 doi: 10.18502/htaa.v9i1.17863

Research Article

Efficient Feature Selection and Hyperparameter Tuning for Improved Speech Signal-Based Parkinson's Disease Diagnosis via Machine Learning Techniques

Deepak Painuli[1*], Suyash Bhardwaj[2], Utku Kose[3]

[1]COER University, Roorkee, India.
[2]Gurukula Kangri Vishwavidyalaya, Haridwar, India.
[3]Suleyman Demirel University, Isparta, Turkey.

*Corresponding Author: Deepak Painuli, COER University, Roorkee, India. Email: deepak.painuli@gmail.com

Received 2024 August 20; Accepted 2024 October 10.

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder that progressively worsens with age, particularly affecting the elderly. Symptoms of PD include visual hallucinations, depression, autonomic dysfunction, and motor difficulties. Conventional diagnostic methods often rely on subjective interpretations of movement, which can be subtle and challenging to assess accurately, potentially leading to misdiagnoses. However, recent studies indicate that over 90% of individuals with PD exhibit vocal abnormalities at the onset of the disease. Machine learning (ML) techniques have shown promise in addressing these diagnostic challenges due to their higher efficiency and reduced error rates in analyzing complex, high-dimensional datasets, particularly those derived from speech signals.

This study investigates 12 machine learning models—logistic regression (LR), support vector machine (SVM, linear/RBF), K-nearest neighbor (KNN), Naïve bayes (NB), decision tree (DT), random forest (RF), extra trees (ET), gradient boosting (GbBoost), extreme gradient boosting (XgBoost), adaboost, and multi-layer perceptron (MLP)—to develop a robust ML model capable of reliably identifying PD cases. The analysis utilized a PD voice dataset comprising 756 acoustic samples from 252 participants, including 188 individuals with PD and 64 healthy controls. The dataset included 130 male and 122 female subjects, with age ranges of 33 - 87 years and 41 - 82 years, respectively.

To enhance model performance, the GridSearchCV method was employed for hyperparameter tuning, alongside recursive feature elimination (RFE) and minimum redundancy maximum relevance (mRMR) feature selection techniques. Among the 12 ML models evaluated, the RF model with the RFE-generated feature subset (RFE-50) emerged as the top performer. It achieved an accuracy of 96.46%, a recall of 0.96, a precision of 0.97, an F1-score of 0.96, and an AUC score of 0.998, marking the highest performance metrics reported for this dataset in recent studies.

Keywords: Medical Diagnosis; Parkinson's Disease; Machine Learning; Data Preprocessing; Feature Selection; GridSearchCV

1. Background

Parkinson's disease (PD) is a neuropathological condition that impairs a human's ability to move (1). It is observed that approximately 10+ million people worldwide are currently afflicted by PD, making it the second-most prevalent neurological condition after Alzheimer's disease (2, 3). Regarding the prognostics, diagnosis, management, and treatment of the disease, the identification of PD-positive people is essential. Parkinson's disease is known to have a variety of early signs, including alterations in writing and speech (4, 5). Numerous studies have demonstrated that this figure will increase in an aging population because it is frequently observed in adults over 60 (6, 7). The PD is typically defined by the deterioration of specific brain cell clusters, which are accountable for generating neurotransmitters including dopamine, acetylcholine, and serotonin (8, 9). Dopamine deficiency causes symptoms like anxiety, depression, weight loss, and vision issues. Poor balance, speech impairment, and tremor are other notable signs of Parkinson's disease that can be found in patients (10-12).

Observations from several recent studies suggest that 90% of people with PD experience speech and vocal issues, including monotone, dysphonia, and hypophonia. Consequently, the deterioration of voice is perceived as the preliminary symptom of PD (13-16). Although there is no known cause or treatment for Parkinson's disease, the availability of numerous medications allows for significant symptom mitigation, particularly in the disease's early stages, which enhances patients' quality of life and lowers the pathology's projected cost (17-19). One of the most notable clinical indications that can help confirm the diagnosis and gauge the severity of Parkinson's disease is a change in the patient's voice. Voice measurement analysis is straightforward and non-intrusive. Consequently, the measurement of speech may be utilized to monitor the development of PD (20, 21).

To track the advancement of PD, a number of vocal tests have been created, such as protracted phonations and flowing speech texts. Since voice signals are affordable and simple to use, monitoring and diagnosis systems have been widely adopted (22, 23). Conventional diagnostic methods could be subject to subjectivity because they rely on the evaluation of motions, which are frequently subtle and hence difficult to describe, potentially leading to misdiagnosis (24, 25). Voice analysis-based studies might be categorized into four major aspect groups: Phonatory, articulatory, prosodic, and cognitive-linguistic. The majority of sustained vowels are used as acoustic material for phonatory investigations, which are related to the glottal source and resonant structures of the vocal tract (26, 27). Studies focused on articulatory aspects are more varied since there are more analytic options available. For example, the features or acoustic measurements analyzed can be taken from various kinds of sound segments. They can be connected to articulator speed or acceleration, the type of segment transitions, or the evolution of formants, among other things (28). The primary focus of prosodic studies is on paralinguistic elements such as pitch fluctuation, syllable rate analysis, and emotional expression in speech signals. Finally, the cognitive-linguistic techniques examine the vocabulary, sentence complexity, phrase construction, and the presence of word repeats, among other manifestations, to analyze abnormalities in cognitive behavior (29, 30).

This study falls under phonatory analysis; hence, it utilizes sustained vowels as acoustic materials due to the following reasons: Sustained vowels are anticipated to produce straightforward acoustic traces that could result in a consistent and trustworthy evaluation of voice quality to some extent. The higher efficiency and lower error rate of ML methods on complex and high-dimensional data problems make them a suitable choice for PD diagnosis tasks (31). As a result, this work makes an effort to first investigate a solely baseline traditional and simple ML-based model and later a fine-tuned ML model for early PD detection using the subject's voice samples.

Considering the advantages of ML-based methods over traditional diagnosis methods, the primary objective of this study was to first investigate solely baseline traditional and simple ML-based models and later fine-tune those ML models using the GridSearchCV method for early detection of PD using the subject's voice samples. Along with that, this study also made an effort to optimize the PD detection task using class imbalance control (CB) as well as the efficient mRMR and RFE-based feature selection (FS) methods.

2. Objectives

Considering the advantages of ML-based methods over traditional diagnostic approaches, the primary objective of this study was to initially evaluate baseline traditional and straightforward ML-based models, followed by fine-tuning these models using the GridSearchCV method for the early detection of PD through analysis of the subject's voice samples. Additionally, this study sought to optimize the PD detection process by employing class imbalance control (CB) and leveraging efficient feature selection methods, including minimum redundancy maximum relevance (mRMR) and recursive feature elimination (RFE).

3. Methods

Figure-1 illustrates the major steps involved in the methodology utilized in this study, which include data collection, data preprocessing, data sampling, pre-evaluation or model development, feature selection phases, followed by post-evaluation, model selection phases, and finally the PD classification phase (32, 33). Standard scaling and normalization methods were employed during the data preprocessing phase, while oversampling methods for data imbalance control were applied before the data sampling step. During the data sampling step, the most promising sampling method, known as "Stratified Random Sampling," was used. This method selects samples from the main dataset for different subsets (training or testing) randomly while preserving the original class ratio.

The pre-evaluation phase was followed by the feature selection step, which involved selecting the best 50, 10, and 5 features using two well-known feature selection techniques: RFE and mRMR. The subsequent phase included the post-evaluation of the best three classifiers on the selected feature subsets using default hyperparameters and tuned hyperparameters obtained through grid search optimization techniques to identify the best classifier for the final PD detection task. Once the best classifier was determined through the above experimentation, it was employed on a test set of the PD dataset in the final stage (34, 35).

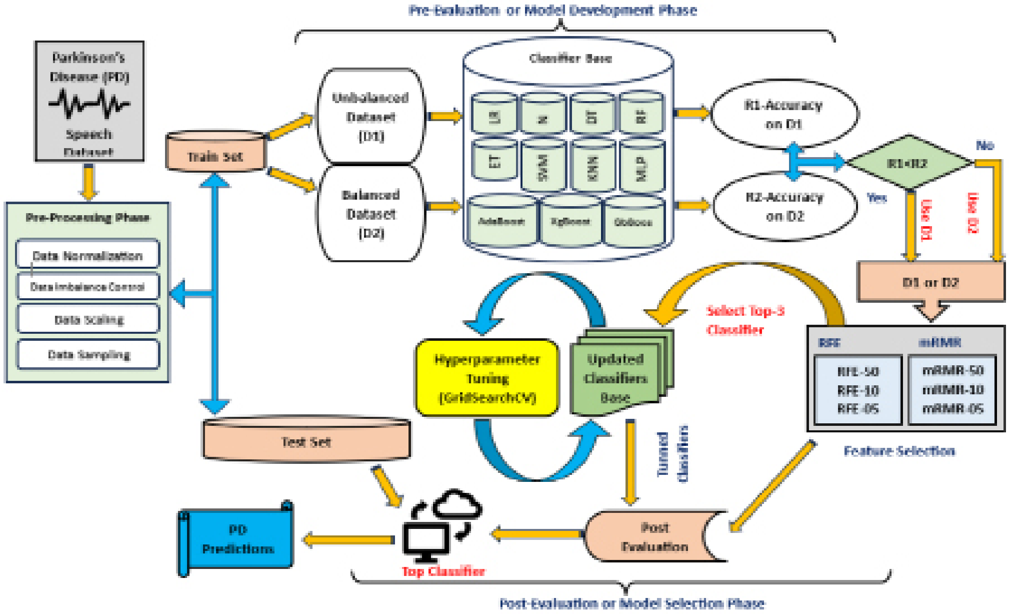


Figure 1. Process flow of proposed method

3.1. Parkinson's Disease Dataset

The gathering of data is the primary step in any classification process. Due to the heterogeneity observed in attributes/features across different publicly available PD speech datasets, this study utilized the current PD speech dataset from the UCI ML repository, which is available in the public domain to the scientific and academic community for experimentation purposes. The PD speech dataset includes voice analysis data for both healthy and PD participants.

As presented in Figure-2, the collected PD dataset comprises 756 samples from 252 subjects (188 PD and 64 healthy). A total of 130 male and 122 female participants, with age ranges of 33 - 87 years and 41 - 82 years, respectively, contributed to the above-mentioned PD dataset. Following a doctor's examination, the sustained phonation of the vowel "/a/" was recorded from each subject three times using a microphone preset to 44.10 kHz, as defined by the protocol designed for recording patient acoustics (36).

