Decision Analytics Journal 10 (2024) 100381

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

An artificial intelligence-based decision support system for early and accurate diagnosis of Parkinson's Disease

Mahesh T.R. [a], Vinoth Kumar V. [b], Rajat Bhardwaj [c], Surbhi B. Khan [d,e,*], Nora A. Alkhaldi [f], Nancy Victor [g], Amit Verma [h]

[a] Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore, India
[b] School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, India
[c] Department of Computer Science and Engineering, ASET, Amity University, Bengaluru, KA, India
[d] Department of Data Science, School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom
[e] Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon
[f] School of Computer Science and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia
[g] School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamilnadu, India
[h] University Centre for Research and Development, Department of Computer Science, Chandigarh University, Gharuan, Mohali, Punjab, India

ARTICLE INFO

Keywords:
Machine learning
Parkinson's Disease
Predictive analytics
Diagnosis and classification
Artificial intelligence
Entropy

ABSTRACT

People with Parkinson's Disease (PD) might struggle with sadness, restlessness, or difficulty speaking, chewing, or swallowing. A diagnosis can be challenging because there is no specific PD test. It is diagnosed by doctors using a neurological exam and a medical history. This study proposes several Machine Learning (ML) algorithms to predict PD. These ML algorithms include K-Nearest Neighbor (KNN), Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting algorithms (XGBoost), and their ensemble methods using publicly available PD dataset with 195 instances. The ML algorithms are used to predict and classify PD using homogeneous XGBoost ensemble techniques with reduced amount of entropy. Synthetic Minority Oversampling Technique (SMOTE) is utilized to handle imbalanced data, and 10-fold cross-validation is employed for evaluation. The results show that the homogeneous XGBoost-Random Forest outperforms other ML methods with 98% accuracy and Matthew's correlation coefficient value 0.93.

1. Introduction

A neurological ailment that affects the central nervous system and is currently becoming more prevalent in an ageing population is Parkinson's disease (PD). Similarly, it affects 1.2 million people in Europe, and by 2030, experts predict that figure will have doubled [1]. To distinguish PD from other neurological conditions and healthy people, as well as to improve PD treatment and follow-up, an accurate diagnosis is necessary. In the early stages of PD, the diagnosis is very difficult [2]. Although several criteria and recommendations have been put forth to help with PD diagnosis, clinical evaluation remains the gold standard for PD diagnosis and symptom monitoring. Clinical evaluation comprises a number of subjective factors and has an accuracy range of 75% to 82% [3].

For PD to be diagnosed, it is crucial to have bradykinesia, stiffness, and resting tremor. These motor characteristics appear when 70% of dopaminergic synapses in the striatum, and 50% of substantia nigra dopaminergic synapses are destroyed [4]. As a result, PD has an insidious clinical onset, and by the time it is diagnosed, brain disease has already progressed significantly. Clinical traits that exist prior to motor symptoms may be helpful in this regard. In addition to accompanying, but frequently coming before the beginning of motor features, non-motor indications including olfactory dysfunction, autonomic symptoms, sleep issues, visual impairment, cognitive decline, or depressive symptoms are becoming more widely acknowledged [5,6]. There is growing interest in exploiting this variety of premotor symptoms to spot PD patients at early stages. This premotor or prodromal phase in PD lasts between 5 and 20 years.

Motor or non-motor symptoms of PD can be distinguished. Tremor, stiffness, bradykinesia, and postural instability are all part of the first category. Loss of taste and smell, sleep disturbances, sexual dysfunction, anxiety, pain, gastrointestinal issues, constipation, swallowing issues, fatigue, depression, hallucinations and psychosis, cognitive impairment, impulse control issues, and dementia are just a few of the non-motor symptoms that can occur [7].

Studies on motor disorders continue to present a variety of clinical difficulties for the scientific community. For instance, the presence of

* Corresponding author at: Department of Data Science, School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom.
E-mail addresses: trmahesh.1978@gmail.com (Mahesh T.R.), pvkumar243@gmail.com (Vinoth Kumar V.), rajatbhardwaj.x@gmail.com (R. Bhardwaj), s.khan138@salford.ac.uk (S.B. Khan), N.alkhaldi@kfu.edu.sa (N.A. Alkhaldi), nancyvictor@vit.ac.in (N. Victor), d_k.t1816@cumail.in (A. Verma).

https://doi.org/10.1016/j.dajour.2023.100381
Received 23 May 2023; Received in revised form 4 December 2023; Accepted 10 December 2023
Available online 13 December 2023

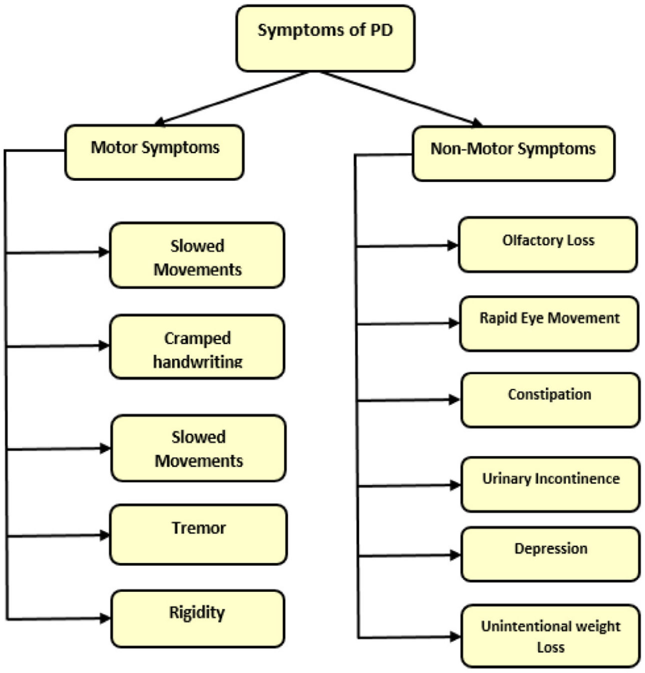


Fig. 1. Parkinson's Disease symptoms.

motor signs is a necessary component of the diagnostic criteria for PD, and the neurologist evaluates movement disorders by visually assessing motor tasks and using semi-quantitative rating scales. The necessity for objective motor evaluation methods is essential for the development of future PD diagnosis procedures [8]. Similar to motor performance, pattern interpretation and analysis appear to be important in early diagnosis [9]. Hence, one of the most difficult chances to develop reliable and impartial biomarkers to identify early PD symptoms is the assessment of movement performance [10]. The symptoms of PD disease as discussed in this study [10] are shown in Fig. 1.

It is commonly known that in the majority of research completed over the previous ten years, ensemble classifiers have outperformed single classifiers. The aim of this proposed study is to provide the early detection of PD using single classifiers as well as homogeneous Extreme Gradient Boosting (XGBoost) classifiers. Homogeneous XGBoost classifiers are formed by mixing an individual classifier with XGBoost.

The summary of the outcomes of this proposed study are as follows:

- Use synthetic minority over-sampling technique to handle imbalanced data in the breast cancer diagnosis dataset.
- Analyze before and after applying the synthetic minority over-sampling technique
- Utilize k-fold cross-validation for increasing performance evaluation.
- Compare the performance of different machine learning algorithms for Parkinson's disease prediction and diagnosis.
- Show homogeneous XGBoost-Random Forest outperforms other classifiers and offers the highest accuracy.

The remaining part of the paper is structured as follows: Section 2 provides review of literature. Section 3 introduces the workflow of the proposed methodology. Section 4 provides the results and discussions along with the comparison to state-of-art existing works and Section 5 concludes the study with future directions.

2. Related work

Machine Learning (ML) techniques are now being used more frequently for the prior identification of PD. As a result, the accuracy of PD prediction has significantly increased when using a variety of data modalities, such as handwritten patterns [5], voice and speech signals [6], various neuroimaging approaches, or biofluids. The authors [11] employed vocal-based PD detection using four classifiers: k closest neighbor, Support Vector Machine(SVM), and Random Forest (RF). Their method has a 94.7% accuracy rate, a 98.4% sensitivity rate, a 92.68% specificity rate, and a 97.22% precision rate. Using a support vector machine produced the best accuracy. By choosing no more than 20 characteristics, the associated computing complexity was further decreased. Although confirming a Parkinson diagnosis is technically difficult, clinicians can identify the condition by examining patients and examining various symptoms. The authors [12] proposed a technique to identify the Parkinsonian gait as well as forecast the severity of the disease based on gait data because gait disturbance is one of the key motor symptoms. The generic classifier was found to perform significantly worse than the sex-specific and age-dependent classifier [13]. In comparison to the generalized classifier's accuracy of 75.76% and the female-specific classifier's accuracy of 83.75%, the old-age dependent classifier's accuracy of 79.55%, was also noticed. The authors draw the conclusion that combining the sex and age information was successful in classifying the samples. In a different classification category, a certain set of traits was shown to be predominating for improved classification accuracy.

This proposed study [14], basically looked at touchscreen typing properties including descriptive statistics (covariance, skewness, and kurtosis) and temporal information to look for indicators of PD motor. This study [15] integrated data from multiple sources, including clinical, imaging, genetics, as well as demographic data, when constructing models for PD prediction. In [16], three widely used ML algorithms namely SVM, RF, as well as neural networks were used to analyze speech acoustics in order to identify PD. The promising outcomes of SVM as well as RF in the early detection of PD have been demonstrated.

