Received: 6 March 2023 | Revised: 6 April 2023 | Accepted: 17 April 2023 | Published online: 17 April 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2023, Vol. 1(2) 92–98
DOI: 10.47852/bonviewJDSIS3202831

BON VIEW PUBLISHING

Voice Biomarkers for Parkinson's Disease Prediction Using Machine Learning Models with Improved Feature Reduction Techniques

Nalini Chintalapudi[1],* , Venkata Rao Dhulipalla[2] , Gopi Battineni[1] , Ciro Rucco[1] and Francesco Amenta[1]

[1]*The Clinical Research Center, University of Camerino, Italy*

[2]*The Research Centre of the ECE Department, V. R. Siddhartha Engineering College, India*

Abstract: As a chronic and life-threatening disease, Parkinson's disease (PD) causes people to become rigid and inactive and have shaky voices. There is an argument that current PD detection techniques are ineffective due to their high latency and low accuracy. To enhance the accuracy of PD identification, voice recordings were used as biomarkers in conjunction with the synthetic minority oversampling technique (SMOTE). Three machine learning (ML) models namely support vector machine (SVM), K-nearest neighbors (KNN), and random forest (RF) were adopted to calculate the prediction accuracy. By applying an unsupervised dimensional reduction method, the generated model eliminates redundant data and speeds up training and testing. Model performance is estimated with three parameters, including accuracy, F1 score, and area under the curve (AUC) values. Experimental outcomes suggested that the RF model outperforms other models with 97.4% of classification accuracy. This type of research aims to analyze patient voice recordings to determine the disease severity.

Keywords: Parkinson's disease, machine learning, SMOTE, accuracy, AUC

1. Introduction

A neurodegenerative disease, Parkinson's disease (PD) affects more than 1% of adults (>age of 50 years) throughout the world (Bloem et al., 2021). In recent years, this problem has become more prevalent as the age of global population has increased. PD causes motor symptoms that include tremors, bradykinesia, stiffness, and others and cognitive dysfunction (Hayes, 2019; Tolosa et al., 2021). In 2040, it is estimated that 12.9 million people will be affected by PD, which is almost twice as many as in 2015 (Feigin et al., 2021). The absence of a precise diagnostic test makes it difficult to diagnose PD and is unfortunate that 25% of people with PD are misdiagnosed and mistreated (Chia et al., 2020; Elsworth, 2020).

When the disease is detected and treated early, mortality rates can be reduced and quality of life can be improved. A limited number of resources and a lack of knowledge make it difficult to detect PD early with improved accuracy (Weintraub et al., 2022). Despite its incurability, early treatment can reduce and make more affordable PD symptoms. To detect PD, conventional approaches like magnetic resonance imaging (MRI) (Hemond & Bakshi, 2018), positron emission tomography (PET) (Weber, 2020), and dopamine transporter (DAT) (Brücke & Brücke, 2021) are available. Despite their significant improvements in PD detection and prevention, these techniques do possess some shortcomings in terms of accuracy. The main disadvantage of imaging scans is their high cost and the time it takes to prepare them.

In recent years, machine learning (ML) has received a great deal of attention due to its ability to produce precise results through the use of novel data taxonomies and algorithms (Esteva et al., 2019; Jayatilake & Ganegoda, 2021). Due to its strong impact on advanced disease diagnostics, scientists turned to ML techniques to accurately predict disease. Scanned images with help of PET, MRI, and DAT can help to do disease prediction using ML models (Ohana et al., 2018; De Wachter et al., 2020). A large dataset makes it impossible for classification algorithms to create reliable models. The classifiers misclassify these datasets based on many redundant or irrelevant features. Since many large datasets have an imbalanced distribution of classes, the classification process tends to favor the majority class.

It is possible to detect PD by analyzing the characteristics of hand drawing and vocalization in the available datasets (Arora et al., 2015; Papadopoulos et al., 2019). It is common for PD patients to experience voice problems at the onset of the disease. A supervised support vector machine (SVM) model was proposed with different feature selection approaches with voice biomarkers dataset and reported 97.57% of accuracy (Aich et al., 2019). A deep belief network-based system was developed for efficient PD diagnosis and reported a testing accuracy of 94% (Al-Fatlawi et al., 2016). Another study proposed the evaluation of PD-related speech samples using a multimodel

*Corresponding author: Nalini Chintalapudi, The Clinical Research Center, University of Camerino, Italy. Email: nalini.chintalapudi@unicam.it

framework in order to classify PD subjects and healthy controls, and reported SVM and regression models produced the highest accuracy of 70% (Ali et al., 2019).