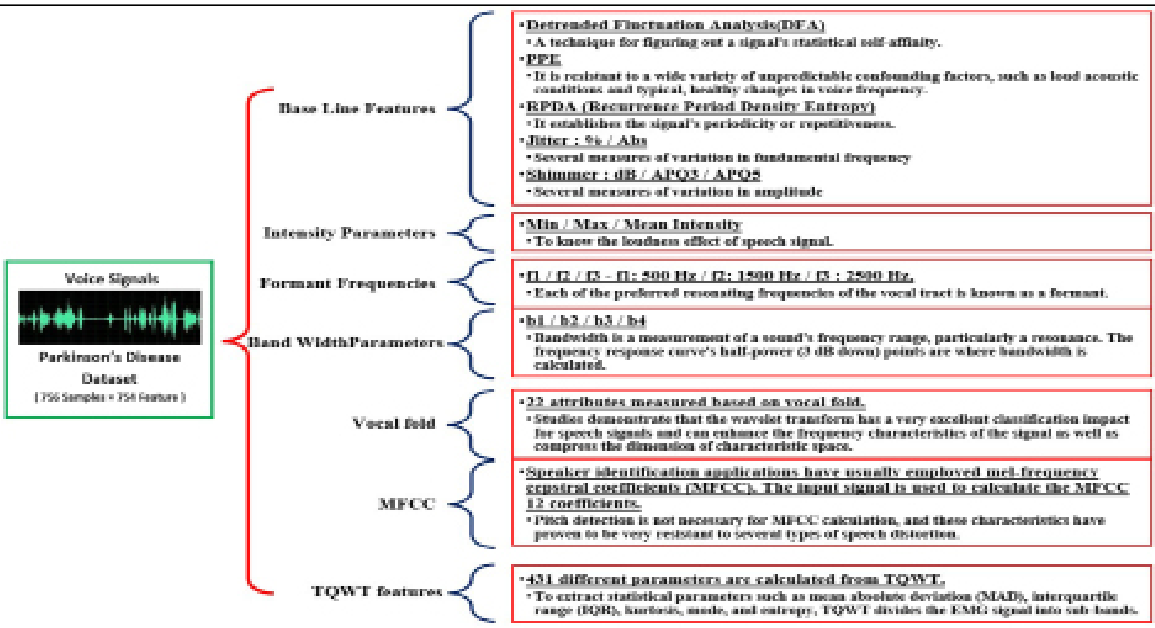


Figure 2. Brief overview of Parkinson's disease (PD) Dataset

The PD dataset contains a total of 754 features, including 753 independent features and 1 dependent feature. These features are categorized into groups such as Time Frequency Features, MFCCs, Vocal Fold Features, Wavelet Transform-Based Features, and TWQT Features. Brief information about the independent features is presented in Figure 2 below (37).

During data exploration, a class imbalance problem was identified in the PD dataset, which was addressed using the random oversampling method. Initially, the dataset consisted of 564 observations in the PD class ("1") and 192 observations in the healthy class ("0"), making it an imbalanced dataset. This imbalance could lead to biased performance by the employed ML model, as the majority class ("1") was three times larger than the minority class ("0").

To mitigate this issue, the minority class was oversampled through random selection to equalize the majority and minority class values (564 observations each), resulting in a balanced PD dataset. A glimpse of the pre- and post-oversampling scenarios is demonstrated in Figure 3.

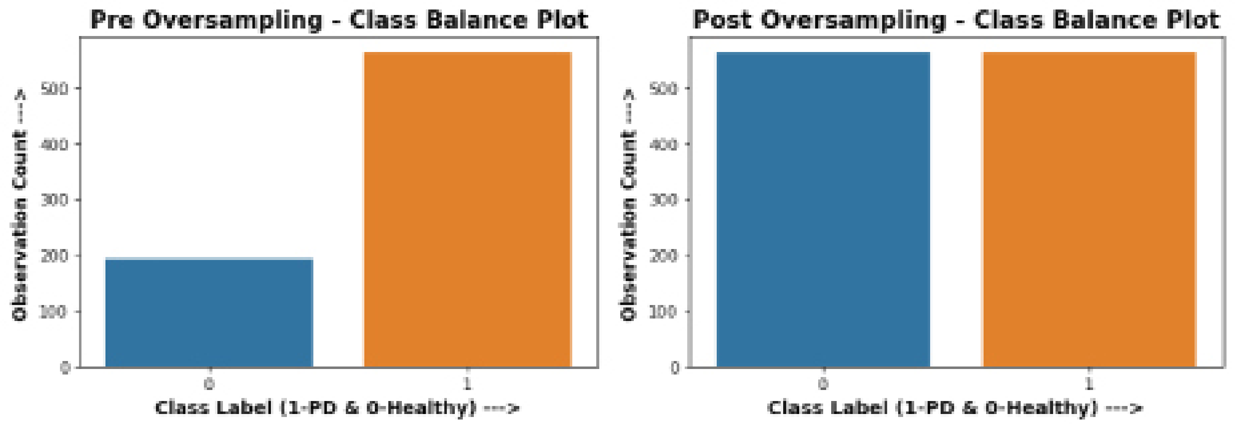


Figure 3. Parkinson's disease (PD) dataset's class balance - pre & post oversampling

3.2. Data Pre-processing

Data preprocessing in this study includes two processes: Data standardization and feature selection. The former is typically a data preparation method required to adjust the feature value scale to a certain range, thereby minimizing the computation required by the ML model when employed. The latter involves a dimensionality reduction process aimed at making the ML model less complex by reducing the number of features considered during model training (38-41).

3.2.1. Data Standardization

Machine learning models work well and converge faster when features are roughly the same size and/or close to being normally distributed. This study utilizes the Standard Scaler method to re-scale feature values to unit variance and standard deviation, without compromising the information present in the features (42). Thus, this leads to normally distributed features. The computation of the Standard Scaler may be described as follows:

$$Z = \frac{(\chi - \mu)}{\sigma}$$

Where:
μ = Training sample's mean; σ = Training sample's Standard deviation; χ = Training sample's value; $_z$ = Standard score / z-scores of training sample.

3.2.2. Feature Selection

As the PD dataset includes 754 features, making it a high-dimensional dataset, this may lead to the development of a complex ML model and could result in the overfitting issue. To avoid overfitting, only non-redundant and highly correlated features that contribute the most to target class prediction need to be included for model training purposes. Thus, following the data standardization step, this study utilizes two well-known FS methods: The RFE and mRMR. The RFE, a wrapper-type FS method, uses all independent features in the train set as an initial point and attempts to obtain a feature subset by recursively removing features one by one until the predetermined number of features are left (43). On the other hand, mRMR, an information theory-based FS method, ranks features based on how relevant they are to the class label and how redundant they are with other features. Computation time consumption is higher in the case of RFE, whereas mRMR is observed to be less time-consuming. Thus, mRMR is a suitable option for high-dimensional datasets, whereas RFE appears to be a better choice for low-dimensional datasets (44, 45).

After implementing the RFE and mRMR feature selection methods on the PD dataset, this study identified three feature subsets of 50, 10, and 5 most relevant features out of the total 754 features. The naming of RFE and mRMR generated feature subsets are RFE-50, RFE-10, RFE-5, and mRMR-50, mRMR-10, and mRMR-5, respectively. A brief description of individual feature subsets is presented in Table 1.

Feature Elimination and Minimum Redundancy Maximum Relevance Feature Selection Method

Table 1. List of Selected Features Subsets via Recursive

Variables	Values
Recursive Feature Elimination	
RFE-5	[“mean_MFCC_2nd_coef”/ “std_delta_delta_log_energy”/ “std_6th_delta_delta”/”tqwt-energy-dec_26”/ tqwt_entropy_log_dec_12”]
RFE-10	[“mean_MFCC_2nd_coef”/”std-delta-log_energy”/”std_delta_delta_log_energy”/”std_6th_delta_delta”/”std_7th_delta_delta”/”tqwt_energy-dec_26”/”tqwt_entropy-shannon-dec_12”/”tqwt_entropy_log_dec_12”/”tqwt-entropy-log-dec_33”/”tqwt-entropy-log_dec_35”]
RFE-50	[“DFA”/”numPeriodsPulses”/”meanPeriodPulses”/”minIntensity”/”mean-MFCC_2nd-coef”/mean-MFCC-5th-coef”/”mean_MFCC-6th-coef”/”std-MFCC_4th_coef”/”std_delta_log_energy”/”std_4th_delta”/”std_6th_delta”/”std_7th_delta”/”std_delta_delta_log_energy”/”std_4th_delta_delta”/”std_6th_delta_delta”/”std_7th_delta_delta”/”std_8th_delta_delta”/”std_9th-delta-delta”/”app-det-TKEO-mean_9_coef”/”tqwt-energy-dec-25”/”tqwt-energy-dec-26”/”tqwt-energy-dec-27”/”tqwt-entropy-shannon-dec-12”/”tqwt-entropy-shannon-dec-25”/”tqwt-entropy-log-dec-12”/”tqwt-entropy-log-dec-17”/”tqwt-entropy-log-dec-18”/”tqwt-entropy-log-dec-25”/”tqwt-entropy-log-dec-27”/”tqwt-entropy-log-dec-32”/”tqwt-entropy-log-dec-33”/”tqwt-entropy-log-dec-34”/”tqwt-entropy-log-dec-35”/”tqwt-TKEO-mean-dec-12”/”tqwt-TKEO-mean-dec-25”/”tqwt-TKEO-std-dec-11”/”tqwt-TKEO-std-dec-12”/”tqwt-TKEO-std-dec-13”/”tqwt-TKEO-std-dec-19”/”tqwt_TKEO_std_dec_20”/”tqwt_medianValue_dec_36”/”tqwt_meanValue_dec_36”/”tqwt_stdValue_dec_12”/”tqwt_stdValue_dec_33”/”tqwt_minValue_dec_17”/”tqwt_maxValue_dec_13”/”tqwt_maxValue_dec_17”/”tqwt_kurtosisValue_dec_17”/”tqwt_kurtosisValue_dec_18”/”tqwt_kurtosisValue_dec_36”]
Maximum Relevance-Minimum Redundancy	
mRMR-5	[“std_9th_delta_delta”/ “app_det_TKEO_mean_1_coef”/ “tqwt_medianValue_dec_25”/ “mean_MFCC_2nd_coef”/ “tqwt_maxValue_dec_5”]
mRMR-10	[“std_9th_delta_delta”/ “app_det_TKEO_mean_1_coef”/ “tqwt_medianValue_dec_25”/ “mean_MFCC_2nd_coef”/ “tqwt_maxValue_dec_5”/ “std_7th_delta_delta”/ “std_8th_delta_delta”/ “tqwt_entropy_log_dec_12”/ “std_6th_delta_delta”/ “tqwt_entropy_log_dec_26”]
mRMR-50	[“std_9th_delta_delta”/”app-det-TKEO-mean-1-coef”/”tqwt_medianValue_dec_25”/”mean-MFCC-2nd-coef”/”tqwt_maxValue_dec_5”/”std_7th_delta-delta”/”std_8th_delta-delta”/”tqwt_entropy_log_dec_12”/ “std_6th_delta_delta”/ “tqwt_entropy_log_dec_26”/ “tqwt_kurtosisValue_dec_36”/”std_delta_delta_log_energy”/ “tqwt_minValue_dec_12”/ “DFA”/ “std_8th_delta”/ “std_10th_delta_delta”/ “tqwt_maxValue_dec_12”/ “std_11th_delta_delta”/ “tqwt_maxValue_dec_11”/ “std_6th_delta”/ “tqwt_kurtosisValue_dec_27”/ “std_9th_delta”/ “tqwt_stdValue_dec_12”/”tqwt-kurtosisValue-dec-34”/ “std_7th_delta”/ “tqwt_stdValue_dec_11”/ “mean_2nd_delta”/ “tqwt_entropy_log_dec_27”/ “tqwt_entropy_log_dec_11”/ “std_10th_delta”/ “locPctJitter”/ “tqwt_minValue_dec_13”/ “tqwt_entropy_shannon_dec_34”/”std_11th_delta”/ “tqwt_entropy_log_dec_13”/ “tqwt_energy_dec_25”/ “f1”/”tqwt_minValue_dec_11”/”std_4th_delta-delta”/ “tqwt_kurtosisValue_dec_26”/ “tqwt_entropy_log_dec_34”/ “tqwt_maxValue_dec_13”/ “std_delta_log_energy”/ “numPeriodsPulses”/ “std_4th_delta”/ “tqwt_entropy_log_dec_16”/ “std_12th_delta_delta”/ “tqwt_stdValue_dec_6”/ “tqwt_energy_dec_26”/ “app_det_TKEO_mean_7_coef”]

