

Voice-Based Detection of Parkinson's Disease through Ensemble Machine Learning Approach: A Performance Study

Iqra Nissar^{1,*}, Danish Raza Rizvi¹, Sarfaraz Masood¹ and Aqib Nazir Mir¹

¹Department of Computer Engineering, Jamia Millia Islamia, New Delhi-110025

Abstract

INTRODUCTION: Parkinson's disease (PD) occurs due to the deficiency of dopamine that regulates various activities of the human body. Researchers have identified that voice is an underlying symptom of PD. Recently, Machine learning (ML) has helped in solving problems of computer vision, natural language processing, speech recognition etc.

OBJECTIVES: This paper aims to analyse the effect of feature type selection i.e. MFCC and TQWT on the efficiency of voice based PD detection system along with the use an ensemble learning based classifier for this task.

METHODS: Hence, in this work, various machine learning models, including Logistic Regression, Naive Bayes, KNN, Random Forest, Decision Tree, SVM, MLP, and XGBoost, have been employed and explored for PD detection purpose. The task of Feature selection was also done using minimum-Redundancy and Maximum-Relevance (mRMR) and Recursive Feature Elimination (RFE) techniques.

RESULTS: The results of the XGBoost with mRMR feature selection, outperformed all other models with a high accuracy of 95.39% and precision, recall and F1-score of 0.95 each, when both MFCC and TQWT features were selected.

CONCLUSION: The results obtained strongly support the use of XGBoost model for the voice sample based detection of PD along with mRMR feature selection technique.

Keywords: Parkinson's disease, speech signals, machine learning, classification, feature selection.

Received on 15 June 2019, accepted on 19 July 2019, published on 23 August 2019

Copyright © 2019 Iqra Nissar *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.13-7-2018.162806

1. Introduction

Parkinson's disease (PD) is a neuropathological disorder which deteriorates the motor functions of the human body [1]. It is the second most common neurological disease seen after Alzheimer's disease [2] and it is estimated that more than one million people are suffering from PD in North America alone [3]. In 1817, PD was termed as shaking palsy by Dr. James Parkinson [4]. Various studies have shown that

this number will rise in an ageing population as it is commonly seen in the people whose age is over 60 [5] [6].

Parkinson's disease is characterized by the degeneration of certain brain cell clusters that are responsible for producing the neurotransmitters that include dopamine, serotonin and acetylcholine. The loss of dopamine's result in the symptoms like anxiety, depression, weight loss and visual problems. The other symptoms that can be seen in the people with Parkinson's disease are poor balance, voice impairment and tremor [7] [8]. Various research studies have shown that 90% of people who suffer

*Corresponding author. Email: iqra.nissar13@gmail.com

from PD have speech and vocal problems [9] which include dysphonia, monotone and hypophonia [10] [11] [12]. Thus, the degradation of voice is considered to be as the initial symptom of Parkinson's disease [13].

The cause and cure of PD are yet unknown [14][15][16] but the availability of various drug therapies offers the significant mitigation of symptoms especially at its earlier stages, thus improving the life quality of patients and also reduces the estimated cost of the Pathology. The analysis of voice measurement is simple and non-invasive. Thus, to track the progression of PD the measurement of voice can be used [17] [18]. For assessing the progression of PD, various vocal tests have been devised which include sustained phonations and running speech texts [19]. The telemonitoring and teliagnosis systems have been widely used as these systems are based on speech signals which are economical and easy to use. Hence, in this paper, there is an attempt to explore a better machine learning based model for an early detection of PD from the voice samples of the subject.

The structure of this paper is as follows: the review of previously done studies on PD detection have been substantiated in section 2. Section 3 presents the proposed methodology employed for PD detection. Section 4 entails the analysis of classification results and discussion. The paper is concluded finally in section 5.

2. Literature Review

From time to time, several notable attempts were done by various researchers for detecting Parkinson's disease. The following is a brief review of some work done for detecting Parkinson disease from voice samples of subjects.