In their proposed approach, the authors of this study [17] utilized decision trees, neural networks, and regression analysis. The performance score of the classifiers was determined using a variety of evaluation schemas. A comparative investigation was also the goal. The classifier using neural networks produced the highest score. The findings of the experiment improved the classification accuracy of neural networks by 92.9%. The authors of [18] used a long short term memory algorithm to identify the Freezing of Gait (FOG), a reliable sign of PD patients who may trip and fall. The premotor or pro-dromal stage in PD should be closely watched in order to ensure early diagnosis of the disorder [19]. Other than the typical motor symptoms, this premotor stage is typically characterized by symptoms including Rapid Eye Movement (REM) loss of olfaction and sleep behavior disorder.

The authors in this study [20] used Artificial neural networks (ANN) and KNN on UCI ML repository and achieved an accuracy of 91.28% for KNN_Multilayer Perceptron (MLP) and 90.76% accuracy with KNN+Bagging. The authors in this [21] employed many ML algorithms on the dataset collected from participants, resulting in an accuracy of 86% with SVM. This study [22] used mPower database and the results showed that highest accuracy achieved from gradient boosted trees. The authors in this employed many ML algorithms on the dataset collected from participants, resulting in an accuracy of 89.3% with SVM [23]. Hidden Markov Models (HMM) technique was implemented in this study [24] to achieve an accuracy of 95.16%. The authors in this study [25] implemented ensemble method to produce an accuracy of 90.6%. Montana D et al. employed SVM on UCI ML repository to achieve an accuracy of 94.4% with 10-fold cross validation [26]. This study [27] used XGBoost to achieve 96% accuracy. An accuracy of 97.57% was achieved using SVM when compared to RF and SVN in this study [28].

Patients with PD have optimism for their prognosis and future results thanks to emerging medicines such new symptomatic medications, creative drug delivery methods, and novel surgical approaches [29]. Current management techniques for both motor and nonmotor symptoms in the various stages of Parkinson disease are presented in this review [30]. Based on current theories and the most recent research, the authors [31] offered an outline of novel techniques in this area. While translational research on PD has made significant progress in many areas, there is still a need for more potent therapy alternatives based on understanding of the fundamental biological processes. Freezing of gait (FOG) greatly impacts the daily life of patients with PD. Elastic net-support vector machine models, which had an accuracy range of 0.69 to 0.78, outperformed all other ML techniques. The key structural morphological features, which were primarily distributed in the left cerebrum, were used to predict FOG using elastic net-support vector machine models [32]. With an average area under the curve (AUC) of 0.92 (95% CI: 0.95 0.01) for the slower progressing group (PDvec1), 0.87 0.03 for moderate progressors, and 0.95 0.02 for the fast-progressing group (PDvec3), the authors of this study [33] were able to make extremely accurate predictions of disease progression five years after initial diagnosis. Among other important biomarkers of interest, serum neurofilament light was found to be a strong predictor of rapid illness development.

The authors [34] discovered persistent spatial clustering of incident PD diagnoses in the U.S. PD incidence estimates varied across our data sources, possibly as a result of population factors (prevalence of genetic risk factors or protective markers) and geographic factors (exposure to environmental toxins), as well as case ascertainment and diagnosis methods. In this work, four alternative ML algorithms were used to assess selected baseline variables divided into three subgroups of clinical, biofluid, and genetic/epigenetic data [35]. Regardless of the machine learning technique employed, models based on clinical variables performed the best and demonstrated better prognosis of cognitive impairment outcome than dementia conversion. In this work, nine different machine learning classifiers were used [36]. The authors found that early-stage PD patients and controls could be distinguished using combined face and speech data with an area under the receiver operating characteristic (AUROC) diagnostic value of 0.85. The ideal diagnostic value (0.90) remained in the validation cohort. We came to the conclusion that combined speech and facial expression biometrics could help distinguish early-stage PD patients from elderly controls. The PD Biomarkers Programme (PDBP) cohort was trained using data from the Parkinson's Progression Markers Initiative (PPMI) and subjected to independent assessment [37]. When assessed on a hold-out PPMI set, 12-month PD total progression was predicted with an F-measure of 0.77, ROC AUC of 0.77, and PR AUC of 0.76. They obtained an F-measure of 0.75, ROC AUC of 0.74, and PR AUC of 0.73 when tested on PDBP.

The symptoms of Parkinson disease can be different from patient to patient. Early sign of the disease can be simple and can go sometimes unnoticed. Sometimes it is difficult to detect whether there is PD disease present in the patient's body. PD if detected in the early stage will be curable, and will be time and cost effective, but there is no effective treatment in the advanced stage.

3. Methodology

A ML framework is suggested in this article for the early detection of PD. Fig. 2 depicts the overall layout of the suggested detection strategy. Training and testing are the two basic phases of PD detection. Initially, the data is pre-processed, feature extraction is done before the ML model is built. These ML models built are then evaluated for PD detection in the testing step.

3.1. Data pre-processing

In this study, the dataset used from Kaggle data bank. The dataset has 24 attributes (including name), thus the dimension has to be reduced before training the data. The target or independent variable is "status" with binary values of 0 and 1. Status values for healthy person and PD person are 0 and 1 respectively. This is a classification problem.

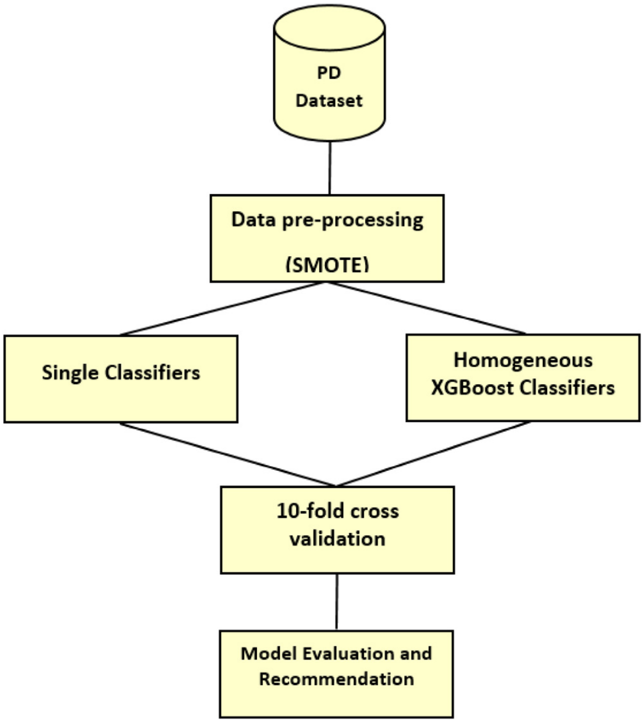


Fig. 2. Proposed methodology.

The purpose of this work is to develop the best ML model to predict the PD so that patient can be treated in the timely manner. Histograms are a common way to visualize the distribution of numerical features. We can create histograms for each numerical feature to see how the data is distributed. Matplotlib in Python has been used to create histograms. By using these visualization and summary statistics techniques, we gain insights into the distribution of numerical features in the PD dataset. Histograms display the frequency or count of data points within each bin or interval. This helps in identifying which values or ranges of values are more common or rare in the dataset. The distribution of the numerical features of the dataset is shown in Fig. 3.

There are 24 attributes including name. The dimension needs to be reduced. Dimensionality is reduced based on the correlation coefficient. The predictors which are highly correlated are dropped. Now, there are 11 features as it is depicted in the Pearson Correlation coefficient graph in Fig. 4.

PD dataset has features like vocal frequency, tremor intensity, muscle rigidity, and gait pattern. These models provide a score indicating the usefulness of each feature in predicting the target variable.

After training the model, we get feature importance scores as:

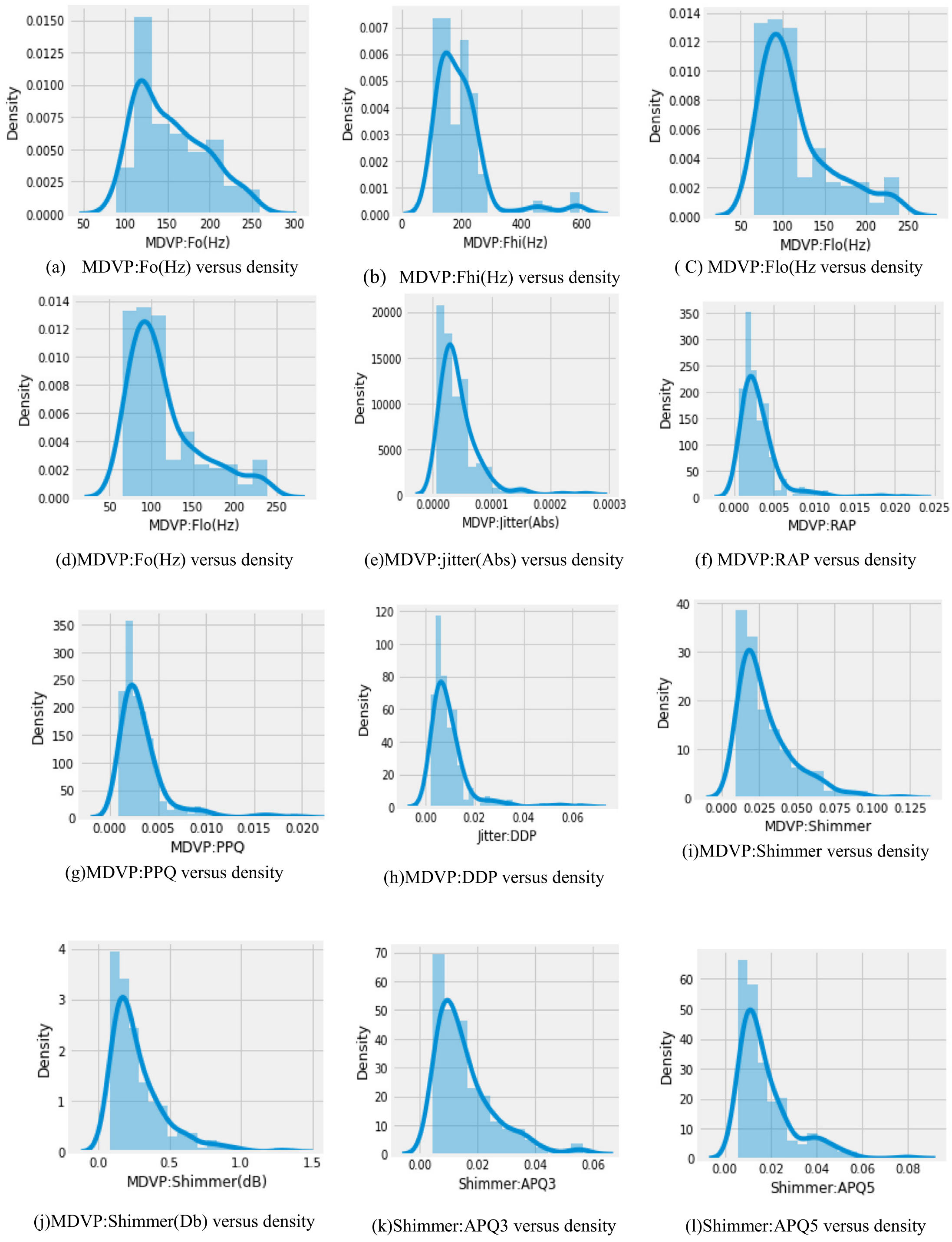
Vocal Frequency: 0.45
Tremor Intensity: 0.35
Muscle Rigidity: 0.15
Gait Pattern: 0.05