Abbreviations: RFE, recursive feature elimination; mRMR, maximum relevance-minimum redundancy.

3.2.3. *Brief Introduction to Utilized Feature Selection Methods*

This section presents fundamental feature selection methods used in this study.

3.2.3.1. Recursive Feature Elimination

The RFE is a popular feature selection (FS) algorithm. Its fundamental goal is to find the subset of attributes that are most pertinent to a specific predictive modeling task. The RFE operates by repeatedly eliminating the dataset's least important features until the desired number of features is obtained. As a backward feature selection process, RFE starts with every feature and gradually eliminates them until the target number of features is obtained. The approach works well for high-dimensional datasets with plenty of features, where it might not be viable to examine every feature subset (46, 47).

3.2.3.2. Minimum Redundancy Maximum Relevance

The mRMR is also a well-known FS algorithm that aims to recognize a feature subset from the complete feature set, consisting of features that are both minimally redundant and highly informative. The algorithm evaluates the redundancy and relevance of each feature and selects a feature subset, including features that demonstrate symmetry between these two aspects (40, 47).

3.3. Model Training and Model Selection

During the model training phase, multiple ML models, as listed in the classifier base mentioned in figure-1, were trained and evaluated on different feature subsets obtained from the RFE and mRMR algorithms. A total of 12 ML classifiers, i.e., LR, SVM (Linear & RBF), NB, KNN, RF, DT, ET, GbBoost, XgBoost, AdaBoost, and Multi-Layer Perceptron (MLP), were utilized during the model training process of the pre-evaluation and model development phase of the proposed methodology. A dramatic decrease in ML model performance witnessed in previous studies due to the leave-one-out subject validation CV (36, 48-50) led the current study to utilize the most promising and well-known 10-fold CV (51-53) for validation of ML models employed throughout this study, ensuring better control of the overfitting issue of trained ML models. All 12 classifiers were employed and evaluated on both the unbalanced PD dataset (D1) and the balanced PD dataset (D2) with their default parameter values and the sampling parameter "random state" preset to the value 42 at this stage. Based on multiple performance metrics, the three best-performing ML models, i.e., RF, AdaBoost, and MLP, were identified for further stages of the current study.

3.3.1. Brief introduction to adopted ML Classifier

3.3.1.1. Random Forest

Random Forest, a well-known ML classifier, is an ensemble learning method used to combine different decision trees in order to increase accuracy and reduce overfitting. To enhance the ensemble's diversity and minimize overfitting, the approach utilizes feature selection and bootstrapping with aggregation (41, 51).

3.3.1.2. Adaptive Boosting (AdaBoost)

AdaBoost is an ML algorithm that combines multiple weak classifiers to generate a strong classifier. The algorithm adjusts the weights of the training set based on the misclassified samples; in other words, it gives higher priority to misclassified samples to increase the likelihood of correct classification by the classifier during the next sampling. The final classifier is a weighted combination of the weak classifiers (45, 54)

3.3.1.3. Multi-Layer Perceptron

The MLP is a type of neural network widely used in machine learning and deep learning. It is a feedforward neural network that can address various supervised learning problems. It operates by initializing weights and biases, propagating input forward through the network while calculating error, and then backpropagating the calculated error to adjust weights and biases. The weights and biases are updated iteratively until the anticipated accuracy is achieved (46, 55).

3.4. Model Optimization

Post-selection of the best ML classifier through the pre-evaluation step, the next stage involves finding the optimal values for different model parameters to achieve optimal prediction efficiency. This process is typically known as hyperparameter tuning or optimization. A hyperparameter is a configuration external to the ML model whose value cannot be inferred from the data. These parameters describe crucial model characteristics, such as complexity and learning rate, and are often used in procedures to assist with model parameter estimation. Searching for the optimal values of a model's hyperparameters can be performed manually or by using tree-based algorithms such as grid search or random search.

Manual searching can be a tedious task, as one model might have multiple hyperparameters with reasonably extensive value spaces (56). Therefore, ML practitioners often rely on automatic hyperparameter searching algorithms like grid search or random search methods. Both methods train and evaluate the model's performance on different combinations of hyperparameter values and return the combination for which the model demonstrates the best classification performance. This study utilized the grid search method for hyperparameter tuning, as it searches parameter values across the complete parameter combination space.

The grid search method was evaluated for various hyperparameter values of the three selected ML models—RF, AdaBoost, and MLP—identified during the pre-evaluation phase of the study (57). The mapped hyperparameter search space and the optimal hyperparameter values obtained for these three models using the grid search method are detailed in Table 2.

Table 2. List of Mapped Hyperparameter Search Space and Optimal Hyperparameters Values Using Grid Search Method

Model Names and Hyperparameter Name	Hyperparameter Search Space	Optimal Hyperparameter Value
RF		
"max_depth"	[3, 5, 7, 9, 11, 15, 20, 30, 50]	15
"n_estimators"	[10, 50, 100, 200, 300,500]	300

"Criterion"	['entropy', 'gini']	entropy
AdaBoost		
"learning_rate"	[0.05, 0.1, 0.2, 0.5, 1]	1
"n_estimators"	[10, 50, 100, 200, 300, 500]	300
"Algorithm"	['SAMME', 'SAMME.R']	SAMME.R
MLP		
"hidden_layer_sizes"	[(10,), (20,), (30,), (50,), (100,), (200,), (300,), (500,)]	500
"Activation"	['sigmoid', 'identity','tanh', 'relu']	relu
"Solver"	['lbfgs', 'sgd', 'adam']	sgd
"Alpha"	[0.0001, 0.05]	0.0001

[z] Abbreviations: RF, random forest; AdaBoost, adaptive boosting; MLP, multi-layer perceptron.

3.4.1. Grid Search Method

Grid Search is a popular hyperparameter optimization algorithm used to find the best hyperparameters or set of hyperparameters for an ML model. Hyperparameters are parameters that are specific to the applied ML model and are not learned during training. Grid Search is a straightforward yet efficient technique for optimizing hyperparameters; however, it can be computationally expensive, particularly when working with large datasets or numerous hyperparameters.

For each hyperparameter, a grid of potential values is defined, and all conceivable combinations of the hyperparameters in the grid are thoroughly searched. The algorithm trains a model on a portion of the data using each combination of hyperparameters and then assesses the model's performance on a validation set. This process is repeated for all potential grid hyperparameter combinations multiple times (31).

3.5. Performance Metrics

Following the model optimization stage, the ML model is implemented, and results are generated as a class or probability. The next stage involves using a test dataset and appropriate performance metrics to assess the model's effectiveness in predicting Parkinson's disease. Various metrics, including accuracy, recall, F-1 score, precision, and the AUC-ROC curve, were employed in this work to evaluate the classification performance. Figure-4 demonstrates the brief formulas of the different performance metrics used in the classification task.