Max A. little et al [15] suggested a novel technique for the classification of subjects into Parkinson diseased and control subjects by detecting dysphonia. In their work, pitch period entropy (PPE) a new robust measure of dysphonia was introduced. The data was collected from 31 people (23 were PD patients and 8 were healthy subjects) which comprised of 195 sustained vowel phonations. Their methodology consisted of three stages; feature calculation, preprocessing and selection of features and finally the classification. For the classification purpose, they used linear kernel support vector machine (SVM). Their proposed model achieved an accuracy of 91.4%.

To separate the healthy subjects from PD subjects, Ipsita Bhattacharya et al [20] used a tool for data mining known as weka. They used SVM, a supervised machine learning algorithm for the classification purpose. Prior to classification, the data preprocessing was done on the dataset. Different kernel values were used to get the best possible accuracy by applying libSVM. The linear kernel SVM produced the best accuracy of 65.2174%, whereas the RBF kernel and polykernel SVM achieved the accuracy of 60. 8696%.

In another work, B.E Sakar et al [12] suggested a model for differentiating the control subjects from the PD

subjects. In their study, the data was collected from 40 subjects (20 were healthy subjects and 20 were PD subjects). From each subject, 26 voice samples were recorded which include short sentences, words, numbers and sustained vowels. For classification, they used SVM and k-nearest neighbor (k-NN). For cross-validation, they used Summarized Leave-One-Out (s-LOO) and Leave-One-Subject-Out (LOSO). The value of 1, 3, 5 and 7 was chosen for k-NN and for SVM, linear and RBF kernel were used. An accuracy of 82.50% was achieved by k-NN and an accuracy of 85% was reported on using SVM classifier.

Achraf Benba et al [21] aimed to separate the people with PD from the control subjects. In their work, the data comprised of 34 sustained vowels, which was collected from 34 people of which 17 were PD subjects. From each subject, 1 to 20 Mel-frequency cepstral coefficients (MFCC) were obtained. SVM with different kernel types was used for classification. LOSO was used as a cross-validation technique. The best accuracy of 91.17% was reported by linear kernel SVM on taking the top 12 MFCC coefficients.

For PD detection, the different speech signal processing algorithms were compared by C.O Sakar et al [22]. In their work, a new feature was introduced called as tunable Q-factor wavelet transform (TQWT). The effectiveness of TQWT outperformed the state-of-the-art speech signal processing methods that were used for the extraction of features in PD detection. On different feature subsets, different classifiers were used and using the ensemble techniques the prediction of the classifiers were combined. It was found that MFCCs and TQWT achieved the highest accuracies and thus are considered as important features in the problem of PD classification. Also, the minimum redundancy- maximum relevance feature selection technique was used as a data preprocessing step. The highest accuracy of 86% was reported by RBF kernel SVM on all feature subsets.

Richa Mathur et al [23] suggested a method for predicting the PD. They used a weka tool for implementing the algorithms to perform preprocessing of data, classification and the result analysis on the given dataset. They used k-NN along with Adaboost.M1, bagging, and MLP. It was observed that k-NN + Adaboost.M1 yielded the best classification accuracy of 91.28%.

A.Yasar et al [24] used artificial neural networks for the detection of Parkinson's disease. The dataset was taken from UCI machine learning repository. Using the MATLAB tool, 45 properties were chosen as input values and one output for the classification. Their proposed model was able to distinguish the healthy subjects from the PD subjects with an accuracy of 94.93%.

From the review above, it may be observed that various ML techniques have been applied in recent research works over voice based PD detection. But it may be observed that in none of these works, the ensemble based ML approaches like the XGBoost were used for model construction, which now have been used in this work. This paper also uses advanced feature selection techniques including the mRMR and RFE which help in getting rid of some lesser wanted features in the samples resulting in an overall efficient

model. The success of proposed machine learning model was also evaluated using various performance metrics like accuracy, precision, recall, sensitivity and specificity. These results were also compared with results obtained from various other ML models which were used in the recently reviewed works to establish the model's efficiency.