Another study used functional MRI and ML to predict optimal deep brain stimulation parameters for PD with 88% of accuracy. Authors mentioned that PD patients with a priori clinically optimized stimulation settings and those who have never received stimulation can be predicted using the model in previously unseen datasets (Boutet et al., 2021). Dimensionality reduction techniques largely help to ease the training dataset to make visualization. To compact and eliminate irrelevant features, a dimensionality reduction technique called principal component analysis (PCA) was applied (Jolliffe & Cadima, 2016). An algorithm for diagnosing PD based on PCA and ML algorithms using vocal features reported 95% of accuracy (Rao et al., 2022). Further, the synthetic minority oversampling technique (SMOTE) (Fernández et al., 2018) resampling was employed to balance the class distribution and broaden the sample range. The detection of PD was achieved with 99% accuracy using cascaded deep learning frameworks and SMOTE approaches (Chintalapudi et al., 2022).

It has been reported that ML-based PD diagnosis has high classification rates, but the literature either used many features which caused the computation time to increase, or the extraction of the features was difficult even when few features were used. As a result, indirect computation time is also high. There is a scarcity of studies on the application of both PCA and SMOTE feature selection techniques. As such we coupled both techniques of PD detection using voice biomarkers by ML techniques. A speech processing technique was used to evaluate and estimate the persistence of PD using recordings of patient voices. The performance of ML models is improved when there are tiny dimensions in the data because they are simpler to compute and visualize. By reducing the number of variables and removing irrelevant data, the model becomes more accurate.

After data preprocessing, the dataset is trained with three different ML classifiers namely k-nearest neighbors (KNN), SVM, and random forest (RF) algorithms. The choice of an ML model depends on the specific problem and the characteristics of the data you have. SVM is a supervised learning algorithm used for classification and regression analysis. It works by finding the hyperplane that maximizes the margin between the two classes and then classifying the data based on which side of the hyperplane it falls on. SVM can be effective when the number of features is small compared to the number of samples. In addition, there is a clear division of authority between the classes. RF is an ensemble learning method that combines multiple decision trees to create a more accurate and robust model. It works by randomly selecting subsets of features and samples to create multiple decision trees and then combining the predictions of those trees to make the final prediction. RF can be effective when dealing with high-dimensional data, and when there is a lot of noise or missing data in the dataset. KNN is a nonparametric algorithm used for classification and regression analysis. It works by finding the KNN to the current data point. It uses the majority vote or weighted average of those neighbors to predict. KNN can be effective when there is no clear functional form that describes the relationship between the features and the target variable. In addition, the data is clustered in certain regions of the feature space. It is often an ideal idea to try multiple models and compare their performance to determine which one works best for your particular situation.

The performance of three adopted models was tested in terms of model accuracy, precision, F1 score, false-positive rate, and recall. In this work, we extrapolate the promise of data resampling techniques in voice recordings during PD diagnosis and produce the ML model performance differences with and without feature reduction techniques.