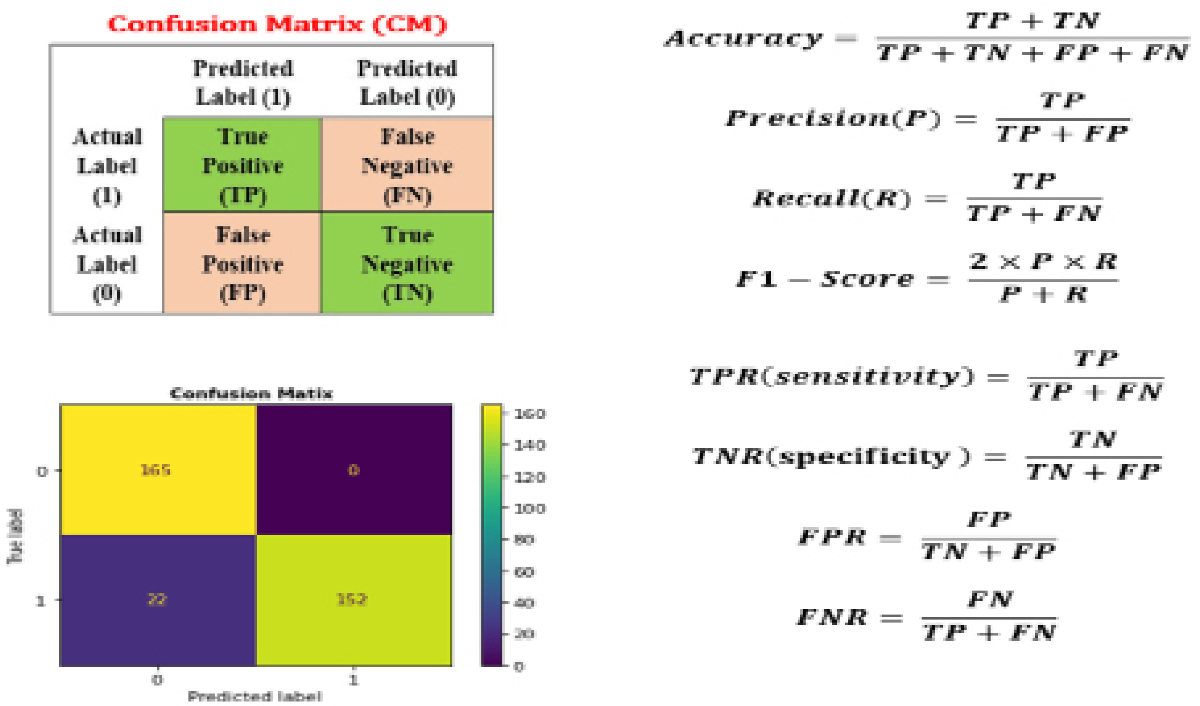


Figure 4. Confusion matrix for binary classification problems

4. Results

As mentioned in Section 3.1, the PD dataset was found to be imbalanced; therefore, the random oversampling technique was first employed to obtain a balanced PD dataset. After oversampling, this study pre-evaluated 12 different ML models mentioned in Table 3 on both datasets, D1 (Unbalanced PD dataset) and D2 (Balanced PD dataset), to observe performance enhancements of the employed ML models. At this stage, all models were employed with their default parameters only. The results of the pre-evaluation are presented in Table 3 and Figure 5.

Table 3. Pre-evaluation Performance of Different Machine Learning Models

ML Model Name	Accuracy		Precision		Recall		F1-Score		AUC	
Dataset	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2
LR	0.84	0.89	0.83	0.9	0.84	0.9	0.83	0.9	0.8601	0.9525
KNN	0.87	0.90	0.89	0.91	0.87	0.9	0.86	0.9	0.8911	0.9679
SVM-linear	0.83	0.88	0.82	0.89	0.83	0.88	0.82	0.88	0.7633	0.8861
SVM-RBF	0.84	0.91	0.86	0.91	0.84	0.91	0.81	0.91	0.7019	0.9086
Gaussian NB	0.76	0.76	0.78	0.77	0.76	0.76	0.77	0.76	0.773	0.7893
DT	0.82	0.92	0.81	0.92	0.81	0.91	0.81	0.91	0.7233	0.899
RF	0.86	0.94	0.86	0.95	0.85	0.96	0.85	0.94	0.9292	0.9955
ET	0.78	0.89	0.77	0.91	0.78	0.89	0.77	0.89	0.7435	0.9277
GbBoost	0.82	0.93	0.82	0.94	0.82	0.93	0.82	0.93	0.9151	0.9966
XgBoost	0.85	0.93	0.85	0.93	0.85	0.93	0.85	0.93	0.9254	0.9992
AdaBoost	0.84	0.94	0.84	0.94	0.85	0.94	0.84	0.94	0.8994	0.9869
MLP	0.86	0.93	0.85	0.94	0.86	0.94	0.85	0.94	0.8968	0.9838

Abbreviations: LR, Logistic regression; KNN, K-nearest neighbor; SVM, support vector machine; NB, naive bayes; DT, decision tree; RF, random forest; ET, extra tree; GbBoost, Gradient boost; XgBoost, Extreme gradient boost; AdaBoost, Adaptive boost; MLP, Multi-layer perceptron.

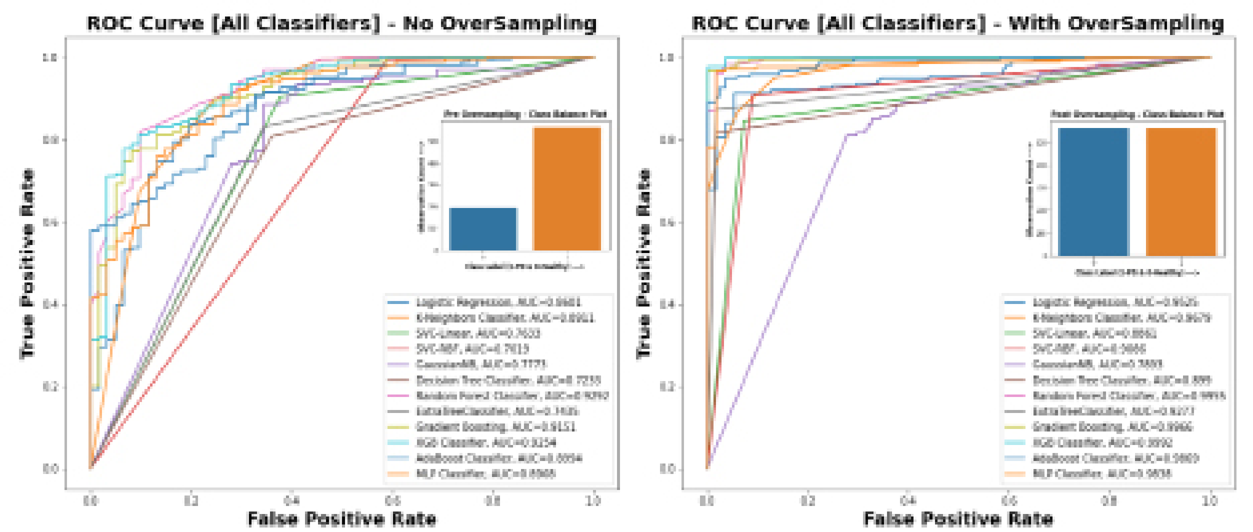


Figure 5. Pre-evaluation performance of different multi-layer (ML) models using AUC-ROC (AUROC) curve

Figure-5 represents the prediction performance of 12 ML models on both balanced (D2) and unbalanced (D1) datasets in terms of the AUC-ROC (AUROC) curve. The higher the area under the curve, the lower the misclassification rate of that ML model.

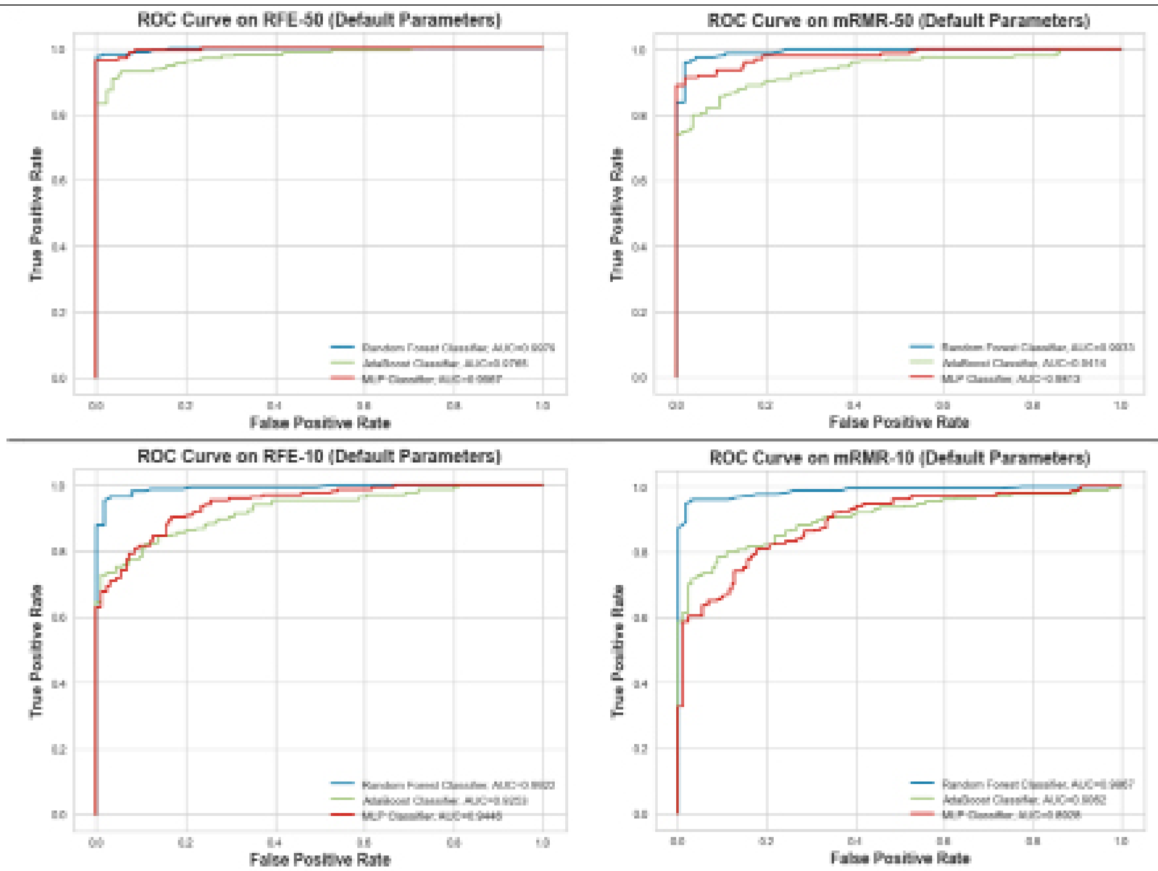
As demonstrated in Table 3 and Figure 5, a performance increase has been observed in almost all ML models in terms of various performance metrics, signifying the fact that ML models perform well when the dataset provided has balanced predictor classes. This ensures unbiased performance of the applied ML models by reducing misclassification counts and increasing correct classifications by the models. Based on the above performance, this study considered the D2 dataset for the forthcoming experimentation process.

Post pre-evaluation phase, this study identified the top three ML models, i.e., RF, AdaBoost, and MLP, for further phases of the current study. Concurrently with the pre-evaluation phase, the feature selection (FS) phase was implemented to identify three feature subsets, including the best 50, 10, and 5 features, using the RFE and mRMR FS methods. Both RFE [RFE-50, RFE-10, RFE-5] and mRMR [mRMR-50, mRMR-10, mRMR-5] feature subsets have been investigated during this study to identify the best FS approach using the selected ML models based on various performance metrics. The list of feature subsets generated by both FS methods is presented in Table 1 under Section 3.2.2. Experimentation results on the above-mentioned subsets by RF, AdaBoost, and MLP models with their default parameters are presented in Table 4 and Figure 6 below.