3. Methodology

The methodology for building a model to detect the Parkinson's disease at its early stage using the machine learning algorithms is presented in figure 1. It consists of the following steps:

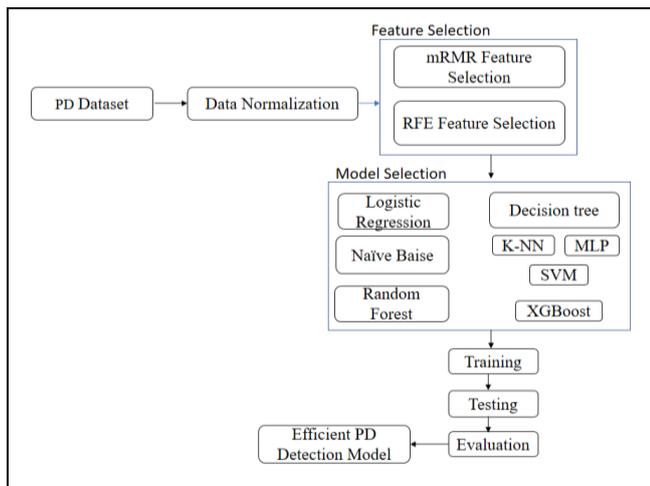


Figure 1. Overview of the proposed framework

The details of the components of the proposed framework have been explained below individually.

3.1. PD Dataset

The initial step towards the classification is the collection of data. For the voice analysis, the data is collected from the UCI, a machine learning repository which contains the voice data for both PD and healthy subjects [22]. The dataset used consists of 756 instances and 754 attributes. It was gathered from 188 PD patients in which 107 were men, 81 were women and the control individuals in which 41 were women and 23 were men. By performing the three repetitions of sustained phonation, the data was collected.

3.2. Data Pre-processing

This step is a combination of two individual processes, namely Data normalization and feature reduction or selection process which have been explained below.

Data Normalization

Data normalization is a data preparation technique that is often applied to datasets while working with most of the machine learning algorithms. It changes the numeric values of columns without losing any information. It is required so

as to re-scale the values of a particular feature in a specific range. In this work, the feature values in the selected dataset were normalized using the Min-Max scaling method between the range (0, 1) as these were of varying ranges. This normalization may be described as follows

$$\text{MinMax}(X) = \frac{x_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

where X is a particular feature represented by a column in the dataset, x_i is a value of this column where i is the number of elements in the column. The minimum value of the column is represented as X_{\min} and the maximum value of the column is X_{\max} .

Feature Selection

In our proposed work, after the normalization of attributes is done, two feature selection techniques namely RFE and mRMR are implemented. The feature selection mRMR [25] ranks the features according to the redundancy with other features and relevance with the class label. The RFE [26] [27] [28] as the name suggests, it recursively removes the features and build the model with the remaining attributes and assesses the model performance. The selected features were trained on different algorithms that result in increasing the efficiency of our proposed model.

3.3. Performance Metrics for Model Evaluation

After the feature selection, the model is implemented and output is produced in the form of probability or a class. The next step is to find out how efficient the model is using test dataset based on some metrics. In our work, to assess the classification performance different metrics like accuracy, recall, precision, F-1 score, and AUC-ROC curve have been used. Choosing the correct metrics to evaluate the machine learning model is very important as it influences how the performance is measured and compared.

Confusion Matrix

A confusion matrix is the most intuitive matrix that is used to find the accuracy and correctness of the models. It is used for the binary class and multiclass classification problems. It describes the performance of classification models in which the truth values are already known. A confusion matrix is a table with two dimensions, one for the actual target value and one for the predicted value. To explain the concept of the confusion matrix, assume the binary classification problem in which classes are 1 and 0 which is shown in figure 2. The actual labels are represented by rows and the predicted labels are represented by columns. The basic terms of confusion Matrix are discussed as follows:

- **True Positive (TP):** It can be defined as a systems ability to correctly classify the instances as positive, which means that if the actual label is 1, then predicted label is also 1. In terms of percentage, it is expressed as the True Positive Rate (TPR), also called sensitivity is

the proportion of correctly predicted positive examples. It is given by:

$$TPR \text{ (sensitivity)} = \frac{TP}{TP+FN} \quad (2)$$

	Predicted label (1)	Predicted label (0)
Actual label (1)	True positive	False negative
Actual label (0)	False positive	True negative

Figure 2. Confusion Matrix for classification problems.