The study uses advanced algorithms like XGBoost and Random Forest, a typical approach for feature selection in such scenarios is Recursive Feature Elimination (RFE).

Suppose the PD dataset has features like vocal frequency, tremor intensity, muscle rigidity, and gait pattern. If we are using RFE with a tree-based model:

The model is trained on the initial set of features and weights are assigned to each one of them. The least important features (based on weights) are pruned from the current set of features.

The model is then retrained on the pruned subset of features. This process is recursively repeated until the desired number of features is reached.



(a) MDVP:Fo(Hz) versus density

(b) MDVP:Fhi(Hz) versus density

(C) MDVP:Flo(Hz versus density

(d)MDVP:Fo(Hz) versus density

(e)MDVP:jitter(Abs) versus density

(f) MDVP:RAP versus density

(g)MDVP:PPQ versus density

(h)MDVP:DDP versus density

(i)MDVP:Shimmer versus density

(j)MDVP:Shimmer(Db) versus density

(k)Shimmer:APQ3 versus density

(l)Shimmer:APQ5 versus density

Fig. 3. Numerical features distribution.

The dimensionality of the dataset is now under control as shown in Fig. 3. There are total 195 records in the dataset. Among them, 147 are positive results and 48 are negative results as shown in Fig. 5. The dataset is imbalanced as it contains 75% positive results and 25% negative results.

When there is an imbalance in a class, the machine learning model has a bias and tends to anticipate the majority class. An oversampling technique called SMOTE is being used to balance the dataset's class distribution. Thus, a new sample is drawn at a location along the line that is drawn between the examples in the features space.

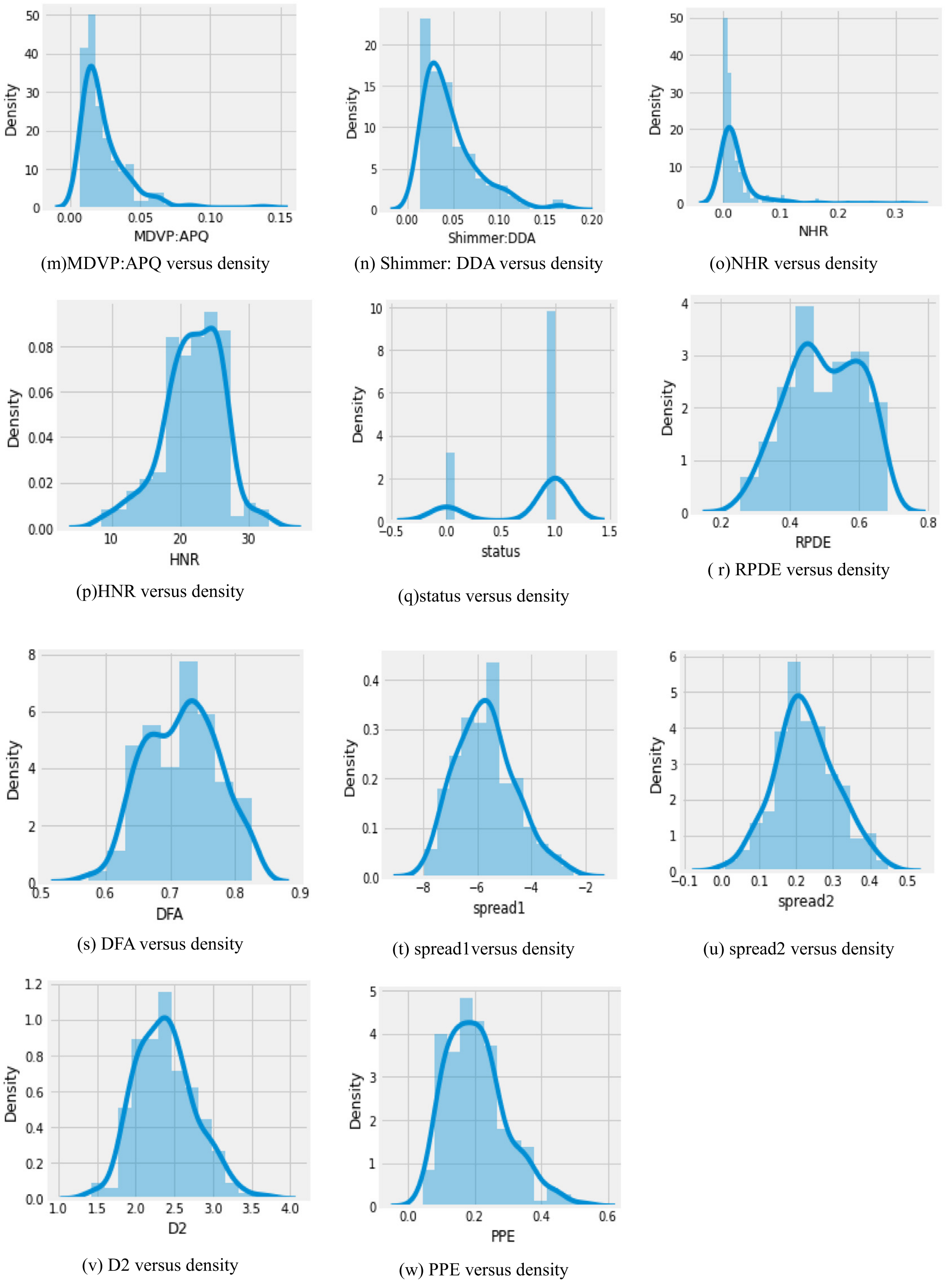
In simple words, the technique uses K Nearest Neighbors to choose a random neighbor and a random example from the minority class. Between two instances in the feature space, the synthetic example is produced. The SMOTE samples are defined as linear combinations of two comparable minority class samples (X *and* X^R) as shown in Eq. (1).

$$S = X + u \cdot (X^R - X) \tag{1}$$

Where, $0 \le u \le 1$; X^R is chosen randomly among five minority class nearest neighbors of X.

X^R and X must be assumed to be independent and to have the same predicted value for the majority of the proofs ($E(\cdot)$) as shown in equation (2) and ($var(\cdot)$) as shown in Eq. (3).

$$(E\left(X_j^{SMOTE}\right) = E(X_j)) \tag{2}$$



(m)MDVP:APQ versus density (n) Shimmer: DDA versus density (o)NHR versus density

(p)HNR versus density (q)status versus density (r) RPDE versus density

(s) DFA versus density (t) spread1versus density (u) spread2 versus density

(v) D2 versus density (w) PPE versus density

Fig. 3. (*continued*).

But smaller variance

$$(var\left(X_j^{SMOTE}\right) = \frac{2}{3}var\left(X_j\right)) \tag{3}$$

SMOTE reduces the variability of the (SMOTE-augmented) minority class while maintaining the expected value for that class [38]. The predicted value for SMOTE samples is identical to that of the initial minority class samples. Although SMOTE increases sample size and decreases variance, it does not significantly alter the difference between sample means [39]. SMOTE avoids adding correlation between several variables. The SMOTE samples and the samples from the minority class that were utilized to create them have a substantial positive correlation.

3.2. ML models

The different ML techniques namely KNN, SVM, RF and XGBoost are used to train the model.

3.2.1. KNN

KNN is non-parametric since the model is distributed from the data and no assumptions are made about the data being investigated.