Table 4. Performance of Random Forest, AdaBoost and Multi-Layer Perceptron Model (Default Parameters) Models on Recursive Feature Elimination and Maximum Relevance-Minimum Redundancy Feature Subsets

ML Model Name	RFE	Accuracy	Precision	Recall	F1-Score	AUC	mRMR	Accuracy	Precision	Recall	F1-Score	AUC
RF	RFE-50	0.9469	0.95	0.95	0.95	0.9979	mRMR-50	0.9292	0.93	0.93	0.93	0.9933
AdaBoost		0.9498	0.95	0.95	0.95	0.9765		0.9204	0.92	0.92	0.92	0.9414
MLP		0.9734	0.97	0.97	0.97	0.9967		0.9292	0.94	0.93	0.93	0.9813
RF	RFE-10	0.9439	0.95	0.94	0.94	0.9922	mRMR-10	0.9233	0.93	0.92	0.92	0.9867
Adaptive Boost (AdaBoost		0.8702	0.89	0.87	0.87	0.9253		0.8555	0.86	0.86	0.86	0.9062
MLP		0.8851	0.89	0.88	0.88	0.9446		0.8348	0.84	0.83	0.83	0.8928
RF	RFE-5	0.8967	0.91	0.9	0.9	0.9785	mRMR-5	0.9026	0.91	0.9	0.9	0.9793
AdaBoost		0.8259	0.83	0.83	0.83	0.8901		0.8377	0.84	0.84	0.84	0.9017
MLP		0.80823	0.81	0.81	0.81	0.8815		0.7876	0.79	0.79	0.79	0.859

Abbreviations: RF, random forest; AdaBoost, adaptive boost; MLP, multi-layer perceptron.



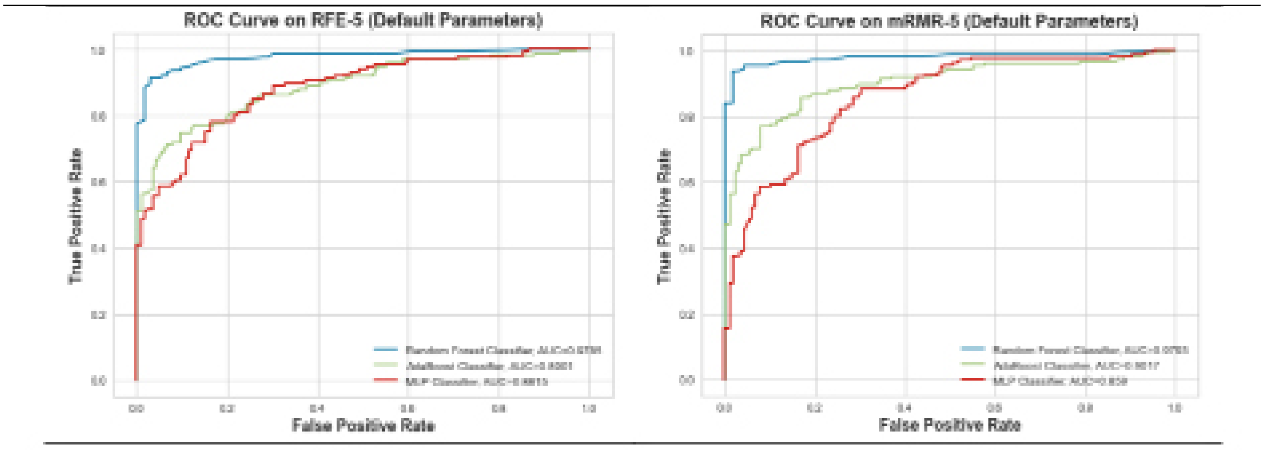


Figure 6. Pre-evaluation performance of random forest (RF), adaptive boost (AdaBoost) and multi-layer perceptron (MLP) (default parameters) models using AUROC curve

Figure 6 represents the prediction performance of RF, AdaBoost, and MLP models on both RFE [RFE-50, RFE-10, RFE-5] and mRMR [mRMR-50, mRMR-10, mRMR-5] feature subsets in terms of the AUC-ROC (AUROC) curve using the models' default parameters before model optimization.

Post pre-evaluation phase, the next phase of this study was hyperparameter tuning. This phase aimed to further improve the performance of selected ML (RF, AdaBoost, and MLP) models by tuning their hyperparameters to optimal values. The GridSearchCV method was employed for model optimization or hyperparameter tuning purposes during the current study. Different search space values for the various models considered were supplied to the grid search method, and through rigorous processing, the optimal hyperparameter values were obtained. A list of model-wise hyperparameter names, supplied search space values, and the optimal hyperparameter values obtained via the grid search method is presented in Table 2.

After obtaining tuned hyperparameters for RF ['criterion': 'entropy', 'max_depth': 15, 'n_estimators': 300], AdaBoost ['algorithm': 'SAMME.R', 'learning_rate': 1, 'n_estimators': 300], and MLP ['activation': 'relu', 'alpha': 0.0001, 'hidden_layer_sizes': (500,), 'learning_rate': 'invscaling', 'solver': 'sgd'] models using grid search, these models were evaluated on both RFE [RFE-50, RFE-10, RFE-5] and mRMR [mRMR-50, mRMR-10, mRMR-5] feature subsets using the tuned hyperparameters for post-evaluation in terms of accuracy, recall, F1-score, precision, and AUC. Evaluation results of the post-evaluation of these models are presented in Table-5, and the prediction performance of RF, AdaBoost, and MLP models in terms of the AUC-ROC (AUROC) curve is shown in Figure 7.

Table 5. Performance of Random Forest, Adaptive Boost and Multi-Layer Perceptron Model (Tuned Parameter) Models on Random Forest and mRMR Feature

ML Model Name	Feature Subsets	Accuracy	Precision	Recall	F1-Score	AUC	Feature Subsets	Accuracy	Precision	Recall	F1-Score	AUC
RF	Full Set	0.9646	0.97	0.96	0.96	0.9969	Full Set	0.9646	0.97	0.96	0.96	0.9969
AdaBoost		0.9557	0.96	0.96	0.96	0.9924		0.9557	0.96	0.96	0.96	0.9924
MLP		0.9439	0.95	0.94	0.94	0.9905		0.9439	0.95	0.94	0.94	0.9905
RF	RFE-50	0.9646	0.97	0.96	0.96	0.9981	mRMR-50	0.9438	0.95	0.95	0.95	0.9929
AdaBoost		0.9646	0.97	0.96	0.96	0.9889		0.9204	0.92	0.92	0.92	0.959
MLP		0.9852	0.99	0.99	0.99	0.9965		0.9174	0.93	0.92	0.92	0.9714
RF	RFE-10	0.9528	0.95	0.95	0.95	0.9923	mRMR-10	0.9204	0.93	0.92	0.92	0.9814
AdaBoost		0.9027	0.92	0.9	0.9	0.9418		0.8997	0.92	0.9	0.9	0.927
MLP		0.9233	0.93	0.92	0.92	0.9658		0.9056	0.92	0.91	0.9	0.9352
RF	RFE-5	0.9145	0.92	0.91	0.91	0.9805	mRMR-5	0.9203	0.93	0.92	0.92	0.9776
AdaBoost		0.8732	0.88	0.87	0.87	0.9191		0.8761	0.88	0.88	0.88	0.9289
MLP		0.8791	0.89	0.88	0.88	0.9137		0.8908	0.9	0.89	0.89	0.9189

Abbreviations: RF, random forest; AdaBoost, adaptive boost; MLP, multi-layer perceptron.

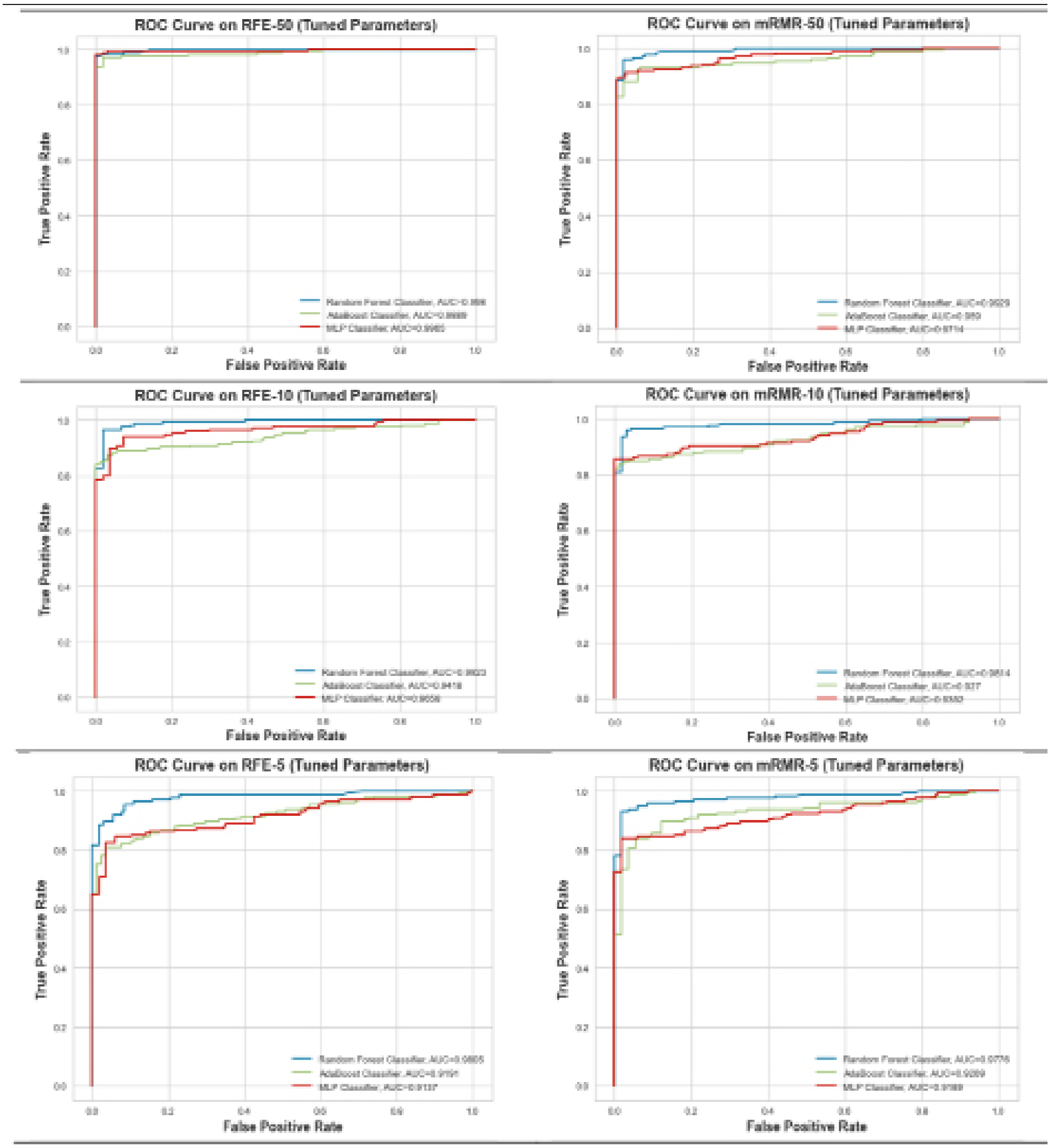


Figure 7. Post-evaluation performance of random forest (RF), adaptive boost (AdaBoost) and multi-layer perceptron (MLP) (tuned parameter) models using AUROC Curve

Figure-7 represents the prediction performance of RF, AdaBoost, and MLP models on both RFE [RFE-50, RFE-10, RFE-5] and mRMR [mRMR-50, mRMR-10, mRMR-5] feature subsets in terms of the AUC-ROC (AUROC) curve using the models' tuned parameters after model optimization.