- **True Negative (TN):** It is also called as specificity and is defined as the ability of a system to correctly classify the examples as negative, which means that if the actual label is 0 then predicted label is also 0. It is represented as True Negative Rate (TNR) in terms of the fraction of correctly predicted negative samples by the model.

$$TNR \text{ (specificity)} = \frac{TN}{TN+FP} \quad (3)$$

- **False Positive (FP):** In this case, the model incorrectly classifies the instances as positive. That is, the model predicts the class label as 1 whose label was originally 0. False Positive Rate (FPR) is represented as the fraction of negative cases which are predicted as the positive ones and is given by:

$$FPR = \frac{FP}{TN+FP} \quad (4)$$

- **False Negative (FN):** It is the system's ability to incorrectly classify the examples as negative which means that for the actual label 1, the predicted label for a class is 0. False Negative Rate (FNR) is the fraction of positive samples that were predicted as negative instances and is given by:

$$FNR = \frac{FN}{TP+FN} \quad (5)$$

- **Precision:** It is defined as the ratio of true positive relevant instances to the total number of retrieved instances. It is given by:

$$\text{precision } (p) = \frac{TP}{TP+FP} \quad (6)$$

- **Recall:** It is also called as the sensitivity and is defined as the fraction of correct positive examples predicted to the total number of positive occurrences.

$$\text{recall } (r) = \frac{TP}{TP+FN} \quad (7)$$

- **F1-score:** Precision and recall are summarized into another metric which is called as F1- score. It represents a harmonic mean of recall and precision.

$$F1 \text{ - score} = \frac{2 \cdot r \cdot p}{r+p} \quad (8)$$

It may also be represented as:

$$F1 \text{ - score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (9)$$

- **Accuracy:** It is the fraction of the number of correct predicted examples to the total number of instances present in the dataset. It is given by:

$$\text{accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \quad (10)$$

For binary classification, it is expressed as:

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

AUC-ROC Curve

It is a probability curve and a performance measurement that is widely used for binary classification problems. This curve tells how much the classification model is able to distinguish one class from the other class. In the ROC curve, the y-axis represents the true positive rate and the x-axis represents the false positive rate. The value of AUC ranges from 0 to 1. The model which has AUC close to 1 has an excellent classification performance, whereas the model which has AUC close to 0 has the worst measure of separability, and when AUC of the model is 0.5, the model does not have separation capacity.

4. Results

We have evaluated the performance of nine machine learning based models including naive Bayes, k-nearest neighbor, logistic regression, multilayer perceptron, random forest, support vector machines (linear and RBF kernel), decision tree, and Extreme Gradient Boost (XGBoost) with RFE and mRMR feature selection techniques. Table 1 shows the precision, F1-score, recall and test accuracies obtained with all feature subsets except tunable Q-factor wavelet transform using RFE and mRMR feature selection techniques. It can be observed that the random forest and decision tree achieved the accuracy of 84.86% when using RFE as a feature selection technique. Among all the models, the XGBoost achieved the highest accuracy of 88.15% with the 0.88 as precision, 0.88 as recall and F1-score of 0.88. The lowest accuracy of 74.34% was achieved by the Naive Bayes classification model.

Table 1. All feature subsets except TQWT using RFE and mRMR.

Models	With RFE feature selection				With mRMR feature selection			
	Precision	Recall	F1-score	Accuracy (%)	Precision	Recall	F1-score	Accuracy (%)
Naive Bayes	0.73	0.74	0.74	74.34	0.75	0.70	0.73	69.73
Logistic regression	0.83	0.84	0.81	84.21	0.82	0.84	0.82	83.55
k-NN	0.81	0.83	0.81	82.89	0.80	0.82	0.81	81.57
Multilayer Perceptron	0.81	0.82	0.81	82.23	0.82	0.83	0.81	82.89
Random forest	0.86	0.85	0.85	84.86	0.82	0.83	0.82	82.89
SVM (linear)	0.80	0.82	0.80	82.23	0.83	0.83	0.83	82.89
SVM (RBF)	0.84	0.84	0.83	84.21	0.84	0.84	0.84	84.0
Decision Tree	0.84	0.85	0.84	84.86	0.86	0.86	0.86	85.52
XGBoost	0.88	0.88	0.88	88.15	0.88	0.89	0.88	88.82

When using the mRMR feature selection technique, the decision tree achieved the accuracy of 85.52% with a precision of 0.86, 0.86 as recall and 0.86 as F1-score. The best accuracy of 88.82% was achieved by the XGBoost classifier with the precision of 0.88, 0.89 as recall and F1-score of 0.88 respectively. The Naive Bayes classifier achieved the accuracy of 69.73% with a precision of 0.75, 0.70 as recall and F1-score of 0.73 respectively which is lowest than all other classifiers. In both of the feature selection techniques, it is observed that the XGBoost classifier outclassed all other classification models with reasonable accuracy.