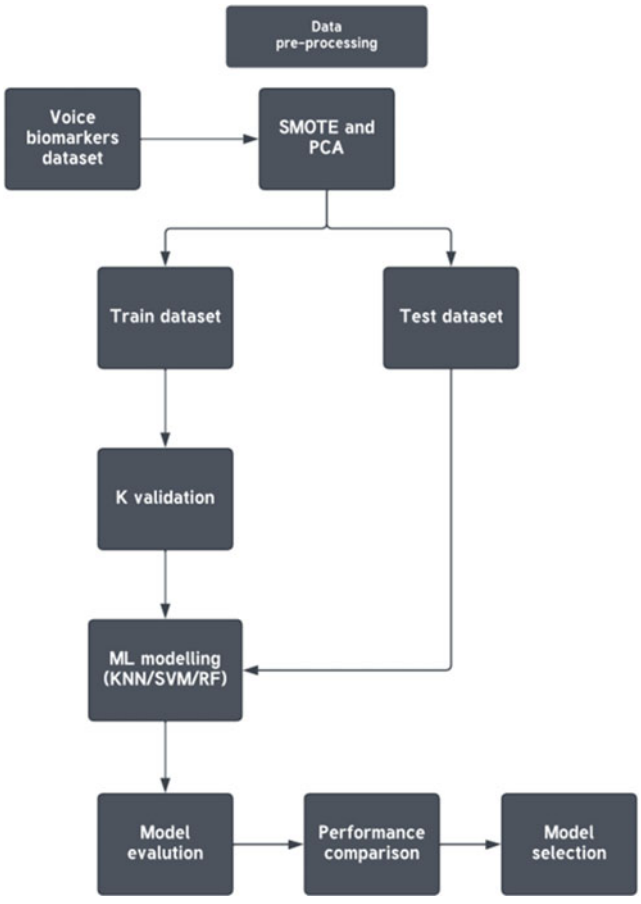
2. Methods

Figure 1 presents the experimental framework of PD classification with feature reduction techniques combined with ML modeling. At first, the PD dataset was adopted from the UCI ML repository. After data collection, PCA and SMOTE techniques were applied as the data preprocessing techniques. After data preprocessing, the resulting dataset is decomposed into training and testing datasets. With k-fold validation, the three-model performance is computed using metrics like accuracy, precision, sensitivity, and F1 score. Ultimately, the comparative analysis is conducted to estimate the optimal model for better PD subject classification.

2.1. Dataset

A dataset from the UCI collection on PD was considered (UCI ML Repository: Parkinsons Data Set n.d.). These data included 195 voice recordings data of 31 PD subjects who were categorized as 23 with PD and the remaining eight are healthy characteristics (HC). There is a special voice measurement applied to each column that pertains to a voice recording of each subject who attended at least

Figure 1
Experimental framework



six speaking sessions. Table 1 presents a vocal feature along with the acoustic features that were extracted. Each column represents a special voice measurement based on voice recordings from at least

Table 1
Voice biomarkers

Attribute	Descriptions
name	ASCII subject name and recording number
MDVP	Average vocal fundamental frequency
MDVP	Fhi(Hz) - Maximum vocal fundamental frequency
MDVP	Flo(Hz) - Minimum vocal fundamental frequency
MDVP:Jitter(%), MDVP: Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP	Several measures of variation in fundamental frequency
MDVP:Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP:APQ, Shimmer:DDA	Several measures of variation in amplitude
NHR,HNR	Two measures of ratio of noise to tonal components in the voice
status	Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE, D2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent
spread1,spread2, PPE	Three nonlinear measures of fundamental frequency variation

six speech sessions. According to biomedical voice measurements, the 31 subjects are classified as healthy or having PD. The status variable was set to healthy (0), and the PD variable to 1. We use this variable as our target or dependent value.

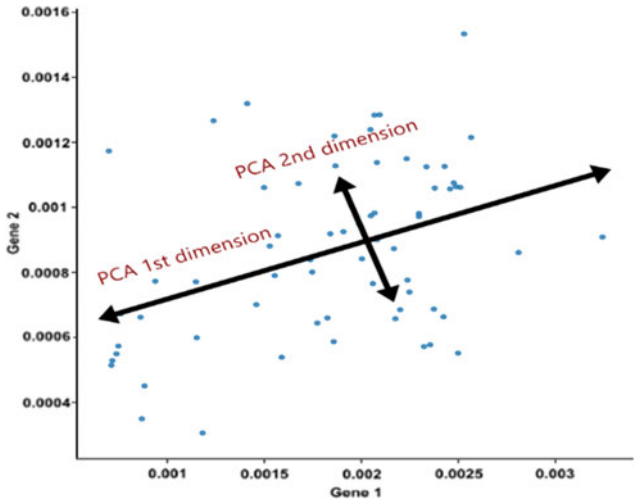
2.2. Feature reduction techniques

2.2.1. Principal component analysis (PCA)

The PCA technique can be used to reduce data and denoise data by preserving the most relevant information while reducing the size of large datasets. By using PCA, we aimed to select the two principal components with the highest variances. The updated variables can be obtained by deleting components with less information. The selection of principal components can be explained in Figure 2. On PCA1, the largest variances are spread out, and on PCA2, they are spread out perpendicular to PCA1 and have the second-highest variances. To put it all together, we examined the absolute values of the eigenvector components corresponding to the k greatest eigenvalues.

PCA is particularly useful for data with high dimensions and highly associated variables and features. Before every variable can contribute to uniform output, it is important to standardize the input variable range. There will be a bias in the results if a variable with a range of 0 to 100 predominates over a variable with a range of 0 to 1. It is therefore necessary to rescale the input variables to keep them on the same scale. The range of continuous input variables must first be standardized to ensure

Figure 2
Selection of principal components



every variable contributes to providing the output uniformly. A variable with a range of 0 to 100 will predominate over a variable with a range of 0 to 1, producing a biased result. Thus, the input variables must be rescaled to be on the same scale such that they are in the same range. Mathematically, it was written as:

$$Z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}, \tag{1}$$

after standardization, all input variables will have the same range.