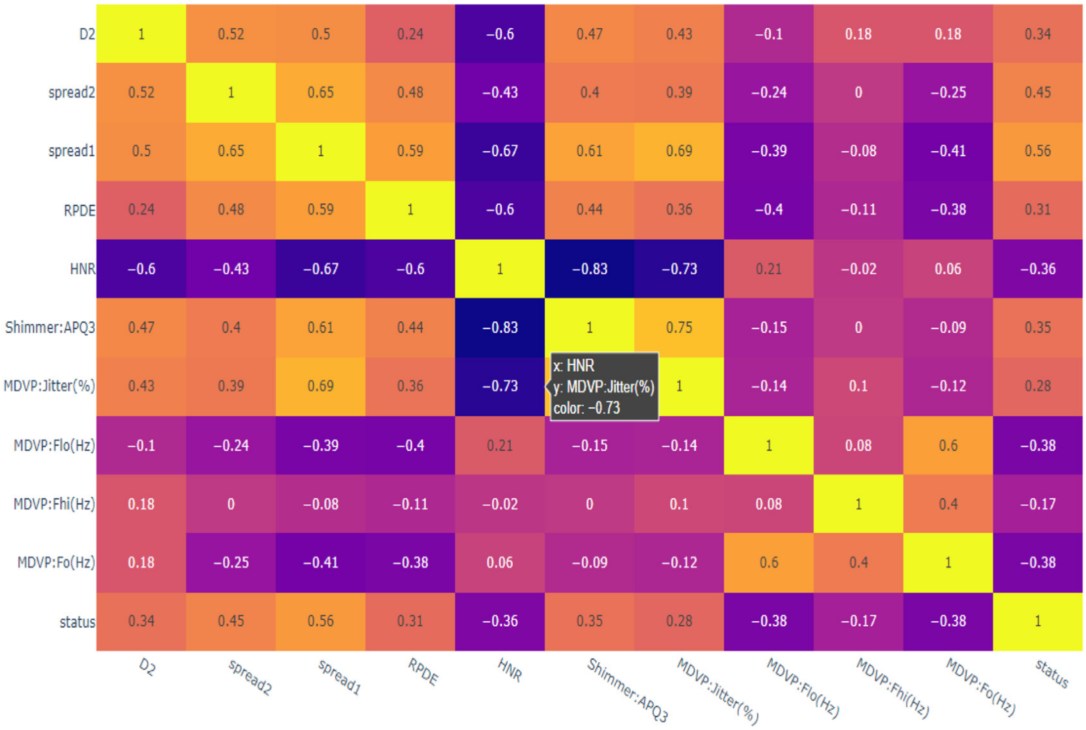


Fig. 4. Parkinson's Disease features correlation plot.

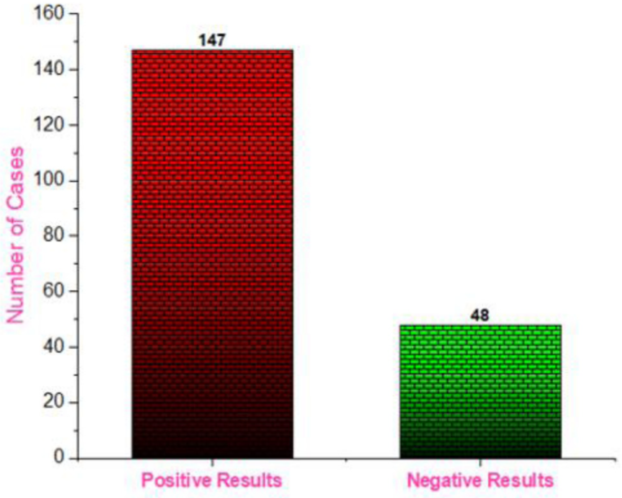


Fig. 5. Positive and negative cases of Parkinson Disease.

Both regression and classification prediction issues can be solved using KNN. Yet, since it performs well across all factors considered when establishing the applicability of a strategy, it is primarily employed in categorization when it comes to medical field challenges. One of the fundamental machine learning algorithms is KNN. A set of input values are used by machine learning models to forecast output values. One of the simplest machine learning algorithms, KNN is primarily employed for categorization [39]. The data point is categorized based on how its neighbor is categorized. Based on the similarity score of the previously stored data points, KNN categorizes the new data points. To get the most out of the model, it is essential to select a suitable value for k in the KNN. Since there are few votes, the model's error rate will be high if K is low, especially for fresh data points [40]. As a result, the model is over fitted and extremely sensitive to input noise.

KNN is useful in situations where data do not match up with theoretical predictions in the real world. But KNN is also a lazy algorithm, meaning it may generate models without using any training data. The testing phase used all training data. As a result, training becomes quicker, and testing becomes slower and more expensive.

3.2.2. SVM

Building models for classification and regression may both be done using the straightforward yet effective SVM technique. Both linearly separable and non-linearly separable datasets can yield excellent results when using the SVM method. The SVM algorithm continues to work its magic even with scant data. The SVM technique is built around the idea of "decision planes", where hyperplanes are employed to categorize a given set of objects [41].

Making a straight line between two classes is how a straightforward linear SVM classifier functions. In other words, the data points on one side of the line will all be assigned to one category, while the data points on the other side of the line will be assigned to a different category. This implies that the number of possible lines is unlimited. Because it selects the optimal line to categorize the data points, the linear SVM algorithm is superior to several other algorithms like KNN. It selects the line that divides the data and is as far from the nearest data points as it may be. Making sense of all the machine learning lingo is made easier by using a 2-D illustration. In essence, we have a grid of data points. We are attempting to group these data points according to the category they belong in, but we do not want any data to be placed in the incorrect category. To keep the other data points apart, we must find a line connecting the two points that are closest to one another. The support vectors you will employ to locate that line are therefore provided by the two nearest data points. That line is called the decision boundary.

3.2.3. RF

Several classification trees are cultivated in RFs. Place the input vector below each tree in the forest in order to categorize a new object from an input vector. Each tree assigns a category, and we refer to this as the tree "voting" for that category [42]. The classification that receives the most votes is selected by the forest. There is no overfitting in random forest. The issue of overfitting does not arise here because they are built from subsets of data and the outcome is based on average or majority rating. RF develops a decision tree from observations that are chosen at random, and then the outcome is determined by majority voting. Here, formulas are not necessary.

The following steps explain the working Random Forest Algorithm:

Step 1: Select random samples from a given data or training set.

Step 2: This algorithm will construct a decision tree for every training data.

Step 3: Voting will take place by averaging the decision tree.

Step 4: Finally, select the most voted prediction result as the final prediction result.

Bagging is the process of generating an alternative training subset via replacement from a sample training dataset. The outcome is decided by a majority vote. In random forest, bagging is sometimes referred to as Bootstrap Aggregation. Starting with any initial random data, the process begins. After arranging, it is divided into Bootstrap Sample samples. Bootstrapping is the name for this procedure. Additionally, each model is trained separately, producing distinct outcomes known as Aggregation. The final stage combines all the findings, and the output that is produced is based on majority voting. The Bagging phase of the process makes use of an Ensemble Classifier.

3.2.4. XGBoost

A strong machine learning tool, that is open-source is XGBoost. It functions by fusing decision trees and gradient boosting, and is expected to assist in the development of better models. Large dataset performance, usability, and speed are all priorities in the design of XGBoost [43]. It does not require parameter optimization or adjustment; therefore, it may be used right away after installation, with no additional settings. Using the weighted quantile sketch algorithm, XGBoost also has the capacity to handle sparse datasets. By maintaining the same level of computational complexity as previous algorithms like stochastic gradient descent, this algorithm enables us to cope with feature matrices that include non-zero elements. Moreover, XGBoost provides a block structure for concurrent learning. It makes scaling up on multicore computers or clusters simple. Moreover, it makes advantage of cache awareness, which lowers memory consumption when training models using sizable datasets. During the computation phase, XGBoost uses disk-based data structures rather than in-memory ones to provide out-of-core processing capabilities. XGBoost is preferred because of its great execution speed [44].

The working of XGBoost is as follows: Let us consider DS as a dataset, which has m features and n number of instances. $DS = \{(x_i, y_i) : i = 1 \ldots .n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}\}$. Let us consider $\hat{y}_i$ as the predicted outcome of an ensemble model created from the Eq. (4).

$$\dot{A}\cdot_i = \varnothing(X_i) = \sum_{k=1}^{K} f_k(X_i), f_k \in \ell \tag{4}$$

Where, K represents number of trees, f_k denotes kth tree. Now, we need to compute the best function by reducing the loss and regularization

$$\mathcal{L}(\varnothing) = \sum_i l(y_i, \dot{A}\cdot_i) + \sum_k \Omega(f_k) \tag{5}$$

Where l denotes the loss function, it is the difference between $\boldsymbol{y_i}$, **the actual output and** $\hat{y}_i$, **the predicted output**

Ω represents a measure of how complexity, the model is, this in fact, helps in avoiding

$$\Omega(f_k) = \Upsilon T + \frac{1}{2}\lambda||w||^2 \tag{6}$$

over-fitting. It is computed using the Eq. (6)

Where T denotes the number of trees of the tree and w denotes the weight of the leaf. Boosting is used in the training model in case of DTs to reduce the objective function [45]. So, tth iteration, a new function is included as shown in equations from (7) to (9).