Table 2 shows the performance analysis using RFE and mRMR feature selection techniques when taking all feature subsets into consideration except the MFCC features. It is evident that the decision tree achieved the test accuracy of 86.84% with a precision of 0.87, 0.87 as recall and 0.87 as F1-score respectively. The XGBoost produced the highest accuracy of 91.44% with a precision of 0.91, 0.91 as recall and F1-score of 0.91 respectively. By mRMR feature selection technique, the XGBoost produced the accuracy of 92.10% with a precision of 0.92, recall of 0.92 and F1-score of 0.92 respectively which is highest than all other classification models. The next highest accuracy was achieved by the decision tree classifier which was 85.52%

with a precision of 0.85, recall of 0.86 and F1-score of 0.84. For the multilayer perceptron, an accuracy of 80.92% was reported with a precision of 0.79, recall of 0.81 and F1-score of 0.80 respectively. While analyzing Table 1 and Table 2, it can be observed that the decision tree produced the accuracy of 86.84% on all feature subsets except MFCC using RFE feature selection technique. However, the performance of XGBoost classifier has significantly increased to almost 91.44%. By excluding the TQWT features, the classification accuracy of XGBoost was 88.82% and by excluding the MFCC features, the classifier produced an accuracy of 92.10% which clearly shows that the classifier performs better on taking the TQWT features into consideration which in turn means that TQWT plays a vital role in detection of Parkinson's disease.

The performance analysis of all feature subsets which includes MFCC and TQWT features using RFE and mRMR feature selection techniques is shown in Table 3. It can be observed that the classification accuracy of the decision tree is 87.50% with a 0.88 as precision, recall of 0.88, and F1-score of 0.88 respectively. The XGBoost classifier achieved the classification accuracy of 92.76% with 0.93 of precision, 0.93 of recall and 0.92 of F1-score respectively. These results are obtained when using the RFE feature selection technique.

Table 2. All feature subsets except MFCC using RFE and mRMR.

Models	With RFE feature selection				With mRMR feature selection			
	Precision	Recall	F1-score	Accuracy (%)	Precision	Recall	F1-score	Accuracy (%)
Naive Bayes	0.84	0.85	0.83	84.86	0.85	0.84	0.85	84.21
Logistic regression	0.83	0.84	0.81	84.21	0.84	0.84	0.81	83.55
k-NN	0.81	0.83	0.81	82.89	0.84	0.84	0.85	84.86
Multilayer Perceptron	0.82	0.82	0.81	82.23	0.79	0.81	0.80	80.92
Random forest	0.85	0.86	0.85	85.52	0.84	0.85	0.84	84.86
SVM (linear)	0.84	0.83	0.84	84.21	0.82	0.83	0.80	82.89
SVM (RBF)	0.81	0.82	0.82	83.55	0.82	0.81	0.82	82.89
Decision Tree	0.87	0.87	0.87	86.84	0.85	0.86	0.84	85.52
XGBoost	0.91	0.91	0.91	91.44	0.92	0.92	0.92	92.10

Table 3. All feature subsets using RFE and mRMR

Models	With RFE feature selection				With mRMR feature selection			
	Precision	Recall	F1-score	Accuracy (%)	Precision	Recall	F1-score	Accuracy (%)
Naive Bayes	0.83	0.86	0.83	85.52	0.89	0.84	0.85	84.21
Logistic regression	0.84	0.84	0.85	84.86	0.85	0.86	0.85	86.18
k-NN	0.82	0.84	0.83	83.55	0.83	0.86	0.83	85.53
Multilayer Perceptron	0.83	0.84	0.81	84.21	0.85	0.84	0.85	84.86
Random forest	0.87	0.87	0.86	86.84	0.86	0.87	0.86	86.84
SVM (linear)	0.83	0.84	0.83	84.21	0.83	0.84	0.83	84.21
SVM (RBF)	0.86	0.86	0.86	86.18	0.89	0.88	0.89	88.15
Decision Tree	0.88	0.88	0.88	87.50	0.89	0.89	0.89	88.81
XGBoost	0.93	0.93	0.92	92.76	0.95	0.95	0.95	95.39

