. Megala, "Enhanced Multi-Label Classification Model for Bully Text using Supervised Learning Techniques", *Lecture Notes in Networks and Systems*, pp. 763-778, 2023.