Correlations are found by computing the covariance matrix. When the covariance is positive, it indicates that the two variables are increasing or decreasing in the same direction, meaning they are correlated. A negative covariance means that the variables are changing in opposite directions. There can be an increase in one variable, while a decrease in the other. To identify the primary components of the input data, we compute the eigenvalues and eigenvectors of the covariance matrix. Information is conveyed by eigenvectors. An eigenvalue is used to express the variance of a principal component. The proportion of information for each primary component can be determined by dividing each eigenvalue by the sum of all eigenvalues.

As a result, we have been able to determine what the differences are between the input datasets. Datasets with highly correlated variables tend to contain more redundant information. An $M \times M$ matrix which is symmetric called the 3×3 covariance matrix contains covariance values linked to the potential pairs:

$$\text{COV}_{\text{Matrix}} = \begin{bmatrix} \text{cov}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\ \text{cov}(y,x) & \text{cov}(y,y) & \text{cov}(y,z) \\ \text{cov}(z,x) & \text{cov}(z,y) & \text{cov}(z,z) \end{bmatrix} \tag{2}$$

signs play an imperative role here. As part of the data retention process, Eigenvectors were constructed as a matrix consisting of eigenvalues but arranged in decreasing order.

2.2.2. Synthetic minority oversampling technique (SMOTE)

The process of making prediction models for datasets with significant class imbalances is known as imbalanced classification. Although performance in the minority class is frequently the most critical in unbalanced datasets, most ML approaches ignore it. Based on the binary classification (healthy: 0 and PD: 1), the dataset contains information on 195 sustained vowel phonations.

The binary values 0 and 1 correspond to 48 (no PD) and 147 (PD) individual records, respectively. The minority class can be oversampled to handle unbalanced datasets, but this will lead to subpar results. A type of data augmentation approach is to create synthesized examples by synthesizing previous ones. Oversampling is used in the SMOTE algorithm to rebalance the original training set (Blagus & Lusa 2013; Fernández et al., 2018). Rather than simply replicating minority class instances, SMOTE introduces synthetic examples. By interpolating between several minority class instances within a defined neighborhood, this new data is generated. The instructions to execute SMOTE methods presented in algorithm 1 are described as:

Algorithm 1: SMOTE algorithm

1. Input: minority class examples i.e., T; N; k.
   a. Output: synthetic minority class samples i.e., (N/100) *T
   b. Variables: Sample array of minority class
2. *If* N < 100
   a. *then* ensure that T minorities are randomly selected
   b. T= (N/100) *T5
   c. N = 100
3. Endif multiples of 100 are assumed to represent SMOTE
   a. N= (int)N/100
4. *For* i = 1 to T do Create an nn-array with the indices of k nearest neighbors for i
5. POPULATE(N,i,nnarray)
6. End for
7. End Function

2.3. ML modeling

Before model training, the given data is decomposed into an 80:20 ratio where 80% of the data is used for training (156 samples) and the remaining 20% (39 samples) for testing purposes. The training dataset is used to train the ML model, and the test dataset is used to predict cases with and without PD using the trained model. Three supervised classification models were adopted to classify HC and PD categories. A brief explanation of each model is further explained.

Support vector machine (SVM): SVMs have supervised learning models that are used for a variety of purposes, including classification, outlier detection, and regression (Cervantes et al., 2020). Due to the way the decision boundary is selected, SVM differs from other classification methods. In N-dimensional space, the decision boundary (hyperplane) distinguishes the data.

K-nearest neighbor (KNN): Based on proximity to the clustering of individual data points, the KNN model makes classifications or predictions about the clustering of individual data points (Triguero et al., 2018). KNN predicts based on the number of nearest neighbors k. During the implementation of KNN, Euclidian distance is the most commonly used proximity metric. It is present as given below:

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{N} (p_i - q_i)^2} \tag{3}$$

where $d(p, q)$ represents a point in a plane.

Random forest (RF): In classification and regression problems, RF is another supervised learning algorithm (Schonlau, Journal, and 2020 2020). There is "n" randomly chosen bushes in a forest with "k" randomly selected bushes. A decision tree will be built for each bush, which will generate an output. Using majority voting or averaging, the final output will be used for classification, which enhances prediction accuracy.