By the analysis of Table 3, it is evident that the performance of XGBoost classifier has significantly increased from 92.76% to 95.39% which means that mRMR feature selection technique has performed better in making a decision for detection. The classification performance of XGBoost is highest among all the classification models in terms of recall, accuracy, precision, and F1-score and the highest accuracy of 95.39% was reported on all feature subsets. This clearly shows that when TQWT and MFCC features are taken into consideration, the highest accuracies have been reported by the classification models which means that both MFCC and TQWT contribute most in the Parkinson’s disease detection problem.

Among all the classifiers applied on the dataset, the XGBoost based model produced significant results for the PD detection in voice samples. The ROC curve of the XGBoost using RFE on all feature subsets is shown in Figure 3 which clearly shows that its value is close to 1 which means it is able to perform the better classification. Figure 4 shows the performance evaluation of all feature subsets using RFE feature selection technique.

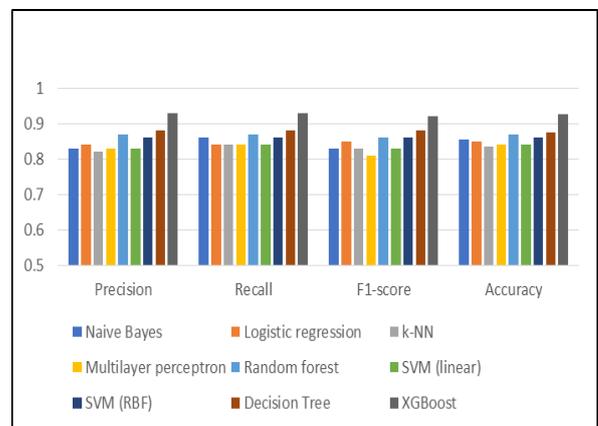


Figure 4. Performance evaluation of all feature subsets using RFE.

Also, the observations made on all feature subsets when using mRMR as the feature selection technique shows that a highest accuracy of 95.39% was produced by the XGBoost classifier with the 0.95 as precision, recall of 0.95, and F1-score of 0.95 respectively.

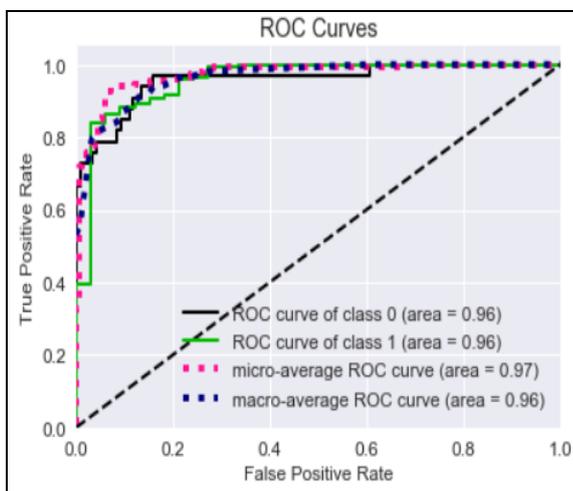


Figure 3. ROC curve of XGBoost using RFE on all feature subsets.

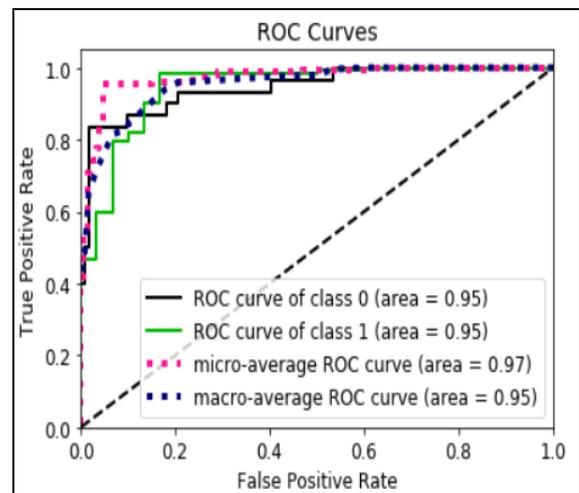


Figure 5. ROC curve of XGBoost using mRMR on all feature subsets.