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \dot{A}\cdot_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{7}$$

$$\mathcal{L}_{split} = \frac{1}{2}\left[\frac{(\sum_{i\epsilon I_L} g_i)^2}{\sum_{i\epsilon I_L} h_i + \lambda} + \frac{(\sum_{i\epsilon I_R} g_i)^2}{\sum_{i\epsilon I_R} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\epsilon I} h_i + \lambda}\right] - \Upsilon \tag{8}$$

Where

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \dot{A}\cdot_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{9}$$

$$and \quad h_i = \delta^2{}_{\dot{A}\cdot(t-1)} l(y_i, \dot{A}\cdot^{(t-1)}) \tag{10}$$

3.3. Performance measures

The effectiveness of single classifiers and their ensemble with XGBoost in diagnosing as well as predicting PD is done in this section. Depending on only the accuracy for the effectiveness, especially when we are dealing with medical dataset is not sufficient. As a result, in addition to accuracy, the effectiveness of the classifier models is evaluated using measures including f1 measure, precision, sensitivity, Matthew's Correlation coefficient (MCC) and specificity. Using the parameters derived from the confusion matrix, namely True Positive (TPs), which predicted PD as true and in reality it is true, True Negative (TNs), which predicted PD as false and in reality it is false. False Positive (FPs), which predicted PD as true and in reality it is false and False Negative (FNs), which predicted PD as false and in reality in true, the effectiveness of the classifier is experimentally assessed [46]. The equations to compute the different performance metrics are shown from (11) to (17).

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{11}$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{12}$$

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \tag{13}$$

$$F1\ \text{Score} = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \tag{14}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{15}$$

Matthew's Correlation Coefficient (MCC), is computed using Eq. (15). MCC is a metric used to assess the effectiveness of a binary classifier for identifying PD in a patient. MCC ranges from −1 to 1, where −1 denotes a binary classifier that is entirely incorrect and 1 denotes a binary classifier that is entirely correct [47].

$$MCC = \frac{TPs * TNs * FPs * FNs}{\sqrt{(TPs + FPs)(TPs + FNs)(TNs + FPs)(TNs + FNs)}} \tag{16}$$

The amount of time needed to complete training or modeling a dataset is known as Time Took to Build the Model (TTBM). It is computed using the Eq. (17).

$$TTDM = Time\ to\ complete\ the\ training\ of\ PD\ dataset\ in\ seconds \tag{17}$$

The average of the discrepancy between the original values as well as the predicted values is called as the mean absolute error (MAE) [48]. It provides a measurement of how far the projections missed the actual output. However, it does not indicate whether the error is under- or over-predicting the data, therefore there is no way to tell which is the case. We use the absolute value of the distances so that negative errors are accounted properly. The model is more accurate the closer MAE is to zero. It is denoted mathematically by the following Eq. (18).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i^{real} - y_i^{pred}| \tag{18}$$

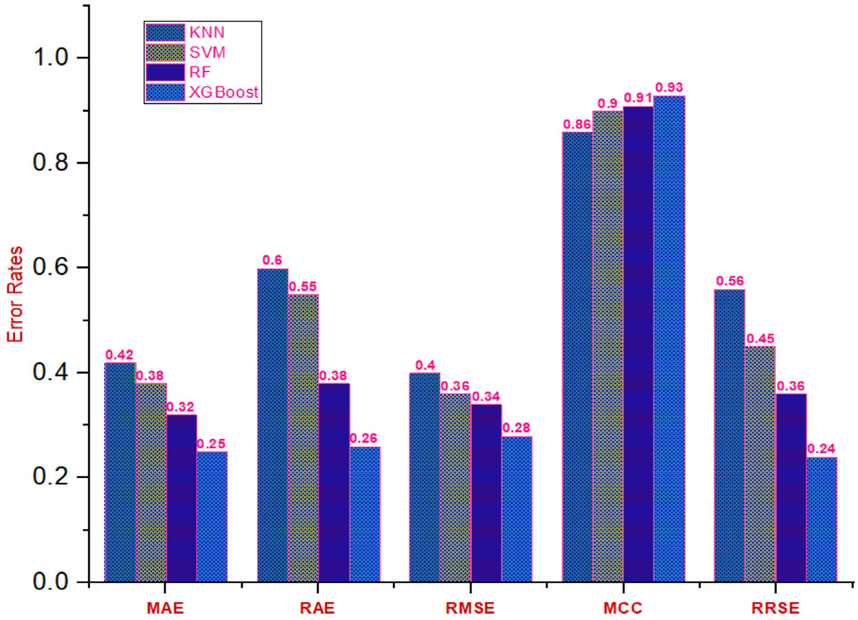


Fig. 6. Error rates of Single classifiers.

The main distinction between Mean Squared Error (MSE) and Mean Absolute Error is that MSE takes the average of the square of the discrepancy between the actual values and the predicted values [49]. The benefit of MSE is that it is simpler to compute the gradient than MAE. MSE is computed using (19).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i^{real} - y_i^{pred})^2 \tag{19}$$

Using root-mean-squared error (RMSE), the MSE error is square-rooted to return it to its original unit while keeping the property of punishing larger mistakes [50]. The model can be deemed to be reasonably accurate in predicting the data if the RMSE values are between 0.2 and 0.5. RMSE is computed as shown in (20).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{real} - y_i^{pred})^2} \tag{20}$$

The total absolute error is normalized by dividing it by the total absolute error of the simple predictor to get the relative absolute error (RAE). Equation shown in (21) evaluates an individual model's relative absolute error $E_{i.}$

$$E_i = \frac{\sum_{j=1}^{n}|P_{(ij)} - T_j|}{\sum_{j=1}^{n}|P_j - T|} \tag{21}$$

Where $P_{(ij)}$ is the value predicted by the individual model i for record of j(of n records); T_i the target value for record j and T is given by the formula (22).

$$\overline{T} = \frac{1}{n}\sum_{j=1}^{n}T_j \tag{22}$$

4. Results and discussions

The error is reduced to the same dimensions as the prediction using the root relative squared error (RRSE). Total square error divided by total square error of a straightforward predictor yields relative square error. Ratios are used to express the RSE output value. It might be between 0 and 1. A realistic model should have a value that is near to zero, while one with a number higher than 1 is not. The formula shown in (23) determines the root of the individual model *j's* relative square error $E_{i.}$

$$E_i = \sqrt{\frac{\sum_{j=1}^{n}(P_{(ij)} - T_j)^2}{\sum_{j=1}^{n}(T_j - \overline{T}^2)}} \tag{23}$$

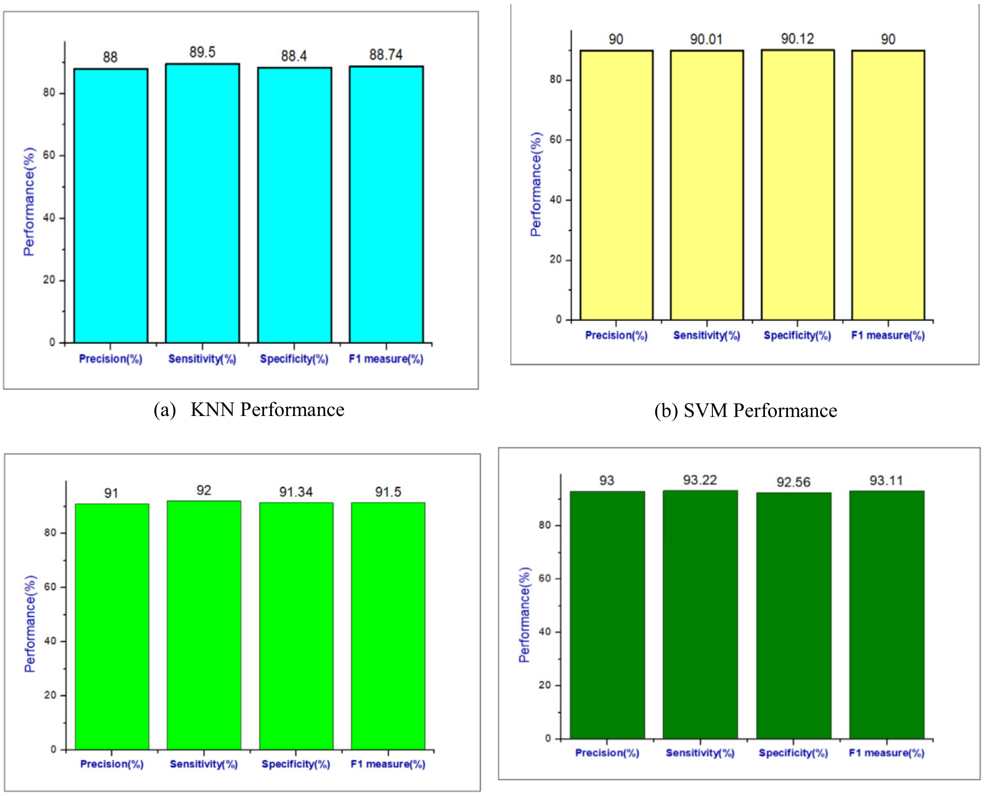
Table 1
Performance of individual classifiers.

Performance metrics	KNN	SVM	RF	XGBoost
TTBM(s)	30.15	22.55	14.66	12.88
Accuracy (%)	87.60	91.12	93.67	94.89
Precision (%)	88.00	90.00	91.00	93.00
Sensitivity (%)	89.50	90.01	92.00	93.22
Specificity (%)	88.40	90.12	91.34	92.56
F1 measure (%)	88.74	90.00	91.50	93.11
MAE	0.42	0.38	0.32	0.25
RAE	0.60	0.55	0.38	0.26
RMSE	0.40	0.36	0.34	0.28
MCC	0.86	0.90	0.91	0.93
RRSE	0.56	0.45	0.36	0.24

Using stratified K-Fold cross-validation, the models are refined and the hyperparameters are tweaked. The method most frequently employed for hyper parameter optimization is grid search. For each hyper parameter, we first create a set of values. The model then chooses the hyperparameter with the highest performance after evaluating the hyperparameters for each possible value. For 10-fold cross-validation, the PD dataset is divided into 10-folds of identical size. After the K-1 group training is over, the classifiers are tested in the remaining time. The performance of the classifiers is also assessed for each k. Lastly, an evaluation classifier based on average performance is created. The performance of individual classifiers against various performance parameters is tracked in Table 1.