2.4. Model validation

We have applied cross-validation (CV) for model validation (Pal & Patel, 2020). The CV is a statistical model that has been used to estimate the model performance and is commonly applied to ML for any problem regarding predictions. This is selected as it is simple to implement and understand and also produces low bias when compared with other models. This is also known as k-fold validation and k represents randomly spilled dataset portions (Pal & Patel, 2020; Vabalas et al., 2019). This work adopted a 10-fold CV technique where the model has initiated with one dataset, further randomly split into 10 portions. Therefore, 90% of the data was used for model training, and 10% was used for testing purposes. Models can be described in terms of the likelihood of error functions for each data point, and discrepancies between expected and actual results can be clarified.

2.5. Performance metrics

Performance metrics can be used to evaluate a model's ability. It is possible to compare model predictions with known values of dependent features in a dataset by using parameters such as accuracy, precision, recall, and receiver operating characteristic (ROC) curve. The prediction outcomes can be evaluated based on the confusion matrix that is presented in Table 2.

Table 2
Confusion matrix representation

	Prediction outcomes	
True class	Predicted with no PD	Predicted with PD
0: (no PD)	True Negatives (TN)	False positives (FP)
1: Patient with PD	False Negatives (FN)	True Positives (TP)

Accuracy: The model accuracy is defined as the ratio of the number of total true positive outcomes and total outcomes. Mathematically it is presented as:

$$\frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{4}$$

Precision: This is also known as the positive (true) prediction value and consider as a fraction of true positives over total positives. Mathematically, it is presented as:

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

Recall: It is also called sensitivity which is a fraction of true positives from positive class predictions. Mathematically, it is presented as:

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

F1 score: This parameter is defined as the harmonic mean of recall and precision and mathematically it was written as shown below:

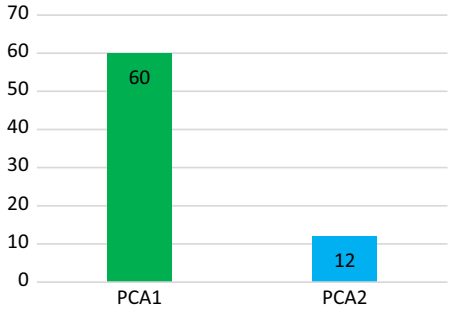
$$\text{F1 Score} = 2 \times \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7)$$

3. Results

Twenty-two derived features from voice recordings of both healthy and ill people are included in the PD dataset. Some of them are redundant or might not have value. Before the model training, standardization was followed by PCA to provide the best results. It is then necessary to balance the data after dimensionality reduction. In the training set, there are 115 records as PD and 41 as HC after the data has been divided into a training set and a testing set. Equalizing PD and HC records will balance the data, and SMOTE was used to accomplish this. As a result of SMOTE oversampling, the labeling of HC and PD was equal to 115.

PD datasets adopted here include 22 voice features, which means they have 22 dimensions. The more dimensional the data, the more redundant information there is, leading to biased results. By retaining as much data as possible, we should be able to identify the qualities that are highly connected among the 22 voice features. Since PCA1 and PCA2 contain the most information, two of the 22 traits were chosen for this study (Refer to Figure 3). The magnitude of the associated values in the eigenvectors indicates the importance of each feature. The maximum variance proof can also be obtained by calculating the covariance matrix of the smaller space. Seventy-two percent of the total information is represented by PCA1 (i.e., 60%) and PCA2 (i.e., 12%).

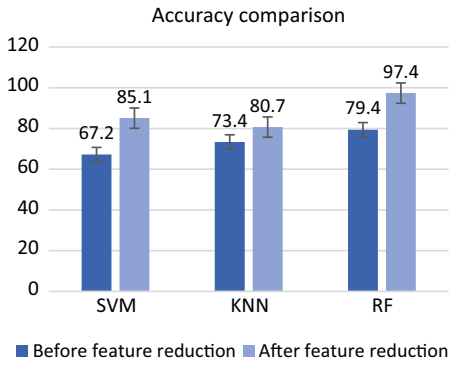
Figure 3
Information percentage in PCA1 and PCA2



Following feature reduction, the dataset was trained using k-fold CV (k = 10), using a different set of data for testing and training each time. For three different models, the confusion matrix for the test dataset can be visualized in Figure 4. My classification ML algorithms are plotted on a confusion matrix to help me see how they perform. A test dataset classification with 39 voice recordings is presented, including 32 with PD (1) and 7 without PD (0).

Model performance comparison was done for the test dataset for both before and after feature reduction techniques that were presented in Figure 5. It is evident that the RF model outperforms others in terms of accuracy. The highest PD classification accuracy was achieved at 97.4% for the RF model with the feature reduction approach followed by SVM (85.1%) and KNN (80.7%).