The relative operating characteristic (ROC) curve of XGBoost using mRMR on all feature subsets is shown in Figure 5 above. The XGBoost has produced excellent classification performance as its area under curve (AUC) value is 0.95 which is nearer to 1. Figure 6 below, shows the performance of classifiers as obtained on all feature subsets using the mRMR feature selection technique.

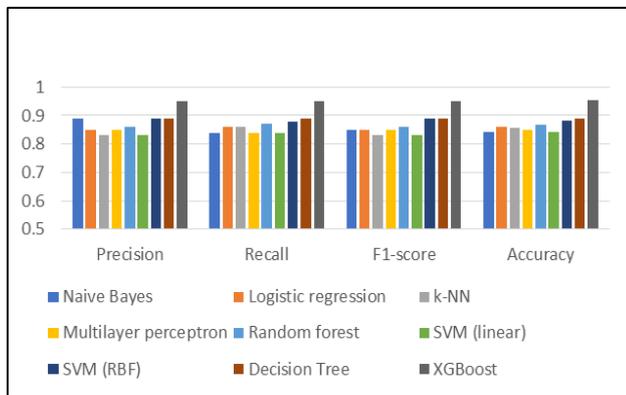


Figure 6. Performance evaluation of all feature subsets using mRMR.

All the models developed in this work (except XGBoost), i.e. SVM (linear), SVM (RBF), MLP and k-NN, were selected as they were developed in the reported works on this problem. As per the results of Table 3, when compared to these models, the proposed XGBoost based model outperforms them all. Even when compared to the actually reported accuracies of these works (see Table 4), the results of the (XGBoost + mRMR) model is significantly better, even when the size of the selected dataset is way bigger than what were used in previous works.

Table 4. Comparative Analysis of various models for Parkinson's disease detection.

Models	Proposed by	Accuracy (%)
SVM (RBF)	Little et al [15]	91.4
Linear SVM	Ipsita et al [20]	65.21
Linear SVM	B.E Sakar et al [12]	85.0
Linear SVM	Achraf Benba et al [21]	91.17
kNN+ Adaboost.M1	Richa Mathur et al [23]	91.28
ANN	A.Yasar et al [24]	94.93
SVM (RBF)	C.O. Sakar et al [22]	86.0
XGBoost	Proposed in this work	95.39

In comparison to most recent work done by Sakar et al. [22] on the very dataset which is also used in our work, the results are comparatively better with a significant improvement in PD detection accuracy of 95.39% against 86.0% achieved in [22]. One of the profound reasons behind the performance on the XGBoost technique in developing an efficient model for the problem, was that it constructs several decision trees and finally aggregates the

predictions made by each decision tree. Another reason which attributed to success of this technique was its regularization feature which greatly assisted in reducing data overfitting problem.

5. Conclusion

Currently, the Parkinson's disease research area is of much significance and its detection at the early stage can make the patient's life better. The recent developments in the methodologies through speech analysis have produced significant results. In our work, the problem of identification of Parkinson's disease is coped through a machine learning approach and different types of machine learning models have been employed for its detection. The main aim of this work is to show the PD diagnosis by analysing the voice signals. From many years, speech processing has an incredible potential in the detection of PD as voice measurements are non-invasive. This work is intended to ascertain and analyse the performance of various classification algorithms. The different classifiers were applied on a voice dataset and various evaluation metrics have been compared based on visualization and statistical analysis. Among all the classifiers, it was found that the XGBoost outperforms the other classifiers in machine learning algorithms. An accuracy of 92.76% was reported by using RFE feature selection technique while the accuracy of 95.39% was reported when using the mRMR feature selection technique on all feature subsets which is higher than all state-of-art methods. Based on the results, the followings may be recommended:

- (i) The Extreme Gradient Boost (XGBoost) technique should be used to develop model for PD detection problems.
- (ii) As the features initially available to the system can be numerous, hence it is highly advisable to apply some feature reduction / selection technique to reduce to the complexity of the detection system.
- (iii) The mRMR technique assisted in achieving better results in our case, hence is strongly recommended for feature selection task.