From Table 1 it is observed that XGBoost has taken only 12.88 s to build the model. The error rates of XGBoost are 0.25, 0.26, 0.28, 0.93 and 0.24 MAE, RAE, RMSE, MCC and RRSE respectively and comparatively better when compared to other models as shown in Fig. 6.

The precision values of KNN, SVM, RF and XGBoost are 88.00, 90.00, 91.00 and 93.00 respectively. The sensitivity values of KNN,



(a) KNN Performance

(b) SVM Performance

(c) RF Performance

(d) XGBoost Performance

Fig. 7. Performance of single classifiers.

SVM, RF and XGBoost are 89.50, 90.01, 92.00 and 93.22 respectively. The specificity values of KNN, SVM, RF and XGBoost are 88.40, 90.12, 91.34 and 92.56 respectively. The f1 score values of KNN, SVM, RF and XGBoost are 88.74, 90.00, 91.50 and 93.11 respectively. Fig. 7 depicts these performance values of single classifiers.

The precision, specificity, sensitivity and f1 score values of XGBoost are comparatively better than any individual classifiers. The accuracy of KNN, SVM, RF and XGBoost are 87.60%, 91.12%, 93.67% and 94.89% respectively. It is observed that XGBoost has outperformed with an accuracy of 94.89% compared to other models and KNN has exhibited the poor performance with 87.60% of accuracy as shown in Fig. 8.

XGBoost, when evaluated as an individual classifier, the performance is found to be better than other classifiers. Further, every single classifier mentioned in Table 1 is ensembled with XGBoost classifier. There are three homogeneous ensemble combinations possible namely, XGBoost-KNN, XGBoost-SVM and XGBoost-RF as shown in Table 2.

XGBoost-RF has taken only 10.22 s to build the model which is lesser than XGBoost single classifier. However, XGBoost-KNN has taken 32.15 s to build the model, whereas KNN classifier had taken 30.15 s, which means the homogeneous XGBoost-KNN has taken 2 s more than that of individual KNN classifier. The MAE, RAE, RMSE, MCC and RRSE values of XGBoost-RF are 0.22, 0.29, 0.30, 0.98 and 0.30 respectively as shown in Fig. 9. This clearly shows that XGBoost-RF has the least error rates and provides almost correct predictions.

The precision values of XGBoost-KNN, XGBoost-SVM and XGBoost-RF are 91.00, 93.00 and 97.24 respectively. The sensitivity values of XGBoost-KNN, XGBoost-SVM and XGBoost-RF are 90.50, 94.00 and 97.56 respectively. The specificity values of XGBoost-KNN, XGBoost-SVM and XGBoost-RF are 91.40, 93.12 and 97.00 respectively. The f1 score of XGBoost-KNN, XGBoost-SVM and XGBoost-RF are 90.75, 93.50 and 97.40 respectively. It is a clear evident that the homogeneous XGBoost-RF classifier has performed better than other classifiers as shown in Fig. 10.

Table 2
Performance of homogeneous XGBoost classifiers.

Performance metrics	XGBoost-KNN	XGBoost-SVM	XGBoost-RF
TTBM (sec)	32.15	20.55	10.22
Accuracy (%)	91.55	94.66	98.00
Precision (%)	91.00	93.00	97.24
Sensitivity (%)	90.50	94.00	97.56
Specificity (%)	91.40	93.12	97.00
F1 measure (%)	90.75	93.50	97.40
MAE	0.41	0.30	0.22
RAE	0.50	0.45	0.29
RMSE	0.30	0.31	0.30
MCC	0.91	0.94	0.98
RRSE	0.46	0.40	0.30

The accuracy of XGBoost-KNN, XGBoost-SVM, and XGBoost-RF are 91.55%, 94.66%, and 98.00% respectively as shown in Fig. 11. It is observed that XGBoost-RF has outperformed with an accuracy of 98.00%% compared to other homogeneous classifiers as shown in Fig. 10 and XGBoost-KNN has the least performance of 91.55%. So, it is a clear evident that out of all the KNN, SVM, RF, XGBoost, XGBoost-SVM, XGBoost-KNN and XGBoost-RF classifiers, the homogeneous XGBoost-RF is recommended as the best model for the prediction PD.

Table 3 depicts the comparison of the proposed study with the existing state-of-art techniques. It is observed that the proposed model provides uses homogeneous XGBoost-Random Forest on Kaggle dataset

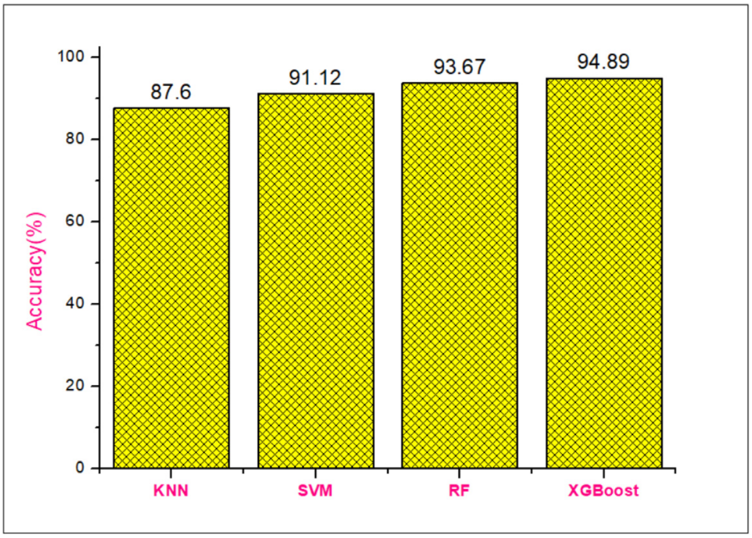


Fig. 8. Accuracy of single classifiers.

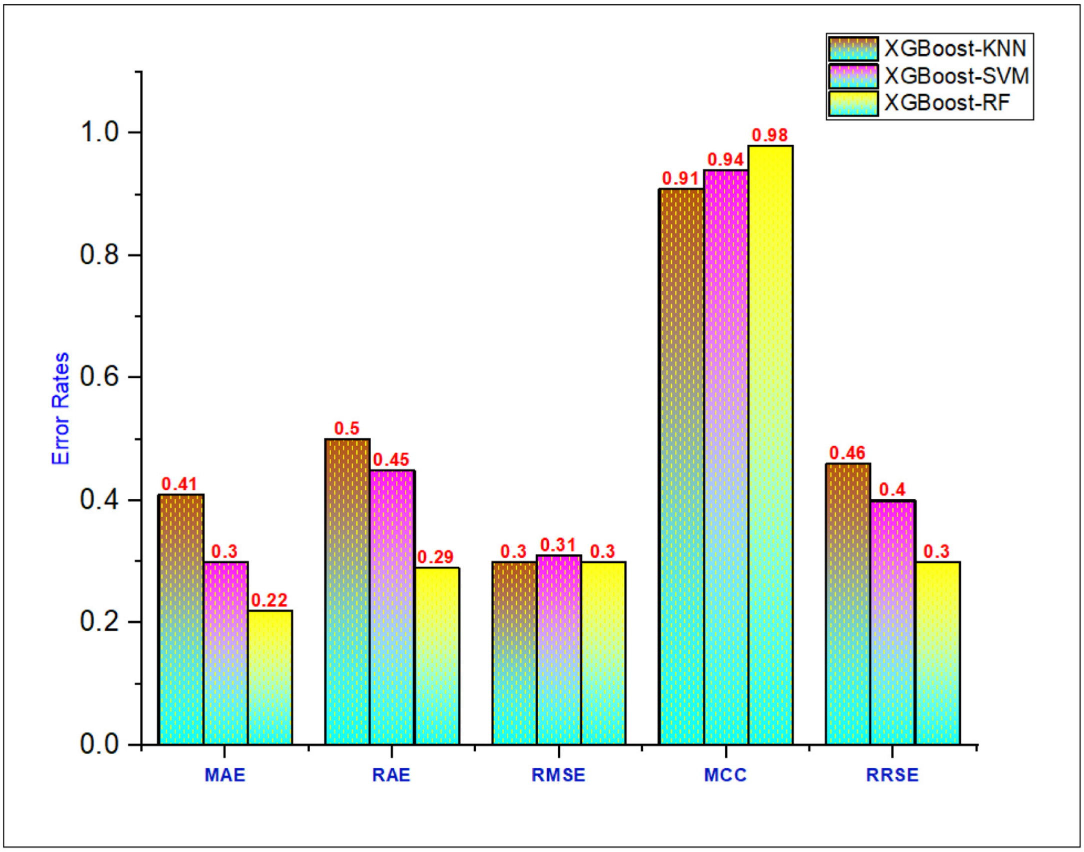
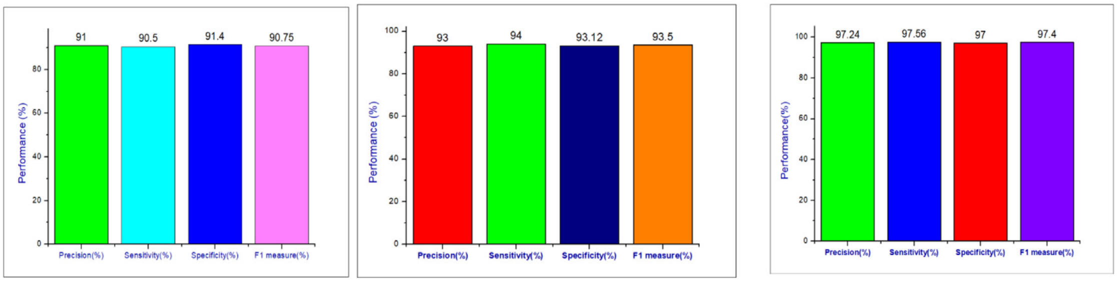


Fig. 9. Error rates of Homogeneous XGBoost classifiers.