Observations from Table 4 show the dominance of the RF model in terms of prediction performance over the other two ML models when evaluation is performed on RFE and mRMR feature subsets using their default parameters. Except for the mRMR-50 subset, in all other five feature subsets, the RF model demonstrated the highest classification accuracy of up to 94.39%, precision of up to 0.95, recall of up to 0.94, F1-score of up to 0.94, and AUC-score of 0.9922. However, in the sole case of the mRMR-50 subset, the MLP model achieved an accuracy of 97.37%. Additionally, it can be observed in Table 4 as well as in

Figure-6 and Figure-8 that the performance of all three models is significantly better when using RFE-generated feature subsets rather than mRMR-generated feature subsets, demonstrating the superiority of RFE feature subsets over mRMR feature subsets for the current study.

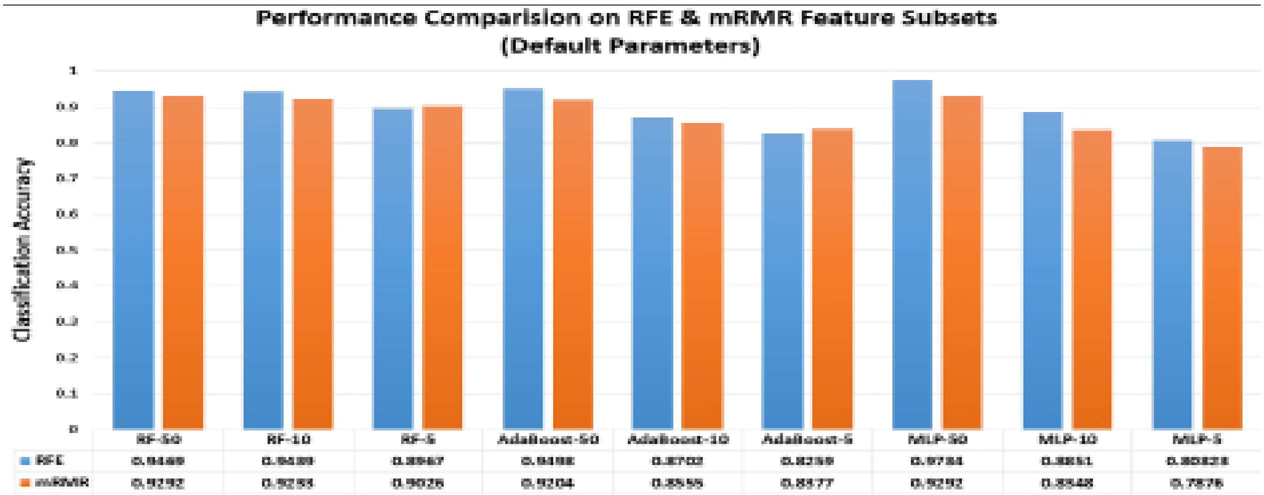


Figure 8. Performance comparison of RFE & mRMR feature subsets using default parameters

Furthermore, observing Table-5 reveals two important findings during the post-evaluation phase of the study. The first observation indicates the dominance of the tuned RF model over the AdaBoost and MLP models, with an improved accuracy of "96.46%", precision of "0.97", recall of "0.96", F1-score of "0.96", and AUC-score of "0.998" on every feature subset generated by the RFE and mRMR feature selection methods, including the full set consisting of all features of the PD dataset.

The second fact that may be noticed using Table-5 and Figures 7 and 9 is that all three ML models (RF, AdaBoost, and MLP) demonstrated significantly improved performance on the RFE-50 and mRMR-50 feature subsets of the PD dataset, even when compared to performance on the full feature set. However, the models' performance on the remaining feature subsets, i.e., RFE-10, RFE-5, mRMR-10, and mRMR-5, was observed to reduce to a certain level. This signifies the better contribution of the 50 features identified by the RFE and mRMR FS methods to PD class prediction rather than the subsets with 10 or 5 features.

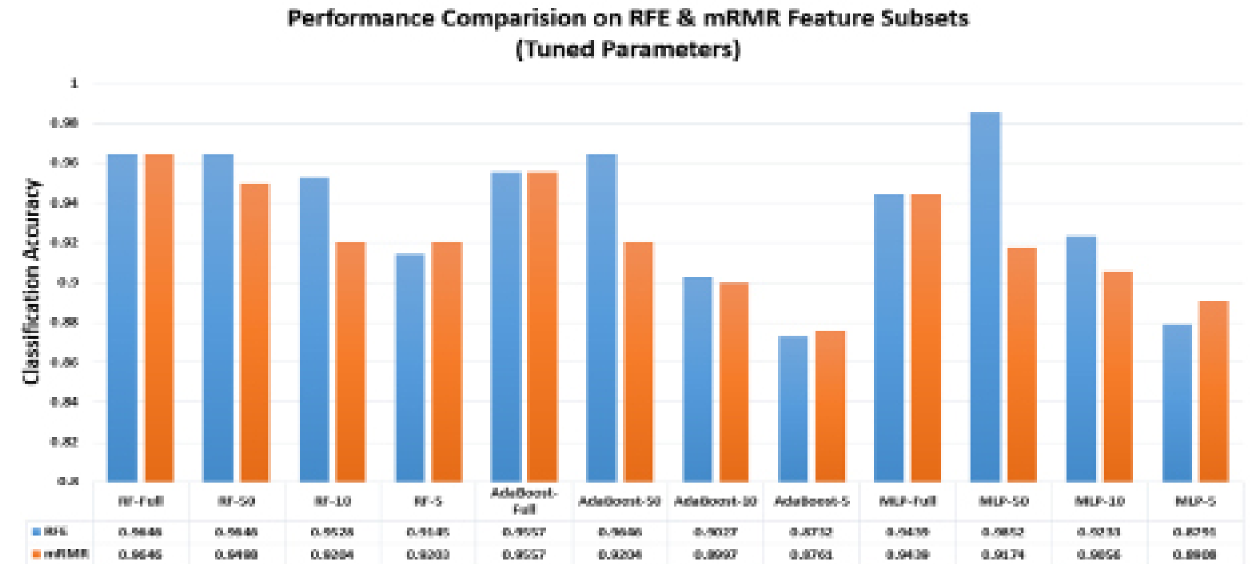


Figure 9. Performance comparison of RFE & mRMR feature subsets using tuned parameters ure 9. Performance comparison of RFE & mRMR feature subsets using tuned parameters

It also represents that the 10 and 5 feature subsets may not be fair representatives of the complete PD dataset due to the higher information loss encountered during feature selection. In contrast to the subsets with 10 or 5 features, the 50-feature subsets showed enhanced performance for all three ML models, even compared to when all features of the PD dataset were utilized. This signifies higher information gain and correlation with the target feature, demonstrating that the identified 50 features are strong candidates for representing the complete PD dataset without losing much of the information originally available in the dataset.

As demonstrated in Table 3, the RF, AdaBoost, and MLP models were identified as the most accurate models for the PD detection task during the pre-evaluation phase of this study. However, during the post-evaluation phase after model optimization, the results demonstrated significantly higher accuracy by the RF-based model when employed with RFE feature subsets, even in almost every test scenario mentioned in Tables 3, 4, and 5. Although the MLP-based model also demonstrated higher accuracy than the RF-based model on the RFE-50 feature subset, for the sake of model generalizability across almost every test scenario, it cannot be reported as the best classifier for the PD classification task.

5. Discussion

The primary objective of the study was to investigate baseline machine learning models and fine-tune them using GridSearchCV for early PD detection based on speech samples. By optimizing feature selection methods and addressing class imbalance, the research aimed to enhance the accuracy of PD diagnosis. The study emphasizes the importance of efficient feature selection techniques and hyperparameter tuning in improving model performance. The experimental results, presented in Section-4, demonstrate the superiority of the RF-based method in PD diagnosis tasks compared to other machine learning models. The findings suggest that efficient class balancing and feature selection, along with complete case analysis-based hyperparameter tuning, play a crucial role in reducing model complexity and improving performance, with RFE showing promising outcomes for high-dimensional datasets like the PD speech dataset used in the study. However, the study acknowledges certain limitations, such as the relatively small sample size of the dataset, which may impact the generalizability of the models. Future research directions include incorporating the neurological conditions of patients, exploring cross-database research, and developing ML models using larger datasets for improved generalization. Despite these limitations, the study provides valuable insights into the application of machine learning techniques for early PD detection.

On further evaluation, when the presented method, referred to as the RF with RFE model (RF+RFE), is compared with previously reported studies on the same PD dataset, as shown in Table-6, it is observed that the diagnostic performance of the RF+RFE model is significantly superior to the models employed in recent research. As presented in Table-6, the RF+RFE model emerges as the most effective among other ML-based models utilized in previous studies on the same PD dataset.

Table 6. Comparative Analysis of Various Models for Parkinson's Disease Detection

Machine Learning Model	Proposed by	Accuracy (%)
Linear SVM	Achraf Benba et al. (58)	91.17
XGBoost	Nissar, Iqra, et al. (59)	95.39
kNN + Adaboost.M1	Richa Mathur et al. (60)	91.28
ANN	Yasar et al. (61)	94.93
SVM (RBF)	Sakar et al. (62)	86
NB+Fisher Score	Pramanik et al. (18)	78.97
AdaBoost + IG	Barukab et al. (2)	90.3
Stack (RF+SVM+KNN)	Joshi (63)	91
KNN+Chi Square	Demir et al. (15)	95.4
SVM + RFE	Solana-Lavalle et al. (64)	94.7
DL Model		
2D-CNN + 1D-CNN	Quan et al. (65)	92
DNN Ensemble	Yuan et al. (66)	95
U-Lossian Network	Maskeliūnas et al. (20)	94.33
CNN +SVM	Khaskhoussy and Ayed (67)	98
ResNet50 + GDABC	Wang et al. (68)	96
RF + RFE	Proposed by this study	96.46

One of the key factors contributing to the RF + RFE model's success is the Random Forest technique's ability to build multiple decision trees and aggregate their predictions. This ensemble approach enhances the model's robustness and accuracy. Another critical factor is the regularization aspect of the RF method, which effectively reduces the risk of overfitting. By incorporating advanced feature selection methods and hyperparameter tuning, the study not only improves the accuracy of PD classification but also increases its efficiency. This progress paves the way for the development of more reliable and practical diagnostic tools for PD in the future.