Figure 5
Model performance chart



Due to the distinct decision trees in the RF, the variance of the RF classifier as a whole is reduced. An RF classifier's final decision is an aggregation of the individual tree decisions, which allows it to generalize effectively. Without overfitting, the RF model is more accurate than others. Further, the RF model predicts 100% true positives, while SVM and KNN predict 0.92 and 0.84, respectively. In addition to the recall, F1 score, and area under the curve (AUC), RF also performed superiorly on other metrics. The performance metrics of the three models are presented in Table 3.

Table 3
Performance measurements

Algorithm	Precision	Recall	F1 score	AUC
SVM	0.92	0.88	0.86	0.91
KNN	0.84	0.91	0.92	0.93
RF	1.00	0.96	0.98	0.96

Figure 4
Confusion matrix outcome (A) SVM, (B) KNN, and (C) RF

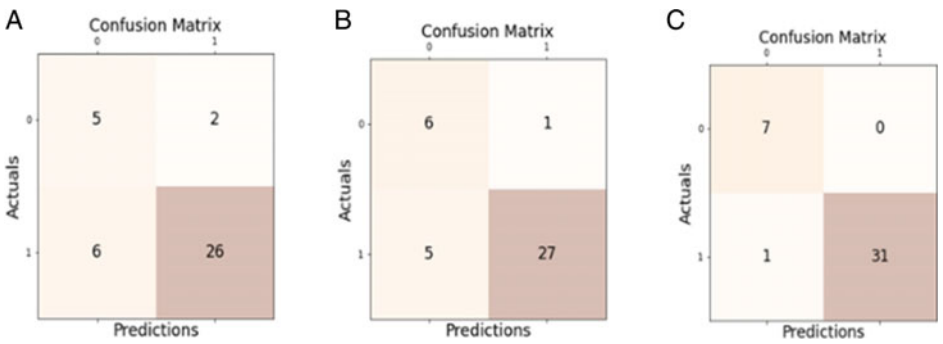
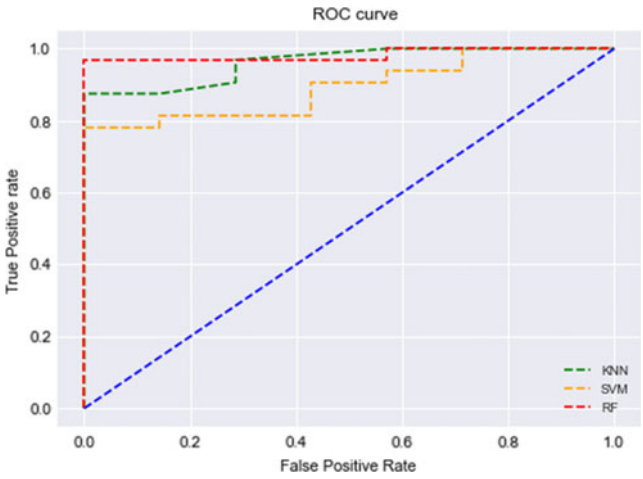


Figure 6
ROC curve outcome



Binary classification is largely based on ROC curves. The ROC curve in Figure 6 shows the RF value approaching near 1 when the ROC curve is visualized.

4. Discussion

People with PD have a dynamic sensory system problem that affects their development. A slight tremor may be the first PD symptom, and diagnosis of this disease should be done at an early condition providing the opportunity for disease treatment and getting adopted with a new lifestyle with positive outcomes. In this work, we proposed a combined approach of PCA and SMOTE. Auditory features have been used to test three supervised ML methods for PD detection. The SVM, KNN, and RF confusion matrices derived from the testing data are used to compare each classification model after feature selection with SMOTE sampling.

A proteomic biomarker, ideally one that accurately represents the disease process, should be available for investigation in diseased tissues, such as those impacted. There is an obstacle that prevents causal or disease-modifying medicines for PD from being developed. To achieve the most favorable clinical outcome, each PD case must be diagnosed and treated individually. With Ml techniques, it is possible to identify complex data patterns, automate data analysis, and make inferences and classifications based on individual patient data. This could be useful for precision medicine for PD.