(a)XGBoost-KNN performance

(b) XGBoost-SVM Performance

(c) XGBoost-RF Performance

Fig. 10. Performance of homogeneous XGBoost classifiers.

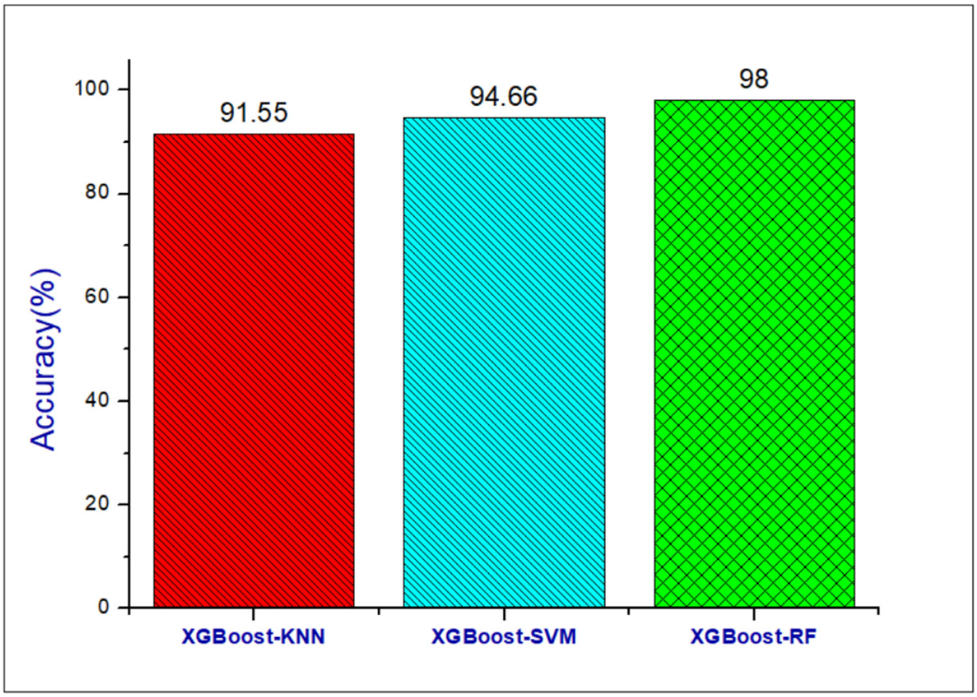


Fig. 11. Accuracy of homogeneous classifiers.

of PD providing 98% accuracy and outperforms compared to other models.

Table 3
Comparison with the existing methods.

Study and year	Algorithms used	Source of data	Outcomes
Sztaho D. et al. (2019) [23]	ANN, SVM, SVM with RBF, DNN, KNN	UCI ML repository	Highest accuracy achieved from SVM with RBF: 89.3%
M.S. Roobini et al. (2022) [27]	XGBoost	UCI ML repository	Accuracy −96%, Matthews parametric statistic (MCC) of 89%.
Ramakrishna, M.T. et al. (2022) [42]	Adaboost-random forest classifier.	UCI ML repository	Accuracy −97.95%,
Proposed model	XGBoost-Random Forest	UCI ML repository	Accuracy −98%

4.1. Implications of the study

The concept of using ML algorithms, particularly XGBoost and ensemble methods, to predict PD has both theoretical and practical implications.

4.1.1. Theoretical implications

- Model Efficacy with Medical Data: The study offers insights into how machine learning algorithms, specifically ensemble methods, perform when applied to medical datasets. The results can contribute to the broader literature on the applications and limits of machine learning in healthcare diagnostics.
- Ensemble Methods: The study suggests that the ensemble of XGBoost and Random Forest outperforms other algorithms, reinforcing the idea that combining predictions from multiple models often results in better performance than using individual models.
- Balancing Imbalanced Data: The use of SMOTE for handling imbalanced data in the study can serve as a case study for the effectiveness of such techniques in real-world applications. The results might offer insights into how oversampling impacts model performance, especially in medical datasets where imbalances are common.
- Evaluation Metrics: The emphasis on Matthew's correlation coefficient, alongside accuracy, underscores the importance of using

multiple evaluation metrics, especially when dealing with medical data where false positives or negatives can have significant implications.

4.1.2. Practical implications

- Enhanced PD Diagnostics: If the model consistently achieves high accuracy and other performance metrics in real-world applications, it could potentially be used as an aid in PD diagnosis, especially in early stages where symptoms might be less pronounced.
- Cost-Effectiveness: Implementing machine learning models for preliminary screening or diagnosis could reduce the number of unnecessary tests or misdiagnoses, leading to cost savings in healthcare.
- Addressing Data Imbalances: The success of SMOTE in this context can guide other researchers and practitioners in handling imbalanced data in their own projects, especially in medical domains.
- Framework for Other Diseases: The methodology used in this study could be adapted and applied to other diseases or conditions, potentially leading to breakthroughs in early diagnosis or risk prediction.
- Informed Treatment Decisions: Understanding the probability or risk of a patient having PD can lead to better-informed treatment decisions, potentially altering the patient's treatment path and improving outcomes.
- Awareness and Education: The study can serve as a basis for educating healthcare professionals about the potential of machine learning in diagnostics. This can lead to a more integrated approach to healthcare, combining traditional methods with advanced technologies.

5. Conclusions

Because Parkinson Disease symptoms overlap with those of other disorders, diagnosing the condition can be challenging. Also, the patient's health is more vulnerable due to ignorance. This frequently results in the disorder being misdiagnosed. Parkinson Disease diagnosis is a gradual process, hence Parkinson Disease cannot be identified in an individual by a single test like an ECG or blood test. Before doing any neurological testing, doctors must review the patient's medical history. A dilemma results from the frequent misdiagnosis of Parkinson's disease brought on by lengthy examinations. Data science and machine learning technologies frequently take advantage of this issue to simplify the diagnosis and care of Parkinson Disease patients. In this decision-making medical system, the different ML algorithms both single classifiers and homogeneous XGBoost classifiers are used for the prediction of Parkinson Disease with minimized entropy. Among all XGBoost-RF came out to be the best with 98% accuracy and Matthew's Correlation Coefficient value of 0.93. Hence, Parkinson's Disease can be predicted at an early stage using the medical history of individuals who exhibit certain Central Nervous System-related symptoms. Since there is currently no known cure for Parkinson Disease, early identification allows for early diagnosis.

Future studies might find it interestingly difficult to apply this approach to bigger datasets and, if possible, to assess it on a broader scale. Several optimization techniques, including Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and GA (Genetic Algorithm), can also be integrated with it. Applying these techniques will enable you to choose the ideal ensemble algorithm settings with accuracy. The future research could focus on collecting more extensive data, including a wider range of PD symptoms and stages. Additionally, experimenting with a broader set of machine learning algorithms are expected to provide more insights and potentially improve the accuracy and generalizability of the findings.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

References

[1] E.R. Dorsey, et al., Projected number of people with Parkinson disease in the most populous nations 2005 through 2030, Neurology 68 (5) (2007) 384–386, http://dx.doi.org/10.1212/01.wnl.0000247740.47667.03.

[2] Antoine Dumortier, Ellen Beckjord, Saul Shiffman, Ervin Sejdić, Classifying smoking urges via machine learning, Comput. Methods Programs Biomed. 137 (2016) 203–213, http://dx.doi.org/10.1016/j.cmpb.2016.09.016, (ISSN: 0169–2607).

[3] Gabriel Solana-Lavalle, Juan-Carlos Galán-Hernández, Roberto Rosas-Romero, Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features, Biocybern. Biomed. Eng. 40 (1) (2020) 505–516, http://dx.doi.org/10.1016/j.bbe.2020.01.003, (ISSN: 0208–5216).

[4] K. Marek, D. Jennings, Can we image premotor Parkinson disease? Neurology 72 (7 Supplement 2) (2009) 21–26.

[5] W.R. Adams, High-accuracy detection of early Parkinson's disease using multiple characteristics of finger movement while typing, PLoS One 12 (11) (2017) 0188226.

[6] B.E. Sakar, M.E. Isenkul, C.O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, O. Kursun, Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings, IEEE J. Biomed. Health Inform. 17 (4) (2013) 828–834.

[7] P. Martinez-Martín, et al., Parkinson symptoms and health related quality of life as predictors of costs: a longitudinal observational study with linear mixed model analysis, PLoS One 10 (12) (2015) e0145310, http://dx.doi.org/10.1371/journal.pone.0145310.

[8] E. Rovini, C. Maremmani, F. Cavallo, How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review, Front. Neurosci. 11 (2017) 555, http://dx.doi.org/10.3389/fnins.2017.00555.

[9] D. Long, et al., Automatic classification of early Parkinson's disease with multimodal MR imaging, PLoS One 7 (11) (2012) 1–9, http://dx.doi.org/10.1371/journal.pone.0047714.