5.1. Conclusion

Early diagnosis of PD is of utmost significance in saving patients' lives by enabling timely intervention. PD speech pattern analysis applications for developing predictive diagnosis and monitoring models have gained considerable attention, with remarkable progress observed in speech analysis methodologies recently. Since speech measurements are non-invasive and speech processing has consistently demonstrated significant potential in identifying PD, this research aims to evaluate the effectiveness of several ML-based classification methods. A speech PD dataset was utilized to apply various classifiers, and several assessment criteria were compared using visualization and statistical analysis.

In this study, the issue of PD diagnosis is approached using an ML technique, employing multiple ML models for detection. By analyzing vocal signals, the primary goal is to demonstrate PD diagnosis using various ML models alongside efficient feature selection (FS) methods (RFE and mRMR) and model optimization via the grid search method. The experimentation findings suggested that the RF model outperformed all other classifiers used in the study. The classification performance of the RF model with the RFE FS methodology reported an accuracy of 96.46%, precision of 0.97, F1-score of 0.96, recall of 0.96, and an AUC-score of 0.998. With the mRMR FS technique, the RF model showed an accuracy of 96.38%, precision of 0.95, recall of 0.95, F1-score of 0.95, and an AUC-score of 0.992, which was superior to all 12 ML methods investigated during this study.

The performance of the current method may rely on the user's presumed deviations in the dataset utilized, the validation method considered, the feature selection methods applied, and the hyperparameter strategy employed during experimentation. A limitation of this study is the sample size of the dataset used to train the model, as the dataset contains only 756 instances. A larger dataset would enhance the model's ability to generalize the PD diagnosis task. Furthermore, this study does not incorporate the neurological conditions of the patients during PD classification, which might be another limitation. Future research could address this by incorporating patients' neurological conditions and conducting cross-database research to validate ML models further.

Key findings suggest that the RF method is highly effective for creating a model for PD diagnosis tasks. The study recommends using efficient feature selection techniques to reduce the complexity of detection systems for high-dimensional datasets like the PD speech dataset used in this research, specifically advocating the use of the RFE method for feature selection tasks, as it helped achieve superior outcomes in this scenario. The RF model's effective accuracy, precision, F1-score, recall, and AUC-score make it a trustworthy tool for diagnosing Parkinson's disease.

Acknowledgments

Author is thankful to all my senior and colleagues who helped me during this article writing and also to authors, whose article we have utilized in this study.

Authors' Contribution:

XX

Conflict of Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Reproducibility: The data that support the findings of this study are openly available in UCI Machine Learning Repository at https://doi.org/10.24432/C5MS4X, reference number (37).

Funding/Support: This work is not funded by any research agency or organization during this study.

References

1. Olanow CW, Stern MB, Sethi K. The scientific and clinical basis for the treatment of Parkinson disease (2009). *Neurol J*. 2009;**72**(21_supplement_4). https://doi.org/10.1212/WNL.0b013e3181a1d44c.
2. Barukab O, Ahmad A, Khan T, Thayyil Kunhumuhammed MR. Analysis of Parkinson's Disease Using an Imbalanced-Speech Dataset by Employing Decision Tree Ensemble Methods. *Diagnostic J (Basel)*. 2022;**12**(12). [PubMed ID:36553007]. [PubMed Central ID:PMC9776735]. https://doi.org/10.3390/diagnostics12123000.
3. de Rijk MC, Launer LJ, Berger K, Breteler MM, Dartigues JF, Baldereschi M, et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurol J*. 2000;**54**(11 Suppl 5):S21-3. [PubMed ID:10854357].
4. Ponce FA, Lozano AM. The most cited works in Parkinson's disease. *Mov Disord*. 2011;**26**(3):380-90. [PubMed ID:21462255]. https://doi.org/10.1002/mds.23445.
5. Langston JW. Parkinson's disease: current and future challenges. *Neurotoxicol J*. 2002;**23**(4-5):443-50. [PubMed ID:12428715]. https://doi.org/10.1016/s0161-813x(02)00098-0.
6. Brakedal B, Toker L, Haugarvoll K, Tzoulis C. A nationwide study of the incidence, prevalence and mortality of Parkinson's disease in the Norwegian population. *NPJ Parkinsons Dis*. 2022;**8**(1):19. [PubMed ID:35236852]. [PubMed Central ID:PMC8891365]. https://doi.org/10.1038/s41531-022-00280-4.
7. Simuni T, Okun MS. Adjunctive Therapies in Parkinson Disease-Have We Made Meaningful Progress? *JAMA Neurol*. 2022;**79**(2):119-20. [PubMed ID:34962569]. https://doi.org/10.1001/jamaneurol.2021.4140.
8. Lageman SK, Donovan EK, Villasenor T, Perrin PB. Exploration of Parkinson's Disease Symptomatology Subtypes From the Caregiver Perspective: Implications for Caregiver Burden, Depression, and Anxiety. *J Geriatr Psychiatry Neurol*. 2022;**35**(5):663-70. [PubMed ID:34587822]. https://doi.org/10.1177/08919887211049146.
9. Li S, Ritz B, Gong Y, Cockburn M, Folle AD, Del Rosario I, et al. Proximity to residential and workplace pesticides application and the risk of progression of Parkinson's diseases in Central California. *Sci Total Environ*. 2023;**864**:160851. [PubMed ID:36526213]. [PubMed Central ID:PMC11121507]. https://doi.org/10.1016/j.scitotenv.2022.160851.
10. Tsanas A, Little M, McSharry P, Ramig L. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests.

Natur Preced J. 2009;**1**. https://doi.org/10.1038/npre.2009.3920.1.