Over the past few years, ML has become increasingly popular for diagnosing PD. It is reported that deep learning models like artificial neural networks help to do PD diagnosis with 94.4% of accuracy (Vabalas et al., 2019). It would be possible to screen large populations of patients for PD using such approaches on an affordable basis. Another study highlighted the gradient-boosting technique for PD diagnosis from voice samples and achieved an AUC of 0.951 (Karabayir et al., 2020). Other experiments showed that boosted decision trees, an ensemble model made from gradient-boosted regression trees, had the highest accuracy score on the data with 91–95%. By filter-based feature detection, it was also determined that spread1, spread2, and PPE are among the strongest nonlinear measures of fundamental frequency variation (Dinesh & He, 2017).

An analysis of the role of cascaded deep learning frameworks in detecting PD with voice biomarkers was explained, and a loss function curve was presented to demonstrate the relevance of good-fitting models for PD identification (Chintalapudi et al., 2022). Our study is in line with the work (Abdullah et al., 2020) where the authors introduced the feature selection method of Least Absolute Shrinkage and Selection Operator (LASSO) with multimodel ML approaches, including RF, SVM, deep neural networks (DNN), and gradient boosting. An accuracy of 96.67% was achieved for binary classification based on prediction over the UCI speech recording dataset. To improve data quality, other researchers have proposed data standardization, multicollinearity diagnosis, and dimensionality reduction techniques (Pramanik & Sarker, 2021). Their model achieved 95.1% of accuracy after they applied KNN, SVM, RF, AdaBoost, and logistic regression.

However, the above-mentioned studies have not sufficiently discussed the quality of true PD classification done by given ML models. Our updated RF model is more comprehensive than the values of previous works, as we can classify 100% of true PD subjects. We also achieved 97.4% classification accuracy with our proposed model with a combination of PCA and SMOTE algorithms. K-fold validation was used to test the model's performance and achieve maximum classification accuracy depending on the AUC parameter. Voice biomarkers are both cheap and easy to obtain compared to MRIs and motion-based diagnosis methods. Based on the findings of these studies, the maximum classification accuracy for PD subjects was achieved by RF with SMOTE oversampling. The models we propose can be implemented in neurological studies by using biomarker datasets.

The application of ML algorithms to PD diagnosis and treatment has the potential to make a significant difference in the future. Our ability to diagnose and manage this disease will become more sophisticated with the growing availability of data and the advancement of algorithms. In the future, by identifying new biomarkers for PD, ML algorithms could help diagnose and track the disease. By analyzing data from wearable devices, an algorithm could identify changes in movement patterns or physiological changes related to the disease.

5. Conclusion

The paper proposed an ML framework for PD detection at its early stages. Three models namely KNN, SVM, and RF were meticulously analyzed for performance and accuracy. In terms of model performance, RF outperforms the rest of the models with 98% of accuracy. This investigation reported an increase in recall value from 0.9 to 0.989 at an early stage. From speech recordings of healthy and PD patients, this model aims to identify the real positives. With the suggested model, recall increased from 0.95 to 0.989. The availability of a large amount of clinical data can further enhance the accuracy of the proposed model, which can assist future PD researchers. It is straightforward to collect samples using this model, since it is noninvasive. PD prediction by periodic voice recordings reduces the patient's hospitalizations.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Abdullah, A. A., Norazman, N. N., Ahmad, W. K. W., Awang, S. A., & Jian, F. W. (2020). Detection of Parkinson's disease (PD) based on speech recordings using machine learning techniques. In *International Conference on Innovation and Intelligence for Informatics, Computing and Technologies*, 1–6.

Aich, S., Kim, H. C., Hui, K. L., Al-Absi, A. A., & Sain, M. (2019). A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of Parkinson's disease. In *21st International Conference on Advanced Communication Technology,* 1116–1121.

Al-Fatlawi, A. H., Jabardi, M. H., & Ling, S. H. (2016). Efficient diagnosis system for Parkinson's disease using deep belief network. In *2016 IEEE Congress on evolutionary computation,* 1324–1330.

Ali, L., Khan, S. U., Arshad, M., Ali, S., & Anwar, M. (2019). A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease. In *2019 International Conference on Electrical, Communication, and Computer Engineering,* 1–5.

Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K. M., Dorsey, E. R., & Little, M. A. (2015). Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. *Parkinsonism & Related Disorders*, *21*(6), 650–653.

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*, 1–16.

Bloem, B. R., Okun, M. S., & Klein, C. (2021). Parkinson's disease. *Lancet*, *397*(10291), 2284–2303. https://doi.org/10.1016/S0140-6736(21)00218-X.

Boutet, A., Madhavan, R., Elias, G. J., Joel, S. E., Gramer, R., Ranjan, M., . . . , & Lozano, A. M. (2021). Predicting optimal deep brain stimulation parameters for Parkinson's disease using functional MRI and machine learning. *Nature Communications*, *12*(1), 3043.