[10] A. Rana, A. Dumka, R. Singh, M.K. Panda, N. Priyadarshi, B. Twala, Imperative role of machine learning algorithm for detection of Parkinson's disease: Review, challenges and recommendations, Diagnostics (Basel) 12 (8) (2022) 2003, http://dx.doi.org/10.3390/diagnostics12082003, PMID: 36010353; PMCID: PMC9407112.

[11] G. Solana-Lavalle, J.-C. Galán-Hernández, R. Rosas-Romero, Auto-matic Parkinson disease detection at early stages as a pre-diagnosis tooby using classifiers and a small set of vocal features, Biocybern. Biomed. Eng. 40 (1) (2020) 505–516.

[12] El Maachi, G.-A. Bilodeau, W. Bouachir, Deep 1d-convnet foraccurate Parkinson disease detection and severity prediction from gait, Expert Syst. Appl. 143 (2020) 113075.

[13] U. Gupta, H. Bansal, D. Joshi, An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement, Comput. Methods Programs Biomed. 189 (2020) 105305.

[14] T. Arroyo-Gallego, M.J. Ledesma-Carbayo, A. Sánchez-Ferro, I. But-terworth, C.S. Mendoza, M. Matarazzo, P. Montero, R. López-Blanco, V. Puertas-Martin, R. Trincado, et al., Detection of motor impairment inParkinson's disease via mobile touchscreen typing, IEEE Trans. Biomed. Eng. 64 (9) (2017) 1994–2002.

[15] I.D. Dinov, B. Heavner, M. Tang, G. Glusman, K. Chard, M. Darcy, R. Madduri, J. Pa, C. Spino, C. Kesselman, et al., Predictive big data ana-lytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations, PLoS One 11 (8) (2016) e0157077.

[16] D. Braga, A.M. Madureira, L. Coelho, R. Ajith, Automatic detectionof Parkinson's disease based on acoustic analysis of speech, Eng. Appl. Artif. Intell. 77 (2019) 148–158.

[17] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, Expert Syst. Appl. 37 (2) (2010) 1568–1572.

[18] A.S. Ashour, A. El-Attar, N. Dey, H.A. El-Kader, M.M.A. El-Naby, Long short term memory based patient-dependent model for fog detectionin Parkinson's disease, Pattern Recognit. Lett. 131 (2020) 23–29.

[19] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artificial Intelligence Res. 16 (2002) 341–378.

[20] R. Mathur, V. Pathak, D. Bandil, Parkinson disease prediction using machine learning algorithm, in: Emerging Trends in Expert Applications and Security, Springer, Singapore, 2019, pp. 357–363.

[21] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263.

[22] J.M. Tracy, Y. Özkanca, D.C. Atkins, R.H. Ghomi, Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease, J. Biomed. Inform. 104 (2020) 103362.

[23] D. Sztahó, I. Valálik, K. Vicsi, Parkinson's disease severity estimation on hungarian speech using various speech tasks, in: Proceedings of the 2019 International Conference on Speech Technology and Human-Computer Dialogue, SpeD, Timisoara, Romania, 2019, pp. 1–6, 10–12.

[24] H. Kuresan, D. Samiappan, S. Masunda, Fusion of WPT and MFCC feature extraction in Parkinson's disease diagnosis, Technol. Health Care 27 (2019) 363–372.

[25] R. Sheibani, E. Nikookar, S.E. Alavi, An ensemble method for diagnosis of Parkinson's disease based on voice measurements, J. Med. Signals Sens. 24 (2019) 221–226.

[26] D. Montaña, Y. Campos-Roca, J. Pérez Carlos, A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease, Comput. Methods Programs Biomed. 154 (2018) 89–97.

[27] M.S. Roobini, Y.R.K. Reddy, U.S.G. Royal, A.K. Singh, K. Babu, Parkinson's disease detection using machine learning, in: 2022 International Conference on Communication, Computing and Internet of Things, IC3IoT, Chennai, India, 2022, pp. 1–6, http://dx.doi.org/10.1109/IC3IOT53935.2022.9768002.

[28] Chen Tianqi, Carlos Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016.

[29] J. Jankovic, E.K. Tan, Parkinson's disease: Etiopathogenesis and treatment, J. Neurol. Neurosurg. Psychiatry 91 (2020) 795–808.

[30] P. Rizek, N. Kumar, M.S. Jog, An update on the diagnosis and treatment of Parkinson disease, CMAJ 188 (16) (2016) 1157–1165, http://dx.doi.org/10.1503/cmaj.151179, Epub 2016 May 24. PMID: 27221269; PMCID: PMC5088077.

[31] E. Dinter, T. Saridaki, L. Diederichs, et al., Parkinson's disease and translational research, Transl. Neurodegener. 9 (2020) 43, http://dx.doi.org/10.1186/s40035-020-00223-0.

[32] Y. Li, X. Huang, X. others Ruan, Baseline cerebral structural morphology predict freezing of gait in early drug-Naïve Parkinson's disease, NPJ Parkinsons Dis. 8 (2022) 176, http://dx.doi.org/10.1038/s41531-022-00442-4.

[33] A. Dadu, V. Satone, R. others Kaur, Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts, NPJ Parkinsons Dis. 8 (2022) 172, http://dx.doi.org/10.1038/s41531-022-00439-z.

[34] A.W. Willis, E. Roberts, J.C. others Beck, Incidence of Parkinson disease in North America, NPJ Parkinsons Dis. 8 (2022) 170, http://dx.doi.org/10.1038/s41531-022-00410-y.

[35] J. Harvey, R.A. Reijnders, R. others Cavill, Machine learning-based prediction of cognitive outcomes in de novo Parkinson's disease, NPJ Parkinsons Dis. 8 (2022) 150, http://dx.doi.org/10.1038/s41531-022-00409-5.

[36] W.S. Lim, S.I. Chiu, M.C. others Wu, An integrated biometric voice and facial features for early detection of Parkinson's disease, NPJ Parkinsons Dis. 8 (2022) 145, http://dx.doi.org/10.1038/s41531-022-00414-8.

[37] H.J. Sadaei, A. Cordova-Palomera, J. others Lee, Genetically-informed prediction of short-term Parkinson's disease progression, NPJ Parkinsons Dis. 8 (2022) 143, http://dx.doi.org/10.1038/s41531-022-00412-w.

[38] J.H. Joloudari, A. Marefat, M.A. Nematollahi, S.S. Oyelere, S. Hussain, Effective class-imbalance learning based on SMOTE and convolutional neural networks, Appl. Sci. 13 (2023) 4006, http://dx.doi.org/10.3390/app13064006.

[39] Nur Ulfa Maulidevi Asniar, Kridanto Surendro, SMOTE-LOF for noise identification in imbalanced data classification, J. King Saud Univ. - Comput. Inform. Sci. (ISSN: 1319-1578) 34 (6, Part B) (2022) 3413–3423, http://dx.doi.org/10.1016/j.jksuci.2021.01.014.

[40] K. Taunk, S. De, S. Verma, A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, in: 2019 International Conference on Intelligent Computing and Control Systems, ICCS, Madurai, India, 2019, pp. 1255–1260, http://dx.doi.org/10.1109/ICCS45141.2019.9065747.

[41] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, Asdrubal Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing (ISSN: 0925-2312) 408 (2020) 189–215, http://dx.doi.org/10.1016/j.neucom.2019.10.118.

[42] M.T. Ramakrishna, V.K. Venkatesan, I. Izonin, M. Havryliuk, C.R. Bhat, Homogeneous adaboost ensemble machine learning algorithms with reduced entropy on balanced data, Entropy 25 (2) (2023) 245.

[43] M. Chen, Q. Liu, S. Chen, Y. Liu, C.-H. Zhang, R. Liu, XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system, IEEE Access 7 (2019) 13149–13158, http://dx.doi.org/10.1109/ACCESS.2019.2893448.

[44] Ting Hu, Ting Song, J. Phys. Conf. Ser. 1324 (2019) 012091, http://dx.doi.org/10.1088/1742-6596/1324/1/012091.

[45] Ahmedbahaaaldin Ibrahem Ahmed Osman, Ali Najah Ahmed, Ming Fai Chow, Yuk Feng Huang, Ahmed El-Shafie, Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, Ain Shams Eng. J. 12 (2) (2021) 1545–1556, http://dx.doi.org/10.1016/j.asej.2020.11.011, (ISSN: 2090–4479).

[46] B.N. Kumar, T.R. Mahesh, G. Geetha, S. Guluwadi, Redefining retinal lesion segmentation: A quantum leap with DL-UNet enhanced auto encoder-decoder for fundus image analysis, IEEE Access 11 (2023) 70853–70864, http://dx.doi.org/10.1109/ACCESS.2023.3294443.

[47] V.K. Venkatesan, M.T. Ramakrishna, I. Izonin, R. Tkachenko, M. Havryliuk, Efficient data preprocessing with ensemble machine learning technique for the early detection of chronic kidney disease, Appl. Sci. 13 (5) (2023) 2885.

[48] P. Zhang, Y. Jia, Y. Shang, Research and application of XGBoost in imbalanced data, Int. J. Distrib. Sensor Netw. 18 (6) (2022) http://dx.doi.org/10.1177/15501329221106935.

[49] Kumar Sanjeev Priyanka, IOP Conf. Ser. Mater. Sci. Eng. 1022 (2021) 012071, http://dx.doi.org/10.1088/1757-899X/1022/1/012071.

[50] S. Jahan, M.D.S. Islam, L. Islam, et al., Automated invasive cervical cancer disease detection at early stage through suitable machine learning model, SN Appl. Sci. 3 (2021) 806, http://dx.doi.org/10.1007/s42452-021-04786-z.