11. Rigas G, Tzallas AT, Tsipouras MG, Bougia P, Tripoliti EE, Baga D, et al. Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. *IEEE Trans Inf Technol Biomed*. 2012;**16**(3):478-87. [PubMed ID:22231198]. https://doi.org/10.1109/TITB.2011.2182616.
12. Gasser T, Wichmann T. Parkinson disease and other synucleinopathies. *Neurobiology of Brain Disorders*. Cambridge: Academic Press; 2023. p. 253-74.
13. 13.Ghane M, Ang MC, Nilashi M, Sorooshian S. Enhanced decision tree induction using evolutionary techniques for Parkinson's disease classification. *Biocybernetic Biomed Eng*. 2022;**42**(3):902-20. https://doi.org/10.1016/j.bbe.2022.07.002.
14. Liu Y, Reddy MK, Penttila N, Ihalainen T, Alku P, Rasanen O. Automatic Assessment of Parkinson's Disease Using Speech Representations of Phonation and Articulation. *IEEE/ACM Transact Audio, Speech, Lang Proc*. 2023;**31**:242-55. https://doi.org/10.1109/taslp.2022.3212829.
15. Demir F, Siddique K, Alswaitti M, Demir K, Sengur A. A Simple and Effective Approach Based on a Multi-Level Feature Selection for Automated Parkinson's Disease Detection. *J Pers Med*. 2022;**12**(1). [PubMed ID:35055370]. [PubMed Central ID:PMC8781034]. https://doi.org/10.3390/jpm12010055.
16. Kepesiova Z, Kozak S, Ruzicky E, Zimmermann A, Malaschitz R. Comparative Analysis of Advanced Machine Learning Algorithms for Early Detection of Parkinson Disease. 2022 Cybernetics & Informatics (K&I). 2022. p. 1-6.
17. Weintraub D, Aarsland D, Chaudhuri KR, Dobkin RD, Leentjens AF, Rodriguez-Violante M, et al. The neuropsychiatry of Parkinson's disease: advances and challenges. *Lancet Neurol*. 2022;**21**(1):89-102. [PubMed ID:34942142]. [PubMed Central ID:PMC8800169]. https://doi.org/10.1016/S1474-4422(21)00330-6.
18. Pramanik M, Pradhan R, Nandy P, Qaisar SM, Bhoi AK. Assessment of Acoustic Features and Machine Learning for Parkinson's Detection. *J Healthc Eng*. 2021;**2021**:9957132. [PubMed ID:34471507]. [PubMed Central ID:PMC8405321]. https://doi.org/10.1155/2021/9957132.
19. Skrabal D, Rusz J, Novotny M, Sonka K, Ruzicka E, Dusek P, et al. Articulatory undershoot of vowels in isolated REM sleep behavior disorder and early Parkinson's disease. *NPJ Parkinsons Dis*. 2022;**8**(1):137. [PubMed ID:36266347]. [PubMed Central ID:PMC9584921]. https://doi.org/10.1038/s41531-022-00407-7.
20. Maskeliūnas R, Damaševičius R, Kulikajevas A, Padervinskis E, Pribuišis K, Uloza V. A Hybrid U-Lossian Deep Learning Network for Screening and Evaluating Parkinson's Disease. *Appl Sci J*. 2022;**12**(22). https://doi.org/10.3390/app122211601.
21. Moro-Velazquez L, Gomez-Garcia JA, Arias-Londoño JD, Dehak N, Godino-Llorente JI. Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomed Signal Proc Control*. 2021;**66**. https://doi.org/10.1016/j.bspc.2021.102418.
22. Suphinnapong P, Phokaewvarangkul O, Thubthong N, Teeramongkonrasmee A, Mahattanasakul P, Lorwattanapongsa P, et al. Objective vowel sound characteristics and their relationship with motor dysfunction in Asian Parkinson's disease patients. *J Neurol Sci*. 2021;**426**:117487. [PubMed ID:34004464]. https://doi.org/10.1016/j.jns.2021.117487.
23. Lamba R, Gulati T, Alharbi H, Jain A. A hybrid system for Parkinson's disease diagnosis using machine learning techniques. *Int J Speech Technol*. 2022;**25**:1-11. https://doi.org/10.1007/s10772-021-09837-9.
24. Baker S, Tekriwal A, Felsen G, Christensen E, Hirt L, Ojemann SG, et al. Automatic extraction of upper-limb kinematic activity using deep learning-based markerless tracking during deep brain stimulation implantation for Parkinson's disease: A proof of concept study. *PLoS One*. 2022;**17**(10):e0275490. [PubMed ID:36264986]. [PubMed Central ID:PMC9584454]. https://doi.org/10.1371/journal.pone.0275490.
25. Bohra A, Vasudevan A, Kutaiba N, Van Langenberg DR. Challenges and Strategies to Optimising the Quality of Small Bowel Magnetic Resonance Imaging in Crohn's Disease. *Diagnostics (Basel)*. 2022;**12**(10). [PubMed ID:36292222]. [PubMed Central ID:PMC9600769]. https://doi.org/10.3390/diagnostics12102533.
26. Duffy JR. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. 3rd ed. Amsterdam: Elsevier 2013.
27. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. *J Speech Hear Res*. 1969;**12**(2):246-69. [PubMed ID:5808852]. https://doi.org/10.1044/jshr.1202.246.
28. Blanchet P, Snyder G. Speech rate deficits in individuals with Parkinson's disease: A review of the literature. *J Med Speech-Lang Pathol*. 2009;**17**:1-7.
29. Ackermann H, Ziegler W. Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. *J Neurol Neurosurg Psychiatry*. 1991;**54**(12):1093-8. [PubMed ID:1783924]. [PubMed Central ID:PMC1014687]. https://doi.org/10.1136/jnnp.54.12.1093.
30. Rana A, Dumka A, Singh R, Panda MK, Priyadarshi N. A Computerized Analysis with Machine Learning Techniques for the Diagnosis of Parkinson's Disease: Past Studies and Future Perspectives. *Diagnostics (Basel)*. 2022;**12**(11). [PubMed ID:36359550]. [PubMed Central ID:PMC9689408]. https://doi.org/10.3390/diagnostics12112708.
31. Painuli D, Bhardwaj S, Kose U. Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. *Comput Biol Med*. 2022;**146**:105580. [PubMed ID:35551012]. https://doi.org/10.1016/j.compbiomed.2022.105580.
32. NeeluJyoti DM, Painuli AD. Fuzzy Expert System to diagnose Psoriasis Disease. *Int J Comput Sci Inform Secur*. 2018;**16**(9):33-8.
33. Gunduz H. An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification. *Biomed Signal Proc Control*. 2021;**66**. https://doi.org/10.1016/j.bspc.2021.102452.
34. Painuli D, Mishra D, Bhardwaj S, Aggarwal M. Machine Learning based Model to combat Covid19. *Int J Information Technol Electric Eng*. 2020;**9**(4):33-40.
35. Kaur S, Aggarwal H, Rani R. Diagnosis of Parkinson's disease using deep CNN with transfer learning and data augmentation. *Multimedia Tools App*. 2020;**80**(7):10113-39. https://doi.org/10.1007/s11042-020-10114-1.
36. Holmes RJ, Oates JM, Phyland DJ, Hughes AJ. Voice characteristics in the progression of Parkinson's disease. *Int J Lang Commun Disord*. 2000;**35**(3):407-18. [PubMed ID:10963022]. https://doi.org/10.1080/136828200410654.
37. Sakar C, Serbes G, Gunduz A, Nizam H, Sakar B. Parkinson's Disease Classification Dataset. UCI Machine Learning Repository; 2018. Available from: https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification.
38. Izonin I, Tkachenko R, Shakhovska N, Ilchyshyn B, Singh KK. A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematic J*. 2022;**10**(11). https://doi.org/10.3390/math10111942.
39. Albert AJ, Murugan R, Sripriya T. Diagnosis of heart disease using oversampling methods and decision tree classifier in cardiology. *Res Biomed Eng*. 2022;**39**(1):99-113. https://doi.org/10.1007/s42600-022-00253-9.
40. Mishra D, D. P. Rule Based Expert System for Medical Diagnosis-A Review. *Int J Eng Technol, Manag, Appl Sci*. 2016;**4**(12):167-72.
41. Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *Neuroimage Clin*. 2017;**16**:586-94. [PubMed ID:28971009]. [PubMed Central ID:PMC5610036]. https://doi.org/10.1016/j.nicl.2017.09.010.
42. Abdelwahed NM, El-Tawel GS, Makhlouf MA. Effective hybrid feature selection using different bootstrap enhances cancers classification performance. *BioData Min*. 2022;**15**(1):24. [PubMed ID:36175944]. [PubMed Central ID:PMC9523996]. https://doi.org/10.1186/s13040-022-00304-y.
43. Ji Y, Li J, Huang Z, Xie W, Zhao D. A Data Dimensionality Reduction Method Based on mRMR and Genetic Algorithm for High-Dimensional Small Sample Data. *Web Information Systems and Applications*. Berlin: Springer 2022. p. 485-96.
44. Signol F, Arnal L, Navarro-Cerdan JR, Llobet R, Arlandis J, Perez-Cortes JC. SEQENS: An ensemble method for relevant gene identification in microarray data. *Comput Biol Med*. 2023;**152**:106413.

[PubMed ID:36521355]. https://doi.org/10.1016/j.compbiomed.2022.106413.

45. Fujita T, Luo Z, Quan C, Mori K, Cao S. Performance Evaluation of RNN with Hyperbolic Secant in Gate Structure through Application of Parkinson's Disease Detection. *Appl Sci*. 2021;**11**(10). https://doi.org/10.3390/app11104361.
46. Painuli D, Mishra D, Bhardwaj S, Aggarwal M. Forecast and prediction of COVID-19 using machine learning. *Data Science for COVID-19*. Cambridge: Academic Press; 2021. p. 381-97.
47. Sakar CO, Kursun O. Telediagnosis of Parkinson's disease using measurements of dysphonia. *J Med Syst*. 2010;**34**(4):591-9. [PubMed ID:20703913]. https://doi.org/10.1007/s10916-009-9272-y.
48. Rao RB, Fung G, Rosales R. On the Dangers of Cross-Validation. An Experimental Evaluation. Proceedings of the 2008 SIAM International Conference on Data Mining. 2008. p. 588-96.
49. Kearns M, Ron D. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. Proceedings of the tenth annual conference on Computational learning theory - COLT '97. 1997. p. 152-62.
50. Wong T-T, Yeh P-Y. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transact Knowledge Data Eng*. 2020;**32**(8):1586-94. https://doi.org/10.1109/tkde.2019.2912815.
51. Zhang X, Liu C-A. Model averaging prediction byK-fold cross-validation. *J Econometr*. 2023;**235**(1):280-301. https://doi.org/10.1016/j.jeconom.2022.04.007.
52. Bergmeir C, Hyndman RJ, Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computation Statistic Data Analys*. 2018;**120**:70-83. https://doi.org/10.1016/j.csda.2017.11.003.
53. Zhang B, Quan C, Ren F. *Study on CNN in the recognition of emotion in audio and images*. Washington: AIA International Advanced Information Institute; 2016.
54. Muntasir Nishat M, Faisal F, Rahman Dip R, Nasrullah SM, Ahsan R, Shikder F, et al. A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms. *EAI Endorsed Transact Pervasive Health Technol*. 2021;**7**(29). https://doi.org/10.4108/eai.13-8-2021.170671.
55. Zhang B, Quan C, Ren F. Study on CNN in the recognition of emotion in audio and images. 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). 2016. p. 1-5.
56. Srivastava S. Genetic Algorithm Optimized Deep Learning Model for Parkinson Disease Severity Detection. [Dissertation]. Dublin: National College of Ireland 2021.
57. Alqahtani A, Alsubai S, Sha M, Vilcekova L, Javed T. Cardiovascular Disease Detection using Ensemble Learning. *Comput Intell Neurosci*. 2022;**2022**:5267498. [PubMed ID:36017452]. [PubMed Central ID:PMC9398727]. https://doi.org/10.1155/2022/5267498.
58. Benba A, Jilbab A, Hammouch A, Sandabad S. Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. 2015 International Conference on Electrical and Information Technologies (ICEIT). 2015. p. 300-4.
59. Nissar I, Raza Rizvi D, Masood S, Nazir Mir A. Voice-Based Detection of Parkinson's Disease through Ensemble Machine Learning Approach: A Performance Study. *EAI Endorsed Transact Pervasive Health Technol*. 2019;**5**(19). https://doi.org/10.4108/eai.13-7-2018.162806.
60. Mathur R, Pathak V, Bandil D. Parkinson Disease Prediction Using Machine Learning Algorithm. *Emerging Trends in Expert Applications and Security*. Berlin: Springer 2019. p. 357-63.
61. Yasar A, Saritas I, Sahman MA, Cinar AC. Classification of Parkinson disease data with artificial neural networks. *IOP Conference Series: Material Sci Eng*. 2019;**675**(1). https://doi.org/10.1088/1757-899x/675/1/012031.
62. Sakar CO, Serbes G, Gunduz A, Tunc HC, Nizam H, Sakar BE, et al. 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*. 2019;**74**:255-63. https://doi.org/10.1016/j.asoc.2018.10.022.
63. Joshi DD, Joshi HH, Panchal BY, Goel P, Ganatra A. A Parkinson Disease Classification Using Stacking Ensemble Machine Learning Methodology. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). 2022. p. 1335-41.
64. Solana-Lavalle G, Galán-Hernández J-C, Rosas-Romero R. Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybernetic Biomedic Eng*. 2020;**40**(1):505-16. https://doi.org/10.1016/j.bbe.2020.01.003.
65. Quan C, Ren K, Luo Z, Chen Z, Ling Y. End-to-end deep learning approach for Parkinson's disease detection from speech signals. *Biocybernetics Biomedic Eng*. 2022;**42**(2):556-74. https://doi.org/10.1016/j.bbe.2022.04.002.
66. Yuan L, Liu Y, Feng H-M. Parkinson disease prediction using machine learning-based features from speech signal. *Serv Orient Comput App*. 2023;**18**(1):101-7. https://doi.org/10.1007/s11761-023-00372-w.
67. Khaskhoussy R, Ayed YB. Improving Parkinson's disease recognition through voice analysis using deep learning. *Pattern Recogn Letters*. 2023;**168**:64-70. https://doi.org/10.1016/j.patrec.2023.03.011.
68. Wang X, Li S, Pun CM, Guo Y, Xu F, Gao H, et al. A Parkinson's Auxiliary Diagnosis Algorithm Based on a Hyperparameter Optimization Method of Deep Learning. *IEEE/ACM Trans Comput Biol Bioinform*. 2024;**21**(4):912-23. [PubMed ID:37027659]. https://doi.org/10.1109/TCBB.2023.3246961.