Though the model works efficiently, this is limited by the richness of the dataset with which it is being trained. The selected dataset, has only 756 instances, hence a dataset with more no of samples would help the model generalize better. The proposed model is thus a reliable model to detect Parkinson's disease due to its efficient precision, F1-score, recall, and accuracy rates.

References

- [1] Olanow, C. W., Stern, M. B., & Sethi, K. (2009). The scientific and clinical basis for the treatment of Parkinson disease (2009). *Neurology*, 72(21 Supplement 4), S1-S136.

- [2] Launer, L. J., Berger, K., Breteler, M. M., Dartigues, J. F., Baldereschi, M., Fratiglioni, L., & Hofman, A. (2000). Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group. Neurology*, 54(11 Suppl 5), S21-3.
- [3] Lang A. E., and Lozano A. M. (1998), "Parkinson's disease—First of two parts," *New. Engl. J. Med.*, vol. 339, pp. 1044–1053.
- [4] Langston, J. W. (2002). Parkinson's disease: current and future challenges. *Neurotoxicology*, 23(4-5), 443-450.
- [5] Van Den Eeden, S. K., Tanner, C. M., Bernstein, A. L., Fross, R. D., Leimpeter, A., Bloch, D. A., & Nelson, L. M. (2003). Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *American journal of epidemiology*, 157(11), 1015-1022.
- [6] Huse, D. M., Schulman, K., Orsini, L., Castelli-Haley, J., Kennedy, S., & Lenhart, G. (2005). Burden of illness in Parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, 20(11), 1449-1454.
- [7] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2009). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE transactions on Biomedical Engineering*, 57(4), 884-893.
- [8] Rigas, G., Tzallas, A. T., Tsiouras, M. G., Bougia, P., Tripoliti, E. E., Baga, D., & Konitsiotis, S. (2012). Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 16(3), 478-487.
- [9] Ho, A. K., Ianse, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural neurology*, 11(3), 131-137.
- [10] Sakar, C. O., & Kursun, O. (2010). Tlediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of medical systems*, 34(4), 591-599.
- [11] Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and hearing Disorders*, 43(1), 47-57.
- [12] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgun, F., Delil, S., ... & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4), 828-834.
- [13] Perez, K. S., Ramig, L. O., Smith, M. E., & Dromey, C. (1996). The Parkinson larynx: tremor and videostroboscopic findings. *Journal of Voice*, 10(4), 354-361.
- [14] Singh, N., Pillay, V., & Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. *Progress in neurobiology*, 81(1), 29-44.
- [15] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1010–1022, 2009.
- [16] Sapir, S., Spielman, J. L., Ramig, L. O., Story, B. H., & Fox, C. (2007). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research*.
- [17] Rahn III, D. A., Chou, M., Jiang, J. J., & Zhang, Y. (2007). Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice*, 21(1), 64-71.
- [18] Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.
- [19] Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., & Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-rhino-laryngology*, 258(2), 77-82.
- [20] Bhattacharya, I., & Bhatia, M. P. S. (2010, September). SVM classification to distinguish Parkinson disease patients. In *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India* (p. 14). ACM.
- [21] Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S. (2015, March). Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In *2015 International conference on electrical and information technologies (ICEIT)* (pp. 300-304). IEEE.
- [22] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., ... & Apaydin, H. (2019). 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*, 74, 255-263.
- [23] Mathur, R., Pathak, V., & Bandil, D. (2019). Parkinson Disease Prediction Using Machine Learning Algorithm. In *Emerging Trends in Expert Applications and Security* (pp. 357-363). Springer, Singapore.
- [24] Yasar, A., Saritas, I., Sahman, M. A., & Cinar, A. C. (2019, November). Classification of Parkinson disease data with artificial neural networks. In *IOP Conference Series: Materials Science and Engineering* (Vol. 675, No. 1, p. 012031). IOP Publishing.
- [25] Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (8), 1226-1238.
- [26] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3), 389-422.
- [27] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- [28] Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212, 353-363