Brücke, T., & Brücke, C. (2021). Dopamine transporter (DAT) imaging in Parkinson's disease and related disorders. *Journal of Neural Transmission*, *58*(12), 1665–1672.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215.

Chia, S. J., Tan, E. K., & Chao, Y. X. (2020). Historical perspective: models of Parkinson's disease. *International Journal of Molecular Sciences*, *21*(7), 2464.

Chintalapudi, N., Battineni, G., Hossain, M. A., & Amenta, F. (2022). Cascaded deep learning frameworks in contribution to the detection of Parkinson's disease. *Bioengineering*, *9*(3), 116.

Dinesh, A., & He, J. (2017). Using machine learning to diagnose Parkinson's disease from voice recordings. In *2017 IEEE MIT Undergraduate Research Technology Conference,* 1–4.

Elsworth, J. D. (2020). Parkinson's disease treatment: Past, present, and future. *Journal of Neural Transmission*, *127*(5), 785–791.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., . . . , & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29.

Feigin, V. L., Stark, B. A., Johnson, C. O., Roth, G. A., Bisignano, C., Abady, G. G., . . . , & Hamidi, S. (2021). Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Neurology*, *20*(10), 795–820.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905.

Hayes, M. T. (2019). Parkinson's disease and parkinsonism. *The American Journal of Medicine*, *132*(7), 802–807.

Hemond, C. C., & Bakshi, R. (2018). Magnetic resonance imaging in multiple sclerosis. *Cold Spring Harbor Perspectives in Medicine*, *8*(5).

Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of Healthcare Engineering*, *2021*.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065).

Karabayir, I., Goldman, S. M., Pappu, S., & Akbilgic, O. (2020). Gradient boosting for Parkinson's disease diagnosis from voice recordings. *BMC Medical Informatics and Decision Making*, *20*(1), 1–7.

Ohana, O., Soffer, S., Zimlichman, E., & Klang, E. (2018). Overuse of CT and MRI in paediatric emergency departments. *The British Journal of Radiology*, *91*(1085).

Pal, K., & Patel, B. V. (2020). Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. In *2020 Fourth International Conference on Computing Methodologies and Communication,* 83–87.

Papadopoulos, A., Kyritsis, K., Bostanjopoulou, S., Klingelhoefer, L., Chaudhuri, R. K., & Delopoulos, A. (2019). Multiple-instance learning for in-the-wild Parkinsonian tremor detection. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 6188–6191.

Pramanik, A., & Sarker, A. (2021). Parkinson's disease detection from voice and speech data using machine learning. In *Proceedings of International Joint Conference on Advances in Computational Intelligence*, 445–456.

Rao, D. V., Sucharitha, Y., Venkatesh, D., Mahamthy, K., & Yasin, S. M. (2022). Diagnosis of Parkinson's disease using principal component analysis and machine learning algorithms with vocal features. In *2022 International Conference on Sustainable Computing and Data Communication Systems,* 200–206.

Tolosa, E., Garrido, A., Scholz, S. W., & Poewe, W. (2021). Challenges in the diagnosis of Parkinson's disease. *The Lancet Neurology*, *20*(5), 385–397. https://doi.org/10.1016/S1474-4422(21)00030-2.

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS One*, *14*(11).

Wachter, S. D., Knowles, C. H., Elterman, D. S., Kennelly, M. J., Lehur, P. A., Matzel, K. E., . . . , & Van Kerrebroeck, P. E. (2020). New technologies and applications in sacral neuromodulation: An update. *Advances in Therapy*, *37*, 637–643.

Weber, W. (2020). Clinical PET/MR. *Molecular Imaging in Oncology*, *216*, 747–764.

Weintraub, D., Aarsland, D., Chaudhuri, K. R., Dobkin, R. D., Leentjens, A. F., Rodriguez-Violante, M., & Schrag, A. (2022). The neuropsychiatry of Parkinson's disease: Advances and challenges. *The Lancet Neurology*, *21*(1), 89–102. https://doi.org/10.1016/S1474-4422(21)00330-6.

How to Cite: Chintalapudi, N., Dhulipalla, V. R., Battineni, G., Rucco, C., & Amenta, F. (2023). Voice Biomarkers for Parkinson's Disease Prediction Using Machine Learning Models with Improved Feature Reduction Techniques. *Journal of Data Science and Intelligent Systems* 1(2), 92–98, https://doi.org/10.47852/bonviewJDSIS3202831